

**PROF. DR. Ş. İSMAİL KIRBAŞLAR
DOÇ. DR. GAMZE SART
ÜMİT DÜLGER**



STRATEJİK BÜYÜK VERİ YÖNETİMİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ

STRATEJİK BÜYÜK VERİ YÖNETİMİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ

PROF. DR. Ş. İSMAİL KIRBAŞLAR
DOÇ. DR. GAMZE SART
MÜHENDİS ÜMİT DÜLGER

İÇİNDEKİLER

ŞEKİLLER LİSTESİ.....	10
SİMGİ LİSTESİ.....	17
KISALTMA LİSTESİ	18
ÖNSÖZ.....	21
1. BÖLÜM	
GİRİŞ	25
2. BÖLÜM	
BÜYÜK VERİ VE BİLEŞENLERİ.....	33
2.1 BÜYÜK VERİ KAVRAMI VE BÜYÜK VERİNİN GELİŞİMİ	38
2.2 BÜYÜK VERİ BİLEŞENLERİ.....	45
2.2.1 Çeşitlilik (Variety)	46
2.2.2 Hız (Velocity)	48
2.2.3 Veri Büyüklüğü (Volume)	49
2.2.4 Doğrulama (Verification)	52
2.2.5 Değer (Value)	54
2.3 VERİ TIPLERİ VE VERİNİN GERÇEK DEĞERİ	56
KAYNAKLAR	63
3. BÖLÜM	
BÜYÜK VERİNİN YÖNETİMİ.....	67
3.1 BÜYÜK VERİNİN GELİŞİM DİNAMİKLERİ	69
3.2 BÜYÜK VERİYE UYUM STRATEJİLERİ	74
3.3 BÜYÜK VERİNİN İNSAN UNSURU	78
3.4 YENİ MODELLER VE BÜYÜK VERİ	80

3.5 VERİ PAYLAŞIMI, GİZLİLİK VE ETİK	84
KAYNAKLAR	92

4. BÖLÜM

BÜYÜK VERİ TEKNOLOJİSİ	95
4.1 BÜYÜK VERİ TEKNOLOJİSİNE GİRİŞ	97
4.2 BÜYÜK VERİ ANALİZ YAZILIMI: HADOOP	102
4.3 BÜYÜK VERİ ANALİZİ TEKNOLOJİSİ	109
4.3.1 Toplama ve Ön İşleme	109
4.3.2 Depolama	110
4.3.3 İşleme	112
4.3.4 Veri Madenciliği	114
4.4 BAZI ÖZEL BÜYÜK VERİ YAZILIMLARI VE ŞİRKETLERİ	116
KAYNAKLAR	118

5. BÖLÜM

BÜYÜK VERİNİN ZORLUKLARI VE FIRSATLARI	121
5.1 BÜYÜK VERİNİN AVANTAJLARI	129
5.2 BÜYÜK VERİ ANALİZİNDEKİ ZORLUKLAR	131
5.2.1 Kapsamlı Veri Kaynakları ve Kötü Veri Kalitesi	135
5.2.2 Büyük Verinin Oldukça Verimli Depolanması	136
5.2.3 Veriyi Verimli Bir Şekilde İşleme	137
5.2.4 Yığın Veri Madenciliği	139
KAYNAKLAR	140

6. BÖLÜM

BÜYÜK VERİ UYGULAMA ALANLARI	143
6.1 FİNANSAL SERVİSLER	150
6.2 PAZARLAMA VE PERAKENDECİLİK	158
6.3 KAMU VE HÜKÜMETLER	161
6.4 EĞİTİM	163
6.5 SAĞLIK	166

6.6 TARIM	171
6.7 ENERJİ.....	173
6.8 HABERLEŞME VE ULAŞIM	176
6.9 BÜYÜK VERİ KAYNAKLARI	177
6.9.1 Kamusal Veri (Public Data)	178
6.9.2 Özel Veri (Private Data).....	179
6.9.3 Çevresel Veri (Data Exhaust).....	180
6.9.4 Topluluk Verisi (Community Data)	180
6.9.5 Kişisel Veri (Self-Quantification Data)	181
KAYNAKLAR	182
 7. BÖLÜM	
BÜYÜK VERİ YÖNETİMİ İÇİN ÖRNEK VAKALAR	185
7.1 VAKA 1: HIZLI ANALİTİK MODELLEME VE SİMÜLASYON KULLANIMI.....	189
7.2 VAKA 2: MÜŞTERİLERİ EŞLEŞTİRME.....	191
7.3 VAKA 3: MÜŞTERİ DAVRANIŞLARININ DEĞERLENDİRİLMESİ.....	193
KAYNAKLAR	196
 8. BÖLÜM	
BÜYÜK VERİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ İÇİN SAHA ARAŞTIRMASI.....	199
8.1 ARAŞTIRMANIN YÖNTEMİ VE KISITLARI	201
8.2 ARAŞTIRMA EVRENİ VE VERİ TOPLAMA	202
8.3 ARAŞTIRMA ETİĞİ	205
8.4 ARAŞTIRMA SONUÇLARI VE DEĞERLENDİRME	205
KAYNAKLAR	214
 9. BÖLÜM	
TARTIŞMA.....	217
9.1 BÜYÜK VERİNİN KULLANIMI.....	219

9.2 BÜYÜK VERİ HAKKINDAKİ GÖRÜŞLER VE TAHMİNLER	223
KAYNAKLAR	228

10. BÖLÜM

VERİ BİLİMİ, YAPAY ZEKA VE MESLEKLERİN GELECEĞİ	231
10.1 VERİ BİLİMCİ KİMDİR VE ROLÜ NEDİR?	234
10.2 BİR VERİ BİLİMCİNİN GELİRİ NE KADARDIR?	239
A. Deneyime Göre Bir Veri Bilimcinin Geliri.....	240
B. İş Tanımına Göre Bir Veri Bilimcinin Geliri	242
C. Sektörlerle Şirket Büyüklüğüne Göre Veri Bilimcinin Geliri.....	243
D. Bölgelere Göre Veri Bilimcinin Geliri	246
E. Eğitime Göre Veri Bilimcinin Geliri.....	247
10.3 BİR VERİ BİLİMCİYE İHTİYAÇ DUYULAN ALANLAR VE İŞ TANIMLARI.....	249
10.4 VERİ BİLİMCİLERİN TEMEL BECERİLERİ VE ÖZELLİKLERİ.....	252
10.5 VERİ BİLİMCİ OLMAK İÇİN GEREKENLER VE EĞİTİM.....	256
10.6 2020 DÜNYA EKONOMİK FORUMU'NUN MESLEKLERİN GELECEĞİ RAPORU VE TEKNOLOJİNİN MESLEKLERE ETKİSİ	259
10.7 SONUÇ.....	264
KAYNAKLAR	265

11. BÖLÜM

BÜYÜK VERİNİN YÜKSEKÖĞRENİM ÖNCESİNDE EĞİTİME ENTEGRE EDİLMESİ.....	269
11.1 Öğretimde Değerler Değişimi	276
11.2 Pilot Ders: Büyük Veri Analitiği.....	278
KAYNAKLAR	284

12. BÖLÜM

YÜKSEKÖĞRENİMDE VERİ BİLİMİ	287
12.1 GİRİŞ	289

12.2 VERİ BİLİMİNİN TEMELLERİ: İSTATİSTİK, BİLGİSAYAR BİLİMİ, MATEMATİK VE MÜHENDİSLİK	293
A. İstatistik	293
B. Bilgisayar Bilimi	294
C. Mühendislik	295
D. Matematik	296
12.3 VERİ BİLİMİ TOPLULUKLARININ GEREKSİNİMLERİ	298
12.3.1 Biyomedikal Araştırma	298
12.3.2 Endüstri	300
12.3.3 Hükümet	303
12.4 ALAN UZMANLIĞI VE VERİ BİLİMİNİN KESİŞİMİNİN İNCELENMESİ	304
12.4.1 Dil	304
12.4.2 Astronomi	306
12.4.3 Tarih	308
12.4.4 İşbirliği ve İletişim	309
12.4.5 Veri Okuryazarlığı ve Ders Tasarımı	310
12.4.6 Vaka Araştırması	311
12.5 SONUÇ	317
KAYNAKLAR	318
13. BÖLÜM	
ABD'DEKİ ÇOK ULUSLU ŞİRKETLERİN KÜRESEL YAPAY ZEKA AR-GE ETKİNLİĞİNİ HARİTALAMA	323
13.1 GİRİŞ	325
13.2 METODOLOJİ VE VERİ	327
13.3 BULGULAR	330
13.3.1 AI Laboratuvarları ve Personelinin Bulunduğu Yer	330
13.3.2 AI Laboratuvarları Ne Tür Ar-Ge Yürüttür?	335
13.4 KONUM KARARLARINI NE YÖNLENDİRİR?	338
13.4.1 Yeteneklere Erişim	339
13.4.2 Pazar Erişimi ve Ürün Uyarlaması	340

13.4.3 Maliyet Tasarrufu.....	341
13.5 SONUÇ.....	342
KAYNAKLAR	343
14. BÖLÜM	
İŞ YERİNDE VERİ BİLİMİ EĞİTİMİ VE ALTERNATİF MEKANİZMALAR..	347
14.1 DEVLETTE VERİ BİLİMİ UYGULAMASI	349
14.2 Teknoloji Sektörü.....	352
14.3 İŞYERİNDE VERİ BİLİMİ EĞİTİMİ: YÖNETİCİ EĞİTİMİ	357
14.4 SONUÇ.....	364
KAYNAKLAR	365
15. BÖLÜM	
SONUÇ.....	371
KAYNAKLAR	386
EKLER	
EK 1. “STRATEJİK BÜYÜK VERİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ” ARAŞTIRMASI GÖRÜŞME REHBERİ.....	393

ŞEKİLLER LİSTESİ

- Şekil 2.1 Bugün kuruluşların kullanabileceği veri hacmi artıyor, analiz edebilecekleri veri yüzdesi ise düşüyor.
- Şekil 2.2 Veri biliminin kısa tarihi
- Şekil 2.3 Büyük veri bileşenleri
- Şekil 2.4 Küresel veri hacmi tahmini
- Şekil 2.5 Büyük verinin en önemli beş bileşeni
- Şekil 2.6 Büyük verinin değer zinciri
- Şekil 2.7 Elde edilen faydalara karşı beklenen faydalar
- Şekil 2.8 Kuruluşlara göre büyük verinin değeri
- Şekil 2.9 Büyük veri analizinin değeri
- Şekil 3.1 Büyük veri modelleme ve yönetim trendleri.
- Şekil 3.2 Dyche etkin “büyük veri” girişimleri için altı ana basamak
- Şekil 3.3 Küresel bilgi depolama kapasitesinin büyümesi ve dijitalleşmesi.
- Şekil 3.3 2015-2017 yılları arasında büyük verinin benimsenmesi
- Şekil 3.5 Dikey sektör tarafından büyük verinin benimsenmesi
- Şekil 3.6 Fiziksel envanter işleme.
- Şekil 3.7 Teknolojilerin, araçların ve tekniklerin yenilik açısından değeri.
- Şekil 3.8 Hayali bir kiralık araba şirketi için veri modeli örneği.
- Şekil 3.9 Büyük veri gizliliğiyle ilgili endişeler.
- Şekil 3.10 Kişisel gizlilikle ilgili 6 tehdit
- Şekil 3.11 Büyük veri etiğinin dört yönü.
- Şekil 3.12 Büyük veri ve fikri mülkiyet hakları arasındaki ilişki.
- Şekil 3.13 “Büyük Veri,” “Etik” ve “Büyük Veri ve Etik” konularındaki araştırma makalelerinin sayısı.
- Şekil 3.14 Büyük veri sektörünün etik sorunlarına karşı olası çözümleri.
- Şekil 4.1 En iyi büyük veri teknolojileri.

- Şekil 4.2 Veri tabanı yönetim sistemlerinin çalışma şekli.
- Şekil 4.3 SQL ve NoSQL veri tabanı mimarileri.
- Şekil 4.4 Büyük veri teknolojisi.
- Şekil 4.5 Hadoop'un avantajları.
- Şekil 4.6 Hadoop ekosistemi.
- Şekil 4.7 Hadoop MapReduce'un Çalışma Biçimi
- Şekil 4.8 HDFS'nin Çalışma Biçimi
- Şekil 4.9 HDFS Mimarisi
- Şekil 4.10 Büyük veri yönetim sistemleri.
- Şekil 4.11 Büyük veri mimarisi.
- Şekil 4.12 NoSQL Data Stores
- Şekil 4.13 Büyük veri işleme gereksinimleri.
- Şekil 4.14 Arama motoru mimarisi.
- Şekil 4.15 Akıllı öneri sistemi mimarisi.
- Şekil 4.16 2020'deki en iyi büyük veri şirketleri.
- Şekil 5.1 Büyük veriyi kullanan Fortune 1000 yöneticilerinin raporu.
Başlayan projeler ve değer bulduğu yerler.
- Şekil 5.2 Sektörlerin büyük veriye adaptasyonu.
- Şekil 5.3 Sektörler arasında ABD'deki şirketler en az 100 Terabayt veri depolar -çoğu 1 Petabayt'tan fazla veriye sahiptir. 1.000'den fazla çalışanı olan firma başına ortalama depolanan veri (Terabayt, 2009).
- Şekil 5.4 Büyük veri fırsatları: 560 işletmenin %50'den fazlası büyük verinin operasyonel verimliliğini artırmada faydalı olabileceğini düşünüyor.
- Şekil 5.5 Gerçek Zamanlı Büyük Veri İşleme
- Şekil 5.6 Bilgi keşif süreci.
- Şekil 5.7 Büyük verinin fırsat sunduğu alanlar
- Şekil 5.8 Büyük veride endüstri trendi
- Şekil 5.9 Büyük verinin zorlukları.
- Şekil 5.10 Farklı veri kaynaklarından gelen verilerin hacim ölçüği.
- Şekil 5.11 Bir büyük veri senaryosu.

- Şekil 5.12 Büyuk veri ve küresel veri kanunları. Karışık bir bölgesel tahmin
- Şekil 5.13 Yetersiz veri kalitesinin neden olduğu sorunlar.
- Şekil 5.14 Bu grafik şirketin mevcut depolama altyapısına dahil edilen Veri Depolama Teknolojilerinin sayısını vermektedir.
- Şekil 5.15 RC-Seq'in veri işleme verimliliğinin simülasyonla değerlendirilmesi.
- Şekil 5.16 Makine öğrenme algoritmaları.
- Şekil 5.17 Veri madenciliği algoritmaları.
- Şekil 6.1 Bir hizmet pazarı büyülüğu olarak gelişmesine göre Kuzey Amerika'da büyük veri, 2012-2025 (Milyon ABD Doları).
- Şekil 6.2 1 dakikada internette neler oluyor?
- Şekil 6.3 Müşteri düşünme biçimimi.
- Şekil 6.4 Bilmeniz gereken bilişim teknolojisinde başlıca 3 trend.
- Şekil 6.5 Sektöre göre büyük veri kullanımı.
- Şekil 6.6 Finans alanında veri bilimin uygulanması.
- Şekil 6.7 2004-2014 çevrimiçi bankacılık dolandırıcılık kayıpları.
- Şekil 6.8 Dolandırıcılık tespit sistemi.
- Şekil 6.9 Büyük verinin finansta kullanımı.
- Şekil 6.10 Önümüzdeki iki yıl içinde hangi hedeflerin ve teknolojilerin önemli olacağını belirten küresel bankacılar.
- Şekil 6.11 Yukarıdaki dijital kanalların edinme, çapraz satış veya elde tutmaya yönelik olup olmadığına dair şekil.
- Şekil 6.12 Pazarlamada büyük veri.
- Şekil 6.13 Perakende alanında büyük veri kullanım ve projeleri.
- Şekil 6.14 Dijital hükümet stratejileri hakkında OECD tavsiyesi
- Şekil 6.15 OECD OURdata Endeksi: Açık, Yararlı ve Yeniden Kullanılabilir Veriler
- Şekil 6.16 Eğitimde büyük veri
- Şekil 6.17 Öğrenci davranışları analizi yaklaşımı.
- Şekil 6.18 Sağlık hizmeti verilerinde büyümeye.
- Şekil 6.19 Sağlık hizmetlerinde büyük veri.
- Şekil 6.20 Sayısallaştırılmış-Kişisel Hareket ve Sosyal Medyanın Büyük Veri ile Sağlık Alanında Kullanılmasına Örnekler

- Şekil 6.21 Veri kümelerinden sağlanan bilgiyle çiftçilerin gelişimi
- Şekil 6.22 Technavio, 2018-2022 yılları arasında tarım sektöründeki küresel büyük veri pazarı hakkında yeni bir pazar araştırma raporu yayınladı.
- Şekil 6.23 Enerji sektöründe büyük verinin sağladığı avantajlar.
- Şekil 6.24 Ulaşımda büyük veri.
- Şekil 6.25 Taşıma sistemleri akıllı servisler entegrasyonu.
- Şekil 6.26 Büyük veri kaynakları.
- Şekil 6.27 Açık kamusal verinin faydaları.
- Şekil 6.28 Topluluk verisi.
- Şekil 6.29 Fitbit projesinden kişisel veriler. Ocak 2009'dan Ekim 2018'e kadar herkese açık kayıt sayısı (kümülatif toplam).
- Şekil 8.1 Örneklem tablosu.
- Şekil 8.2 NVivo programına görüşmelerin yüklenmesi.
- Şekil 8.3 Nodes ve Tree Nodes kodlama görünümü.
- Şekil 8.4 Nodes ve Tree Nodes kodlamalarının incelenmesi.
- Şekil 8.5 Analizin genel çerçeve modellemesi.
- Şekil 8.6 Analizin detay modellemesi.
- Şekil 8.7 NVivo kelime sorgulamaları.
- Şekil 8.8 NVivo'da grafik desteği örnek gösterimi.
- Şekil 8.9 Görüşme analiz tablosu.
- Şekil 9.1 Büyük veri sektörünüzde kullanılıyor mu?
- Şekil 9.2 Büyük veri kurumunuzda kullanılıyor mu?
- Şekil 9.3 İş dünyasında büyük veri yönetiminin yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir?
- Şekil 9.4 Büyük veri ve veri yönetiminin sınırının olacağını tahmin edebiliyor musunuz?
- Şekil 9.5 Sizce büyük veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir?
- Şekil 10.1 Nesnelerin İnternetinin ölçeklendirilmesi.
- Şekil 10.2 İhtiyaç duyulan becerilere dair tablo.
- Şekil 10.3 Veri bilimcilerinin iş kategorisine göre ortalama taban maaşlarının karşılaştırılması.

- Şekil 10.4 O'Reilly Maaş Veri Bilimi Maaş Raporu.
- Şekil 10.5 Sektörler ve ortalama gelir.
- Şekil 10.6 Şirket büyülüğüne göre ortalama gelir.
- Şekil 10.7 ABD'deki bölgelere göre maaş ortalaması.
- Şekil 10.8 Veri biliminde ihtiyaç duyulan beceriler
- Şekil 10.9 Sektörlerde veri beceri setlerine duyulan ihtiyaç.
- Şekil 10.10 ABD'li işverenlerin veri bilimci istihdamında eğitime bakışı.
- Şekil 10.11 Dijital öğrenme yaklaşımı kavramlarının altı anahtarı.
- Şekil 10.12 Türkiye'de veri bilimi alanında yüksek lisans eğitimi veren üniversiteler.
- Şekil 10.13 Üniversite kapsamında yer alan merkezler ve laboratuvarlar.
- Şekil 10.14 2025 yılına kadar benimsenmesi muhtemel teknolojiler (İncelenen şirketlerin payına göre).
- Şekil 10.15 İncelenen şirketlerin payına ve seçilen sektörlerde göre 2025 yılına kadar benimsenmesi muhtemel teknolojiler ve endüstriler
- Şekil 10.16 Sektörler arasında talebi artan ve azalan ilk yirmi iş kolu
- Şekil 10.17 Yarının işlerinde kümelenen yeni roller.
- Şekil 10.18 2025 için en iyi 15 beceri
- Şekil 11.1 Concordia International School'daki eğitim yönelimi.
- Şekil 11.2 Büyük veri yöneticisi açık pozisyon eğilimi. İlan verilen tüm BT işlerinin bir oranı olarak iş başlığında büyük veri yöneticisini gösteren iş ilanları.
- Şekil 11.3 Büyük veri yöneticisi maaş eğilimi. Büyük veri yöneticisine atıfta bulunulan işlerde verilen 3 aylık ortalama maaş
- Şekil 11.4 Büyük Veri Yöneticisi Maaş Histogramı. 6 aydan 10 Aralık 2020'ye kadar büyük veri yöneticisine atıfta bulunulan işler için maaş dağılımı.
- Şekil 11.5 Üniversiteden İş Dünyasına Giderkenki Doğal Süreç.
- Şekil 11.6 Z Kuşağının En Çok Kullandığı Sosyal Medya Uygulamaları
- Şekil 11.7 Veri bilimci ile ilgili infografik.
- Şekil 12.1 Büyük veri.
- Şekil 12.2 2020'ye kadar 1.000.000 Doldurulmamış İş

- Şekil 12.3 Mühendisliğin Endüstri 4.0'a uyarlanması.
- Şekil 12.4 Matematik ve büyük veri arasındaki ilişki.
- Şekil 12.5 Büyük veri ile Biyomedikal İlerlemeler.
- Şekil 12.6 Büyük veri uzmanlığında insan kaynağı.
- Şekil 12.7 ABD Eyalet Hükümeti İstihdamı: 2009-2016
- Şekil 12.8 Dört alanın her birinde öngörülen yıllık depolama ve bilgi işlem ihtiyaçları veri yaşam döngüsü boyunca sunuluyor.
- Şekil 12.9 Lise sonrası öğrencilerin bilgisayar bilimleri istatistiklerini ve programlama becerilerini geliştirirken astronomide alan uzmanı olmalarını beklemek bir zorluktur.
- Şekil 12.10 Veri okuryazarlığı.
- Şekil 12.11 EScience Enstitü'nün veri bilimi döngüsü, alan bilimiyle ilgili araştırmaları metodolojik gelişmelere bağlamak için bir araç olarak eğitim ve öğretimi içerir.
- Şekil 12.12 Veri bilimi yaşam döngüsünden esinlenen, pilot bir Veri Bilimi 100 dersinin müfredatı.
- Şekil 13.1 Şirketlere göre herkese açık AI laboratuvarlarının sayısı.
- Şekil 13.2 AI laboratuvarlarının ve personelinin şirketlere göre küresel dağılımı.
- Şekil 13.3 ABD dışındaki konumlarına göre ABD'ye göre ABD MNC AI laboratuvarlarının ve personelinin yüzdesi.
- Şekil 13.4 AI laboratuvarlarının konumlarına göre küresel dağılımı.
- Şekil 13.5 Şirketlere göre farklı yerlerdeki AI laboratuvarlarının sayısı.
- Şekil 13.6 Şirketlerin kümülatif olarak 500'den fazla AI personeline sahip olduğu yerler.
- Şekil 13.7 Ar-Ge kategorisine göre AI laboratuvarlarında yapılan Ar-Ge türleri.
- Şekil 13.8 Ülkeye veya bölgeye göre AI laboratuvarlarındaki Ar-Ge türleri.
- Şekil 13.9 Şirketlere göre laboratuvarlarda AI Ar-Ge türleri.
- Şekil 14.1 STEM İşlerinde Öngörülen Artışların Yüzdesi: 2010-2020.
- Şekil 14.2 New York Üniversitesi İş Zekası için Veri Madenciliği dersinde, veriler aracılığıyla ele alınabilecek bir problemin formülasyonunu vurgulayan yinelemeli bir veri bilimi sürecini izler.

- Şekil 14.3 Azure Açık Veritabanı Servisi: Bloklar kurma.
- Şekil 15.1 Türkiye'de büyük veri için altyapısal yeterlilik var mıdır?
- Şekil 15.2 Büyük veri yönetimine olan ihtiyaç, size göre altyapısal mı, yoksa yönetimsel bir konu mudur?
- Şekil 15.3 Büyük verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz?
- Şekil 15.4 Sizce büyük veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır?
- Şekil 15.5 Türkiye'de büyük veriden faydalanan seviyeleri ile ilgili neler söylemek istersiniz?
- Şekil 15.6 Sizce Türkiye'de Google, Facebook, Amazon vb. gibi büyük veriden ciddi anlamda faydalanan benzeri firmalar var mıdır?
- Şekil 15.7 Büyük verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz?
- Şekil 15.8 Büyük verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz?
- Şekil 15.9 Büyük veri yatırımlarınız üzerinde ne kadar etkili olacaktır?
- Şekil 15.10 Büyük veri ile ilgili ABD Beyaz Saray anketi.

SİMGELİSTESİ

- EB** : Eksabayt, 1018
- EUR** : Euro, Avrupa Birliği Para Birimi
- GB** : Gigabayt, 109
- KB** : Kilobayt, 103
- M** : Milyon
- MB** : Megabayt, 106
- PB** : Petabayt, 1015
- TB** : Terabayt, 1012
- USD** : Amerikan Doları, ABD Para Birimi
- YB** : Yottabayt, 1024
- ZB** : Zettabayt, 1021

KISALTMA LİSTESİ

AB	: Avrupa Birliği
ABD	: Amerika Birleşik Devletleri
API	: Application Programming Interface - Uygulama Programlama Arayüzü
BI	: Business Intelligence - İş Zekası
BLOB	: Binary Large Object - İkili Büyük Nesneler
BSP	: Bulk Synchronous Parallel - Paralel Yığın Senkronizasyon
BT	: Bilişim Teknolojileri
CAP	: Consistency, Availability, Partition Tolerance - Tutarlılık, Ulaşılabilitirlik, Bölünebilme Toleransı
CAQDAS	: Computer Aided Qualitative Data Analysis - Bilgisayar Destekli Nitel Veri Analizi
CEO	: Chief Executive Officer - İcra Kurulu Başkanı
CCTV	: Close Circuit Television - Kapalı Devre Televizyon Sistemi
CPU	: Central Process Unit - Merkezi İşlem Birimi
CRM	: Customer Relationship Management - Müşteri ilişkileri Yönetimi
DAG	: Directed Acyclic Graph - Döngüsüz Grafik Diyagramı
DI	: Digital Inventory - Dijital Envanter
EMR	: Electromagnetic Radiation - Elektromanyetik Radyasyon
ER	: Entity Relationship - Varlık İlişki
ERP	: Enterprise Resource Planning - Kurumsal Kaynak Planlaması
FTC	: Federal Trade Commission - Federal Ticaret Komisyonu
GFS	: Google File System - Google Dosya Sistemi
GPS	: Global Positioning System - Küresel Konumlama Sistemi
GSM	: Global System for Mobile Communication - Küresel Mobil İletişim Sistemi

HANA	: High Performance Analytic Appliance - Yüksek Performanslı Analitik Araç
HDFS	: Hadoop Distributed File System - Hadoop Dağıtık Dosya Sistemi
HPA	: High-Performance Analytics - Yüksek Performanslı Analiz
IP	: Internet Protocol Address - İnternet Protokol Adresi
I/O	: Input/Output - Giriş/Çıkış
IDC	: International Data Corporation - Uluslararası Veri Kurumu
IoT	: Internet of Things - Nesnelerin İnterneti
IT	: Information Technology - Bilgi Teknolojileri
MIKE 2.0	: Method for an Integrated Knowledge Environment - Entegre Bilgi Çevresi Yönetimi
NFS	: Network File System - Ağ Dosya Sistemi
NIH	: The National Institute of Health - Ulusal Sağlık Enstitüsü
NoSQL	: Not Only SQL - Yalnızca SQL değil - İlişkisel Veri Tabanı Yönetim Sistemi
ODS	: The Operational Data Store - Operasyonel Veri Depolama
PI	: Physical Inventory - Fiziksel Envanter
POSIX	: Portable Operating System Interface for Unix - UNIX için Taşınabilir İşletim Sistemi Arabirimi
RAID	: Redundant Array of Independent Disks - Bağımsız Disklerin Artıklıklı Dizisi
RAM	: Random Access Memory - Rastgele Erişimli Bellek
RFID	: Radio Frequency Identification - Radyo Frekansı ile Tanımlama
ROI	: Return On Investment - Yatırı Getirisi
SLA	: Service Level Aggrement - Hizmet Seviye Anlaşması
SQL	: Structured Query Language - Yapılandırılmış Sorğu Dili
SSD	: Solid State Drive - Katı Hal Sürücü
UN	: The United Nations - Birleşmiş Milletler
USPTO	: United States Patent and Trademark Office - ABD Patent Ofisi
VP	: Vice President - Başkan Yardımcısı

ÖNSÖZ

2 1. yüzyıl ve Endüstri 4.0 ile birlikte insanlık neredeyse günlük yaşamdan iş yaşamına, eğitimden ülke ekonomilerine kadar yaşamın birçok alanında adeta bambaşka bir dünyaya adım attı. Özellikle Covid-19 salgınından sonra, neredeyse hayatımıza çevirmişi olarak internette yaşamaya başladık; eğitimlerden konserlere, sosyal medya kullanımındaki artıştan konferanslara, ticaretten lojistiğe kadar her şey internete taşındı. Salgınla birlikte teknolojideki ve bilişim dünyasındaki gelişme daha da hızlanarak bir çığ gibi büyümeye başladı. Bu da beraberinde çok fazla kişi ve kaynaktan çok dağınık ve çok büyük oranda “veri” akışı sağladı. Bu veri akışı ise artık bireylerin sağlık durumundan alışkanlıklarına, beğenilerine, tüketim alışkanlıklarına, bulundukları coğrafi konuma kadar bilgi sağlar. Bunun yanı sıra bireyler de kendi elliyeyle bir saniyede birçok bilgiyi paylaşmaktadır.

Dolayısıyla günümüz dünyasında, sosyal medya ve diğer kaynaklardan gelen bir veri patlaması yaşanıyor. Ticari ve hükümetlere ait kuruluşlar ve kurumlar da bu verileri kullanmak üzere dikkatli bir şekilde toplar ve anlamlı bir biçimde yeniden kullanmak için depolamaya çalışır. Fakat bu kadar büyük orandaki veriyi kullanmak özel araçlar ile teknikler gerektirmektedir ki bu da “büyük veri”nin hem şu an hem de gelecekte hayatımızın önemli bir parçası olacağına işaret etmektedir. Bu bağlamda büyük veri, adından da anlaşılacağı gibi, çeşitli yazılım araçlarının yardımıyla toplanması, yönetilmesi veya işlenmesi zor olan büyük miktarda veriyi ifade eder. Büyük veri, bunlardan daha fazla yararlanılabilecek faydalı içgörüler elde etmek için tahmini kullanıcı davranışları veya diğer gelişmiş veri analitiği gibi çeşitli teknik ve teknolojilerin kullanılmasını gerektirmektedir.

Sistemler ve bilgisayarlar arasındaki bağlantı ve büyük miktarda veriyi analiz etme yeteneği, herhangi bir insan müdahalesi olmadan bilinçli kararlar verebilen akıllı makinelerin varlığını mümkün kılmıştır. Nesnelerin İnterneti (IoT), uzun yıllardır öğeleri birbirine bağlamaktadır, ancak büyük veri aracılığıyla verilerden elde edilen değer, bu kavramı da yeni bir düzeye taşıdı: Sistemlerin İnterneti.

Endüstri 4.0, gerçek dünyanın sanal dünya ile birleşimidir. Bu dijital devrim, otomatik öğrenme sistemlerini beslemek için büyük veri ve yapay zekadan (AI) yararlanan teknolojiyle desteklenmektedir. Günümüz pazarındaki üreticiler, üretim mükemmeliğine ulaşmak için tüm temel işlevsel alanlarda verilerin derlenmesi, analizi ve paylaşılması yoluyla iş zekası elde etmeye çalışmayı amaçlamaktadır.

Nesnelerin interneti ve bugünün üretim sistemleri tarafından üretilen bilgi miktarı, eyleme geçirilebilir fikirlere dönüştürülmelidir. Bu nedenle büyük veri, toplanan bilgileri sınıflandırarak farklı biçimlerde şirketlerin operasyonlarını iyileştirmeye yönelik yardımcı olacak sonuçlar sağlar. Bu sonuçlardan ilki kendisini depo süreçlerini iyileştirmede gösterir. Sensörler ve taşınabilir cihazlar sayesinde şirketler, insan hatalarını tespit ederek, kalite kontrolleri yaparak ve optimum üretim veya montaj rotalarını göstererek operasyonel verimliliği artırrarak daha yüksek oranda verim ve yarar sağlayabilir. Bir diğer sonucu da bir şirket ya da kuruluşun bir krizle karşılaşmadan, bir dar boğaza girmesinin önüne geçmesidir. Büyük veri, fazladan hiçbir maliyet olmaksızın performansı etkileyebilecek değişkenleri tanımlayarak üreticilere sorunu tanımlamada yol gösterici olabilir. Başka bir sonucu da talepleri tahminine dayalı olarak belirlemektir. Önceki verilerin ötesinde müşteri tercihleri ve trendlerle harici olaylar yoluyla faaliyetin görselleştirilmesi sayesinde daha doğru ve anlamlı tahminlerde bulunulabilir. Bu da bir şirket ya da kuruluşun ürün portföyünü değiştirmesine ve optimize etmesine olanak tanır. Bunların dışında bakım konusunda da öngörü sağlamaktadır. Veri beslemeli sensörler modellerdeki arızaları göstererek bir arıza gerçekleşmeden önce makinenin çalışmasındaki olası arızaları tespit etmeye yardımcı olabilir. Sistem, zamanında yanıt verebilmesi için ekipmana bir uyarı göndererek arızanın önüne geçebilir.

Bu noktalardan hareketle uluslararası alanda ve ülkemizde teknoloji alanında giderek yüksèlecek bir kavram olarak “büyük veri” karşımıza çıkmaktadır ve bu kavramın endüstriyel alanda yapılabilecek yatırımlar, problem çözümü, süreç iyileştirme, müşteri memnuniyeti, satış politikası ve kuruluş stratejileri oluşturmada kullanılarak

fayda sağlama yönünde yeniliklerle etkisi kendisini daha da göstermektedir. Büyük verinin kullanılmasıyla elde edilecek analizler sadece endüstriyel alanda değil, özellikle e-ticaret, finansal hizmetler, kamu hizmetleri, eğitim, sağlık olmak üzere birçok alanda uygulanarak her alanda daha çok fayda sağlama imkânı gerçeklesecektir.

Büyük veri çağında yalnızca şirket ve kurum stratejileri değizmeyecek, bunun yanı sıra büyük veri gelecekte yeni meslekler için yetişmiş insan gücüne duyulan ihtiyacı da gözler önüne serecek. Teknoloji ve otomasyon bazı mesleklerin sonunu getirecek olması bir tarafa veri analisti, veri uzmanı gibi bu bilgiyi analiz edecek ve yorumlayacak yeni iş kolları oluşturacaktır. Şirketler, kurum ve hükümetler bu noktada insan gücüne ihtiyaç duyarak bu alanda insan yetiştirmeye yönelik eğitimlere yönelecektir. Dolayısıyla eğitim dünyasında da “büyük veri” kendine yer bulmaya başlayacaktır.

Kısaca bu çalışmayla hem dünyada hem de Türkiye’de “büyük veri” ve bunun yatırımlarla şirketler, eğitim ve ticaret üzerindeki etkilerinin araştırılması hedeflenmektedir. Dünya ve ülkemizde büyük verinin durduğu yer ile birlikte büyük verinin önemüzdeki yıllarda nereye evrileceğinden bahsedilmektedir. Bu çalışma sürecinde bu esere katkıda bulunarak destek olan değerli meslektaşlarımıza teşekkürü bir borç biliriz.

1. BÖLÜM

GİRİŞ



Veri sözcüğü, (İngilizce data) “verilen şeyler” anlamına gelen Latince “datum” sözcüğünün çoğulundan türemiştir. Veri bilimi ise geniş bilgi koleksiyonlarının toplanması, hazırlanması, analizi, görselleştirilmesi, yönetimi ve korunmasıyla ilgili ortaya çıkan bir çalışma alanı demektir. Veri bilimi adı en güçlü şekilde veritabanları ve bilgisayar bilimi gibi alanlarla ilişkili gibi görünse de matematiksel olmayan beceriler de dahil olmak üzere birçok farklı alandaki beceriyle ilgilidir. İnternet, “veriye dayalı uygulamalar” ile doludur, örneğin hemen hemen her e-ticaret uygulaması veri odaklı bir uygulamadır. “Veri bilimi” ile kastedilen yalnızca verileri kullanmak değildir. Bir veri uygulaması, değerini verinin kendisinden alır ve sonuç olarak daha fazla veri oluşturur. Bu sadece veri içeren bir uygulama değil, bir veri ürünüdür. Veri bilimi veri ürünlerinin oluşturulmasını sağlar.

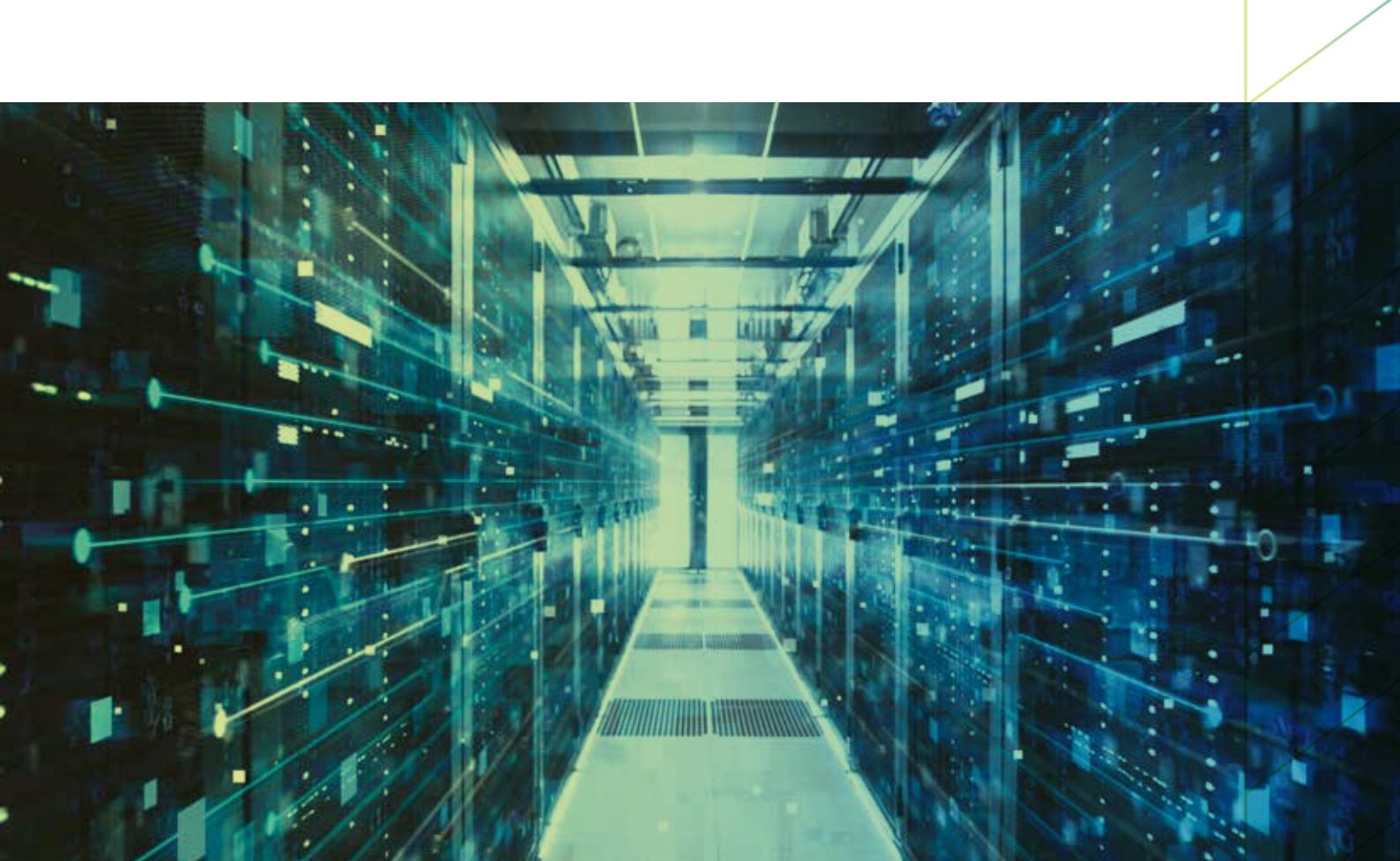
Yaşadığımız çağda, özellikle daha önce belirttiğimiz gibi Covid-19 salgınıyla internet ve teknolojiden uzak bir yaşamı tasavvur etmek neredeyse imkansız bir hale gelmiştir. Şu anda bile Facebook hesabımıza giriş yapmak, Instagram'a fotoğraf yüklemek veya Amazon ya da herhangi bir e-ticaret sitesinde birbirinden farklı ürün ve hizmetlere bakıp incelemek adeta günlük rutinimizin bir parçası haline gelmiş durumda. Whatsapp'ta mesajlarımızı görmediğimizde ya da e-postamıza bakmadığımızda gün içinde bir şeyler eksikmiş gibi hissediyoruz. Teknoloji meraklısı dünyanın ve dünya gündeminin Facebook, Whatsapp, Twitter, vb. gibi çevrimiçi sosyal medya tarafından yönetildiğini görüyoruz. Peki, sosyal medya veya çeşitli kurumsal uygulamalar tarafından günlük olarak üretilen peta ve eksabaytlarca verinin boyutunu ve

çeşitliliğini hiç düşündünüz mü? Wikipedia'ya göre her gün çeşitli çevirmiçi uygulamalar tarafından 2,5 eksabayt veri oluşturuluyor. Bu kadar büyük miktarda veriyi yönetmek ve oynamak oldukça zor hale geliyor. Bilişim dünyası daha biz “büyük veri” kavramına yabancıken büyük oran ve çeşitlilikteki veri ile çalışıyordu lakin bu konu Operasyonel Veri Depolama (ODS/The Operational Data Store), veri ambarı, iş analizi ve veri analizinden büyük ve önemli ölçüde farklı olarak başka bir yerde konumlanmaktadır.

Büyük veri, tek bir teknoloji değil, şirketlerin eyleme geçirilebilir içgörüler elde etmesine yardımcı olan eski ve yeni teknolojilerin bir kombinasyonudur. Bu nedenle büyük veri, gerçek zamanlı analiz ve reaksiyona izin vermek için çok büyük miktarda farklı veriyi doğru hızda ve doğru zaman çerçevesinde yönetme yeteneğidir. Diğer taraftan büyük veri derken kast ettiğimiz aşina olduğumuz veya bildiğimiz süreç veya araçların kullanılmasıyla işlenemeyecek veya analiz edilemeyecek bilgiler için geçerlidir. Bugün şirket, kuruluş ve hükümetler giderek daha fazla büyük veri sorunuyla karşı karşıyadır. Zengin bir bilgiye erişimleri söz konusu olmasına rağmen en ham haliyle ya da yarı yapılandırılmış veya yapılandırılmamış bir biçimde oldukları için bunlardan nasıl değer elde edeceklerini, hatta saklamaya deşip dezmeyeceğini bile bilemezler. Şirketler, herhangi bir şeyi depolayabildikleri ve tarihte hiç olmadığı kadar çok veri ürettikleri bu ortamda bu tür zorluklarla yüzleşirler. Bugünün işletmeleri potansiyel içgörülere her zamankinden daha fazla erişime sahipken bu potansiyel altın madeni veri yığınları arttıkça işletmenin işleyebileceği veri yüzdesi hızla düşer.

Yaşadığımız bu teknoloji çağında orta ölçekli bir KOBİ bile 1 TB'lık veriyi çok az bir sürede üretebilmektedir. Bu bilgiler, yani veriler birçok kaynaktan gelerek geniş bir çeşitlilik ve büyük bir hızla sağlanmaktadır. Dolayısıyla veri yerine, “büyük veri”

Teknolojinin yaygınlaşması, herkes tarafından ulaşılabilir olması ve kullanım biçiminin dönüşmesi üretilen ve işlenen verinin büyüklüğünün de devasa olarak artmasını beraberinde getirmiştir.



Veri merkezi

olgusuya karşı karşıya kaldık. Büyük veri, bir işi günlük bazda dolduran hem yapıilandırılmış hem de yapılandırılmamış büyük hacimli verileri tanımlayan bir terimdir. Ancak önemli olan verinin miktarı değil, bu elde edilecek ve anlamlandırılacak verilerle kuruluş ve şirketlerin ne yaptığıdır. Daha iyi kararlara ve iş alanında stratejik hareketlere yol açan içgörüler için bu büyük veriler analiz edilerek büyük veriyi kullanmak günümüzde şirketler tarafından emsallerinden daha iyi performans göstermesi için yaygın hale geliyor. Coğu sektörde hem mevcut rakipler hem de yeni girişimler, rekabet etmek, yenilik yapmak ve değer elde etmek için analiz edilen verilerden kaynaklanan stratejileri kullanıyor ve gelecekte de daha da fazla kullanacaktır. Bununla beraber her ne kadar bilişim dünyasının bildiği bir alan olmasına rağmen bu verinin anlamlı ve işe yarar bir hale gelmesi için işlenmesi, yönetilebilmesi gerekmektedir ki bunun için de farklı yazılımlar, algoritmalar ve teknolojilere ihtiyaç duyulmaktadır.

Büyük veri basit teknik bir konu değil, ayrıca bir yönetim konusudur; yalnızca teknik yönetim gerektirmeyen yeni veri dünyası ve sınırlıdır, fakat potansiyel kullanımı ve gücünün nasıl dizginleneceği üzerine kararlara yoğunlaşmıştır.

Yüksek hızda üretilen veri beraberinde verinin aynı hızda alınması, anlaşılması, işlenmesi ve yorumlanması konusunda birtakım sorunlara da neden olmaktadır. Veriler bir çok farklı formattan oluşmaktadır; sesler, videolar, metinler, resimler, dolayısıyla bu büyük veri setlerinin entegrasyonu oldukça zorlu ve karmaşıktır. Bu entegrasyon sırasında analiz, veri iyileştirme, yakalama, paylaşım, arama, görselleştirme, bilgi gizliliği ve depolama gibi karşılaşılabilen çeşitli zorluklar mevcuttur. Büyük verinin ana öğeleri verileri geleneksel ilişkiye dayalı veri tabanına kıyasla yeni yöntemlerle ele almayı gerektirmektedir. Büyük veri yönetimini doğru bir şekilde gerçekleştirmek kararların daha güvenli bir şekilde alınmasını sağlayacaktır. Belirttiğimiz gibi, büyük verinin entegrasyonunu ve süreç sırasında karşılaşılabilen zorluklara değinecek olursak şayet, bunlar arasında veri yönetiminin belirsizliği, büyük veri yetenek açığı, verileri büyük bir veri yapısına sokma, veri kaynakları arasında senkronizasyon, büyük veriden yararlı bilgiler elde etme, hacim, beceri kullanılabilirliği, çözüm maliyeti vb. unsurlar sayılabilir.

Büyük verinin işlenmesi esnasındaki zorlukları yanında avantajları da bulunmaktadır. Büyük verinin etkin kullanımı ekonomileri dönüştürmeye yardımcı olarak yeni bir üretken büyümeye dalgasına yol açmaktadır. Büyük veriden ziyade faydalı ve değerli bilgilerden yararlanmak, mevcut işletmelerin temel rekabet stratejisidir. Yeni rakipler büyük verileri işlemeye kritik becerilere sahip çalışanları çekerilmelidir. İşletmeler, büyük veriyi kullanarak artan operasyonel verimlilik, bilinçli stratejik yönlendirme, iyileştirilmiş müşteri hizmetleri, yeni ürünler ve yeni müşteriler ve pazarlar dahil olmak üzere birçok avantaj elde eder, hele de küreselleşen ve gittikçe küçülen bir dünyada. Avantajları ve dezavantajları göz önüne alınarak değerlendirildiğinde,

yeni veri yönetim teknik ve stratejileri geliştirilmektedir. Dolayısıyla bu yeni büyük veri analizi teknoloji ve stratejileri daha eski teknolojilerle entegre edildi. Bu yüzden büyük veri sadece “yazılım” açısından değil, bütünsel bir perspektiften değerlendirilmelidir.

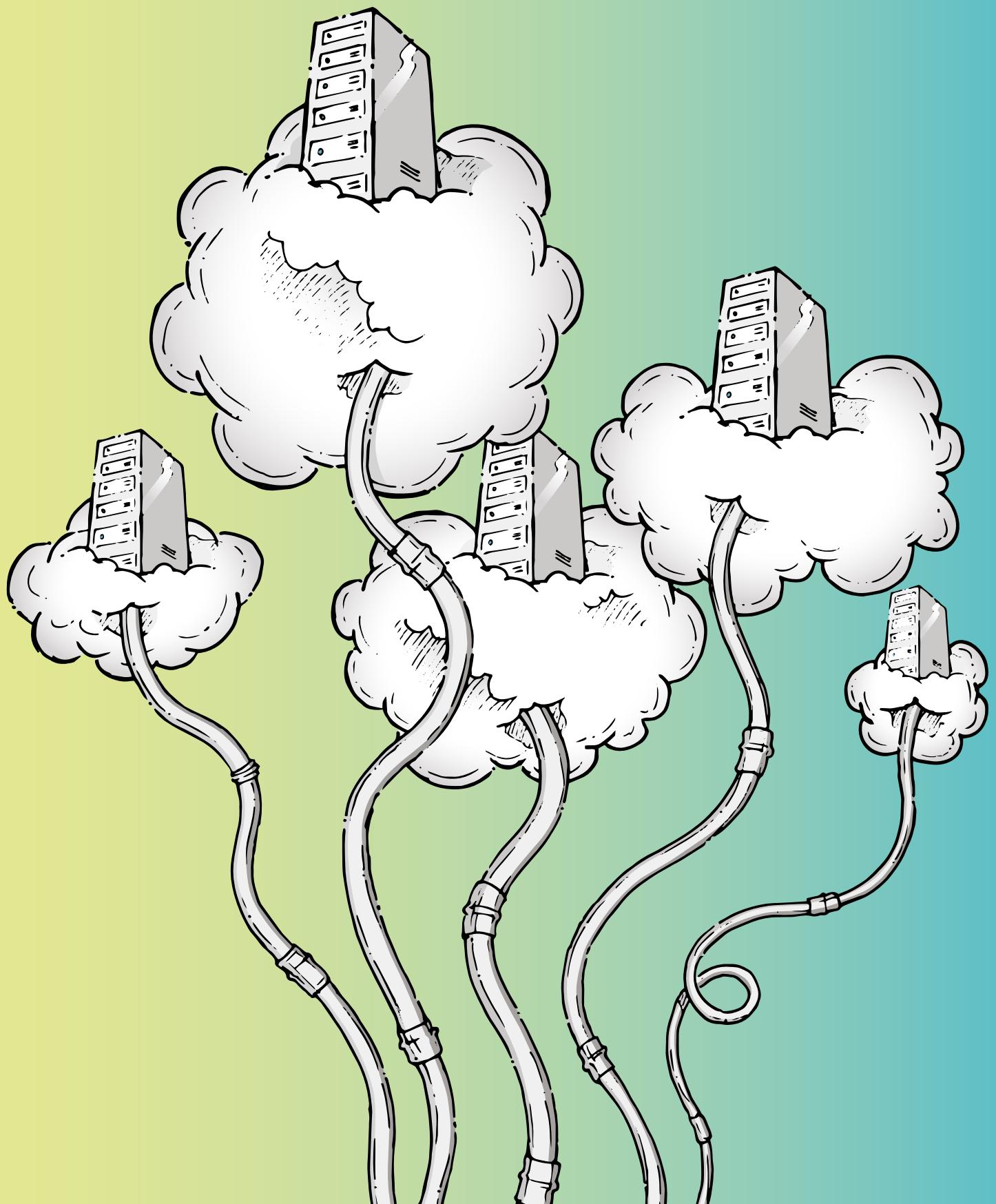
Basitçe ifade etmek gerekirse, büyük veri sayesinde yöneticiler işlerini ölçebilir ve bu nedenle işlerini radikal bir şekilde bilebilir ve bu bilgiyi doğrudan gelişmiş karar alma ve performansa çevirebilir. Yöneticiler veriye dayalı karar alarak daha iyi karar alma eğilimindedirler. Dolayısıyla yöneticiler bu gerçeği benimsemek zorundadır. Kendi alanlarında yöneticiler alan uzmanlığını veri bilimi ile nasıl birleştireceklerini çözebilirlerse, şirketlerinin rakiplerinin önüne geçeceği yadsınamaz bir gerçek. Fakat konunun karmaşık ve teknik olması, yönetilmesinde zorluklar ve alanda yeterli insan gücünün bulunmaması büyük veri kullanımını sınırlırmaktadır. Diğer taraftan olası faydalı kullanımını ve bu gücün nasıl dizginleneceği üzerine kararlara yoğunlaşmıştır. Potansiyel müşterilerin nasıl kazanılacağı üzerine daha çok çalışılmış kararları oluşturmada şirketlere imkân vereceği için birçok eleştirmen pazarlama sektöründe büyük verinin potansiyeli hakkında yazılar yazmıştır.

Bilhassa işletme yöneticileri büyük veri ile uğraşırken bu karmaşık gerçekleri göz ardı edebilirler. Bundan dolayı bu çalışma daha öncesine göre gitgide artan, daha büyük ölçeklerde tanımlanan büyük veriye işletme yöneticilerinin, müdürlerinin, bilşim dünyasındaki profesyonellerin nasıl adapte olacağını, bunun üstesinden nasıl geleceğini veya bununla nasıl yüzleşeceğini göstermeyi de amaçlamaktadır. Görüşülen kişiler bu araştırma için örnek bir temsilci olarak seçildi. Yarı yapılandırılmış röportajlar büyük verinin, kurumların yapısını veya iç operasyonlarını ne derece değiştirdiği ve büyük verinin varlığına nasıl adaptasyon sağlandığının sorulduğu bu kişiler ile yürütüldü.

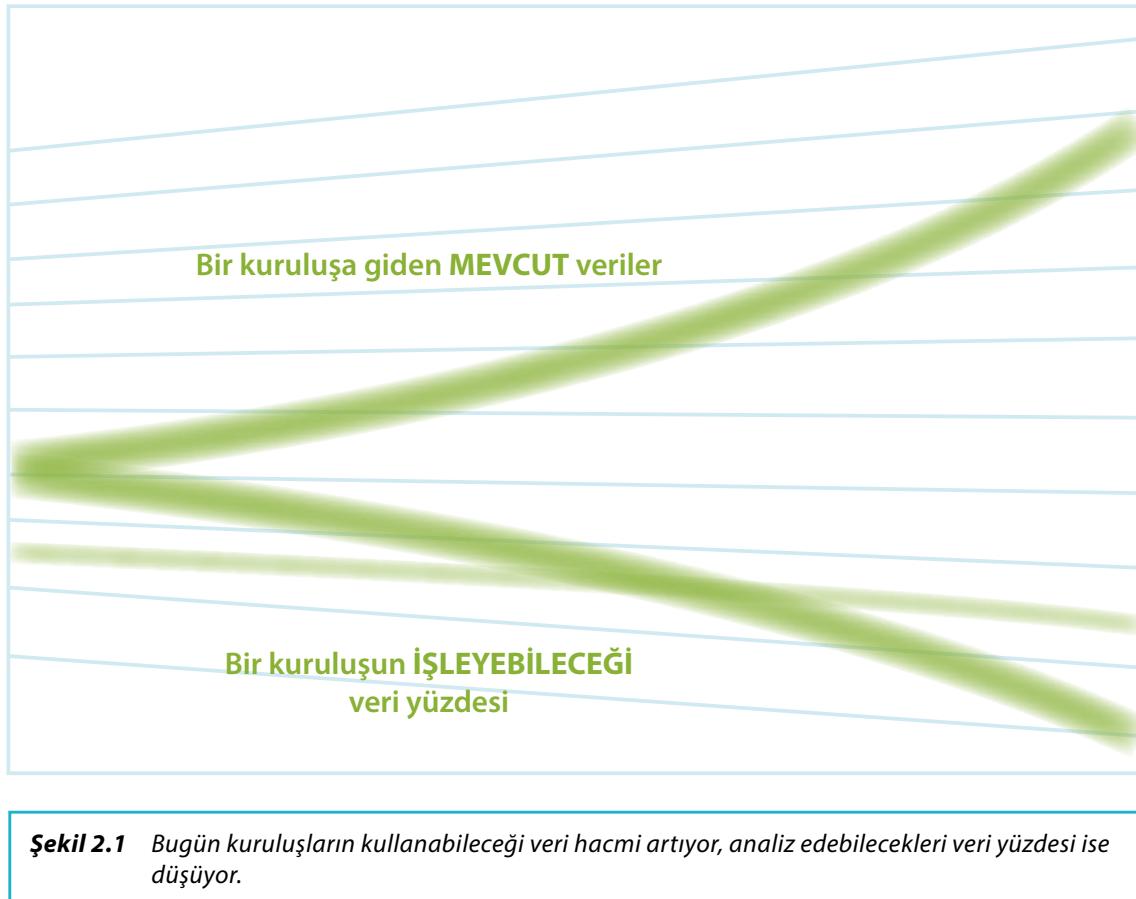
Olgusal bir yaklaşımın kullanıldığı bir çalışma olarak büyük veri ile ilgili katılımcıların deneyimleri, yargı ve görüşleri de dikkate alınarak dünyada “yeni,” Türkiye’de “daha yeni” olan stratejik “büyük veri” ve bunun yatırımlar üzerindeki etkilerinin araştırılması hedeflenmiştir.

2. BÖLÜM

BÜYÜK VERİ VE BİLEŞENLERİ



Kırk yıldır ERP (Enterprise Resource Planning-Kurumsal Kaynak Planlaması) büyük işletmelerin önemli bir ögesi haline gelmiştir. Bu zamanlarda şirketler ve üreticiler adına önem arz eden konular üretikleri ürünler ve bunları müşteriye ulaştırmaktı ve bu sistem dolayısıyla 2020'lerden önce daha çok ön plandaydı. Ancak tüm veriler ERP sistemlerinden geçmemektedir. Büyük veriyi ve ERP sistemleri tarafından toplanan bilgileri yönetmek zor bir görev haline gelebilir. Veri işleme için birçok karmaşıklığa yol açabilir. Bu nedenle büyük veri araçları, veri kaynaklarını ayırabilecek ve bunları işleyebilecek şekilde tasarılmıştır. Veri işleme tamamlandıktan sonra da kullanılmak üzere kullanıcı dostu bir forma dönüştürülür. ERP, kendi içinde büyük veriler için bir veri bankası görevi görür. Hem büyük veri hem de ERP paralel olarak birbirini besleyebilirler. ERP sistemlerinin değeri böylelikle büyük veri entegrasyonu ile büyük ölçüde arttırılabilir. Sosyal medya platformları, sensör ağları ve mobilite büyük veri ve ERP'nin birlikte işleyişini büyük ölçüde etkilemektedir. Şu anda çok sayıda işletme ve kurum müşteri tercihlerini anlamak için sosyal ağ sitelerinden yararlanmaktadır. Büyük verinin kullanılabilirliğiyle işletmeler müşterilerinin hoşlandıkları ve hoşlanmadıkları şeylerin neler olduğunu anlamaya dönük bir içgörü elde edebilir. Bu nedenle işletmeler, müşterilerine



Şekil 2.1 Bugün kuruluşların kullanabileceği veri hacmi artıyor, analiz edebilecekleri veri yüzdesi ise düşüyor.

çok çeşitli seçenekler için özelleştirilmiş öneriler sunabilir ve bu da gelirin artmasını sağlayabilir. Büyük verilerin uygulanması, satışların artmasıyla sonuçlanmaya bile, kullanıcı etkileşimiini artttırmaya yardımcı olabilir. Diğer taraftan daha çok ürün ve müşteriye ulaşımına odaklanan sistem bir noktada doygunluğa ulaşarak değişen dünya ile beraberinde yeni sorunları ortaya çıkardı. Yapay zeka ve makine öğrenimi gibi gelişmiş teknolojilerle müşteriyi daha iyi anlamak ve özel öneriler sağlamak için büyük verilerden yararlanılabilir. Ürünün ne tür müşterilere ulaşacağı sorusu gündeme geldi; bu tüketici kitle kimlerden oluşuyordu?

Bu noktada CRM sistemleri ortaya çıkmıştır. CRM'nin (Customer Relationship Management-Müşteri ilişkileri Yönetimi) bir şirketin mevcut ve gelecekteki müşterileriyle etkileşimiğini içermektedir. Tüm müşteri temas noktalarının birleşimidir. Daha iyi müşteri hizmeti sunmak üzere bir önkoşul haline gelmiştir. CRM her zaman verileri içermektedir, fakat bu verilerin çoğu iletişim bilgileri, en son kişiler, satın

alınan ürünler vb. gibi yapılandırılmış verilerdir. Geçmişte çoğu zaman bir CRM sistemi bekentileri karşılayamıyordu çünkü aslında CRM sadece müşteri ilişkilerini yönetmekle ilgiliydi. Müşteriye hizmet vermek daha iyi sonuçlar verecektir, ancak sadece ilişkiyi yönetmekten çok daha fazlasını gerektirmektedir. CRM'nin ortaya çıkışı beraberinde artık iş dünyasında bambaşka bir dönem başlamıştır; müşteriye göre ürün üretimi ki bu bakış açısı günümüzde önemini giderek daha da artttırmaktadır. Bu durum da artık yatırımcılar ve girişimciler için ve tabii medya platformları için yeni faydalara sağladığından bu kuruluşların büyük veriye daha çok odaklanmalarına neden oluyor. Ayrıca gelişen teknolojiler de büyük veri ve sistemlerinin yükselişine öne yak olmaktadır, örneğin bulut barındırma çözümleri uygun fiyatından dolayı genele yayılmış ve veri işleme konusundaki ekonomik dengeleri temelden değiştirmiştir.

Büyük veri teriminden de anlaşılacağı gibi, kuruluşlar büyük miktarda veri ile karşı karşıyadır. Bu verileri nasıl yöneteceğini bilmeyen kuruluşlara etkisi yadsınamaz. Ancak doğru teknoloji platformuyla bir işletmeyi, müşterileri ve pazarı daha iyi anlamak için neredeyse tüm verileri (veya en azından kendileri için yararlı olan verileri tanımlayarak daha fazlasını) analiz etme fırsatı vardır. Bu da günümüzün tüm sektörlerindeki işletmelerinin karşı karşıya olduğu mevcut muammaya yol açar. Kuruluşun kullanabileceği veri miktarı artarken işleyebileceği, anlayabileceği ve analiz edebileceği veri yüzdesi azalmakta, dolayısıyla Şekil 2.1'de görülen kör alanı oluşturmaktadır. O kör alanda harika bir şey de olabilir veya hiçbir şey de olmayabilir.

Şu an bile, anbean veriler gelmeye devam ediyor ama tüm bu veriler geleneksel sistemlerde depolanamaz. Verilerdeki bu büyük artışın en büyük nedeni yapısal olmayan verilerdir ve yapısal olmayan verilere olan bakış açısı da değişmeye başlamıştır. Gerek bilişim dünyası, gerek şirketlerde bugüne dek yapısal olmayan verilerin çoğunu hiçbir değeri olmadığına dair genel bir yargı bulunmaktaydı. Fakat özellikle de Covid-19 sonrası e-ticaret sitelerindeki artış ve bu sitelerdeki tıklanma oranı önüne alınarak elde edilen başarı ve memnuniyet bu yargının ne kadar geçersiz olduğunu gözler önüne sermiştir. Dolayısıyla artık her türden verinin saklanarak analizinin yapılması ve veri madenciliğinde sürecin önemini daha da artttırmıştır.

Yukarıda üzerinde durduğumuz noktalar doğrultusunda kısaca özetlemek gerekiyor; büyük veri artık yaşamımızın ayrılmaz bir parçası olmuş ve olmaya daha da devam edecektir. Dolayısıyla bu bölümde, genel olarak veri kavramına bakış, büyük veri

nedir, gelişimi ve veri yönetimi, zorlukları ve sunduğu fırsatlar, kuruluşların büyük veriye uyum sağlamaası, insan unsuru, verinin değeri ve yeni modeller gibi konulara degeinilecektir. Büyük veri ile ilgili kavramlar ve tanımlamalar olabildiğince açılığa kavuşturulmaya çalışılacak ve bu yeni olgunun tam olarak ne anlam ifade ettiği ortaya konulacaktır.

2.1 BÜYÜK VERİ KAVRAMI VE BÜYÜK VERİNİN GELİŞİMİ

Geçmişten günümüze kadar gelen süre içerisinde bilgi günden güne artmıştır, hatta son yıllarda katlanarak ölçülemez bir boyuta ulaşarak varlığını sürdürmektedir. Bunun sonucunda yararlı ya da yararsız ayırt etmeksiz bir “bilgi yığını” meydana gelmiştir. Artan miktar ve çeşitlilikteki bilgi ise birçok yazılım şirketinin bu konuya yönelmesine ve bu konuya yönelik Ar-Ge çalışması yapmasına neden olmuştur. Bu gelişme de karşımıza “big data” yani büyük veri kavramının yaşamımıza girmesine neden olmuştur.

“Big data” yani “büyük veri” terimi 1990’ların başından beri kullanılmaktadır. Son yıllarda, verilerin üretildiği hacim ve hız insanın kavrayış ölçütünün ötesinde değişmiştir. Dünyadaki toplam veri miktarı 2013 yılında 4.4 zettabayttı. 2020 yılından sonrası için hızla 44 zettabayta, yani 44 trilyon gigabayta çıkması bekleniyor. Günümüzün en ileri teknolojileriyle bile tüm bu verileri analiz etmek neredeyse olanaksızdır. Dolayısıyla karmaşık ve bu aşırı miktardaki veriden dolayı büyük veri bilgisayar dünyasına ilk olarak 2005 yılında O'Reilly Media'dan Roger Magoulas tarafından getirilmiştir.

Her ne kadar “büyük veri” terimi internet ve teknolojinin gelişmesiyle daha bilinir hale gelse de 1970’lerde “büyük veri” ile ilgili araştırmalar bulunmaktadır. Bu tarihlerde IBM’de matematikçi olan Edgar F. Codd “ilişkisel veri tabanı” için kendi çerçevesini sunmuş, bu tarihe kadar araştırma ve geliştirme ile akademik dünyada kendine yer bulmaya başlayan kavram bilgisayarların ilk ticari kullanımıyla kendine yer bulmaya başlamıştır. Büyük veri teriminin bugün kullanıldığı şekliyle ilk kez ifade edilişi ise yazar Erik Larson’ın Harpers Magazine için kaleme aldığı bir makalede yer almaktadır: “Büyük veriyi koruyanlar, bunu tüketicinin yararı için yaptıklarını söylüyor.”

"Big data," yani "büyük veri" terimi, karmaşıklığı ve boyutundan dolayı geleneksel veri yönetim teknikleri ile yönetilememesi ve işlenememesinden dolayı, çok büyük miktarda veriyi tanımlamak için bilgisayar dünyasına ilk olarak 2005 yılında O'Reilly Media'dan Roger Magoulas tarafından tanıtılmıştır.

1990'larda bugün anladığımız şekliyle internetin ortaya çıkmasıyla dünya daha küçüllererek büyük veri için sonsuz bir kaynak oluşturmaya başlamıştır. Bunun yanı sıra R. J. T. Morris ve B. J. Truskowski yazdıkları kitapta bu tarihlerin dijital depolamanın kâğıttan daha uygun maliyetli hale geldiği yerin burası olduğunu belirtmektedir.

2000'li yıllarda ise hem uygulamada hem de yaynlarda büyük veri kendine yer bulmaya başladı. Yaşadığımız bu çağda ise büyük veri kavramı birçok alandaki uygulamalarda farklı biçimlerde değerlendirilmektedir.

Veri, yani kişisel bilgi parçaları tarih boyunca toplanmış ve kullanılmıştır. Son günlerde değişen şey ise veri toplama, depolama ve analiz etmede belirgin bir şekilde artan yeteneklere sahip dijital teknolojilerdeki ilerlemelerdir. 1880 yılında ABD'de farklı yaş, cinsiyet, ev hane sayısı, etnik köken, doğum tarihi, medeni durum, meslek, sağlık durumu, eğitim ve doğum yerini içeren demografik bilgilerin toplandığı 50 milyon insanın nüfus sayımı gerçekleştirildi. Tüm bu bilgiler bir elde toplandı, incelendi ve devlet arşivlerine, kütüphaneler ve üniversitelere depolanmak için gönderildi. İlk toplamadan sonra düzgün bir şekilde tablolara dökülmesi ise yedi sekiz yıl sürdü.

1890 yılında Sayım Bürosu bir yıl içinde tablo haline getirilebilen makine-okuma-lı zimba kartlarını benimseyerek veri toplama metodlarını yayınladı. En son dönemde ABD sayımı 2010 yılında yapıldı ve sayım bürosu halk ile anket yapmak için coğrafi bilgi sistemleri, sosyal medya, videolar, akıllı karakter tanıma sistemleri ve gelişmiş veri işleme yazılımları dahil yeni gelişen bir dizi teknoloji kullandı.

BİLGİSAYAR BİLİMI

- Liebniz-İkili Mantık

- Turing makinesi
- Bilişim teorisi
- Weiner & Sibernetik
- Von neumann mimarisi

- Sıralama & arama algoritmaları - Dijkstra, Kruskal, Shell Sort
- Sezgisel analiz (Heuristics) - Benzetimli tavlama

VERİ TEKNOLOJİSİ

- Catography
- Astronomik grafikler

- William Playfair
- Charles Minard
- Florence Nightingale

- Çıkarılabilir disk sürücülerı
- İlişkisel DBMS

GÖRSELLEŞTİRME

- Analiz
- Logaritma
- Newton-Raphson

- Optimizasyon metodu
- Fourier ve diğer dönüşümler
- Matris ve genelleştirme
- Öklit dışı geometri

- Orduya başvurular, Üretim, İletişim

- Networkler
- Atama problemleri
- Otomasyon
- Programlama

MATEMATİK/OR

- Olasılık
- Korelasyon
- Bayes teoremi

- Modern istatistikin Teorik Temeli
- Hipotez, DOE
- Matematiksel istatistikler

- Bayesian metodu
- Zaman serileri metodu (Box Cox, Survival vb.)
- Stokastik yöntemi

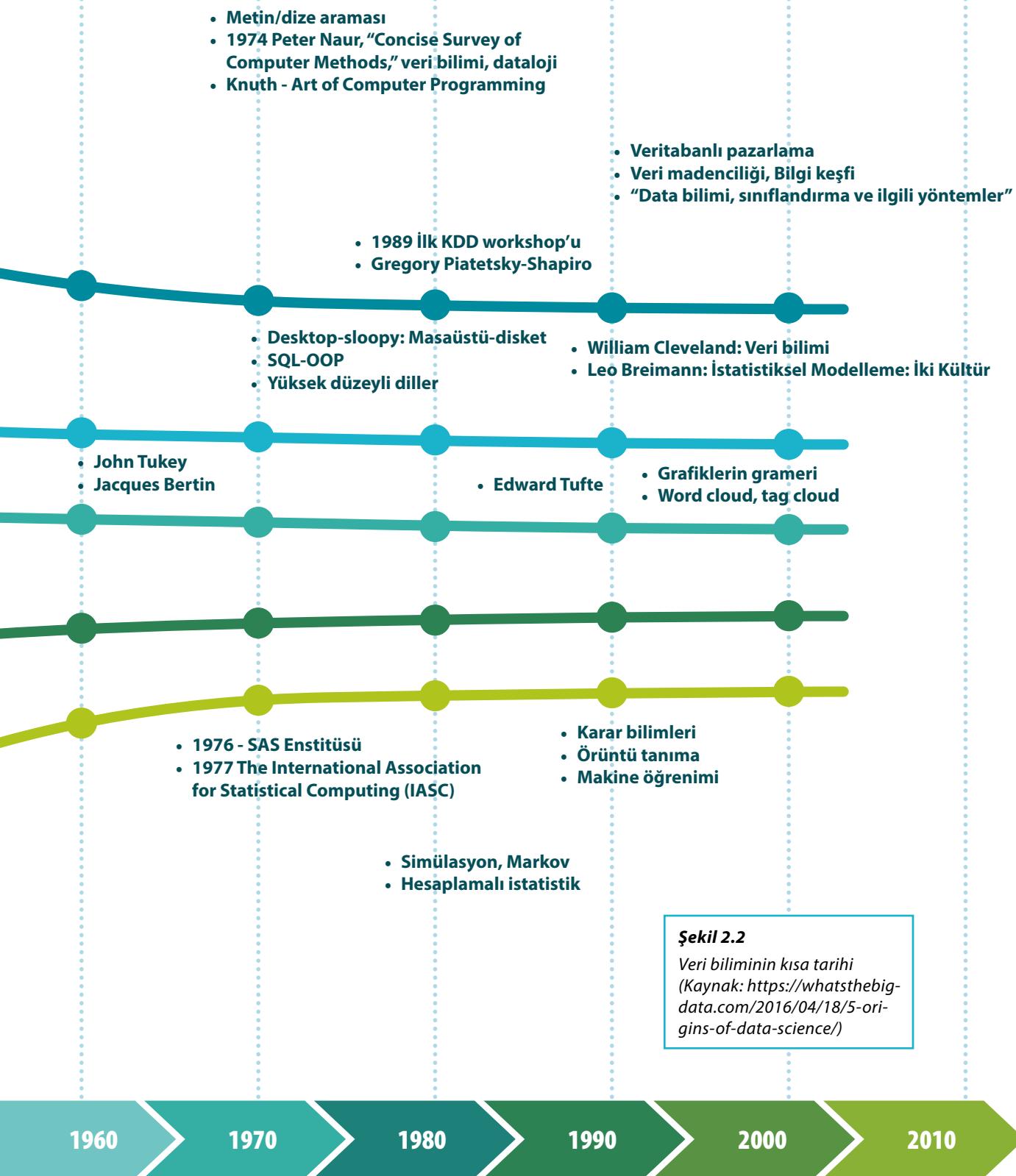
İSTATİSTİK

1880'lərden öncəsi

1800-1900

1900-1940

1940-1960



Bugün büyük veri tek veri depolarının (veri tabanları ya da veri ambarları) çok ötesinde veri setlerini ifade etmek için kullanılmaktadır. Geleneksel veri yönetim ve işleme araçları tarafından işlemek için çok büyük ve çok karmaşıktır. Büyük veri işlemel veriler, sosyal medya, kurumsal içerik, sensörler ve mobil cihazlar gibi bilgileri kapsayabilir.

MIKE 2.0'a (Veri Yönetimi için Açık Kaynak Standardı) göre, büyük veri birbirıyla etkileşimli, çok büyük, karmaşık ve bağımsız veri setleri ve boyutu tarafından tanımlanmaktadır. Ek olarak büyük verinin önemli bir yanı geleneksel veri yönetimi ile ele alınamaması geçerlidir.

O'Reilly Strata Konferansı'ndaki Ed Dumbill'e göre, büyük veri "geleneksel veritabanı sistemlerinin işlem kapasitesini aşan veri" olarak açıklanabilir. Veri çok büyütür, çok hızlıdır, sizin veritabanı mimarinizin yapılandırmasına sığdırılamaz. Bu veriden değer elde etmek adına onu işlemek için alternatif yollar seçmelisiniz.

Daha basit bir tanımlamayla büyük veriyi işletme süreçleri için belirli metodlar ve teknikler çok büyük, çok karmaşık, yapılandırılmış, organize edilmiş, depolanmış ve işlenmiş farklı veri setlerinden oluşmaktadır ifadesini dikkate alabiliriz.

Büyük verinin dünyada birçok farklı tanımı vardır, fakat her şirket kendisinin önem verdiği şekilde tanımlamaktan yanadır. Her lider şirket rekabet avantajı sağlamak için bir kavram tanımlamak zorundadır.

Büyük veri toplumsal medya paylaşımları, ağ günlükleri, bloglar, fotoğraf, video, günlük dosyalar vb. gibi değişik kaynaklardan toplanan tüm verinin anlamlı ve işlenebilir biçimde dönüştürülmüş haline denir.

Pennsylvania Üniversitesi Ekonometri profesörlerinden Francis X. Diebold'un özet çalışmasında (2012) ifade ettiği üzere "büyük veri" kavramı, 1990'ların ortalarından itibaren üretim yönetimi, bilişim, istatistik, ekonometri gibi farklı alanlarda analiz edilen veri miktarının büyüklüğe dikkat çekmek için kullanılmaya başlanmıştır. Kavramın akademik literatüre girişi ise S. M. Weiss ve N. Indurkhya'nın "Predictive Data Mining: A Practical Guide" (1998) başlıklı bilişim sistemleri ve F. X. Diebold'un "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting" (2000) ekonometri konulu çalışmaları sonrasında mümkün olmuştur. Akademik literatürde "büyük veri" kavramının çağrıstdığı "algılayıcılardan ve bilimsel araçlardan büyük hacimde, yüksek çeşitlilikte ve hızda gelen

Büyük veri toplumsal medya paylaşımıları, ağ günlükleri, bloglar, fotoğraf, video, günlük dosyalar vb. gibi değişik kaynaklardan toplanan tüm verinin anlamlı ve işlenebilir biçimde dönüştürülmüş haline denir.

verilerin toplanması, saklanması, temizlenmesi, görselleştirilmesi, analiz edilmesi ve anlamlandırılması” eylemidir (Gürsakal, 2013, s. 21).

En basit tanımıyla ise büyük veri (big data), gerek insan gerekse makineler tarafından sayısal olarak kodlanmış her türden kurumsal veri ile internet ve sosyal medya paylaşımları aracılığıyla ortaya çıkan kişisel verilerin anlamlı ve işlenebilir biçimde dönüştürülmesi durumudur. Bu kapsamda halen yapılandırmamış enformasyon olarak duran ve degersiz (çöplük) olarak algılanan veriler arasında, sosyal medyanın kamuya açık API’leri (Application Programming Interface-Uygulama Programlama Arayüzü) aracılığıyla sayısal platformların kullanıcılarından elde ettikleri tüm bilgilere ulaşılabilmekteidir. Uygulama programlama arayüzü anlamına gelen API, herhangi bir uygulamanın belli işlevlerini diğer uygulamaların da kullanabilmesi için oluşturulmuş bir modüldür. Örneğin Google AdWords Uygulama Programlama Arayüzü (API), teknik konularda birikimi olan büyük reklam verenler ile üçüncü tarafları temsil eden geliştiriciler için tasarlanmıştır. Bunlar arasında ajanslar, arama motoru pazarlamacıları ve birden çok müşteri hesabını veya büyük kampanyaları yöneten diğer çevrimiçi pazarlama uzmanları yer almaktadır. Google AdWords API’sı geliştiricilere doğrudan Google AdWords sunucusuyla etkileşim kuracak uygulamalar oluşturma olanağı vermektedir; reklam verenler ve üçüncü taraflar da bu uygulama sayesinde büyük veya karmaşık AdWords hesaplarını ve kampanyalarını daha etkin ve yaratıcı biçimde yönetebilmektedir.

Olageldiği gibi, ilişkisel veri tabanlarında tutulan, yapısal verinin dışında kalan, son dönemlere dek çok da kullanılmayan yapısal olmayan veri yığınıdır. Artık yıkılmış olan yaygın bilişimci inanışına göre yapısal olmayan veri degersizdi, ancak

2012 rakamları ile dünyada günlük 2,5 kentilyon (1019) bayt veri üretilmektedir.

büyük veri bize bir şey gösterdi. O da günümüzdeki bilgi çöplüğü diye adlandırılan olgudan muazzam derecede önemli, kullanılabilir, yararlı bilgiler elde edileceğidir, yani çöplükten hazine çıkışmasına neden olan yegâne sistemdir.

Büyük veri web sunucularının günlükleri, internet istatistikleri, sosyal medya yayınları, bloglar, mikrobloglar, iklim algılayıcıları ve benzer sensörlerden gelen bilgiler, GSM operatörlerinden elde edilen arama kayıtları gibi büyük sayıda bilgi-den oluşuyor.

Büyük veri doğru analiz metodları ile yorumlandığında şirketlerin stratejik kararlarını doğru bir biçimde almalarına, risklerini daha iyi yönetmelerine ve inovasyon yapmalarına imkân sağlayabiliyor.

Şirketlerin çoğu, halen geleneksel (konvansiyonel) veri ambarı ve veri madenciliği yöntemleriyle elde ettikleri verilerden yola çıkarak karar almayı sürdürüyor. Ancak tüketici eğilimlerini dinamik şekilde öngörebilmek, büyük veriyi analiz edebilmekten ve bu analizlere göre hareket edebilmekten geçiyor.

Büyük veri geleneksel veri tabanı araçları ve algoritmaları ile işlenmesi zor olan bu devasa verinin oluşturulması, saklanması, akışı, analiz edilmesi gibi birçok konuyu içeren bir terim olarak karşımıza çıkmaktadır.

Aynı zamanda büyük veri kendi içinde bir kavram ve bir konu olmaktan çıkararak veri bilimi, veri entegrasyonu, analiz ve bilgi inovasyonu gibi mevcut kullanımlara bir ek değer olmaya doğru ilerlemektedir.

Veriler klasik veri tabanlarının kaldırılamayacağı büyülükte olduğu gibi, verinin büyümeye hızı da bir bilgisayar veya bir veri depolama ünitesini aşmaktadır. 2012 rakamları ile dünyada günlük 2,5 kentilyon (1019) bayt veri üretilmektedir. Bu çapta büyük veriyi işleme, transfer etme gibi işlerin tümüne “büyük veri (big data)” adı verilmektedir. İşte, bu nedenlerle birçok alanda büyük veri bir ihtiyaç olarak karşımıza çıkmaktadır.



Şekil 2.3 Büyük veri bileşenleri. (Kaynak: <http://andressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vs-volume>)

2.2 BÜYÜK VERİ BİLEŞENLERİ

Büyük veri platformları sayısal ağlar üzerinde farklı, ayırsız, aykırı olanları kategorize etmeye çalışırken diğer yandan da sanal ortamlara daha çok verinin eklenmesini sağlayarak maliyetleri düşürmeye çalışmaktadır. Bu kapsamda büyük verinin 5V olarak adlandırılan unsurları önem kazanmıştır (Gürsakal, 2013; Ege, 2013; Wikipedia, 2012).

Yani büyük veri platformunun oluşumunda beş bileşen vardır. Bu bileşenlere İngilizce ifadelerinin baş harfleri kullanılarak kısaca “5V” denilmektedir. Bunlar “variety (çeşitlilik),” “velocity (hız),” “volume (veri büyülüğu),” “verification (doğrulama)” ve “value (değer)”dir. Bu bileşenleri şu şekilde inceleyebiliriz (Şekil 2.3).

2.2.1 Çeşitlilik (Variety)

Çeşitlilik, oluşturulan farklı veri yapılarını ifade eder (De Mauro vd., 2015). Veri yapılandırılmış, yapılandırılmamış ya da yarı-yapılardırılmış olabilir ve bu üç veri tipinin hepsi sık ve yoğun bir şekilde birbiriyle yer değiştirebilmektedir. Yapılandırılmış veriler veri tabanlarında depolanan büyük verinin sadece %20'lik bir kısmını oluşturmaktadır. Kullanıcılar tarafından internet üzerinde, sosyal ağlarda, fiziksel algılama cihazlarındaki veriler dinamiktir ve yapılandırılmamıştır. Mesela internetteki gelişmeyle birlikte Facebook, Twitter ve Instagram gibi sosyal medyanın kolaylaşmasıyla resimler (JPEG, GIF, 3D), videolar, ses formlarında veri üretimi sağlandı ve hashtag kullanımını artırdı.

Üretilen ve veri tabanlarında depolanan verinin yüzde 80'i yapısal (yapılardırılmış) değildir ve üretilen her yeni teknoloji farklı formatlarda veri üretebilmektedir. Telefonlardan, tabletlerden, bütünsel devrelerden gelen türlü çeşitlilikte “veri tipi” ile uğraşılması gerekmektedir. Bir de bu verilerin farklı dillerde, Non-Unicode olabileceğini düşünürseniz, bütünsel olmaları, birbirlerine dönüşmeleri de gereklidir.

Yapılardırılmış verilere örnek olarak hava tahmini veya gözetleme amacıyla uydulardan üretilen görüntüler, gelgit/dalga hareketleri, yer dönüş modelleri vb. gibi bilimsel araştırma amaçları için oluşturulan veriler, güvenlik ve trafik için oluşturulan videolar (Hurwitz ve diğ., 2013) gösterilebilir. Örneğin araba sensörleri hız, motorun sıcaklığı ve yakıt seviyesi gibi çeşitli veriler üretir. Akıllı telefonlardaki Global Konumlandırma Sistemi (GPS) gibi sensörler, coğrafi konum hakkında veri üretir (Davis, 2015). Tarımsal alanda kullanılan sensörler topraktaki su seviyeleri, toprağın verimliliği, gerekli olan böcek ilaçı ve böcek ilaç seviyesi ile bitki büyümeye modeli verileri gibi çeşitli veriler üretir (Herring, 2001). Nesnelerin İnterneti’nin (Internet of Things-IoT) bir uygulaması giyilebilir cihazlardır ki sensörler kişinin uyku düzeni, kalp atışı, yürüme hızı, yiyecek alımı, tükettiği su miktarı ve yakılan kaloriyle ilgili çeşitli veriler üre-



Giyilebilir cihazlarda IoT

tir. E-ticarette ise müşteri tarafından web sitesinde harcanan zaman, tarama/göz atma modeli ve satın alma geçmiş gibi eylemler analiz amacıyla veri olarak saklanır.

Buradan yola çıkarak sosyal medya, sensör verileri, CRM dosyaları, dokümanlar, imajlar, videolar vb. olmak üzere akınıza gelebilecek bütün verileri, kaynakları hayal edin. Bunların tamamını ilişkisel bir veri tabanında saklamak, bildiğimiz bir dosya sistemi üzerinde bile saklamak pek mümkün değildir ve maliyetlidir. Büyük veriyi bir fırsat olarak görerek yararlanmak isteyen işletmeler hem ilişkili hem de ilişkili olmayan her tür veriyi analiz edebilmelidir: Bu veriler; metin, sensör verileri, ses, video, işlemsel ve daha fazla veriyi kapsamaktadır. Başarısı ise hem geleneksel hem de geleneksel olmayanların yer aldığı çeşitli veri türlerinden içgörüler elde etme becerisine dayanacaktır.

2.2.2 Hız (Velocity)

Buradaki hız verilerin üretildiği, depolandığı, analiz edildiği ve işlendiği hızı belirtmektedir (Al Nuaimi ve diğ., 2015). Hız büyük veri analizlerindeki yarı-gerçek zamanlı ve gerçek zamanlı işleme üzerindeki yüksek gereksinimleri ifade etmektedir. Geleneksel veri ambarı ve iş zekâsı için gerçek-zaman gereksinimleri daha düşüktür.

Büyük veri söz konusu olduğunda veriler çok yüksek hızda üretilir ve bu nedenle analiz edilmesi zordur. Yüksek hızlı veri üretiminin birkaç örneği vardır: New York menkul kıymetler borsasının her oturumunda yaklaşık bir terabiyat ticari bilgi elde edilir (Mobertz, 2013). Ortalama olarak Google tarafından her saniye 40.000'den fazla arama sorgusu işlenmektedir (Firican, 2017). Hava kalitesi izleme sensörleri her dakika verileri algılayarak gönderir ve akıllı sayaçlar her bir cihazın dakika dakika ayrıntılı enerji tüketimini elde etmek için kullanılır (Gary Barnett, 2014). Araştırma amaçlı kullanılan teleskopların olduğu astronomi, biyolojik araştırmada üretilen genom dizileri, hava tahmini uyduları, ev uygulamaları ve akıllı şehirlerle bağlantılı IoT gibi alanların tümünde çok yüksek hızda veri üretilir. Akıllı telefonlar ve sensörler gibi cihazlardaki gelişmeler hızlı veri üretimini artırır (Gandomi ve Haider, 2015).

Mobil uygulamalar aracılığıyla cep telefonlarından yüksek oranda üretilen veriler satıcıya müşterinin demografik konumu ve satın alma modeli gibi değerli bilgiler sağlar. Böylece satıcı bu bilgileri analiz ederek doğru müşteri için doğru ürünü hedefleyerek satış değerini artırabilir. Buna örnek verecek olursak bir kullanıcı Amazon veya Netflix gibi gelişmiş bir içerik web sitesini kullandığında açılan reklamlardır. Bu tür reklamlar kullanıcıların web davranışlarını yakalayarak, analiz ederek ve depolayarak söz konusu kullanıcı için seçilir ve görüntülenir (W. Vorhies, 2014). Bu yüksek hızda oluşan veriler hacimli verilere dönüşür. Facebook ve Twitter gibi sosyal medyadan gelen veriler her saniye büyük bir hızla üretilir (Dijcks, 2012).

Büyük verinin üretilme hızı çok yüksektir ve gittikçe artmaktadır. Daha hızlı üretilen veri ise, o veriye uygulanacak işlem sayısının ve çeşitliliğinin de aynı hızda artması sonucunu doğurmaktadır.

2.2.3 Veri Büyüklüğü (Volume)

Üretilen verinin çok hızlı artmasından dolayı, dünya çapında veri hacmi çok devasa boyutlara ulaşmaktadır. Her yerdeki mobil cihazlar ve kablosuz veri sensörleri her dakika veri üretmekte ve toplu veri alışverişleri milyarlarca internet servisi arasında her saniye oluşmaya devam etmektedir.

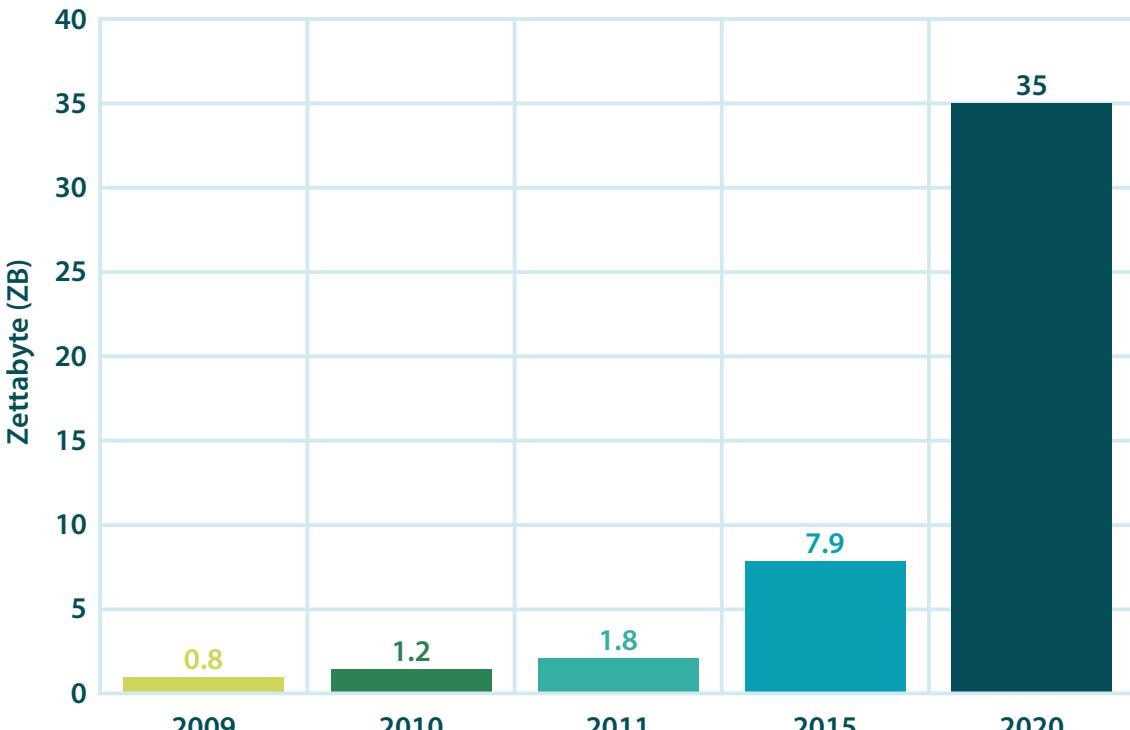


Petrol kuyularında kullanılan sensörler yılda 10 eksabayt veri üretir

Uluslararası Veri Kurumu IDC, 2011'de dünyadaki verilerin her iki yılda bir ikiye katlandığını iddia etmiştir. Yine IDC istatistiklerine göre, 2020'de ulaşılacak veri miktarının 2009'un 44 katı olacağı öngörülmektedir.

Örneğin 2000 yılında dünyada 800.000 petabayt (PB) veri depolanmıştır. Bugün ise yaratılan verilerin çoğu hiç analiz edilmiyor. Yalnızca Twitter her gün 7 terabayttan (TB) fazla veri üretiyor, Facebook 10 TB ve bazı kuruluşlar yılın her gününün her saatinde terabaytlarca veri üretiyor. 2020 yılında ise yaklaşık 40 trilyon gigabayt veri (40 zettabayt) üretileceği öngörmektedir (W. Vorhies, 2014). Wal-Mart her saat 2,5 petabayt müşteri işlem verisi üretir (Zaslavsky ve dig., 2012), 30 dakikalık bir süre içinde, tek bir jet motoru 10 TB veri üretir (Dijcks, 2012), uçaklardaki sensörler her yıl 2,5 milyar terabayt veri üretir ve otonom araçlar 2 petabayt veri üretir (Rijmenam, 2013); petrol kuyularında kullanılan sensörler yılda 10 eksabiyat veri üretir; The Square Kilometer Array teleskopu günlük 1 Eksabiyat veri üretir; UPS, her gün 16 petabayt veri depolayarak 16,3 milyon paketi izler ve yakalar ve 39,5 milyon takip isteğine yanıt verir (W. Vorhies, 2014).

Bu büyük miktarda veri meteoroloji, genomik, fizik, simülasyonlar, biyoloji, çevre bilimi, gerçek zamanlı bilgi toplayan sensörler (Su ve dig., 2016), web verileri (müşteri web davranışları), metin verileri (e-posta, haberler, belgeler) (Mukherjee ve Ravi, 2016), akıllı telefonlar, e-ticaret (Thiyagarajan ve Venkatachalam, 2014), sosyal ağlar (Facebook, LinkedIn, Instagram), web günlükleri, uydu görüntüler, yayın ses akışları, bankacılık işlemleri, web sayfası içerikleri, hükmümet belgelerinin taranması, GPS (Küresel Konumlandırma Sistemleri), finansal piyasa verileri (Pai, 2017), askeri gözetim, doğal afet ve kaynak yönetimi, özel sektör, perakende, radyo frekansı ile tanımlama teknolojisi, çağrı detay kaydı, ge-



Şekil 2.4 Küresel veri hacmi tahmini (Kaynak: Uluslararası Veri Kurumu, IDC, 2011)

nomik, biyojeokimyasal, gelişmiş sağlık sistemi ve borsalar (Uddin ve diğ., 2014) dahil olmak üzere çeşitli alanlarda üretilir.

Uluslararası Veri Kurumu IDC, 2011'de dünyadaki verilerin her iki yılda bir ikiye katlandığını iddia etmiştir. Yine IDC istatistiklerine göre, 2020'de ulaşılacak veri miktarının 2009'un 44 katı olacağı öngörmektedir.

Şu anda kullanılan, "büyük" diye adlandırdığımız kapasiteleri ve "büyük sistemleri" düşünüp bunların 44 kat büyülüklükte verilerle nasıl başa çıkacaklarını hayal etmek gerekmektedir. Bu devasa artışlar depolama için önemli sorunlar oluşturmaktadır. Veri arşivleme, işleme, bütünlendirme, saklama vb. teknolojilerinin bu büyülüklükte veri hacmi ile nasıl başa çıkacağıının kurgulanması elzem bir ihtiyaç olarak ortaya çıkmaktadır. 2010'lu yıllarda dünyadaki üretilen veri miktarının % 40 oranında artmasına rağmen toplam bilişim harcamaları sadece yılda % 5 oranında artmaktadır (bkz. Şekil 2.4).

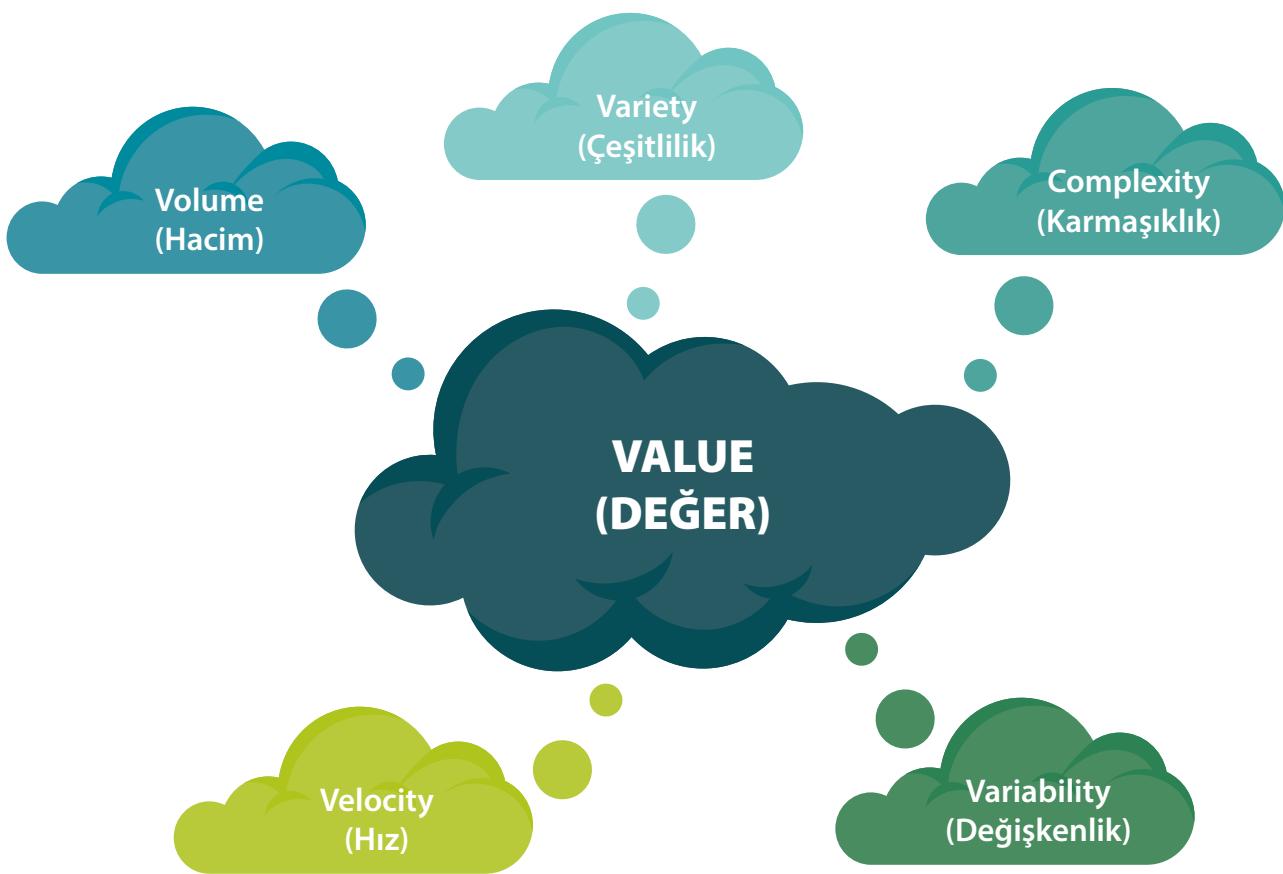
2.2.4 Doğrulama (Verification)

Gandomi & Haider (2015), IBM'in "doğrulama"yı büyük verinin dördüncü V'si olarak ortaya attığına dikkat çeker. Uddin ve diğerleri doğrulamayı "verilerin doğruluğu" olarak tanımlar (Uddin ve dig., 2014, s. 3). Veriler güvenilir değilse, bu tür verileri analiz etmek değerli sonuçlar vermeyecektir. Verilerin kalitesi veri analizinin doğruluğunu belirler. Örneğin sensörler çeşitli alanlarda (IoT, hava kalitesi izleme, hava tahmini, astronomi ve çeşitli diğer amaçlar için) kullanılmaktadır (Anuradha & Ishwarappa, 2015). Bu sensörlerden üretilen veriler doğru değilse, verilerin analiz edilmesi karar vermede güvenilir bir sonuç oluşturmaya- caktır.

Doğrulama verilerdeki önyargıları ve anormallikleri ifade eder (Owais ve Hussein, 2016; Yassin, 2014). Büyük veriler toplanır, depolanır ve analiz edilir, çünkü kuruluşlara yargıda bulunma, tavsiye verme ve karar verme konusunda değerli bilgiler sağlar (Chen ve dig., 2014). Geleneksel iş zekâsı çözümlerinde veriler genellikle bir veya daha fazla iş sisteminden alınır. Çıkarılan veriler yüksek düzeyde yapılandırılmıştır ve güvenilir bir kaynak sistemden gelir. Elde edilen veriler temizlenir ve veri ambarı veya data mart gibi bir havuza entegre edilir. Büyük veri durumunda veriler çeşitli kaynaklardan alınır ve hem yapılandırılmış hem de yapılandırmamış olabilir.

İnsan tarafından oluşturulan veriler (e-postalar, videolar, fotoğraflar, metin mesajları vb.) ve makine tarafından oluşturulan verilerin (günlük dosyalar, sensör verileri vb.) tümü analiz için dahil edilir. Kuruluşların müşteri e-postaları, tweetler vb. yapılandırılmış verilerle iş sistemlerinden gelen müşterilerinin bilgileri gibi yapılandırılmış verileri bir araya getirmelerinin gerektiği senaryoyu düşünün. Dış veri kaynakları doğrulanmayabileceğinden veya verilerin kendisi bir tweet veya e-posta olabileceğinden eksik veya yanlıltıcı olabilir. Kuruluşlar karar alma süreçlerinde analiz için kullanılan veri kaynaklarının belirsizlik düzeyini hesaba katmalıdır.

Örneğin bir işletmenin yeni bir ürünü piyasaya sürtüğünü ve ürünün müşteriler tarafından nasıl alındığına dair gerçek zamanlı bir analiz yapmak istediğini düşünün. Duyarlılık analizi belirli bir ürün veya hizmet hakkında müşteri görüş-



Şekil 2.5 Büyük verinin en önemli beş bileşeni. (Kaynak: <http://mis.sadievrenseker.com/2014/06/buyuk-veri-big-data/>)

lerini toplamak ve analiz etmek için kullanılabilir. Toplanan veriler çoğunlukla çeşitli sosyal medya platformlarından sağlanan tweetler, yorumlar ve gönderiler gibi yapılandırılmamış verilerdir. Bu tür gerçek zamanlı analiz başarı için kilit bir faktör olsa da veri kaynaklarının çeşitliliği ve sosyal medya verilerinin üretildiği hız, verilerde belirli bir belirsizlik ve eksikliğe yol açar. Veri analizine dayalı herhangi bir karar vermeden önce, verilerin belirsizlik, eksiklik ve güvenilirlik gibi doğrulama sorunları ele alınmalıdır.

Dolayısıyla bu bilgi yoğunluğu içinde verinin akışı sırasında “güvenli” olması da bir diğer bileşendir. Akış sırasında doğru katmadan, olması gerektiği güvenlik seviyesinde izlenmesi, doğru kişiler tarafından görünebilir veya gizli kalması gerekmektedir (bkz. Şekil 2.5).

2.2.5 Değer (Value)

Bir işletmedeki geleneksel veri statik ve arşivsel analizler için kullanılabilirken etkileyen büyük veri daha çok gelecek trendleri ve önemli kararların alınmasını etkileyen önemli bir faktördür. Bununla birlikte büyük veri daha düşük değer yoğunluğuna sahiptir. Örneğin bir video kayıt cihazından sürekli gözetim verileri üretilebilir, fakat bu görüntülerin sadece birkaç saniyesi aslında faydalı olabilecektir.

Gandomi ve Haider tarafından belirtildiği gibi, Oracle büyük verinin tanımlayııcı bir özelliği olarak “değer”i tanıtmıştır (Gandomi & Haider, 2015). Büyük veri bir amaç için saklanır ve analiz edilir ve bu amaç kurumla toplanan verilerin türüne bağlıdır. Örneğin bir yoğun bakım ünitesindeki bir kişinin sürekli izlenmesinden elde edilen veriler hayat kurtarıcı amaçlarla izlenir ve analiz edilirken, yürüme hızı ve uyku düzeni gibi giyilebilir bir cihazdan üretilen veriler sağlıklı bir yaşam tarzı için izlenerek analiz edilebilir. Veriler amaçlarına göre saklanır ve analiz edilir. Anuradha ve Ishwarappa (2015) büyük verinin değerine, analizden sonra toplanan verilerin hedeflenen amaca ne ölçüde katkıda bulunabileceği yönelik atıfta bulunur. Kaisler tarafından belirtildiği gibi, veri değeri karar vermede verilerin yararlılığının ölçüsüdür (Kaisler ve diğ., 2013). Ayrıca değerin, “veriler iş ihtiyaçları için değerli bilgiler içeriyor mu?” sorusunun yanıtını vermesi gerektiği savunulmaktadır (Wu ve diğ., 2016, s. 7). Anuradha ve Ishwarappa (2015), değeri büyük verinin en önemli yönü olarak kabul eder, çünkü toplanan büyük veri bir değer taşıymiyorsa, o zaman faydasız olarak kabul edilir. Thiagarajan ve Venkatachalapathy (2014), değeri büyük veri işleminin istenen sonucu olarak tanımlamaktadır.

Şekil 2.6

Büyük verinin değer zinciri (Kaynak: https://www.researchgate.net/figure/The-big-data-value-chain-2_fig3_308391010)

Üretim

- Sonu olmayan ölçek
 - Veri çeşitliliği
 - Alan/zaman korelasyonu
 - Gürültü

Hazırlık

- Birleştirmek
- Temizlemek
- Elemek

Büyük verinin değerine ilişkin örnekler bakacak olursak bankacılık sektöründe dolandırıcılık tespiti ve yasa dışı ticaret için büyük veri analizi kullanılır; ABD Gıda ve İlaç Dairesi’nde (FDA) büyük veri gıdaın neden olduğu hastalıkları ve rahatsızlıklarını tespit etmek ve anlamak için kullanılır; perakende, ticaret ve toptan satışta büyük veri envanter analizi için ve müşterilerin alışveriş modellerini anlamak için kullanılır; ulaşım sektöründe büyük veri rota planlama ve trafik kontrolü ile değer sağlar (Sivakumar, 2015). Wamba ve diğerleri degeri, “Büyük verinin elde etme ve dönüştürme yoluyla ekonomik açıdan değerli içgörüler ve/veya faydalalar ürettiği kapsam” olarak tanımlar (Wamba ve diğ., 2015, s. 240).

Bu yüzden en önemli bileşen ise büyük verinin bir değer oluşturmasıdır. Bütün yukarıdaki bileşenlerle tarif edilen büyük verinin veri üretim ve işleme katmanlarınızdan sonra kurum için bir artı değer meydana getiriyor olması gereklidir. Karar veriş süreçlerine anlık olarak etki etmesi, doğru kararı vermede hemen el altında olması gerekmektedir.

Örneğin sağlık konusunda stratejik kararlar alan bir devlet kurumu anlık olarak bölge, il, ilçe vb. detaylarda hastalık, ilaç, doktor dağılımlarını görebilmelidir. Hava Kuvvetleri bütün uçucu envanterindeki taşıtlarının anlık yerlerini ve durumlarını görebilmeli, geriye dönük bakım tarihçelerini izleyebilmelidir. Bir banka kredi vereceği kişinin, sadece demografik bilgilerini değil, yemek yeme, tatil yapma alışkanlıklarını dahi izleyebilmeli, gerekirse sosyal ağlarda ne yaptığı dahi görebilmelidir.

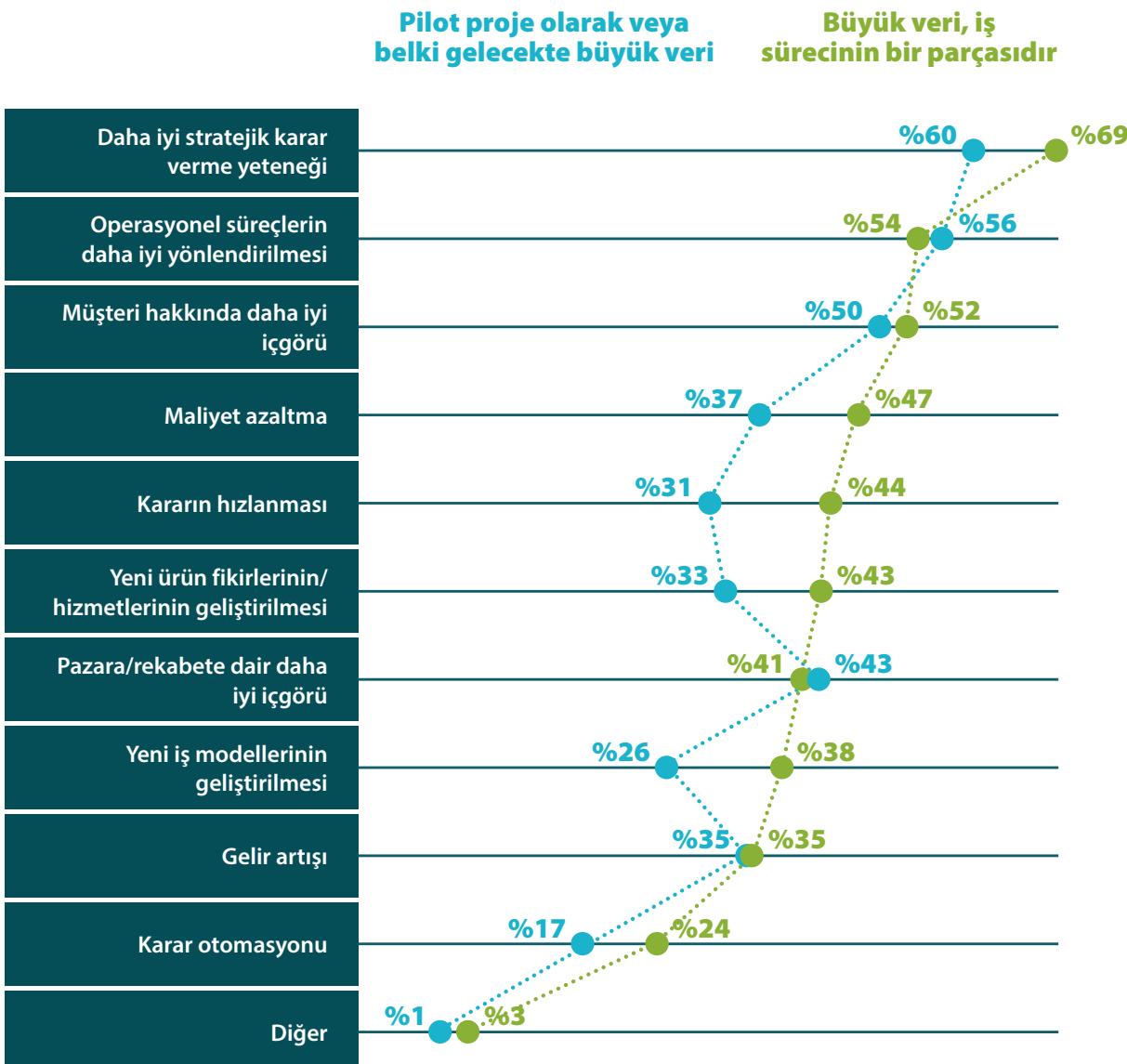


2.3 VERİ TİPLERİ VE VERİNİN GERÇEK DEĞERİ

Veri girişimleri şu anda iş dünyasında oldukça yükselen bir değer olduğundan, Omer Trajman (San Francisco Bay Area) verinin “somut ve gerçek değerinin” ne olduğunu araştırmaya çalıştı. İlk olarak şunu ifade etmektedir: Veri patentler ya da fikri haklar gibi aynı şekilde soyut varlıklardır. Fakat çoğu firma veri depolaması ile karşılaşıldığında, onların muhasebe yöntemleri tarafından kanıtlandığı gibi, bir değer olması için, onu basit bir şekilde değerlendirdir (diğer bir deyişle veri değeri onu depolama maliyeti ile iç içe geçmiştir). Trajman “bu düşünce şekli verinin gerçek değerini gizler” der, çünkü “depolama herhangi veri içersin ya da içermesin, sabit maliyete sahiptir,” verinin kendisi “ek maliyete sahip veri içerikleri olan gerçek değeri ve depolamayı” içerir (2013, s. 73).

Gerçek verinin değerini daha iyi anlamak için, Trajman ilk önce veriyi sınıflandırmayı araştırmaktadır. Bir alışverişin değerini kaydetmek için kullanılan veri (örneğin satın alma) işlem verisidir. Bu olay etrafında üretilen olgu “arızı veri” olarak adlandırılmıştır. Veri sıkılıkla yapılandırılmış ve yapılandırmamış kategoriler içerisinde sınıflandırılmasına rağmen, tüm veri aslında bir yapıya sahiptir.

Hurwitz ve diğerleri (2013, s. 30) “yapılandırılmamış” veya “yapılanmamış” adının gerçeklikten geldiğini açıklığa kavuşturmuştur, o belirli bir formatı takip etmez. Makine üretimi yapılanmamış veri örneği olarak radar ya da sonar (deniz radarı) verisi, gözetleme veya trafik video ve fotoğrafları, bilimsel veri araçları ve uydu görüntüleri tarafından toplanmaktadır. İnsan üretimi yapılanmamış veri örnekleri olarak Youtube ve Flickr benzeri web sitelerinden içerikler, metin mesajları ve konum bilgisi, şirket içi e-posta ve diğer haberleşmeler, sosyal medya verileri, mobil veriyi alıntılarlar. Aksine “tanımlanmış bir boy ve şeke” sahip yapılandırılmış veri dünyadaki toplam verinin yaklaşık % 20’sine tekabül etmektedir. Kaufman ve diğerleri algılayıcı veriyi kapsar (örneğin RFID - radyo frekansı ile tanımlama-, GPS -küresel konumlama sistemi-, tıbbi aygıtlar ve akıllı sayaç okumalar), web günlükleri verisi, satış noktası verisi (Trajman’ın işlemsel olarak adlandırdığı) ve makine üretimi bu tip verinin örnekleri olarak finansal veri, giriş verisi, insan üretimi yapılandırılmış veri örnekleri olarak web sitelerinden tıklama akış verisi ve oyunla ilişkili veri (oyuncuların yaptığı hareketler vb.).



Şekil 2.7 Elde edilen faydalara karşı beklenen faydalardır ($n = 94/335$). (Kaynak: <https://bi-survey.com/big-data-value>)

Toplam verinin sadece % 20'sinin yapılandırılmış veri olması dikkat edilmesi için önemlidir, göz ardı edilmemelidir ve pazara giren yapılandırılmamış veri miktarının büyümesi kadar önemli olabilir.

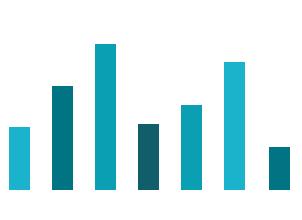
Trajman'ın analizine geri dönersek, rastlantısal veri depolanırken nadiren bir satın almayla ilişkili isim, adres ve kredi kartı şeklindeki kişisel profil verilerin haricindekiler düzenlenmiş ve analiz edilmiş işlemsel veri şeklidir. Bununla bir-

likte genelde gözden kaçan ikinci tip rastlantısal veri “daha dinamiktir” ve sadece müşterilerin nihai satın almalarını ve satın alma yapmadan önceki tıklanan ve incelenen ürünleri değil, dijital servislerin örneklerini içermektedir (Trajman 2013, s. 70). Yıllardır çok başarılı bir internet perakendecisi olan “Amazon.com” sadece müşterilerin son satın almalarını ve almadan baktıkları ürünleri değil, web sitesine gelen ziyaretçiler ve müşteriler için “kişiselleştirilmiş” önerileri derleyen rastlantısal verinin bu “veriyi sömüren” modelini kullanmaktadır.

Trajman rakiplerinin ve sanayi denklerinin yapmadığı şekilde “veri kullanarak bir rekabet avantajı elde eden şirketlerin diferansiyel değerini” dikkate alarak değer olan “verinin bir makroskopik resmini alma” girişiminde bulunmaktadır. İlk olarak “verinin değerini” hesaplar: Verinin nasıl kullanıldığı ve yaşam döngüsündeki yeri üzerinde temellenen farklı değerlere sahip olabilirken, veri yazıldığı anda kesin bir bedele sahiptir, silindiği ve tahrip olduğu zamana kadar bile, depolama gibi bir maliyet doğurur. “Verinin içsel değerinin tanımlanmasının önemini anlaşılması için” Trajman der ki, “bizim sadece yakalama ve depolamanın maliyetini dikkate almamız gereklidir,” 2012’de veri kaybının yaklaşık maliyeti, kayıp ya da çalıntı her müşteri kaydı için 200 doları aşmıştır. Elbette eğer adres ya da kredi kartı bilgisi gibi müşterilerin hassas bilgileri kırılarak (hack) veya bir şekilde birileri tarafından çalınmış olursa, veri sizıntısı başlıca bir güvenlik riskidir, bu da ciddi tanıtım, yasal ve hatta şirket ya da kurumlar için hisse senedi problemlerine neden olabilir.

Sony Playstation, ağları üzerinde bununla mücadele eden bir şirket örneğidir. PlayStation Network erişiminin kesilmesi olayında, gerçekleştirilen hacker saldırısı sebebiyle Sony’nin PlayStation Network ve Qriocity servislerine erişimin kesilmesi ve çevrimiçi hizmetlerden faydalanan yaklaşık 77 milyon PlayStation 3 ve PlayStation Portable kullanıcısının hesap bilgilerinin çalınması hadisesi yaşanmıştır. 17 ile 19 Nisan 2011 tarihleri arasında gerçekleştirilen saldırının ardından Sony, 20 Nisan’da PlayStation Network erişimini kesmek zorunda kaldı. 4 Mayıs’da ise şirket tarafından yapılan açıklamaya yaklaşık 77 milyon kullanıcının kişisel tanımlayıcı bilgilerinin çalındığı kabul edildi. Yaşanan bu olay, “en büyük veri hırsızlığı olaylarından biri” olarak tarihe geçti. 26 Nisan’da Sony, hizmetlerin bir hafta içinde eski haline doneceğini belirtse de 14 Mayıs’ta yayınlanan PlayStation 3 güncellemesiyle birlikte kullanıcıların şifrelerini yenilenmesi istendi. Bu sırada çevrimiçi

Kuruluşların % 80'inden fazlası
şunları söylüyor:



İnsanların mücadele ettiği yer:

01010
10011
01101



Yeni Başlayanlar
%28
Veri niteliği

İleri
%30
Büyük veri hacmi

Yetkin
%29
Nitelikli insan gücü



Şekil 2.8 Kuruluşlara göre büyük verinin değeri. (Kaynak: <https://www.resourceondemand.com/the-value-of-big-data/>)

hizmetlerin açılması için yapılan çalışmalar halen sürdürülmemekteydi. 23 gün süren kesintinin ardından ise hizmetler yeniden kullanıma girdi. Servislerin açılmasının ardından Sony, saldırıldan etkilenen kullanıcılar için “tekrar hoş geldiniz” adı altında çeşitli hediyeler yayınladı. Oyunlar da dahil olmak üzere birtakım içeriklerin yer aldığı bu hediyeler, kullanıcıların belli bir süreliğine ücretsiz olarak erişimine sunuldu (https://tr.wikipedia.org/wiki/PlayStation_Network_erişiminin_kesilmesi, 2015).

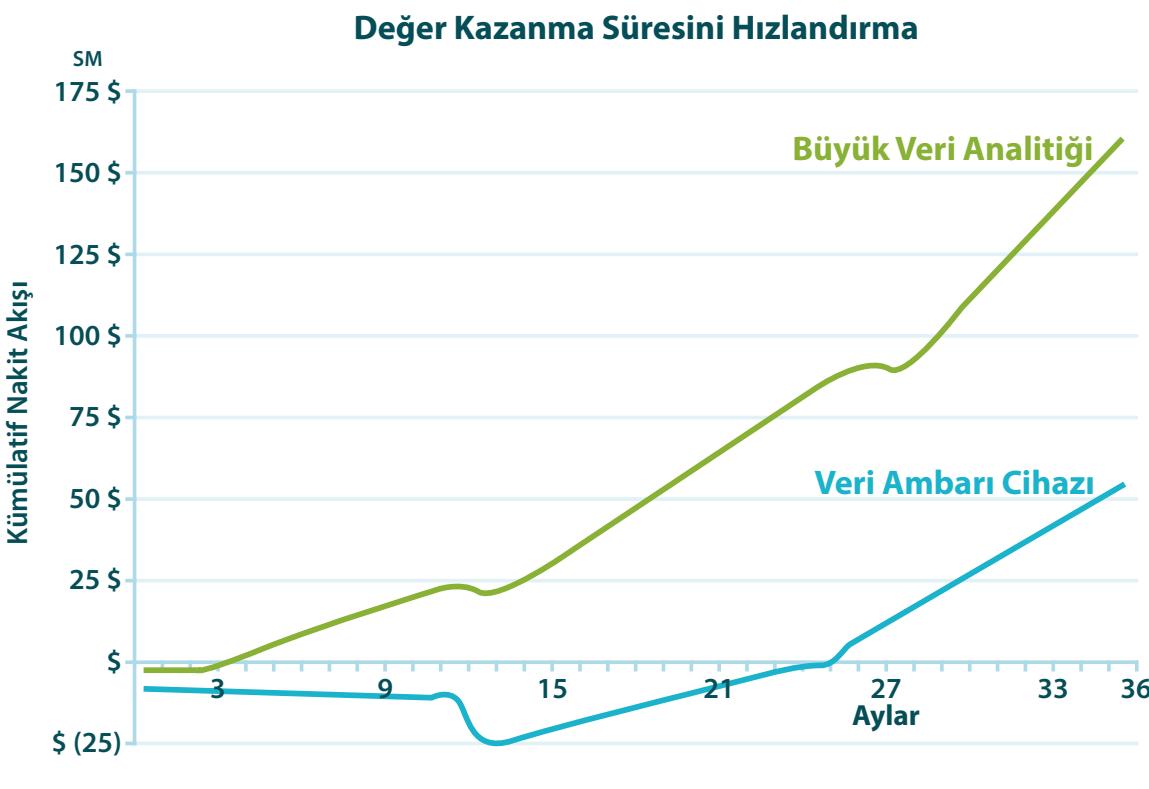
Toplam verinin sadece %20'sinin yapılandırılmış veri olması dikkat edilmesi için önemlidir, göz ardı edilmemelidir ve pazara giren yapılandırılmamış veri miktarının büyümesi kadar önemli olabilir.

Bundan dolayı verinin potansiyel değeri özünde onun maliyetine de bağlıdır; daha değerli bir veri bir şirket içindir, dahası eğer bu veri kaybolur ya da tahrip olur ise o değerini kaybedebilir.

Bir örnek olarak Trajman şu senaryoyu sunar: Acxiom Şirketi 2012 Mayıs ayında yıllık 1,13 milyar dolar gelir oluşturdu. Bunun nedeni, dünyada diğer şirketlerde olmayan veriye erişime sahip olmasıydı. Bundan dolayı Trajman şunu sorar, “Acxiom'un verisine sahip olmak için bir kurum ne ödeyebilir? Bu özel verilerin hepsine sahip olan bir organizasyonun” değeri ne olabilir? Masraflardan sonra, Acxiom'un 2012'de net geliri 85 milyon dolardır: Trajman şunu sormaktadır: Bunun ne kadarı, bu verinin özel olmasından dolayıydı?

Bu tekelcilik faktörüne ek olarak Trajman şöyle der; rastlantısal veri işlemsel veriden daha hızlı oranda büyümeye rağmen, rastlantısal veri üzerine parasal veri koymak daha zor bir işlemidir ve bundan dolayı bilim ve medyanın ilgisini daha fazla alıyor. Fakat Google ve Facebook gibi büyük internet şirketleri, kendi geri dönüşleri için hemen hemen net pozitif bir sonuç sağlayan depolanan rastlantısal veriyi hesaplamıştır ve kabul etmiştir. Trajmana göre, bu şundan dolayı olabilir: Bir satın almayı veya içerisindeki başvuru vergilerini telafi etmek için kullanıldığı zaman işlemsel veri yolunda daha çok değere sahiptir, rastlantısal veri bir rekabet avantajı kazanmak ve bir işletmenin operasyonel etkinliğini artırmak için müşterileri korumak ve çekmek için kullanılabilir. Rastlantısal verinin bu yönü birçok büyük veri girişimlerinin merkezindedir.

Paul Kent, Radhika Kulkarni ve Udo Sglavo'nın büyük verinin değerini araştırmasında görünen şudur ki büyük firmaların sıkılıkla başa çıkmak zorunda olduğu durum, farklı ürün grupları ve çeşitli müşteri portföy grupları ve zıt kutuplarla başa



- Gerçek iş vaka analizi ve gerçek dünya deneyimlerinden elde edilen tahmini nakit akışı verileri
- Başa baş 4 aya karşı 25
- 3 yıllık Proje Değeri 158 milyon dolara karşılık 54 milyon dolar

Şekil 2.9 Büyük veri analizinin değeri (Kaynak: <https://www.semanticscholar.org/paper/Big-Data-Analytics-with-Hadoop-to-analyze-Targeted-Gupta/f31ec0a3ffea29ed90276fc3fae04ae3d-882ceef/figure/0>)

çıkmak etkin bir biçimde birleştirmeye yardımcı olabilir. Büyük veri bir işletme ekibinin üyeleri tarafından önerilen yeni programlar veya çeşitli kârlılık senaryolarını hızlı ve etkin bir şekilde belirleyen algoritmaları çalıştmak için kullanılan toplanmış veriyi kullanan yüksek performanslı analizleri (HPA: High-Performance Analytics) çalıştmak için kullanılabilir. Seigel gibi, onlar da şuna inanırlar; büyük veri “kesin olmayan, göreceli bir terimdir” ve ayrıca şunu da ima ederler; veri miktarının büyümesiyle başa çıkacak bir organizasyonun kapasitesi bilişim departmanının konfor bölgesini geçti. Cukier ve Schonberger'in işaret ettiği gibi, ayrıca belirtmektedirler ki birçok şirket büyük veri miktarlarına sahip iken, verinin potansiyel değerine bakan şirketlerden bazıları bu veriden etkin bir şekilde nasıl faydalanağını biliyor.

“Büyük veri” kullanan yüksek performans analizlerinin değeri, sonucta ölçümllerin taşıyabilme hızına dayanır ve bundan dolayı bu hız karar vermede yöneticilere yardımcı olabilir. Bu nedenle üst yönetim üyeleri arasında karar periyodları ya da çikmazların üstesinden gelmede özel kullanımına sahiptir.

“Büyük veri” kullanan yüksek performans analizlerinin değeri, sonucta ölçümllerin taşıyabilme hızına dayanır ve bundan dolayı bu hız karar vermede yöneticilere yardımcı olabilir. Bu nedenle üst yönetim üyeleri arasında karar periyodları ya da çikmazların üstesinden gelmede özel kullanımına sahiptir. Kent ve diğerlerine göre, hatta bu bir takım arasında yaratıcılığı teşvik edebilir, çünkü “eğer bir daki ka içerisinde bir yanıt almak için yüksek performanslı analizlerini kullanabilirsen, hesaplamalar vasıtasıyla daha çok boş zamanla, birçok soru sorar, birçok alternatif araştırır, birçok senaryo oluşturur ve daha iyi sonuçların peşinden koşarsın.” Kent ve diğerlerine göre, yüksek performanslı analizler birçok avantaja sahip “dağıtımlı hesaplamaya” yakından bağlıdır. Bunlar esneklik ve maliyet avantajları (çoklu görev ve çoklu izleri koruma, farklı kullanıcı grupları, müşteriler ve uygulamaların daha kolay olması) içerir ve aynı zamanda hâlâ firmalar hedef müşterilerine yardım ederken, bunlar altyapı maliyetlerini önemli miktarda düşürebilir.

Kent ve arkadaşlarına göre, yardıma ek olarak daha fazla özelleştirilmiş müşteri pazarlaması ve reklam kampanyaları oluşturan büyük veri, firmaların var olan müşteri portföyüne tam anlamıyla anamasına da yardımcı olabilir. Örneğin veri tabanındaki analizler olası vadesi geçmiş bir müşteri hesabının en sonunda nasıl ödeyeceğini kararlaştırmasına yardımcı olabilir ve birileri hiç ödemeyebilir ve büyük olasılıkla suçludur. Yine zaman burada bir şirket rekabetin üstesinden gelme konusunda farklılık yapabileceğii kritik bir elemandır. Kent, Kulkarani ve Sglavo şunu belirtir: Belli tekke analizler yüksek performanslı analizlerin zirvesidir. Daha hızlı kararlar ver-

mek, uygun tüm verilerden daha kesin cevaplar kazanmak, daha fazla güven kurmak, ölçülebilir güvenilir analiz altyapısı, yöneticilerin daha yaratıcı ve etkin kararlar için beyin firtınası yapabilmelerine olanak sağlayan değerli boş zamanı şirketlere verir. Sık olarak alıntı yapılan “öngörülü analizlere” ek olarak ayrıca hatırlanmasının önemli olduğu şu konuya da işaret ederler, büyük veri girişimleri aynı zamanda “tanımlayıcı analizler” ve “kuralçı analizler” oluşturmaya yardımcı olabilir.

KAYNAKLAR

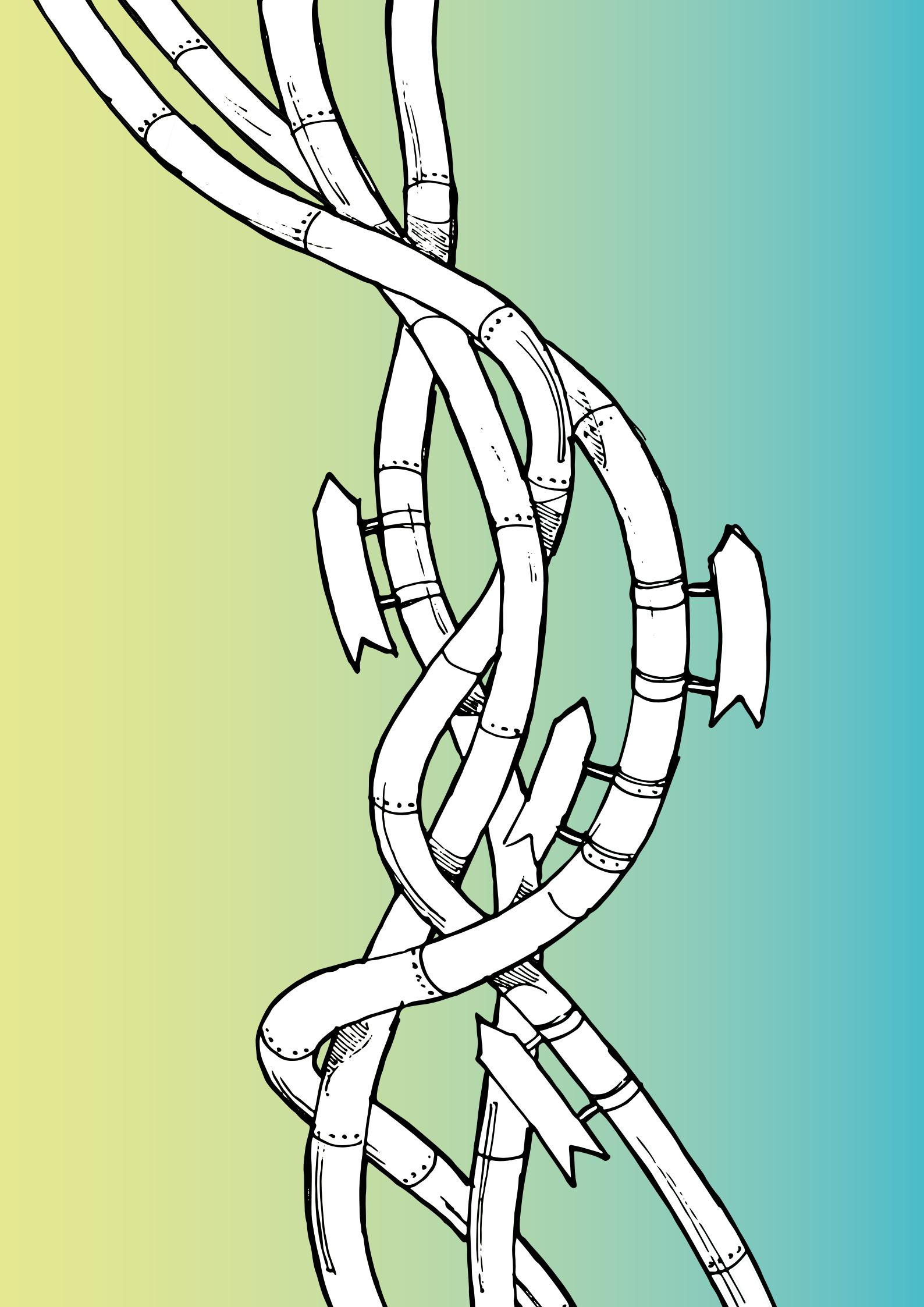
- Aghaei, S., Nemaakhstbh, M. A., & Farsani, H. K. (2012). Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. *International Journal of Web & Semantic Technology*, 3(1), 1-10.
- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 25.
- Andrea, De Mauro, Greco, Marco, Grimaldi, Michele, “What is Big Data? A Consensual Definition and a Review of Key Research Topics,” 4th International Conference on Integrated Information, Eylül 2014.
- Anuradha, J., & Ishwarappa. (2015). A brief introduction on Big data 5Vs characteristics and Hadoop Technology. *Procedia Computer Science*, 48, 319-324.
- Biehn, N. (2013). The missing Vs in big data: viability and value. *Wired. Innovation Insights Community*, 5.
- Bourgeois, J., Pugmire, L., Stevenson, K., Swanson, N., & Swanson, B. (2006). The Delphi method: A qualitative means to a better future. URL: [Http://www.Freequality.org/documents/knowledge/Delphimethod.pdf](http://www.Freequality.org/documents/knowledge/Delphimethod.pdf) (Citirano 2. 11. 2011).
- Cartledge, C. (2016). How Many Vs are there in Big Data?, 1-4. Retrieved from <http://clc-ent.com/TBDE/Docs/vs.pdf>
- Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.
- Chen, M., Mao, S., Zhang, Y., & Leung, V. C. M. (2014). Big Data, 59-79. <https://doi.org/10.1007/978-3-319-06245-7>
- Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4), 2233-2243.

- Dalkey, N. & O. H. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3).
- Dijcks, J. (2012). Oracle: Big data for the enterprise. Oracle White Paper.
- Firican, G. (2017). The 10 Vs of Big Data. <https://upside.tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- Gary Barnett. (2014). Managing Data in the Internet of Things. <http://hotdesks.org/docs/Managing-Data-in-the-Internet-of-Things.pdf>
- Gürsakal, Necmi, İstatığın ABC'si, Dora Yayıncılık, 2013.
- Hurwitz, J., Alan N., Fern H., ve Marcia K., 2013, Big Data for Dummies, John Wiley & Sons, ISBN: 978-1-118-50422-2.
- <https://whatsthebigdata.com/2016/04/18/5-origins-of-data-science/>
- <http://andressilvaa.tumblr.com/post/87206443764/big-data-refers-to-5vs-volume>
- <http://mis.sadievrenseker.com/2014/06/buyuk-veri-big-data/>
- https://www.researchgate.net/figure/The-big-data-value-chain-2_fig3_308391010
- <https://bi-survey.com/big-data-value>
- <https://www.semanticscholar.org/paper/Big-Data-Analytics-with-Hadoop-to-analyze-Targeted-Gupta/f31ec0a3ffea29ed90276fc3fae04ae3d-882ceef/figure/0>
- <https://www.resourceondemand.com/thevalue-of-big-data/>
- https://tr.wikipedia.org/wiki/PlayStation_Network_erişiminin_kesilmesi, 2015
- Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 995-1004). IEEE.
- Kent, Paul, Radhika Kulkarni, Udo Sglavo (2013). Big Data and Business Analytics, Auerbach Publications.
- Mobertz, L. (2013). The Four V's of Big Data [INFOGRAPHIC]. Retrieved from <https://blog.dashburst.com/infographic/big-data-volume-variety-velocity/>
- Mukherjee, S., & Ravi, S. (2016). Big Data - Concepts, Applications, Challenges and Future Scope. *International Journal of Advanced Research in Computer and Communication Engineering*, 5(2).
- Owais, S. S., & Hussein, N. S. (2016). Extract Five Categories CPIVW from the 9V's

- Characteristics of the Big Data. International Journal of Advanced Computer Science and Applications, 7(3), 254-258.
- Sivakumar. (2015). How Top 10 Industries Use Big Data Applications. <http://www.datascienceassn.org/content/how-top-10-industries-use-big-data-applications>
- Su, X., Pattnaik, K., & Prasad Mishra, B. S. (2016). Introduction to Big Data Analysis, 1-20. http://link.springer.com/10.1007/978-3-319-27520-8_1
- Thiyagarajan, V. S., & Venkatachalamathy, K. (2014). Isolating Values From Big Data With The Help of Four V's. International Journal of Research in Engineering and Technology, 4(1), 132-135.
- Trajman, Omer (2013). Big Data and Business Analytics. Auerbach Publications.
- Uddin, M. F., Gupta, N., & Khan, M. A. (2014). Seven V's of Big Data understanding Big Data to extract value. In Proceedings of 2014 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1).
- Uluslararası Veri Kurumu, IDC, 2011
- Viktor Mayer-Schönberger, Kenneth Cukier (2013). Big Data, John Murray.
- Vorhies, W. (2014). How many V's in big data? The characteristics that define big data. Data Science Central.
- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics, 165, 234-246.
- Wu, C., Buyya, R., & Ramamohanarao, K. (2016). Big Data analytics= machine learning + cloud computing. arXiv Preprint arXiv:1601.03115.
- Yassin, A. T. (2014). Analyzing 6Vs of big data using system dynamics. The 2nd Scientific Conference of the College of Science.

3. BÖLÜM

BÜYÜK VERİNİN YÖNETİMİ

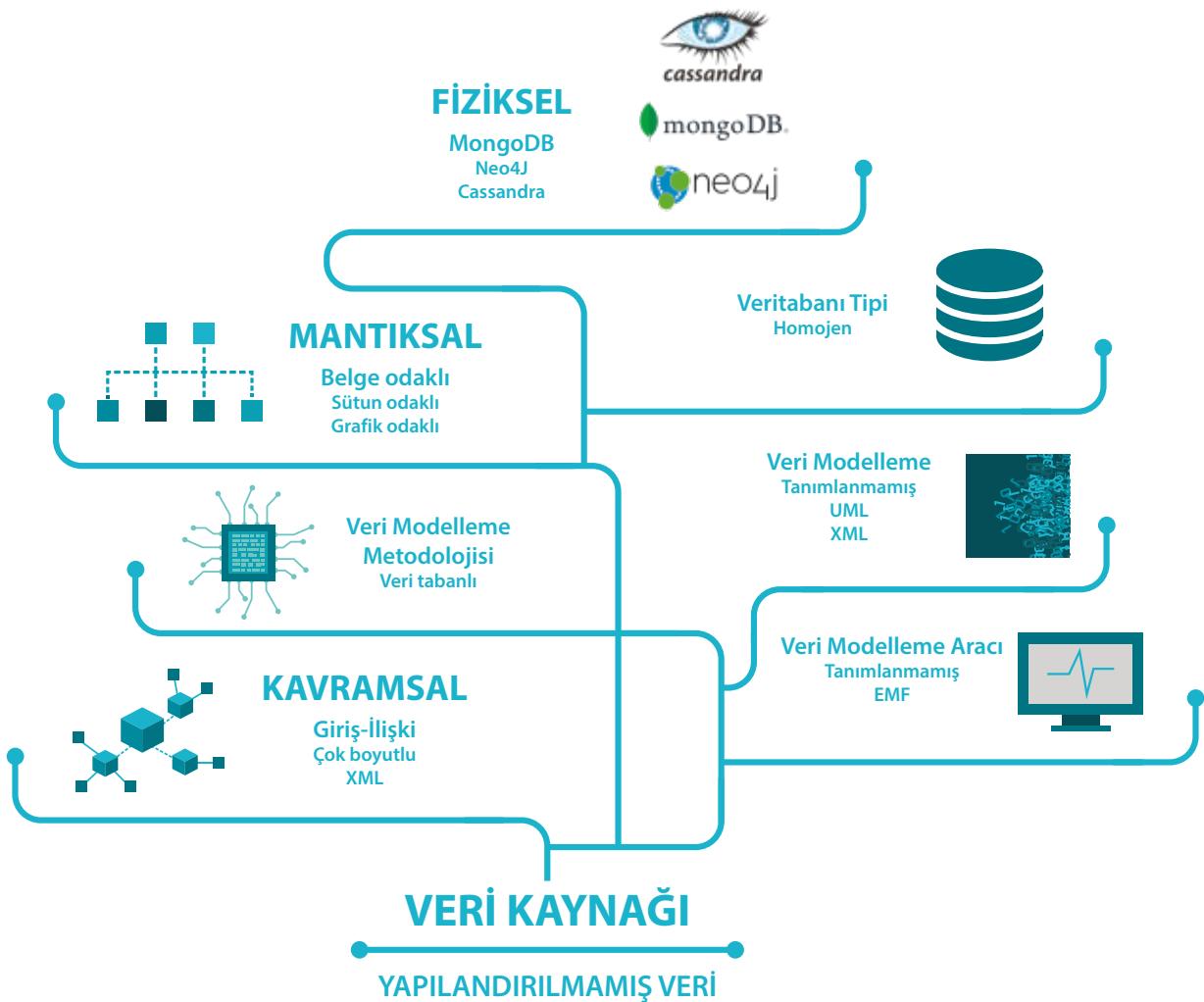


3.1 BÜYÜK VERİNİN GELİŞİM DİNAMİKLERİ

Büyük veri son yıllarda bilgisayar biliminin önemli bir dalı haline gelerek bu alanındaki araştırmalar ve yenilikler hiç durmadı. Büyük veri hızla gelişerek hayatımızı gittikçe daha derinden etkilemeye başladı. Öyle ki büyük veri mevcut ve geleceğin araştırma sınırlarından biri oldu. Gartner “Top 10 Strategic Technology Trends For 2013” ve “Top 10 Critical Tech Trends For The Next Five Years” listeleri yayınladı ve büyük veri her ikisine de girdi.

Judith Hurwitz, Alan Nugent, Dr. Fern Halper ve Marcia Kaufman’ın ifadelerine göre, büyük veri analizi, “Her veri yönetim dalgası gereklilikten ortaya çıkmıştır ve özgün bir problemin çözümüdür.” İlk dalgada veri plansız, “düz tablolar”da depolanmıştı ve şirketler veriden değer çıkarmak için detaylandırılmış program içeren “kaba kuvvet” kullanmak zorundaydı. Bu Öge-İlişki (Entity-Relationship, ER) modeli tarafından desteklenen veri üzerinde dayatılan yapı olan ilişkisel sisteme yol açmıştır. Bu, soyutlama düzeyini artırarak verinin kullanılabilirliğini artırdı. İlişkisel veri tabanları için pazar hâlâ günümüze kadar güçlü kalmıştır.

Verinin hacmi arttıkça, veri ambarları gelişti, onların hız ve erişilebilirlik programlarına yanıt vermek için veri pazarları geliştirildi. Bunlar elle işlenen veriler için oldukça yeterliydi, fakat yapılandırılmamış ya da yarı-yapılardırılmış veri artarken ikili büyük nesneler (BLOB) ve nesne veri tabanı yönetim sistemleri hızla çoğaldı. Bu durum adreslenebilir parça setleri için “yapılardırılmamış veri ile başa çıkmak için birleşik bir yaklaşım” işletmelere ve analistlere sağlanan, depolanmış olan yapılandırılmamış veriye izin verdi (Hurwitz ve dig., s. 13). Bu gelişmeler dikkate degerdır, çünkü geçmişin aksine, bugün elde edilebilir birçok veri yapılandırılmamış şekildedir.



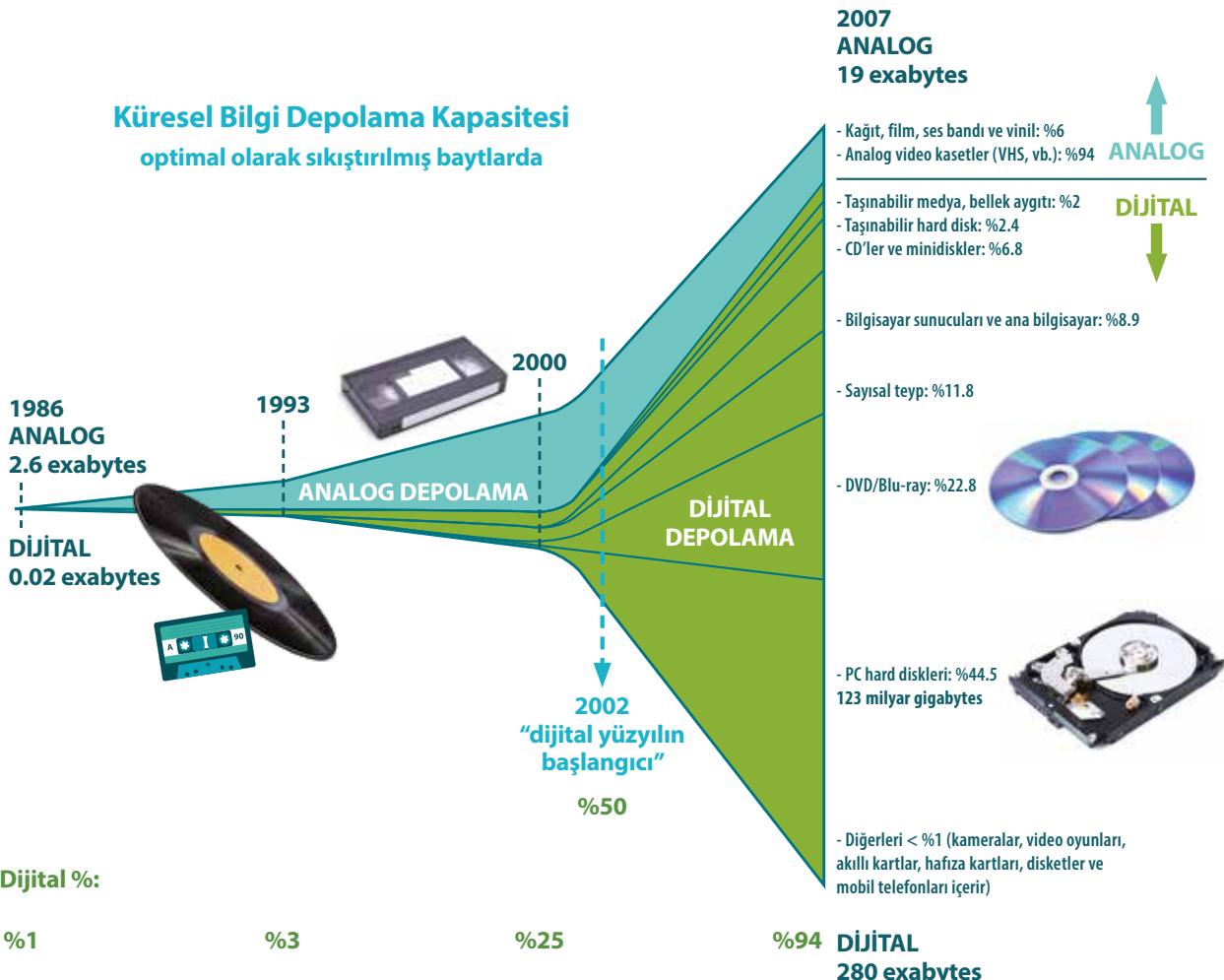
Şekil 3.1 Büyük veri modelleme ve yönetim trendleri. (Kaynak: <https://www.mdpi.com/2071-1050/12/2/634/htm>)

Büyük veri analiz tekniklerinin gelişiminin son safhası sosyal medya ve web 2,0, artan veri sanallaştırma ve bulut programlama ile internetin yükselişinde yatkınlıkta. Özellikle son gelişme, artan erişim hızı ve maliyet verimliliğinin yanı sıra, bir fiziksel konumda tüm verileri depolamak için ihtiyaç duyulan şirketlerin kurulmasıdır. Aynı zamanda insan tarafından üretilen veri miktarı yüksek oranlarda artarak gelişikçe, mantıklı bir şekilde veriyi yönetecek şirketlere duyulan ihtiyaç da önem kazanmaktadır. Ayrıca “modern büyük veri” girişimlerinin birçok teknolojik ihtiyacı vardır ve bu ihtiyaçların hepsi önceki veri yönetim dalgaları tarafından yapılan gelişmelere kısmen dayanmaktadır.



Şekil 3.2 Dyche etkin “büyük veri” girişimleri için altı ana basamak

Dyche etkin “büyük veri” girişimleri için altı ana basamak tanımlar. Bunlar veri toplama, ham veri setlerinin kullanılabılır hale getirilmesi için bir makine ya da bir bilgisayar tarafından işlenmesi, veri yönetimi (sınıflandırma, tanımlama ve heterojen kaynaklardan gelen verinin açıklanması), toplanabilen verideki oranların ölçülmesi, sağlama, yani “işlemin orijinal gerekliliği ile veri kullanılmasındaki sonucu” ve hemen bu veri için depolama çözümlerini adreslemedir.



Şekil 3.3 Küresel bilgi depolama kapasitesinin büyümesi ve dijitalleşmesi. (Kaynak: https://en.wikipedia.org/wiki/Big_data)

Ek olarak büyük veri girişimleri için teknik gereksinimler “geniş veri işlem basamaklarıyla başa çıkmak için çok yüksek çıktıayı” desteklerken, “basit sorguları, kısa zamanda yürütme ve veri yakalama için öngörülen kısa sürede teslim etmek” ve “çok yüksek işlem hacimlerin üstesinden gelmek” için gerekli altyapı kabiliyetlerini içermelidir. Bununla beraber “orijinal depolama yerinde işleyebilmeli ve sürdürmelidir,” (Oracle, 2012). Ayrıca güvenlik endişeleri ve potansiyel hassas veri toplanan çok miktarda veriye özgüdür, bu yüzden veri için “güvenli çerçeve” sağlanması da fevkalade bir endişe olmalıdır (Shuster, 2012).

Daha önce belirtildiği gibi, bugün yeni teknolojiler büyük veri girişimleri için ihtiyaç duyulan altyapıyı oluşturmak için önceki ilerlemelerden gelişti. 2005 yılında Hadoop adında bir proje doğdu. Hadoop, büyük veri alanında çok önemli bir teknolojidir. MapReduce programlama modelini kullanarak büyük verilerin dağıtılmış depolanması ve işlenmesi için bir yazılım çerçevesi sağlar. Hadoop Dağıtılmış Dosya Sistemi özgün depolama kümesinde kalan veriye izin verirken, geniş hacimli verinin işlenmesi ve organizasyonuna müsaade eden bir teknolojidir (çok geniş miktardaki verinin hareketi hem zaman tüketimi hem de zorluk olduğu için büyük veri girişimleri mümkün olduğu kadar ondan kaçınmaya uğraşır). Dünyadaki birçok ülke ve bazı araştırma enstitüleri Hadoop üzerinde bazı pilot projeler yürüttü ve bir dizi sonuç elde etti.

2011'de EMC "Bulut Büyük Veriyle Buluşuyor" üzerine küresel bir zirve düzenledi ve aynı yılın Mayıs ayında McKinsey ilgili bir araştırma raporu yayınladı. Dijital evrenin önumüzdeki on yıl içinde 35 zettabayt bilgi içermesini bekliyorlardı. EMC, gelen büyük verilerin nasıl depolanacağı, yönetileceği ve bu veriler üzerinde nasıl hareket edileceği konusundaki görüşlerini tanımlayan "EMC Büyük Veri Yığını" adını verdiği şeyi tanıttı.

Aynı yılın Aralık ayında, Çin Sanayi ve Bilgi Teknolojileri Bakanlığı Nesnelerin İnterneti için 12. Beş Yıllık Kalkınma Planı'nı yayınladı. Çin gelecekte akıllı endüstri, akıllı tarım, akıllı lojistik, akıllı ulaşım, akıllı şebeke, akıllı çevre koruma, akıllı güvenlik, akıllı tıbbi bakım ve akıllı ev için finansal desteği artıracak. Büyük verinin ilk uygulamasını temsil eder.

Örneğin Amerikan çokuluslu yazılım şirketi Oracle, 2012 yılında büyük veri gereksinimleri ile başa çıkabilen ürünlerinden bazılarının nasıl olduğunu gösteren bir tanıtım belgesi yayınladı. NoSQL veri tabanları sık sık geniş veri miktarları edinmek ve depolamak için kullanılmaktadır.

Bu yazılım ve benzer diğer birçok programlar 1970-1990'ların programlama ve yönetim teknolojisindeki daha önce bahsedilen ilerlemelerden inşa edilmiştir ve 2012 ve 2015 yılları arasında, Birleşmiş Milletler de dahil olmak üzere dünyanın dört bir yanındaki birçok hükümet ve şirket, büyük verinin gelişimini teşvik etmek için bir dizi ilgili fikir veya eylem ana hatları yayınladı. Bundan sonra büyük veri yüksek hızlı bir gelişme aşamasına girdi ve büyük veri endüstrisinin 13. Beş Yıllık

Kalkınma Planı 2017'de Çin'de doğdu, bu da büyük verinin yaygın ve hızlı bir şekilde dünya çapında kullanılmaya ve geliştirilmeye başlandığı anlamına gelmektedir.

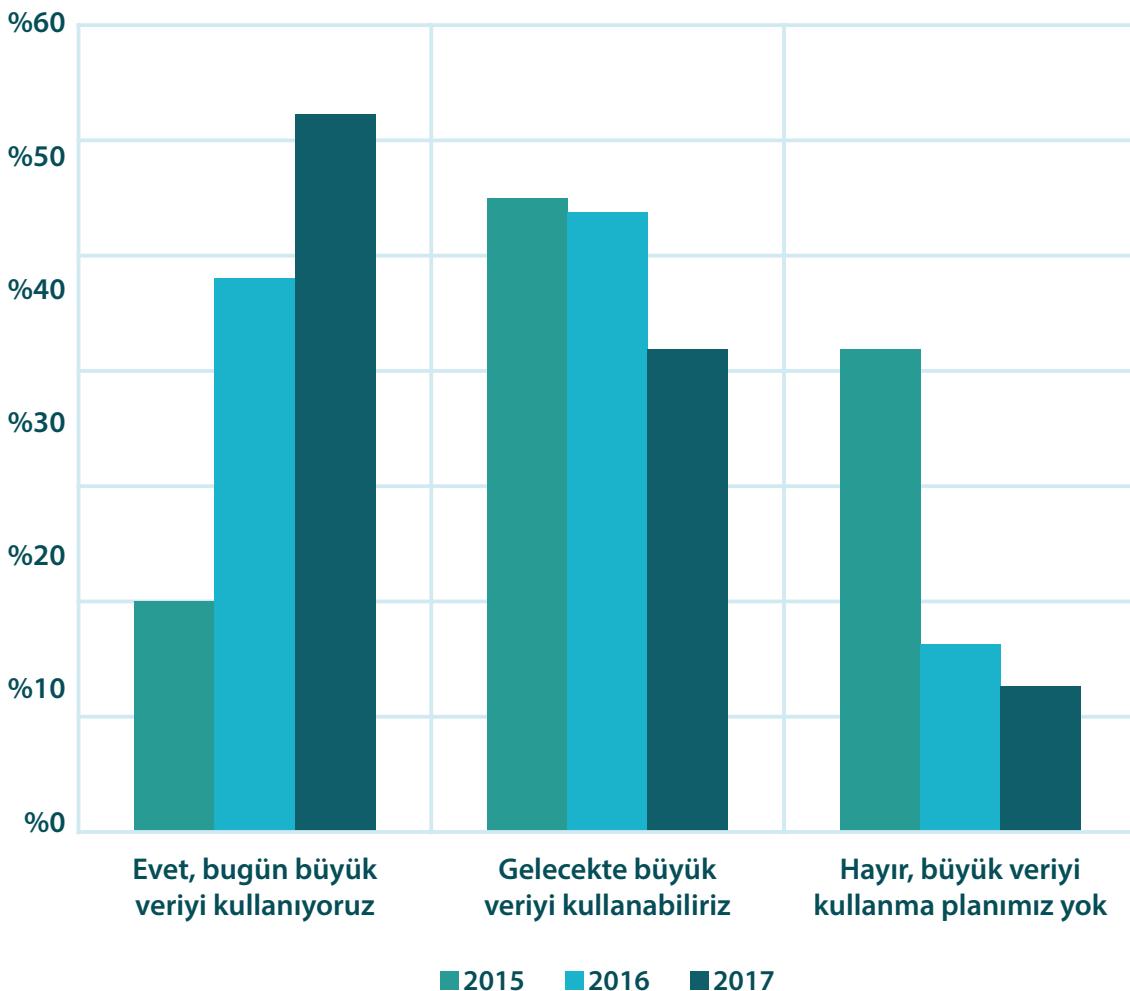
3.2 BÜYÜK VERİYE UYUM STRATEJİLERİ

Tim Suther, Bill Burkart ve Jie Cheng (2013) şu şekilde ifade etmiştir; veri ve “büyük veri” değerli olmasına rağmen durum aşırı basitleştirilmiş olmamalıdır. Şirketler büyük veri girişimleri, içerik gizliliği, yasal ve yazılım endişeleri ile gelen ölçülebilir zorluklarla büyük veriden gelecek potansiyel kazançları denelemek için birkaç “somut adım” atmaya ihtiyaç duyar.

Suther ve diğerleri “veri çok büyük değerin kilidini açar” diye belirtir, fakat özellikle verinin türü hayatı şekilde önemli bir değişkendir. Örneğin en önemli soru müşterilerin ne aldığı, harcama alışkanlıklarıyla ilgili midir? Onlar ne istediklerini söylerler? Ne araştırırlar? Müşterilerin sosyal medya aktivitelerindeki veri? Ya da internette gezerken bıraktıkları dijital kırıntılar? Bu sorulardan birisinin sihirli “veri mermisi” olmasından ziyade, aslında sonuçta önemli olan bu farklı faktörlerin bir kombinasyonudur. Bu zor olsa bile “liderler müşterileri hakkında tüm uygun perspektiflerden rafine anlayışta çok boyutlu olarak düşünmelidir.” Bunun nedeni müşteri davranışını tam olarak açıklayan bir veri yoktur, çünkü her veri noktası veya değişkeni, kendi konumunda değerliken, ayrıca eksiklikleri de vardır (örneğin bir müşteri kendi eliyle deneyene kadar yeni bir elektronik alet istediğini bilmeyebilir; müşterilerin cinsiyet bilgileri önemlidir fakat bazen yanlışdır vb.)

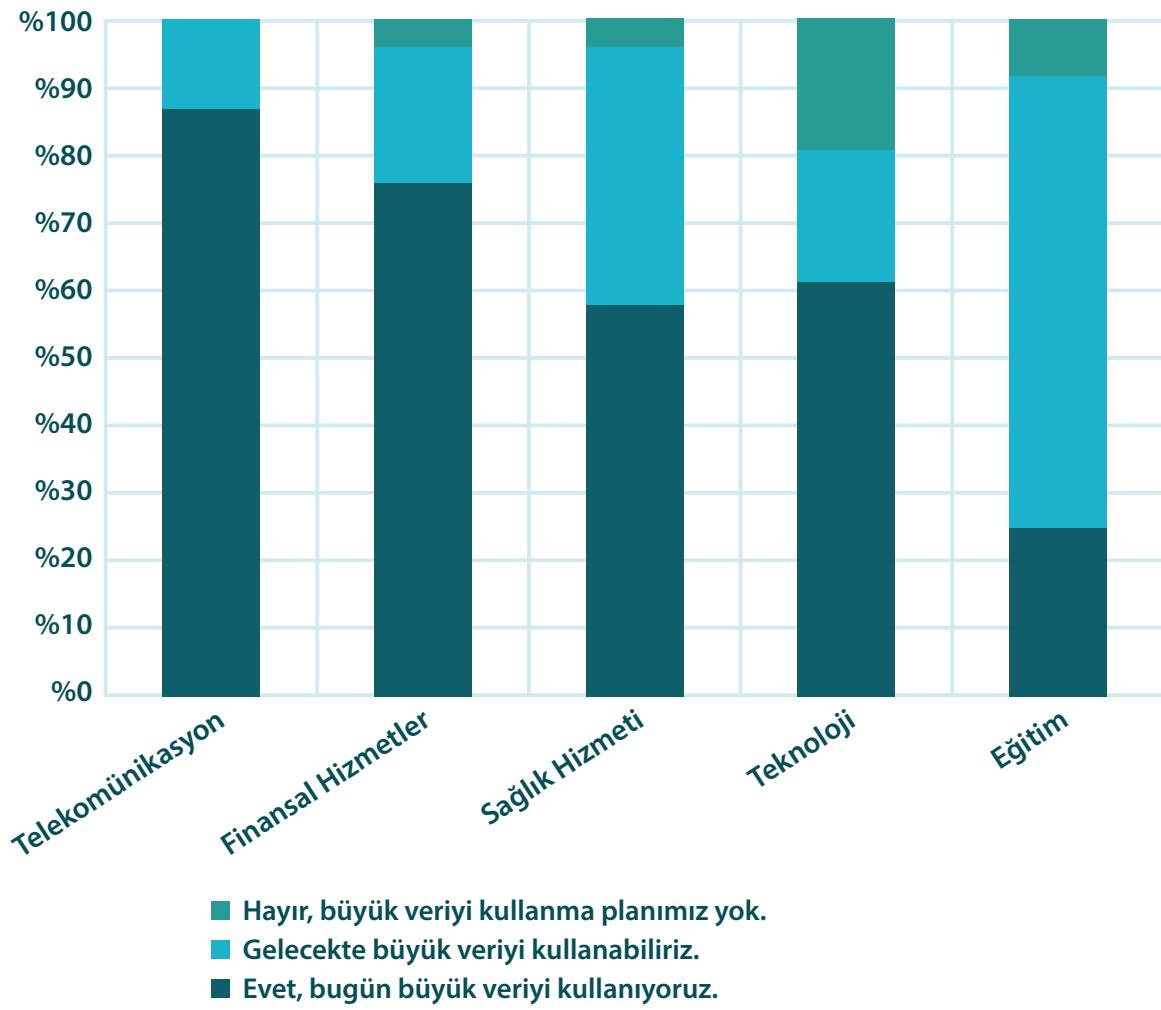
Bundan dolayı Suther ve diğerleri şu şekilde ifade eder, bir şirket “kendi müşteri portföyünü büyüterek” büyük veri girişimleriyle gerçek başarıyı sağlayabilir. Bu “çok boyutlu anlayış” başarılı bir girişimin “dayanağını oluştururken,” Suther ve diğerlerine göre birçok firmanın gözden kaçıldığı anahtar nokta şudur: Kendi var olan müşteri profiline yetersiz yatırım yaparlar.

İfade edildiği gibi, şirketler “su anki müşteri ilişkilerinin değerini ve yatırım kararları yürütmek için onu kullanmanın değerini” hesaplamalıdır. Sonra, “tersüz” çalışarak amaç şu andaki var olan kendi müşterilerinin en kârlı alt kümesine benzeyen potansiyel müşterileri eşleştirmek olacaktır. Bu şekilde müşteri portföyünün değeri katlanarak artacaktır.



Şekil 3.4 2015-2017 yılları arasında büyük verinin benimsenmesi (Kaynak: 2017-Dresner Danışmanlık Hizmetleri. <https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/?sh=1f6e15f539a1>)

Son olarak Suther ve diğerleri şunu söylemiştir, özellikle veri sağlayabilmesine benzer, büyük potansiyel güç önüne alındığı zaman büyük veri ile uğraşırken ahlaki davranış hayatıdır. “Bu ticari zorunlulukları aşan veri müşteriler için kullanılmalıdır, onlara başka bir şey yapmak için kullanılmamalıdır. Bu hızlı puanlara değil, güven tabanlı ilişkiler kurarak uzun dönemli optimizasyon anlamına gelir” şeklinde ifade etmişlerdir. Bu prensipleri takip etme karşılığında, Suther ve diğerlerinin belirttiği gibi şirketler üç önemli karşılık umabilir: Pazarlama performansında ölçülebilir gelişmeler, müşteri portföylerinin artan değeri ve daha anlaşılabilir ve korunabilir fiyat. Büyük veri yanlış tanıtımı düzeltmek için özel potansiyele sahiptir.



Sekil 3.5 Dikey sektör tarafından büyük verinin benimsenmesi (Kaynak: 2017-Dresner Advisory Services. <https://www.forbes.com/sites/louiscolombus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/?sh=1f6e15f539a1>)

ComScore tarafından yapılan bir çalışmaya göre, çevrimiçi reklam % 80'ini aşan oranda istenilen hedefe ulaşmada başarısız olmaktadır. Bu durum kısmen, örneğin yaş ve cinsiyet gibi aşırı geniş bir genelleme alanına dayanmak yerine, müşterilerin bireysel alım uygulamaları ve profillerindeki nüans özelliklerinin başarısızlığından dolayı ortaya çıkabilir (Comscore 2009). Eğer şirketler büyük veri girişimlerini “daha akıllı hedefleme ve daha iyi yönetim” içeresine tutturmak için kullanırlarsa, Suther ve diğerlerine göre, pazarlama yatırım geri dönüşlerinin %15-30, müşteri brüt kâr marjlarının %10-15 arttığını ve %7 iyileştirilmiş fiyatlandırma

görebilirler. Çünkü veri miktarı büyürken genelde her yeni veri parçasının marginal değeri azalır. Daha fazla araştırma için kullanılan bilgisayarlar ve analitik işlemler etkin bir şekilde büyüyen veri miktarını faydalı hale getirmek ve başa çıkmak için tek yöntemdir.

Bu bağlamda 2017 Büyük Veri Analitiği Pazar Araştırması coğrafyalar, işlevler, organizasyon boyutu ve dikey sektörleri kapsayıcı bir çalışma gerçekleştirmiştir. Bu noktada raporlama, gösterge panoları, gelişmiş görselleştirme son kullanıcı “self servis” ve veri ambarı, iş zekâsı için stratejik olan ilk beş teknoloji ve girişimdir. Büyük veri Dresner Danışmanlık Hizmetleri’nin şu anda izlediği 33 temel teknoloji arasında 20. sırada yer alıyor. Büyük veri analitiği nesnelerin interneti (IoT), doğal dil analitiği, bilişsel iş zekasından (BI) daha stratejik bir öneme sahiptir.

Telekom ve finansal hizmetler sektörleri en hızlı uyumu sağlarken 2015’te %17 iken bugün şirketlerin %53’ü büyük veri analizi kullanıyor. Telekom ve finansal hizmetler en aktif erken uyum sağlayanlardır; teknoloji ve sağlık hizmetleri büyük veri analitiğine önem veren üçüncü ve dördüncü sektörlerdir. Eğitim sektörü ise 2017 sonuna yaklaşırken en düşük uyum oranına sahiptir.

Veri ambarı optimizasyonu 2017’deki en önemli büyük veri analitiği kullanım örneği olarak kabul edilir ve bunu müşteri/sosyal analiz ve öngörücü bakım izler. Veri ambarı optimizasyonu tüm yanıt verenlerin % 70’i tarafından kritik veya çok önemli olarak kabul edilir. Nesnelerin internetinin (IoT) günümüzde büyük veri analizi için en düşük öncelikli kullanım örnekleri arasında yer alması ise oldukça ilginç ve ironiktir.

Diğer taraftan Spark, MapReduce ve Yarn, günümüzün en popüler üç yazılım çerçevesidir. Ankete katılanların% 30’undan fazLASI, Spark’ı büyük veri analizi stratejileri için kritik olarak görüyor. MapReduce ve Yarn, ankete katılanların yüzde 20’sinden fazLASI için “kritiktir.”

Katılımcılar tarafından en çok tercih edilen büyük veri erişim yöntemleri arasında Spark SQL, Hive, HDFS ve Amazon S3 bulunuyor. Ankete katılanların %73’ü Spark SQL’ı analistik stratejileri için kritik olarak görüyor. Yanıt verenlerin %30’undan fazLASI Hive ve HDFS’yi de kritik olarak değerlendiriyor. Amazon S3, büyük veri erişimini yönetmek için yanıtlayan beş kişiden biri için kritik önem taşiyor. Aşağıdaki grafik, büyük veri erişim yöntemlerinin dağılımını göstermektedir.

ComScore tarafından yapılan bir çalışmaya göre, çevrimiçi reklam %80'ini aşan oranda istenilen hedefe ulaşmada başarısız olmaktadır.

Daha sonra görüleceği gibi, örüntü tanıma büyük veri yönetimi programlarının hayatı bir parçasıdır. Bununla beraber yöneticiler ve firmalar büyük veri yönetimini yalnızca bilgisayarlara ve uygulama programlarına bırakmazlar, halen insan analizi ve girişine ihtiyaç duydukları yerler vardır.

3.3 BÜYÜK VERİNİN İNSAN UNSURU

Daqing Zhao'nun (Dir Advanced Analytics at Macys.com) belirttiği gibi, neticede bilgisayarlar analistlerin komut ve talimatlarına uydukları için bir bilgisayar sonuçta “sadece onu kullanan analist kadar iyidir.”

Büyük veri kaynakları, özellikle de yapılandırılmış veriler, tamamen büyük ölçüsünden dolayı değil, ayrıca belgesi ve sınıflandırılmış bilgiden dolayı sıkılıkla bunaltıcı olabilir. Eğer “ham” şekilde gelirse, henüz bir etiketlemesi yoktur. Kalitesini sağlamak için veri, doğruluk ve “arılık” için kontrol edilmiş olması gereklidir. Zhao, “veri sadece onları analiz etme için kullanılan çabanın miktarı kadar temizdir,” der. O uzun ve zahmetli fakat sonuçta gerekli olan veri doğruluma ve kalite hata ayıklama sürecini gelir sağlama sürecine benzetir. İnsan unsuru da cevaplanması gereken sorular için hazır olmalıdır. Model bağımlı değişken ve bağımsız belirleyicileri tarafından yapılandırılmış olacaktır. Bunların tümü ilgili kimselerle analistler tarafından bilgisayarların içerişine yerleştirilir.

Gelişen müşteri servisleri açısından “servis kalitesinin” nasıl ölçüleceği bir örnektir. Ayrıca bir web sitesine yapılan tıklamalar, ziyaret frekansı, işlem miktarı veya kullanım süresi de bir ölçü olabilir. Sonra gün, konum, yaş ve cinsiyet gibi birtakım değişkenler belirleyicilerdir. Analistler (ve dolayısıyla insan kaynakları) büyük veri girişimlerinin hâlâ önemli bir parçasıdır, çünkü bilgisayar algoritmaları eğer model-



deki birtakım değişken öngörü yapabiliyorsa karar verebilir, fakat bir bilgisayar eğer kritik öngörüyü gözden kaçırıyor ise kendisi üzerinde fark edemez. Diğer taraftan eğer veri analiz modeli kullanımında herhangi bir problem varsa, bilgisayarın (bağımsız olarak) bilmesinin bir yolu yoktur. Sonuçta Zhao belirttiği gibi, “bilgisayar modelleri insanlar tarafından tasarılmak ve yönetilmek zorundadır.”

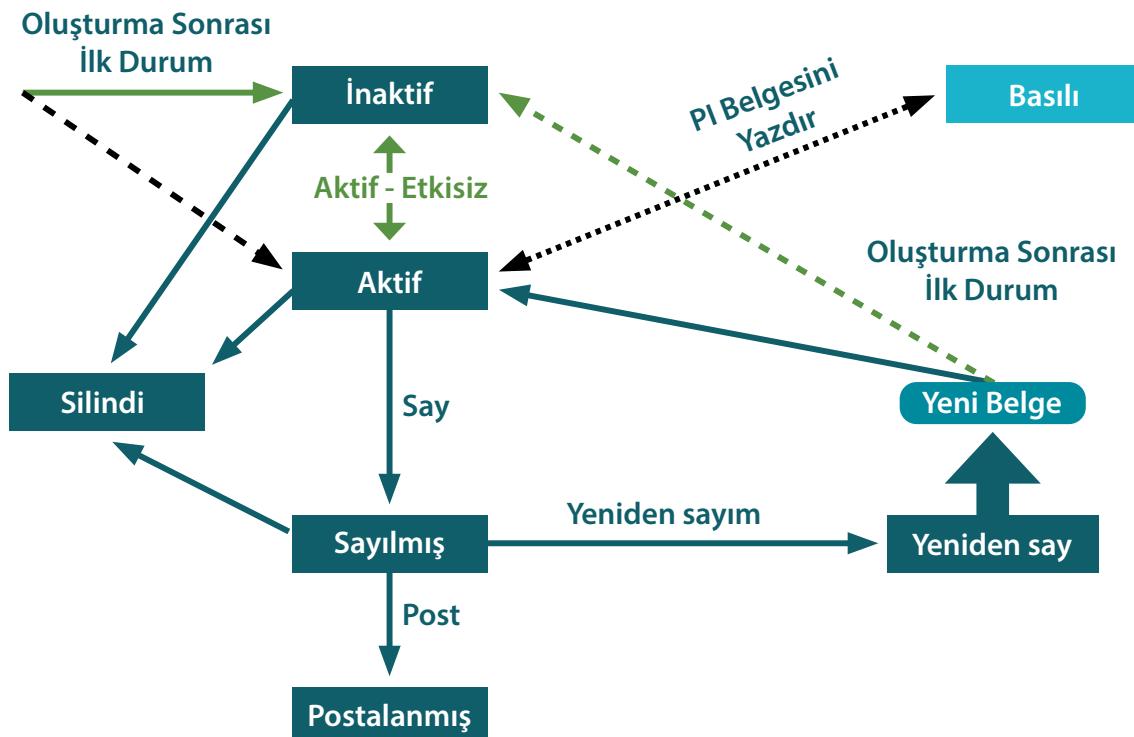
Zhao ayrıca şuna da işaret eder ki modeller sadece uzun bir periyod süresini aşan tekrarlanabilir kalıplar için yapılandırılabilir. Gittikçe daha fazla model inşa edilirken bazıları çok kısa “yaşam süresine” sahiptir ve sadece birkaç hafta içerisinde performansında gerileme olur. Bunun nedeni, öngörü modelleri yapılandırılmak için uzun bir süre alır ve büyük veri ile birlikte öngörücülerin birçok sayısı ve ölçüsü olabilir. Çünkü potansiyel olarak “binlerce değişken” olabilir. Zhao yine şunu ifade etmektedir ki “Kullanıma hazır modelleme ve örnek dışı test etme benzeri teknikleri kullanan, bir nebze otomatik olarak modeller yapılandırmaya yardımcı olan modelleme metodolojisine güvenmemiz gereklidir. Sağlanan veri yüksek kalitededir (çünkü daha yüksek kalite veri daha yüksek modeller sağlar) ve depolama modeli de dikkate alınması gereken önemli faktörlerdendir.

3.4 YENİ MODELLER VE BÜYÜK VERİ

Daniel Conway ve Diego Klabjan “veri bir şirket için etkili bir biçimde bilgiye dönüştürüldüğü zaman, yeniliği teşvik üzerinde pozitif etkilere sahip olacaktır,” demektedir. Veri “uykuda ve pasif” iken, uygun bir şekilde analiz edilip kullanıldığı zaman, “bilgi”nin bir şekli olarak değerli bir emtia olur. Özellikle kullanımındaki (göreceli) ölçü ve maliyetinden dolayı “büyük veri” değerinin maksimum olması için çeşitli bileşenler içeresine uygun bir biçimde entegre edilmiş olması gereklidir. Ayrıca şuna da işaret etmişlerdir: Varlıkların gelecekteki kaldırış etkilerinin nasıl olacağı bilinmediği zaman verinin değerini ölçmek zordur. Bunun için büyük veriyi anlamak için birçok farklı bileşeni açıklamaktadırlar (2013, s. 132).

İlk olarak “yeni bir doğal kaynak şekli” olarak büyük veriyi açıklamak için IBM’den alıntı yaparlar. Doğal bir kaynakta olduğu gibi, tam anlamıyla kullanılmadan önce çıkarılmış, birleştirilmiş ve çeşitli derecelerde işlenmiş olmaya ihtiyaç duymaktadır. Ayrıca dünyadaki geleneksel kaynakların seçiliп çıkarılmasıyla ayırtırılarak ve kullanılarak veri zamanla daha etkin olur (geçen on yılda önemli derece büyүyen bir konsept olarak “büyük verinin” olma nedenin bir parçasıdır). Bununla birlikte doğal kaynaklardan farklı kısırlığı ima eden bir terimdir, büyük veri miktarı sadece büyümektedir ve kurumların karşılaşışı zorluk da aslında bu veriyi etkili bir biçimde nasıl kullanacaklarıdır. Büyük veriyi incelemenin bir başka yolu dijital envanter (digital inventory-DI) olarak ona bakmaktadır. DI etrafında depolama ve taşımının maliyeti fiziksel envanterden (physical inventory-PI) daha az iken, büyük DI halen altyapı ve bant genişliğindeki dar boğazlara maruz kalır, bu yüzden şirketler aynı zamanda “ekonomik gerçeklerin” bu engelinden haberdar olmayı gerektiren büyük veri girişimlerini üstlenmektedirler (2013, s. 133).

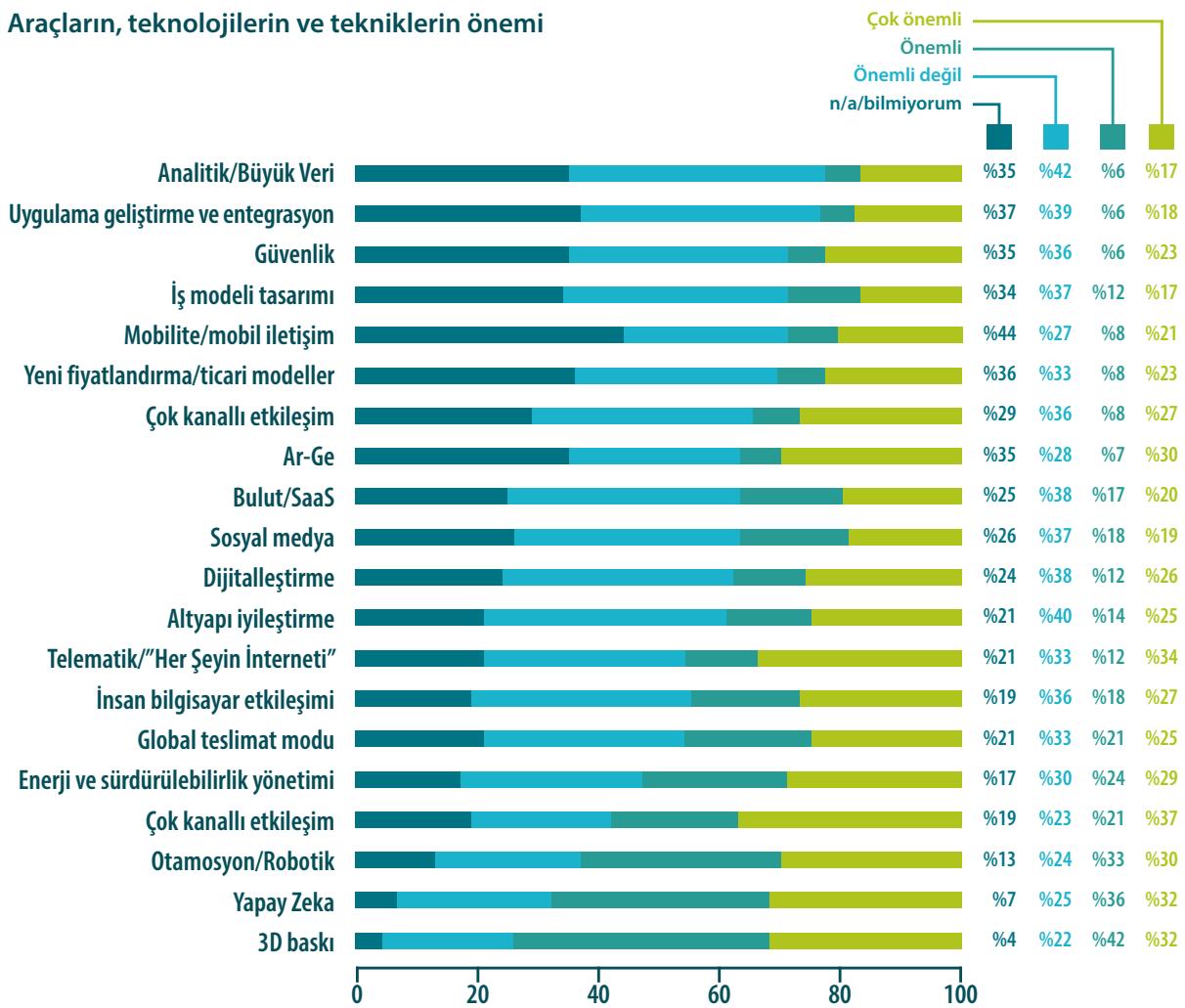
Nihayet Kent ve meslektaşları gibi, Conway ve Klabjan büyük verinin yeniliği teşvik etmede oynayabildiği rolü araştırmaktadır. Yenilik takım işbirliği yoluyla olduğu kadar bireysel kaynaklarla birlikte her ikisinin üzerinde olusabilir ve bir karar süreci “aktiviteler seti, icra emir vb. seçilecek yolları belirleyen karar kuraları”dır. Önemli biçimde, büyük veri her elemanda ve bu sürecin basamaklarında bir role sahip olabilir.



Şekil 3.6 Fiziksel envanter işleme. (Kaynak: https://help.sap.com/saphelp_scm70/helpdata/en/4cb4e7d94c1f1921e1000000a15822b/content.htm?no_cache=true)

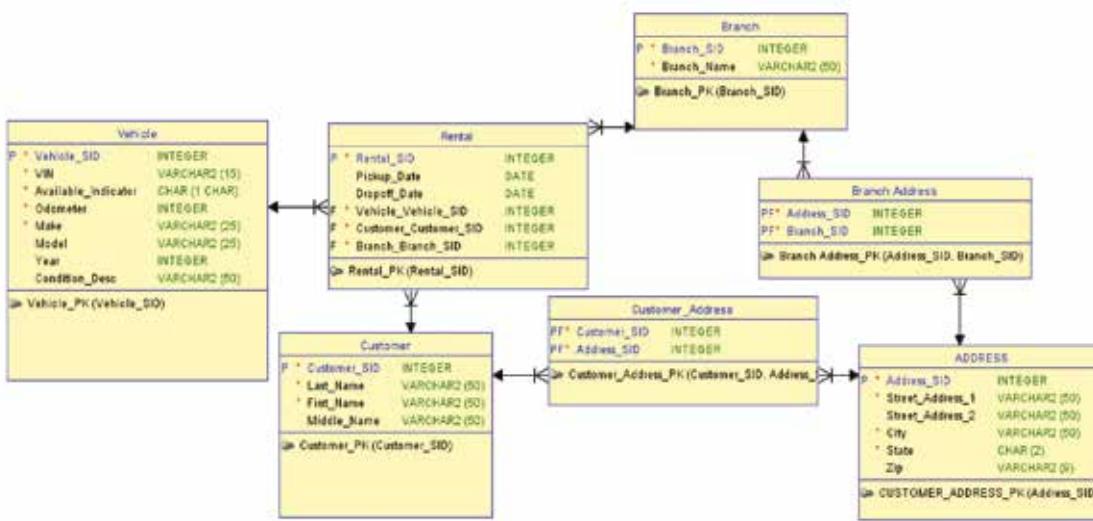
Ian Thomas ayrıca şunu belirtir; çoklu karar alma süreçleri içerisinde büyük veri bir organizasyon için değerini ve etkisini maksimize etmek için en iyi yoldur. Büyük veri ile ilgili yapılan teknoloji seçimlerinin birçok şirket ve organizasyon için zor olabileceğini ifade eder, çünkü “çözümler ile pazardaki çok çeşitli satıcıları birbirine karşılaştırmak kolay değildir,” (Thomas, 2013, s. 155). Şuna da dikkat çekmiştir; sadece teknoloji yönüne odaklanmaktan başka, organizasyonlar için hatırlanması gereken önemli nokta da büyük veri sistemleri ve girişimlerinin gerçek amacı işletmenin nasıl çalışacağı ile ilgili “birçok büyük ve küçük karar verenler için her gün veriye güvenmek için kurumlarda insanlara olanak vermektedir.” Bu bütünsel yaklaşım nihayetinde bir şirket için daha faydalı olabilir.

Thomas şu şekilde devam etmektedir, bir büyük veri projesi değerlendirildiği zaman, bu verinin kullanıcıları karşılaşılan üç şartta sahip olmaya ihtiyaç duymaktadır: Anlamlı verileri yorumlayabilmek için doğru beceri ve yardıma sahip olmalı, verinin güvenilir olduğuna inanmalı ve gerekli detayların doğru seviyelerine kolayca



Şekil 3.7 Teknolojilerin, araçların ve tekniklerin yenilik açısından değeri. (Kaynak: <https://www.consultancy.uk/news/1946/mobile-and-big-data-analytics-key-tools-for-innovation>)

erişebilmeli. Bu üç şartın hepsi karşılanmalıdır, eğer sadece ikisi karşılanırsa veri israf edilmiş olacaktır. Bundan dolayı şirketler başarılı bir büyük veri projesinin üç içeriği olan “insanlar,” “sureç” (Thomas tarafından belirlenen “sıradan bir izleme ön çalışması, veri kalitesi sorunlarını raporlama ve standartları ve tanımları savunma ile ilgilenme” (Thomas, s. 156) ve “teknolojiden” emin olmalıdır. Aslında büyük veri girişimlerini doğru bir şekilde uygulamak ve yorumlamak için çok yönlü ve iyi eğitimli bir takım gereklidir. Thomas'a göre, tam olarak bölünmez bir bütünlük için mühendislik ve işletmenin her ikisini de içeren personel potansiyeli tam artan bir firmanın sahip olduğu veri sağlamdır.



Şekil 3.8 Hayali bir kiralık araba şirketi için veri modeli örneği. (Kaynak: https://www.researchgate.net/figure/An-example-of-a-Data-Model-for-a-fictitious-Rental-Car-company_fig1_277890728)

Nihayetinde Hurwitz, Nugent, Halper ve Kaufman teorik düzeylerin ötesinde gerçek bir işe büyük veri analizlerinin pratik bir şekilde nasıl uygulanacağını birkaç örnekle aktarmıştır. İlk olarak net ve açık analizler için sürecin başında şirketlerin ihtiyaçlarını ve amaçlarını değerlendirmeleri gerekmektedir. Örüntü tanıma birçok veri analiz programının anahtar unsuruudur.

Bir örnek olarak araba kiralama şirketi örneğini aktardılar. Şirket aynı yüksek ek maliyetlere sahip olmayan gelişen şirketlerden güçlü bir rekabet deneyimliyordu (Hurwitz ve diğ., 260). Şirket onların müşteri hizmetleri ve cevap verme yeteneklerindeki zayıf noktayı tespit etti. Online bir anket oluşturdu, şikayetler ve önerilerle ilgili e-posta ile müşterilerini cesaretlendirdi. Geleneksel olarak bununla birlikte bir yönetici bu e-postaları veya yanıtları ayrı ayrı okumaktaydı ve elbette ek olarak bireysel kişinin takdirine bağlı e-postaların sınıflandırılması ve analizi yavaş adımlarla olmaktadır. Bir metin analiz programı halbuki çok daha hızlı şekilde bakabilirdi, ayrıca daha hızlıca dokunan bu coğrafi veri bir yer veya bölümdeki problemleri tanımlamaktadır. Diğer örnek de “etkileşimin erken aşamalarında” dolandırıcılığı daha etkin bir biçimde ortadan kaldırmak için birçok veri kaynağından gelen çaprazlama potansiyel ile dolandırıcılık iddialarına yardım edebilen bir şir-

ket olmasını sağladı. Bundan dolayı bu verinin yine doğru zamanda analiz edilen veri “düzeltici faaliyet için çok geç olmadan önce, müşteri memnuniyetsizliği veya potansiyel ürün kusurunu tanımlamaya” nasıl yardımcı olabileceğinin bir örneğidir (Hurwitz ve diğ., s. 256).

Sonuç olarak bu bölüm göstermiştir ki her derde deva olmak yerine, “büyük veri” (kesin olmayan ve hatta ilişkisel bir terim olarak bile) zorlukları olmasına rağmen bir şirket için birçok potansiyel kullanımlara ve faydalara sahiptir.

Kurumlar bir büyük veri projesi açmadan önce uygun insan ve teknolojik kaynaklara sahip olmayı sağlaması almanın ihtiyacı duyarlar ve daha sonra mümkün olduğu kadar şirketin birçok alanının içerisinde geniş veri kaynaklarını entegre etmek için at gözlüğünü çıkarmaya çaba harcarlar.

3.5 VERİ PAYLAŞIMI, GİZLİLİK VE ETİK

Toplumsal farkındalığımız artarak yeni bilgi toplama ve analiz şekillerinin olduğu bir zaman diliminde yaşıyoruz. Büyük veri devrimi insanlarınla ilgili olan birçok veriyi kaydediyor. Hepimizin kişisel bilgi gizliliği ile ilgili var olan yönetim kuralları bulunmakta, fakat yeni veriden elde edilen yeni bilgi akışları, yeni kullanım şekilleri ve yeni kararlar bu kuralları eksik hale getirmektedir. Bu yeniçağın red dedilemez faydalarını kurban etmeden bu yeni araçların toplumsal maliyetlerini düzenleyen yeni kurallara ihtiyacımız vardır.

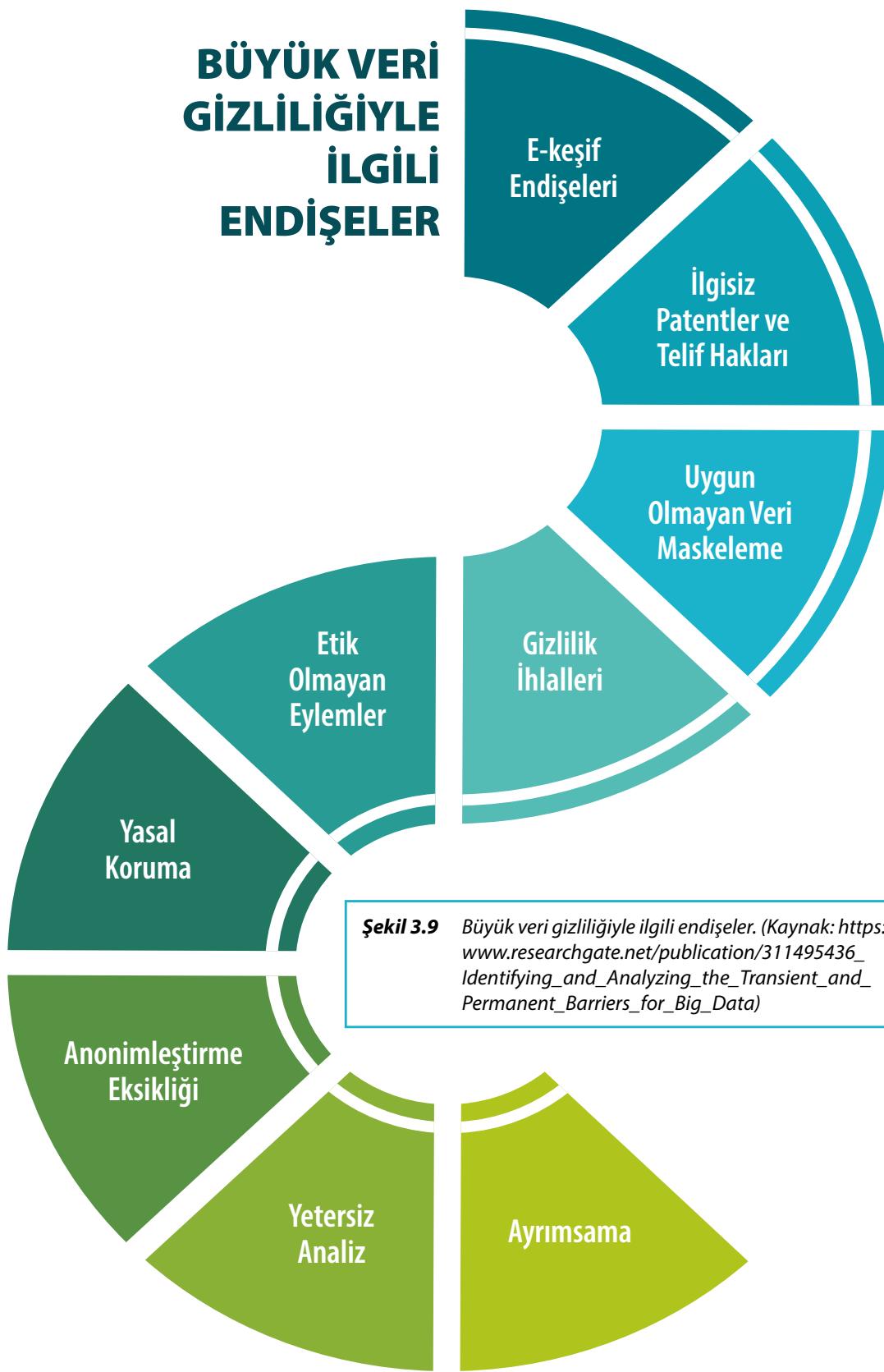
Fakat bu yeni kuralları şekillendirmede bize hangi değerler rehberlik edecek?

Gizlilik, şeffaflık ve kimlik değerleri belki de bu konudaki başlangıç noktası olacaktır.

Genelde “gizlilik” başlığı altında kişisel gizlilik problemlerini düşünürüz. Fakat büyük veri devriminin “gizliliğin ölümü”nü işaret etmesi gerekmez. Aksine şu anki bilgi akış yönetim kurallarının yerine, daha korunan sırlar düşünürüz ve tanımlarız. Bu durum büyük verinin gerçek gizlilik konusunu odak haline getirmektedir. Bizim veri akışını düzenleyen kurallara ihtiyacımız vardır ve bunun anlamı şudur; kişisel verinin toplanması gizlilik tartışmalarının sonu değil, başlangıcı olmalıdır.

Son dönemde teknoloji şirketlerinden bazıları güven restorasyonu için çağrıda bulundular. 9 Ekim 2013’te Başkan Barack Obama ve Kongre’ye açık mektup ya-

BÜYÜK VERİ GİZLİLİĞİYLE İLGİLİ ENDİŞELER



yinlayan Apple, Google, Microsoft, Facebook, Yahoo, LinkedIn, Twitter ve AOL gözetim ve düzenleme reformu için bir çağrıda bulundu. Microsoft Genel Danışmanı Brad Smith şunu yazmıştır: “İnsanlar güvenmedikleri teknolojiyi kullanmak istemeyecektir. Hükümetler bu güven riskini oluşturdukları ve onu tamir etmek için yardımcı olmalıdır.” Yahoo’ nun CEO’su Marissa Mayer de “Hükümetlerin gözetim aktiviteleriyle ilgili son günlerdeki ilişkiler kullanıcılarımızın güvenini sarstı, ABD Hükümeti için dünya çapında halkın güvenini yeniden sağlamak için tam zamanıdır,” dedi. Evet, bu teknoloji sağlayıcıları reform için çağrıda bulunmaktadırlar. Çünkü eğer müşterilerinin kendileri ile paylaştıkları özel bilgilerin artık güvenilirliği olmazsa, müşteriler onların hizmetlerine karşı olan güvenlerini yitireceklerdir ve bu da firmaları korkutmaktadır.

Gizlilik gibi şeffaflık konusu da güveni besleyen bir unsurdur. Hükümetlerin bilgi şeffaflığı devlet kademeleri, özgür bir basın ve bireysel haklar arasında bir denge kurar ve kontrol eder, anayasal denetim içinde kritik bir rol oynar. Finansal raporların şeffaflığı, özellikle para ve hisse senetleriyle ilgili yatırımcıların isteğini artırır. Örneğin Avrupa Birliği Veri Koruma Yönergesi halihazırda şeffaflığın korunmasını sağlamaktadır. Ira Rubenstein, Doc Searls ve arkadaşları da şöyle bir geleceği açıklar: “İlave şeffaflık korumaları müşterilerin kişisel verileri üzerinde kontrole izin veren yeni iş modellerini destekleyecek veri taşınabilirliğine izin verecektir.”

Şeffaflık doğal şekilde açıklık ve gizlilik arasında bir gerilim içerir. Bu gerilim paradokslara neden olabilir. Devlet ve şirketlerin hassas şeffaflık sırları ulusal güvenlik ve ticaret sırları gibi konulara önemli zararlar verebilir. Çok az şeffaflık güven eksikliğine ve beklenmedik sorumlara yol açabilir. Şeffaflık da bir risk taşımaktadır ki dikkatsiz ifşalar zarar verici gizlilik ve gizlilik ihlali gibi beklenmedik sonuçlara neden olacaktır.

Büyük veri kimlikle ilgili olarak da çok daha derin bir şekilde düşünmemizi gerektirir. Kimlik de gizlilik gibi tanımlanması zor fakat aynı derecede korunması hayatı bir unsurdur. Kimlik birçok anlama gelebilir, özgün bir kişiye özgün bir ismi çağrıştırabilir. Gerçekten de bütün kimlik yönetimi ve koruma endüstrilerinde bu kimlik koruma türü vardır. Kimlik ayrıca hukuksal bir kanıt olarak bir şeyin başka birisi ya da bir şeyle aynı olup olmadığı anlamına da gelebilir. Filozoflar uzun dö-



Şekil 3.10 *Kişisel gizlilikle ilgili 6 tehdit.* (Kaynak: <https://blogs.sap.com/2017/04/04/grc-tuesdays-part-one-big-data-privacy-risks-and-the-role-of-the-gdpr/>)

nem bunu tartıştı ve bu tarzdaki kimliği tanımlamaya uğraştılar. Tüketiciler olarak, kimliklerimiz artan bir şekilde büyük veri ve şirketleri kontrolünde şekil almaya başlıyor. Birçok bakımdan ise, bu kontrole bizim ihtiyacımız var ve bunu istiyoruz.

Evet, bugün büyük veri ve diğer yeni bilgi teknolojileri ile ilgili etikleri düşünmeye ve garanti altına almayı ihtiyacımız vardır. Bu teknolojiler “doğal” ve önceden kararlaştırılmış değildir; insan seçimlerinin ürünüdür ve insanı değerleri etkileyecektir. Sahip olmak istediğimiz toplum türünü şekillendiren bu teknolojilerden emin olmak zorundayız. Peki, bunu nasıl yapmalıyız? Hukuksal olarak buna yeni hukuk kuralları oluşturarak adım atabiliz. Aslında halihazırda zaten birçok veri işleme yönetimi kuralı bulunmaktadır. Fakat bunlar bizi yeterli derece korumasa bile, hâlâ amacına uygundur. Bununla ilgili olarak hukuksal konular uzmanlarınınca tartışılmaya ve yeni alternatifler ortaya atılmaya devam etmektedir.



Şekil 3.11 Büyük veri etiğinin dört yönü. (Kaynak: <https://www.youtube.com/watch?v=PsC9CMgyTxY>)

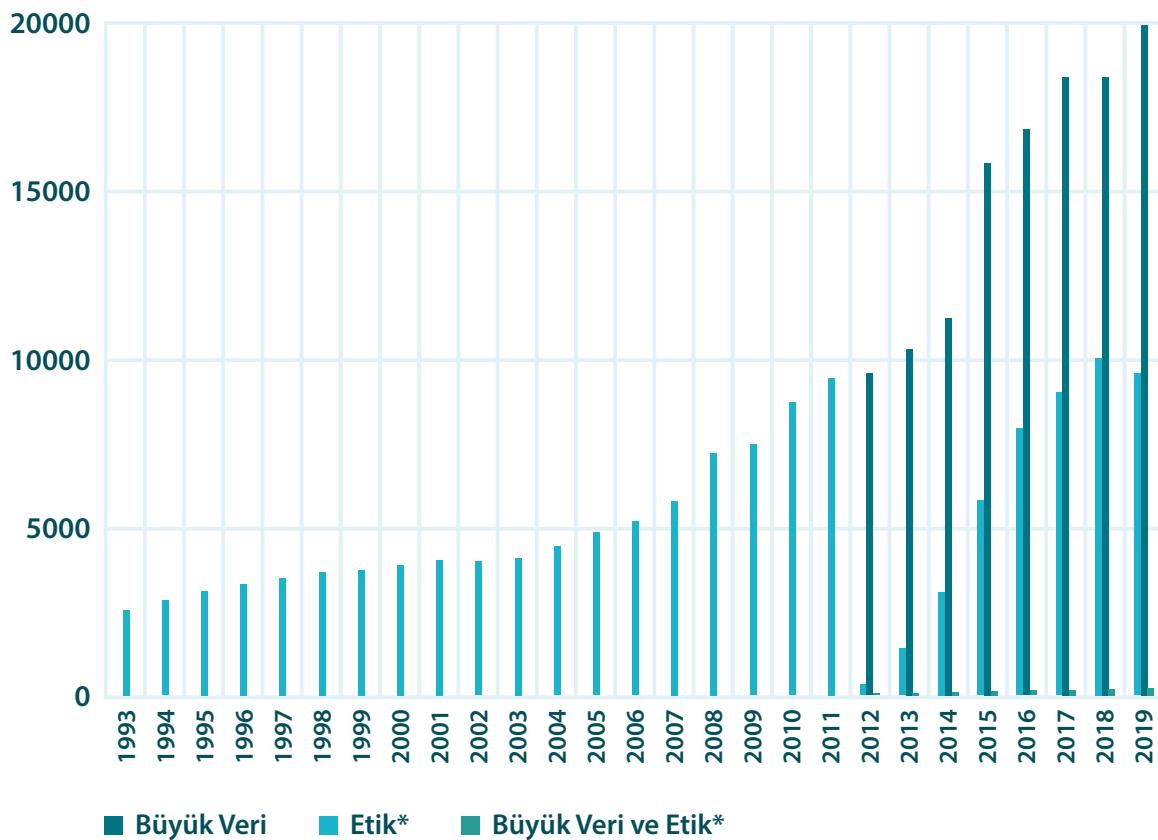
Şunu diyebiliriz ki Licklider'in öngörüsüyle “insanlık tarihinin en heyecan verici ve en yaratıcı entelektüel” döneminde yaşıyor olabiliriz. Diğer yeni bilgi teknolojileri gibi, büyük veri insanlık için keşif ve inovasyon çağında şaşırtıcı imkânlar sunar. Savunma konusunda yapılacaklar için büyük veri kullanımında devlet yetkililerine izin vermemiz gerekmektedir. Büyük veriden daha iyi faydalananmak için şirketlerle bilgilerimizi paylaşmak isteriz. İşte, bu gibi sebeplerle geniş çaplı bir şekilde büyük veri ile ilgili düşünmemiz gerekmektedir. Ayrıca büyük verinin yeni imkânlarını değerlendirmek için oluşan gizlilik, şeffaflık ve kimlik gibi önemli sosyal değerleri korumak adına gizlilik değerlerini, normlarını ve hukuki korumaları geliştirmeliyiz.

Büyük veri gizlilik, güvenilirlik ve kimlik için kesin bir tehdittir, fakat bu hukukun olduğu anlamına gelmez. Toplumsal bilgi akışını yöneten kurallar hem ana esastır hem de bir şekilde kaçınılmazdır. Bilgi devrimi boyutu ne tür kurallar istedigimizle ilgili daha çok hayal gücü ve daha geniş bir şekilde düşünmemiz anlamına gelir. Toplumsal bilgi akışındaki kişisel bilgiyi garanti eden kuralları geliştirmeye ihtiyacımız vardır.



Şekil 3.12 Büyük veri ve fikri mülkiyet hakları arasındaki ilişki. (Kaynak: <https://www.allerin.com/blog/the-sensitive-relationship-between-big-data-and-iprs>)

Geçerli bilgi teknolojileri altyapılarında, ağ bağlantısı gibi servis sağlayıcılar sağladıkları hizmetin niteliği ve kalitesini tanımlayan SLA (Service Level Agreement-Hizmet Seviye Anlaşması) ile bağlıdır. SLA sorumluluğu kısıtlamak için önemlidir. Ağ bağlantıları ve web hizmetlerinin üstel olarak genişlemesi, sözleşme ile uyumluluğun izlenmesi ve onaylanması açısından SLA uygulama otomasyonundaki teknolojik gelişmeler önemliydi. Buna karşılık büyük veri paylaşım anlaşmaları gayri resmi, yetersiz yapılanmış el yordamıyla uygulanan ve soyutlanmış işlemlere bağlı kalır (Koutroumpis ve Leiponen, 2013). Bu durum özellikle kamu kaynaklarıyla integrasyonu için özel veriye erişemeyen sosyal bilim ve yönetim araştırmaları adına önemli bir engeldir. Haklar konusundaki çalışmasında Sart; katılımcılar araştırma sonuçlarının yayımımasını isterken hak sahipliliğinin de korunmasını talep etmektedirler. Burada katılımcıların karşı karşıya kaldıkları önemli bir durum da ortaya çıkmaktadır. Pek çok araştırmacı akademik kariyerinde ilerleyebilmesi için gerekli yayını yapmak zorunda kalırken pek çok fikirlerinin korunamaması nedeniyle patent ve lisanslama haklarını kaybetmektedir (Sart, 2013).



Şekil 3.13 "Büyük Veri," "Eтик*" ve "Büyük Veri ve Eтик*" konularındaki araştırma makalelerinin sayısı. (Kaynak: <https://www.nature.com/articles/s41599-020-00638-0>)

Veri paylaşım anlaşmalarının aynı zamanda açık veri, erişim kontrolü, hak yönetimi ve veri kullanım kontrolü için anonimleştirme içерerek veri koruma ve gizliliği için mekanizmalar ile bağlı olması gerekmektedir. Farklı kaynaklardan elde edilen veri sonucuya bireysel kimliklerden atfedilen konular dikkatli bir şekilde göz önünde bulundurulacak ve açık bir şekilde kabul edilip izin verilmiş olacaktır. Bunu yapmak muhtemelen bize veri paylaşımı ve veri hakları için bir model düzenlemeye izin verecek ve bu gelecekte evrensel yararlar sağlayarak büyük veri işbirlikleri tanımlayabilecektir.

Örneğin fikri mülkiyet hakları ile ilgili yapılan bir araştırmada katılımcıların %91'i elde edilen gelirin ne şekilde paylaşılacağına ilişkin net, şeffaf ve anlaşılır düzenlemelerin olmasının araştırmalardaki motivasyonu artıracağını ve çok daha fazla katılımcının bu tür aktivitelerde yer alacağını vurgulamaktadır (Sart, 2013). Bununla birlikte her geçen gün büyük veri ve etik üzerine sayısız makale yayınlanarak bu konuya olan vurgu daha da artmaktadır.

Büyük verinin temel etiklerini geliştirerek potansiyel zararlarına yenilmeden yeni teknolojinin faydalarından yararlanmaya çaba göstermeliyiz. Büyük veri etikleri herkes için gereklidir.

Konu Türü	Sorunun Nedeni	Olası Çözüm	CIO'lar ve CDO'ları Tarafından Karşı Karşıya Kalınan
Tedarik Zinciri Kaynak Bulma ve Kullanım Sorunları	Veri Yönetimi		
	Üst kaynakların ve alt müşterilerin davranışlarından sorumlu olmayan firmalar	Firmanın daha büyük tedarik zincirindeki rolünü gösterme Tedarik zinciri bilgilerinin makine tarafından okunabilir bildirimini politika yapıcılar, raporlar ve gizlilik savunucuları için erişilebilir hale getirme	Yukarı akış kaynaklarının ve alt müşterilerin/verilerin kullanımlarının tanımlanması ve sahipliğinin alınması Veri yönetimi uygulamalarılarındaki bilgilerin uzmanlara ve acemilere açık olduğundan emin olma
Negatif Haricilik Olarak Gözetim	Tedarik zinciri görünmüyordu	Tedarik zincirinin veri yönetimi uygulamalarını görür hale getirme	Veri ortaklarına, ister yukarı akış kaynaklarına ister alt kullanıcılarla açıklamayı engelleyen gizlilik anlaşmaları yapmama
	Veri İşlemi		
Tüketicisi Bilgisine Yönelik Yıkıcı Talep	Kişisel olarak tanımlanabilir bilgileri (PII) toplayan, depolayan veya kullanan firmalar tarafından yakalanmayanların başkalarına zarar vermesi	Gözetimi en aza indirme Daha fazla veri yasal süreci ile gözetim maliyetini içselleştirme	Takibi tüketiciye görür hale getirme (Sektör) PII alan ve elinde tutan firmalar için ek veri yasal süreci gerekliliği
	Veri Bütünlüğü		
	İkincil veri ticareti pazarı, birincil tüketici odaklı pazardan daha düşük kalite gereksinimlerine sahiptir	PII'yi işlerken veya satarken bir veri bütünlüğü uzmanı kullanma	(Endüstri) Büyük Veri kaynakları ve müşterileriyle ortaklık yapan projeler için veri bütünlüğü uzmanı veya yönetim kurulu
	İkincil pazar, birincil pazar (tüketiciler) tarafından görülmez	İkincil pazardaki etkinliği düzenleyiciler ve tüketiciler için görür hale getirme	İfşa yoluyla büyük veri için ikincil pazarda ortak olmanın getirdiği ek riskleri hesaba katma ve iletme

Şekil 3.14 Büyük veri sektörünün etik sorunlarına karşı olası çözümleri. (Kaynak: https://www.researchgate.net/figure/Possible-Solutions-to-the-Big-Data-Industries-Ethical-Issues_tbl2_273772472)

KAYNAKLAR

- 2017-Dresner Danışmanlık Hizmetleri. <https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/?sh=1f6e15f539a1>
- 2017-Dresner Advisory Services. <https://www.forbes.com/sites/louiscolumbus/2017/12/24/53-of-companies-are-adopting-big-data-analytics/?sh=1f6e15f539a1>
- “Comscore Releases “The 2009 U.S. Digital Year in Review,” https://www.comscore.com/Insights/Press-Releases/2010/2/comScore-Releases-2009-U.S.-Digital-Year-in-Review?cs_edgescape_cc=US
- Conway, D. ve Diego K., 2013, Innovation Patterns and Big Data, Big Data and Business Analytics, In: Liebowitz, J., 8, CRC Press, Boca Raton, ISBN: 978-1-4665-6579-1, 131 - 147.
- <https://www.mdpi.com/2071-1050/12/2/634/htm>
- https://www.researchgate.net/figure/An-example-of-a-Data-Model-for-a-fictitious-Rental-Car-company_fig1_277890728
- https://www.researchgate.net/publication/311495436_Identifying_and_Analyzing_the_Transient_and_Permanent_Barriers_for_Big_Data
- <https://blogs.sap.com/2017/04/04/grc-tuesdays-part-one-big-data-privacy-risks-and-the-role-of-the-gdpr/>
- https://help.sap.com/saphelp_scm70/helpdata/en/4c/b4e7d94c1f1921e1000000a15822b/content.htm?no_cache=true
- https://en.wikipedia.org/wiki/Big_data
- <https://www.allerin.com/blog/the-sensitive-relationship-between-big-data-and-iprs>
- <https://www.consultancy.uk/news/1946/mobile-and-big-data-analytics-key-tools-for-innovation>
- <https://www.youtube.com/watch?v=PsC9CMgyTxY>
- <https://www.nature.com/articles/s41599-020-00638-0>
- https://www.researchgate.net/figure/Possible-Solutions-to-the-Big-Data-Industrys-Ethical-Issues_tbl2_273772472
- Hurwitz, J., Alan N., Fern H., ve Marcia K., 2013, Big Data for Dummies, John Wiley & Sons, ISBN: 978-1-118-50422-2.
- Oracle Corporation White Paper, 2012, Oracle: Big Data for the Enterprise, <http://www>.

oracle.com/us/products/database/big-data-for-enterprise-51_9135.pdf, [Ziyaret Tarihi: 5 Mart 2015].

Sart, G., 2013, Fikri Mülkiyet Hakları Hakkında Üniversitedeki Araştırmacıların Algısı, 8. Ulusal Eğitim Yönetimi Kongresi (Okul Geliştirme Sürecinde Öğretmen ve Yönetimin Niteliği: Politika ve Uygulamalar, İstanbul, 7-9 Kasım 2013, 101-103.

Suther, Tim, Bill Burkart, Jie Cheng (2013). Big Data and Business Analytics. Auerbach Publications.

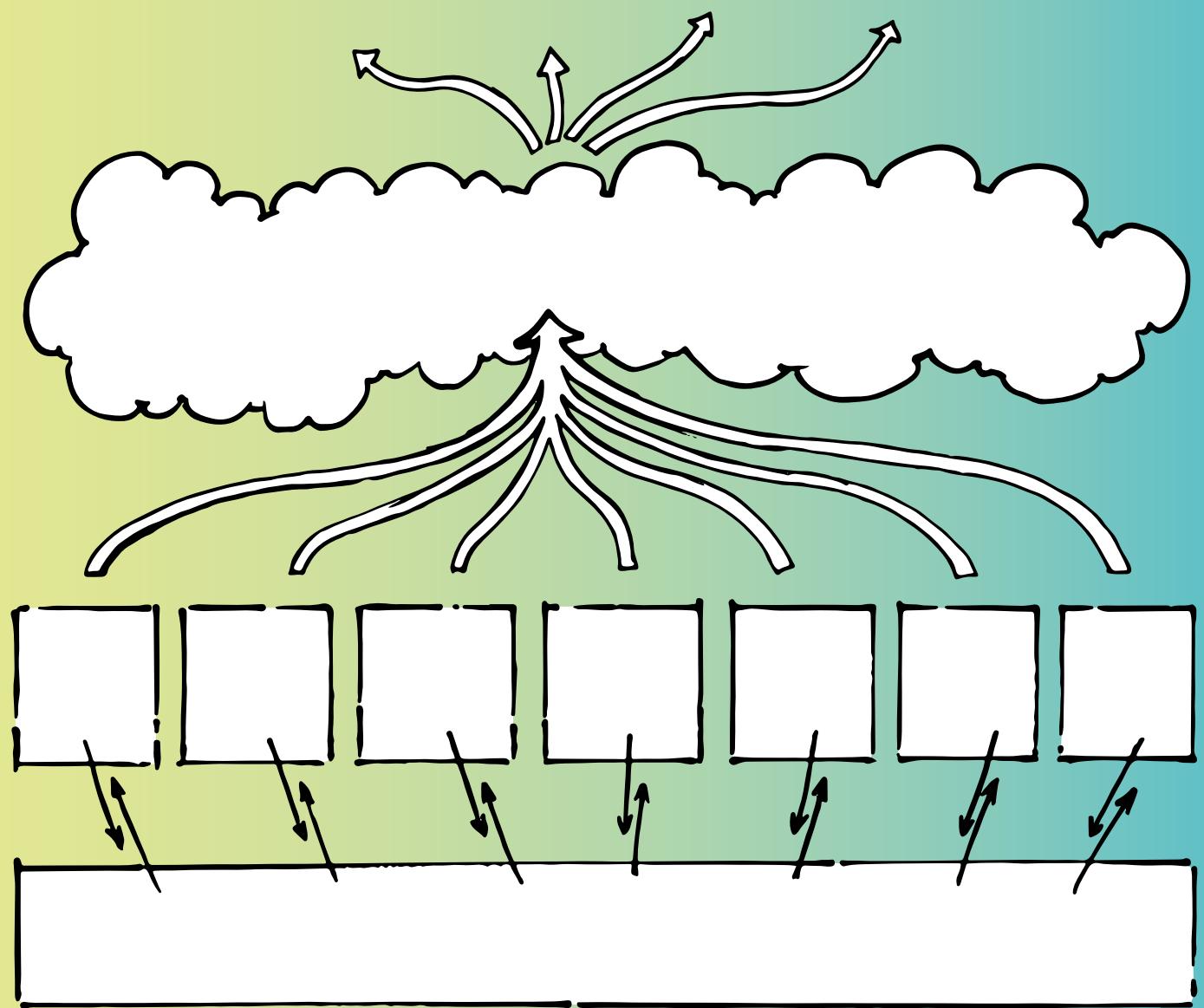
Thomas, I., 2013, Putting Big Data at the Heart of the Decision Making Process, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.

White House Report, 2014, Big Data: Seizing Opportunities, Preserving Values, Executive Office of the President, https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf, [Ziyaret Tarihi: 4 Mayıs 2015].

Zhao, D., 2013, Frontiers of Big Data Business Analytics: Patterns and Cases in Online Marketing, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.

4. BÖLÜM

BÜYÜK VERİ TEKNOLOJİSİ



4.1 BÜYÜK VERİ TEKNOLOJİSİNÉ GİRİŞ

Bugünkü anlamıyla büyük veri kavramının kökleri 2004 yılına dayanmaktadır, yani 2004 yılından beri dünyanın en büyük teknoloji devlerinin yatırım yaptığı ve kullandığı bir kavramdır. Büyük veri yeni bir olgu, yeni bir kavramdır ve bünyesinde açık kaynak kodlu yazılımlar barındırır. Bu yazılımların çekirdeğini Hadoop ismini verdigimiz bir dizi yazılımlar grubu kapsar. Bu yazılımlardan bazılarını şöyle listeleyebiliriz; Hadoop Core, HDFS (Hadoop Distributed File System-Hadoop Dağıtık Dosya Sistemi), Hive (Data Warehouse-Veri Ambarı), HBase, ZooKeeper, Oozie, Mahout, Sqoop, Cloudera Manager. Büyük veri bu ve benzeri araçlarla analiz edilerek kullanılabilir, sonuçlar elde edilmeye çalışılır.

“Büyük veri” terimi geleneksel veri sistemleri için çok büyük olan veri setlerinin toplanması, yönetimi ve analizini mümkün kılan teknolojileri açıklamak için kullanılmaktadır (Dumbill, 2012). Son zamanlarda iş dünyası yeni karar alma olanaklarına imkân tanıyan büyük veri araçlarını tartışmaktadır (Provost ve Fawcett, 2013). Var olan veri tabanı sistemlerinin sınırlarını ele almak için büyük veri teknolojileri çok büyük ölçekte paralel hesaplama yaklaşımlarını kullanmaktadır. Dağıtık veri işleme uzun bir geçmişi sahiptir (Provost ve Kolluri, 1999).



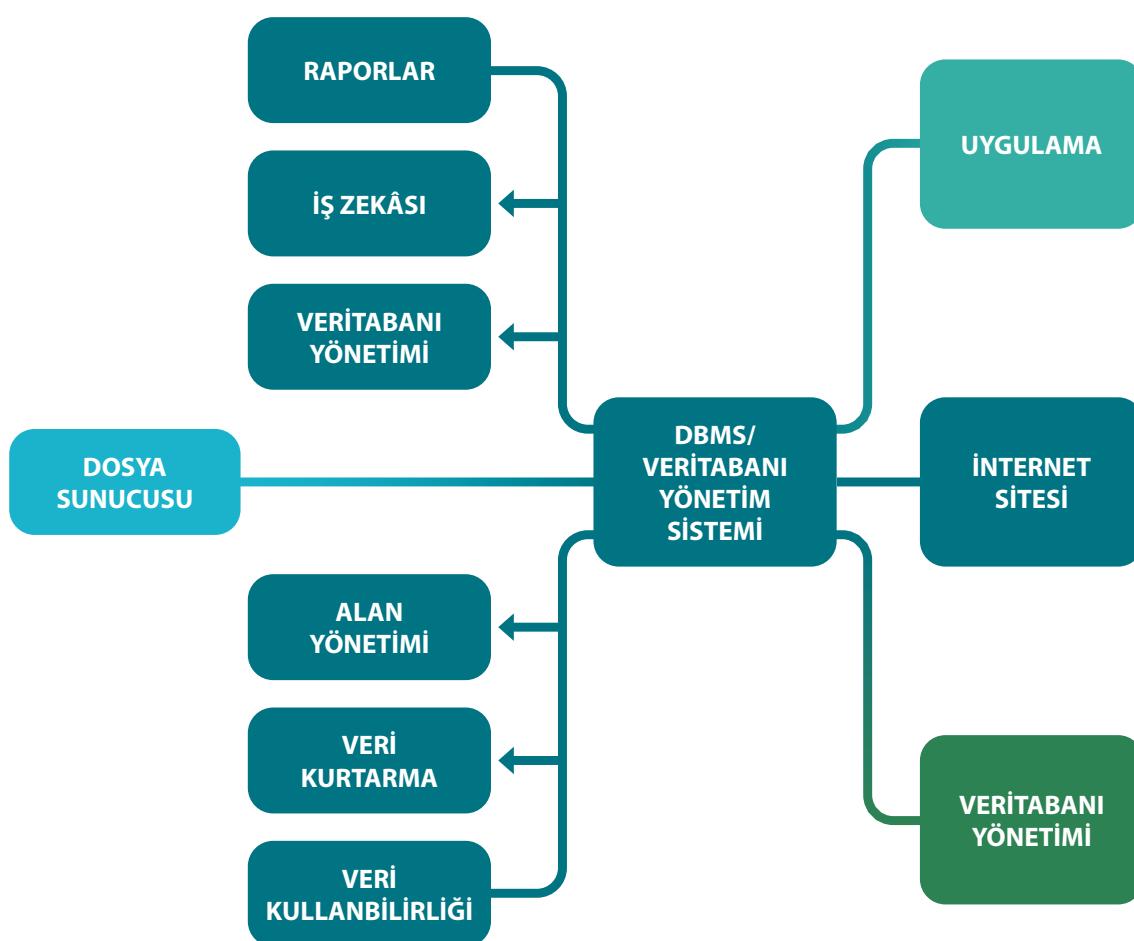
Artan Büyük Veri Teknolojileri

01	Apache Hadoop	09	Apache Storm
02	Apache Spark	10	Apache Hive
03	MongoDB	11	Apache Pig
04	Cassandra	12	Presto
05	Apache Kafka	13	Apache Flink
06	QlikView	14	Apache Sqoop
07	Qlik Sense	15	Rapidminer
08	Tableau	16	KNIME

Şekil 4.1 En iyi büyük veri teknolojileri. (Kaynak: <https://techvidvan.com/tutorials/big-data-technologies/>)

Son yıllarda veri toplama oranı ve ölçüği veri toplama teknolojilerindeki yeniliklerle gelirlerini artırdı. Yeni veri teknolojilerinin en son dönem dalgasının kaynakları 2004 yılında dağıtılmış veri işlemeyi desteklemek için büyük veri algoritmaları kullanmaya başlayan Google çalışanlarının izlerini taşıyabilir. Büyük veri analizleri için yaygın bir şekilde kullanılan yazılım platformu Apache Hadoop Map/Reduce iskeletinden çıkarılmış, Java programlama dilinde uygulanmıştır ve özgür bir şekilde açık kaynak altında dağıtılmaktadır. Bu açık kaynak projesi Hadoop kümelenme arayüzünün, iletişimiminin ve işlem akışının farklı bölümlerini ele alan Cassandra, Pig, Hive, HDFS gibi birçok alt projeye sahiptir. Büyük veri altyapısı bilgisayar kümeleri üzerindeki veri çevresi ve bu yazılımın uygulamasını gerektirmektedir. Çünkü büyük veri işlemlerini desteklemek için gereken hem donanım hem yazılım şirketleri için mümkündür. Buradaki en masraflı kısım büyük veri sistemlerini uygularken kurulum, bakım ve işi kolaylaştıran uzman edinimidir.

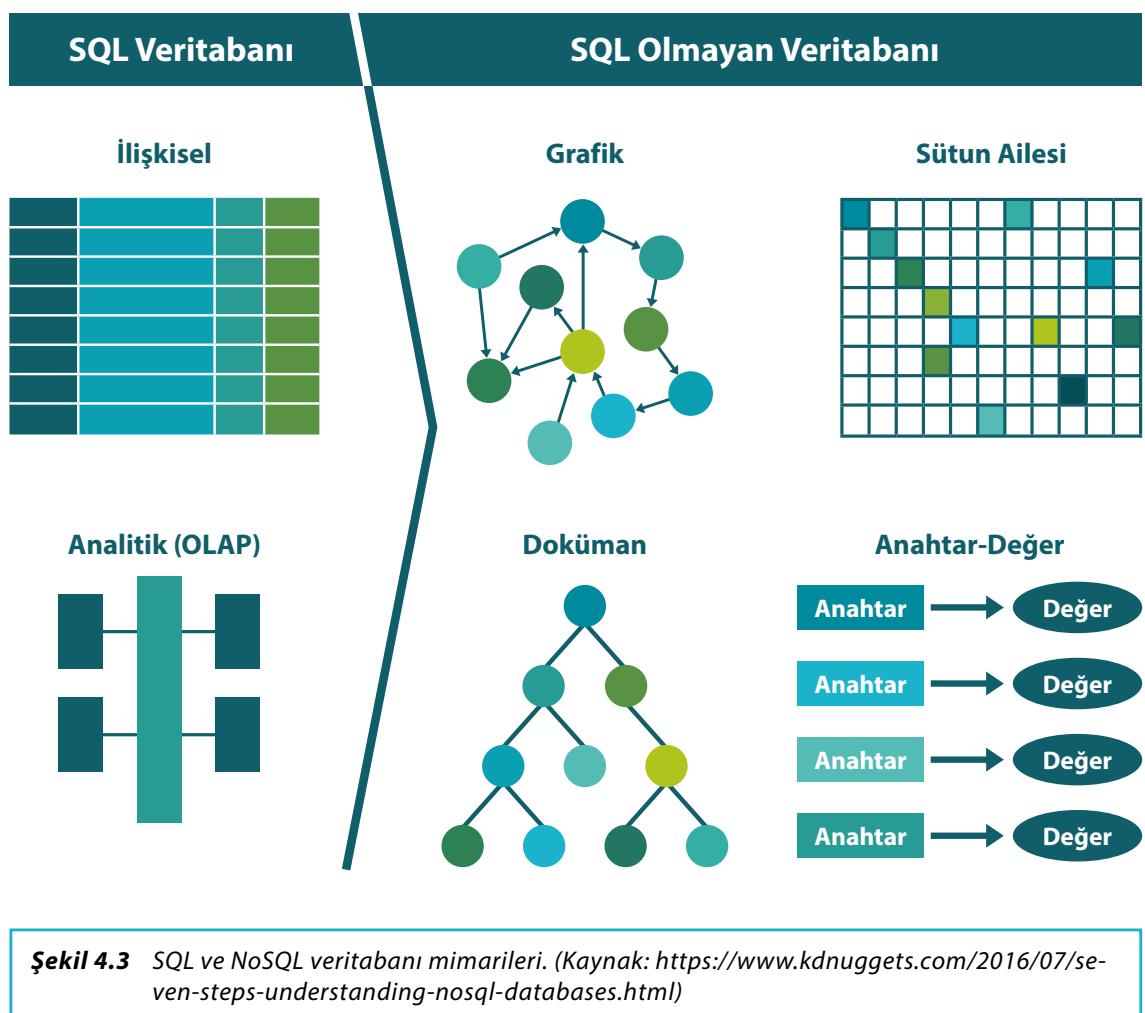
Büyük miktarda verinin analiz edilmesi için oluşturulan raporların çalışma süresi çok vakit almaktadır. Bu sorunun ortadan kalkması için yazılım ve donanım üreticileri bir araya gelerek çözümler üretmeye başladılar. SAP'ın vermiş olduğu bir örnek bize bir fikir verebilir. Elimizde iki milyar müşteri hareket verisi olsun. Biz bu veri içinde zamanında ödeme yapmamış olan müşterileri sorgulayan bir rapor oluşturalım. Bu rapor normalde 70 kusur dakika sürerken SAP'ın HANA çözümü ile 13 saniyeye inerek 350 kat hızlanabiliyor. Bu, ölçeklenebilir platformun büyük miktarda veriyi gerçek zamanlı şekilde analiz etmesini sağlıyor.



Şekil 4.2 Veri tabanı yönetim sistemlerinin çalışma şekli. (Kaynak: <https://www.zeltisinfotech.com/products-solutions/database-management-system.html>)

Eğer programcı bilgiyi hard disk değil de RAM üzerinden alıp işleyebilirse bu hız sağlanabiliyor. SSD'ler hard disk dünyasına yeni bir boyut açtı. Ama büyük veriler için disk teknolojisi yetersiz kalmaktadır. Sizin veriyi diskten alıp RAM üzerinde tutmanız ve yeri geldiğinde oradan veriyi alıp işlemeniz gerekiyor. Örnek vermek gerekirse Google'in veri indekslerini bildiğimiz veri tabanı yönetim sistemi ile tutmaya kalkması pek mümkün değildir. Büyük veri ile veri tabanı sistemleri ve uygulama geliştirme dilleri değişmektedir. Hadoop ve NoSQL veri tabanları günümüzde öne çıkmaktadır.

Yazılım ve donanım şirketleri artık beraber çözümler üretmeye başladı. Bugün büyük veri konusuna yazılımsal olarak en fazla çalışan firmalar arasında IBM, donanımsal olarak ise EMC görünümektedir. Büyük şirketlerin büyük veri için yapmış



Büyük Veri Teknolojisi

GÖRSELLEŞTİRME & ANALİZ	
HESAPLAMA	
DEPOLAMA	
DAĞITIM & VERİ AMBARI	

Şekil 4.4 Büyük veri teknolojisi. (Kaynak: <https://dashbouquet.com/blog/web-development-guide-for-freelance-skills-to-land-a-project-in-2019>)

olduğu çözümlerden de kısaca bahsedersek SAP HP, Fujitsu ve IBM ile beraber geliştirdiği HANA çözümü tanıttı. SAP'nin In-Memory teknolojisi olan HANA (High Performance Analytic Appliance) çözümü ile veri ana bellekte tutuluyor ve alınan raporlar hızlı ve çevrimiçi bir şekilde olabiliyor.

Microsoft HP ile oluşturdukları kutu çözümleri ile özellikle SQL sunucusu hızlandırmayı düşünüyor. Oracle EMC ile geliştirdikleri çözüm ile özellikle sanal makine üzerinde çalışan SAP sistemlerini hızlandırıyor. Ayrıca SUN ile işbirliği yapıyor. EMC bu alanda en fazla yatırım yapan şirketlerdendir. Isilon, OneFs ve Scale-Out Nas çözümlerini örnek olarak gösterebiliriz.

IBM ise bu alanda en öncü şirkettir. Netezza ve IBM DB2 Analytics Accelerator ürünlerini bu kapsamında düşünebiliriz. IBM bu alanda müşterilerine en üst seviyede büyük veri çözümleri sunan firmalardan biridir. Ayrıca Thomson Reuters firması da büyük veri alanında çalışmalar yapmaktadır. Hatta 4 Eylül 2015 tarihli blog yazısında şu soruyu sormaktadır: "How will we feed 9 billion people in 2050?" yani "2050 yılında 9 milyar insanı nasıl besleyeceğiz?" Bununla birlikte bu konuda

<http://9billionbowls.com/> adındaki sitede bir rapor hazırlamıştır. Bunu bir problem olarak ifade ederek yanıtını "... The answer is data," yani "Cevap veridir," şeklinde vererek büyük veri konusundaki projesini ortaya koymuştur.

Artık veri hacimlerini kontrol altına alan yeni teknolojiler ile yüksek hızlı ve riden faydalananarak kuralçı analizler ve analistik teknolojiler ile kademe kademe genişleyen çeşitli veri tiplerinin analiz edilmesi aracılığıyla büyük veri yeni bir standart olmaktadır. Ya da başka bir deyişle büyük veri her şeyin ayrılmaz bir parçası haline gelecektir.

4.2 BÜYÜK VERİ ANALİZ YAZILIMI: HADOOP

Apache Hadoop hızlı gelişen bir büyük veri işleme platformudur. Cloudera'nın mimarı ve Apache Yazılım Vakfı'nın Başkanı Doug Cutting tarafından geliştirilen Apache Hadoop web'in patlamasından oluşan veri için gereklilikten ortaya çıktı ve veriyi ele alan geleneksel sistemlerin çok ötesinde büydü. Hadoop başlangıçta



Şekil 4.5 Hadoop'un avantajları. (Kaynak: <https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/>)



Doug Cutting, yaklaşık on yıl önce Yahoo'da çalışırken Apache Hadoop'un oluşturulmasına yardımcı oldu ve burada projenin açık kaynak kodlu bir ezici güç ve büyük veri hareketinin sembolü olmasına yardım etti. 2009 yılında Cutting, Cloudera'ya baş mimar olarak katıldı ve Hadoop'un haberini verdi. Apache projesine liderlik etmeye ve teknolojiyi ana akım kuruluşlar tarafından tüketilebilir hale getirmeye devam ediyor.

Doug Cutting, San Francisco, California, 2017.

(Kaynak: <https://www.facesofopensource.com/doug-cutting/>)

Google tarafından yayınlanan veri çığını ele alan taslak makalelerden esinlenmişti ve o zamandan beri filen yüzlerce terabiyat, hatta petabayt veriyi depolama, işleme ve analiz için standart halini aldı.

Apache Hadoop %100 açık kaynaktır ve temel olarak veri depolama ve işlemenin yeni şekli olarak öncülük etmiştir. Veriyi depolamak ve işlemek için pahalı, kişiye özel farklı sistemlere güvenmek yerine, Hadoop ucuz bir şekilde çok büyük devasa miktarda verinin dağıtılmış paralel işlenmesine izin verir. Büyük veri analiz sunucuları hem depolama hem işleme üzerine sınırsız ölçekler sunabilir.

Her gün çok daha fazla verinin olduğu hiper-bağılı dünyamızda, Hadoop'un çığır açan avantajları şu anlama gelmektedir: İşletmeler ve organizasyonlar işe yaramaz kabul edilen verideki değeri şimdi bulabilirler.

Hadoop farklı sistemlerden her tipteki veriyi işleyebilir; yapılandırılmış, yapılandırmamış, günlük dosyaları, resimler, müzikler, iletişim kayıtları, e-postalar. Farklı veri tipleri bağlantısız sistemlerde saklansalar bile, daha önceden bir şemaya ihtiyaç duymadan Hadoop kümeleri içerisinde bunları depolamak mümkündür.

Tüm veri kullanılabilir yapılarak Hadoop düzenlenmemiş ilişkiler saptamak için destek sağlar ve yanıtları açığa çıkarır. Ayrıca Hadoop'un miras sistemler üzerindeki maliyet avantajı veri ekonomilerini yeniden tanımlıyor. Eski sistemler, belirli bir iş yükleri için iyiken büyük verinin ihtiyaçları için tasarlanmamıştı ve bugünün çok büyük veri setleri için çok pahalıdır.

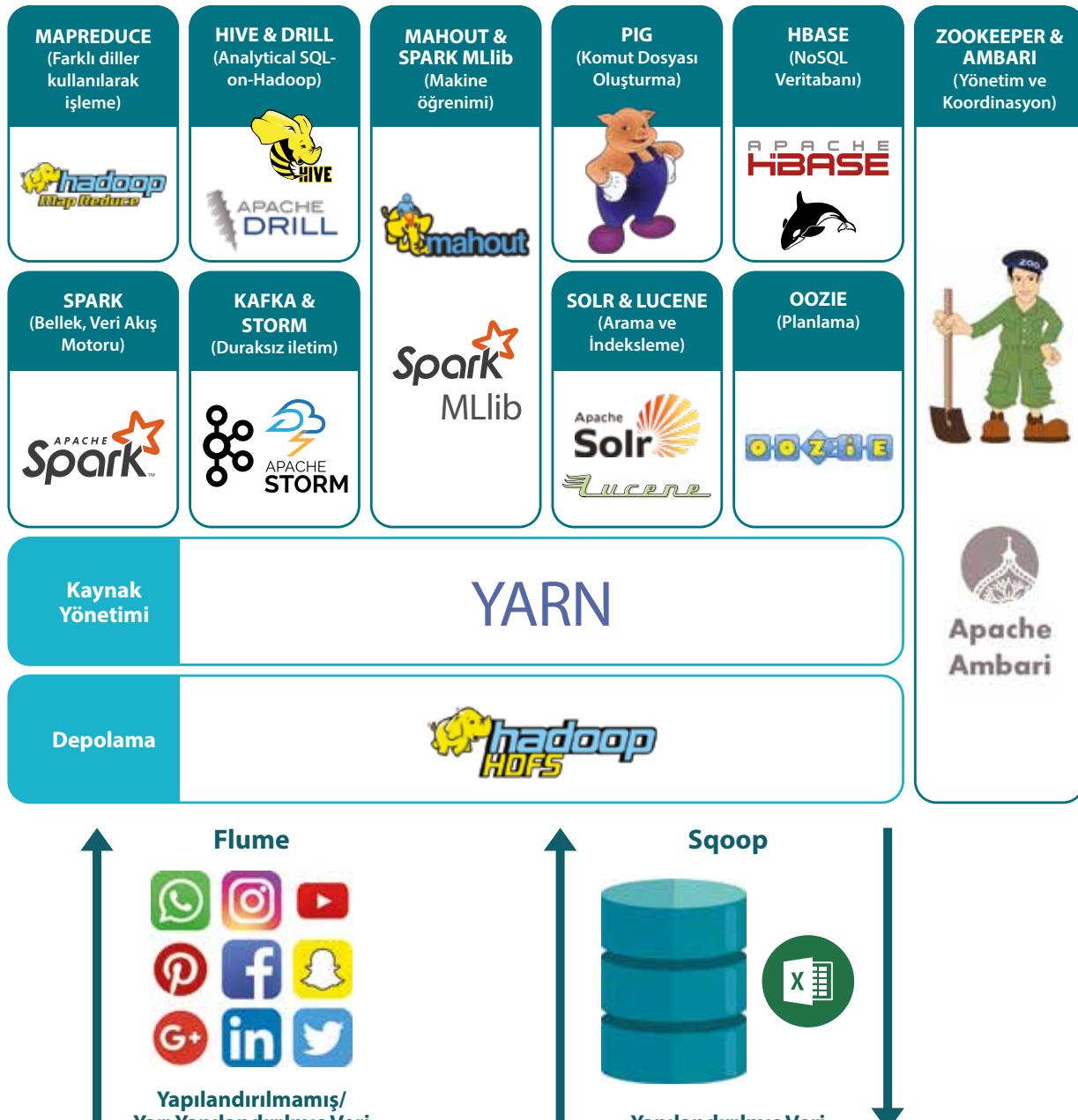
Hadoop, sıradan sunuculardan oluşan küme (cluster) üzerinde büyük verileri işlemek amaçlı uygulamaları çalıştırın ve Hadoop Distributed File System (HDFS) olarak adlandırılan bir dağıtık dosya sistemi ile Hadoop MapReduce -büyük veri setleri ile yapılacak işlemlerin birden fazla iş birimine dağıtılmasını sağlayan yöntem- özelliklerini bir araya getiren Java ile geliştirilmiş açık kaynaklı bir kütüphanedir.

Hadoop projesi büyük verileri işleme konusundaki diğer projelere bir çatı görevi görüyor. Hadoop projesinin altında Avro (veri dizileştirme sistemi), Cassandra (yüksek erişilebilir, ölçeklenebilir NoSQL veritabanı), HBase (Hadoop üzerinde çalışan, büyük veriler için ölçeklenebilir, dağıtık NoSQL veritabanı), Hive (büyük veriler üzerinde iş zekâsı sistemi), Mahout (ölçeklenebilir yapay öğrenme ve veri madenciliği kütüphanesi), Pig (paralel hesaplamlar için yüksek düzeyli bir veri akışı dil ve yürütme kütüphanesi), ZooKeeper (dağıtık uygulamalar için yüksek ölçekli koordinasyon uygulaması) projeleri geliştiriliyor.

Hadoop Pig, Hive ve Zookeeper gibi bir Apache projeleri ekosistemi tarafından desteklenmektedir. Bu Hadoop'un değerini artırmakta ve kullanılabilirliğini genişletmektedir. Fiyat-etki, ölçeklenebilirlik ve elverişli mimarisinden dolayı, Hadoop göze çarpan özelliklerinin etkisiyle büyük ölçüde bilgi işlem dinamiklerini ve ekonomisini değiştirmektedir.

Daha yalın bir dille anlatmak gerekirse Apache Hadoop; HDFS ve MapReduce bileşenlerinden oluşan açık kaynak kodlu bir yazılımdır.

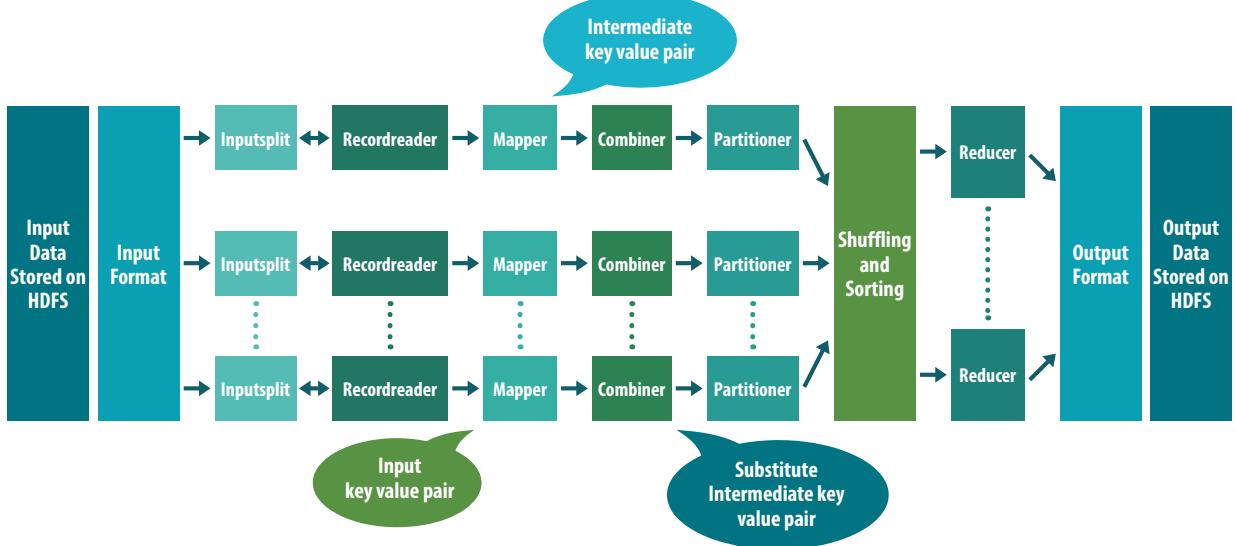
MapReduce: Bir kümedeki (cluster) düğümlere iş atayan ve anlayan iskeletir. 2004 yılında Google tarafından tanımlanmıştır ve binlerce düğüm arasındaki iş yüklerini dağıtmayı mümkün kılmıştır. “Daha küçük alt problemler içerisindeki problemi ayırmak” stratejisi üzerine temellendirilmiştir ve SQL yoluyla ve SQL-tabanlı BI (iş zekâsı) araçları içerisinde ortaya çıkabilir. MapReduce operasyonel olarak büyük veri setleri ile yapılacak işlemlerin birden fazla iş birimine dağıtılmasını sağlayan yöntemdir. Buradaki büyük veri setleri birkaç bilgisayarın bile işleyemeyeceği kadar verinin toplandığı setlerdir. Bu setler üzerindeki işlemler serisi çeşitli birimlere dağıtılmakta ve sonra çıktıları birleştirilip sonuç üretilmektedir. Adından da anlaşılacağı üzere iki katmandan oluşmaktadır: Map ve Reduce. Map katmanı; isim-değer çiftlerini girdi olarak alır, yapılacak işlemi gerçekleştirir. Ürettiği sonuç



Şekil 4.6 Hadoop ekosistemi. (Kaynak: https://www.bogotobogo.com/Hadoop/BigData_hadoop_Ecosystem.php)

listesini ise girdideki isim ile birlikte çıktı olarak verir. Reduce katmanı ise Map katmanındaki sonuç listelerini toplar ve tek sonuca indirgeme işlemini yapar. Bir den fazla iş birimi Map işiyle uğraşırken, bir sonraki katman sonuçları toplayıp tek sonuç haline getirir.

Hadoop MapReduce'un Çalışma Biçimi



Şekil 4.7 Hadoop MapReduce'un Çalışma Biçimi. (Kaynak: <https://techvidvan.com/tutorials/how-mapreduce-works/>)

Hadoop Dağıtık Dosya Sistemi (HDFS): Veri depolama için bir Hadoop kümesi içindeki tüm düğümleri yayan bir Apache açık kaynak dağıtık dosya sistemidir. Büyüyük bir dosya sistemi oluşturmak için birçok yerel sistem düğümü üzerindeki dosya sistemlerini birbirlerine bağlar. HDFS düğümlerin başarısız olacağını varsayar, bu yüzden birden fazla düğüm üzerinden veri çoğaltarak güvenilirliği sağlar.

HDFS sayesinde sıradan sunucuların diskleri bir araya gelerek büyük, tek bir sanal disk oluşturur. Bu sayede çok büyük boyutta birçok dosya bu dosya sisteminde saklanabilir. Bu dosyalar bloklar halinde (varsayılan 64MB) birden fazla ve farklı sunucu üzerine (varsayılan 3 kopya) dağıtılarak RAID diskler arasında veri kopyalama veya paylaşımı için birden fazla sabit diski kullanarak yapılan veri depolama tasarısı benzeri bir yapıyla yedeklenir. Bu sayede veri kaybı önlenmiş olur. Ayrıca HDFS çok büyük boyutlu dosyalar üzerinde okuma işlemi (streaming) imkânı sağlar, ancak rastlantısal erişim (random access) özelliği bulunmaz.

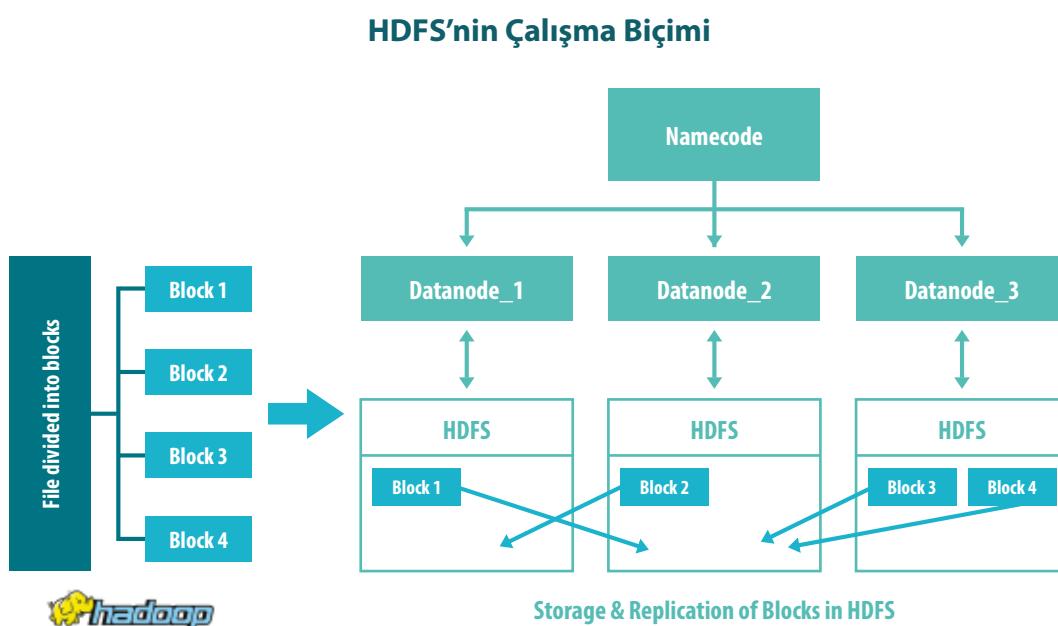
HDFS; NameNode ve DataNode süreçlerinden (process) oluşmaktadır:

NameNode ana (master) süreç olarak blokların sunucular üzerindeki dağılımdan, yaratılmasından, silinmesinden, bir blokta sorun meydana geldiğinde yeniden

oluşturulmasından ve her türlü dosya erişiminden sorumludur. Kısacası HDFS üzerindeki tüm dosyalarlarındaki bilgiler (metadata) NameNode tarafından saklanır ve yönetilir. Her kümede yalnızca bir adet NameNode olabilir.

DataNode ise, işlevi blokları saklamak olan işçi (slave) süreçtir. Her DataNode kendi yerel diskindeki veriden sorumludur. Ayrıca diğer DataNode'lardaki verilerin yedeklerini de barındırır. DataNode'lar küme içerisinde birden fazla olabilir.

Hadoop MapReduce ise HDFS üzerindeki büyük dosyaları verileri işleyebilmek amacıyla kullanılan bir yöntemdir. İstediğiniz verileri filtrelemek için kullanılan Map fonksiyonu ve bu verilerden sonuç elde etmenizi sağlayan Reduce fonksiyonlarından oluşan program yazılıktan sonra Hadoop üzerinde çalıştırılır. Hadoop Map ve Reduce'lerden oluşan iş parçacıklarını küme üzerinde dağıtarak aynı anda işlenmesini ve bu işler sonucunda oluşan verilerin tekrar bir araya getirilmesinden sorumludur. Hadoop'un gücü, işlenen dosyaların her zaman ilgili düğümün (node) yerel diskinde okunması ile ağ trafiğini meşgul etmemesinden ve birden fazla işi aynı anda işleyerek doğrusal olarak ölçeklenmesinden geliyor diyebiliriz.



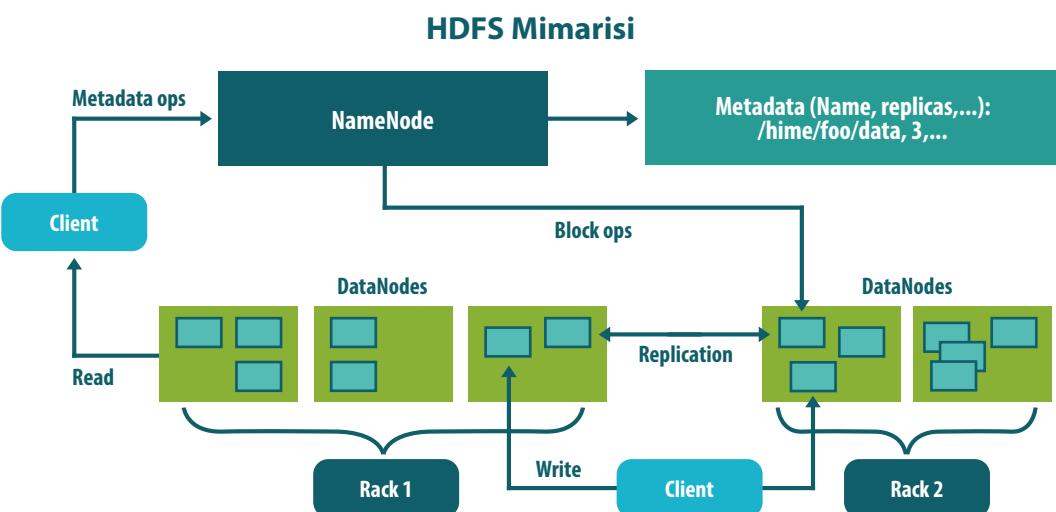
Şekil 4.8 HDFS'nin Çalışma Biçimi. (Kaynak: <https://factspan.com/hadoop-distribution-file-system/>)

MapReduce, JobTracker ve TaskTracker süreçlerinden oluşur. JobTracker yazılan MapReduce programının küme üzerinde dağıtılarak çalıştırılmasından sorumludur. Ayrıca dağıtılan iş parçacıklarının çalışması sırasında oluşabilecek herhangi bir problemde o iş parçacığının sonlandırılması ya da yeniden başlatılması da JobTracker'ın sorumluluğundadır. TaskTracker, DataNode'ların bulunduğu sunucularda çalışır ve JobTracker'dan tamamlanmak üzere iş parçacığı talep eder. JobTracker, NameNode'un yardımıyla DataNode'un yerel diskindeki veriye göre en uygun Map işini TaskTracker'a verir. Bu şekilde verilen iş parçacıkları tamamlanır ve sonuç çıktısı yine HDFS üzerinde bir dosya olarak yazılarak program sonlanır.

Hadoop şu anda Yahoo, Amazon, eBay, Facebook, LinkedIn gibi birçok lider firmada büyük verileri analiz etmek amacıyla kullanılıyor.

IBM'de veri analizi yaparken entegre sistem ile birçok imkân sunmaktadır. IBM Big Data Platformu olarak sunduğu hizmetler temel olarak Hadoop sistemi, akışkan veri işleme ve veri ambarlarından oluşmaktadır.

<http://wiki.apache.org/hadoop/PoweredBy> sitesinden Hadoop kullanan tüm kuruluşlara ve verdikleri hizmetlere ulaşmak mümkündür. Hadoop projesini geliştiren, bu konuda eğitim ve danışmanlık hizmetleri sunan Hortonworks, Cloudera, MapR gibi firmalar mevcuttur.



Şekil 4.9 HDFS Mimarisi. (Kaynak: <https://factspan.com/hadoop-distribution-file-system/>)

ŞİRKET	SİSTEM
IBM	Apache Hadoop, InfoSphere
Cloudera	CDH, Cloudera Standard, Cloudera enterprise
Oracle	Oracle Big Data Appliance
Google	BigTable
Yahoo!	Sherpa
Amazon	SimpleDB
Microsoft	Dryad
Facebook	apache Cassandra
Hypertable	HyperTable
ASF	Apache CouchDB

Şekil 4.10 Büyük veri yönetim sistemleri. (Kaynak: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7293298, 2015)

4.3 BÜYÜK VERİ ANALİZİ TEKNOLOJİSİ

Tipik bir büyük veri işleme sistemi toplama ve ön işleme, depolama, analiz, madencilik ve değer uygulama aşamalarını içermektedir.

Şekil 4.11'de bir büyük veri sisteminin mimarisi gösterilmektedir. En alttaki veri kaynakları tabakasında işletmelerden, endüstri kuruluşlarından, internet ve internete bağlı aygılardan veriler gelir. Bunun bir üstündeki veri toplama tabakasında toplanan veri ön işleme sürecinden geçirilir. Bu ön işleme süreci veri temizleme ve heterojen veri işleme sürecini içerir. Sonraki veri depolama katmanında yapılandırılmış, yapılandırılmamış ve yarı-yapılandırılmış veri depolanmakta ve yönetilmektedir. Veri işleme katmanında ise veri analiz edilir ve veri madenciliği yapılır ki kullanıcılar yaygın telekomünikasyon ve internet servisleri gibi hizmetleri analiz eder.

4.3.1 Toplama ve Ön İşleme

Büyük miktarlardaki yığın verinin biçimini değiştirme masraflıdır ve buna veri toplamanın zorluğu eklenir. Geleneksel veri toplama araçlarının modası geçmiştir ve çoğu



Şekil 4.11 Büyük veri mimarisi. (Kaynak: http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

internet şirketi kendi büyük veri toplama sistemlerine sahiptir. Bu sistemlerin bazı örnekleri Apache Chukwa, Facebook Scribe, Cloudera Flume ve LinkedIn Kafka'dır.

Temizleme ve ayıklama teknolojileri ağlardaki zararlı, gereksiz ve kullanışız yığın veriyi temizlemek ve analiz için kaliteli veriyi ortaya çıkarmada kullanılır. Hadoop veri temizleme, dönüştürme ve yüklemeyi iyileştirmek ve paralel verinin ön işleme sürecini geliştirmek için kullanılmaktadır. Büyük veri toplama ve ön işleme teknolojileri çok kaynaklı ve çok modlu verinin kalitesini optimize etmek için tasarlanmıştır. Bu teknolojiler çoklu veri entegrasyonu için gerekli zorunlulukları yerine getirir. Bilgi içerisindeki yüksek kaliteli veriyi dönüştürür, farklı kaynaklardan elde edilen verinin kalitesini kontrol eder ve veri analizi için temeli oluştururlar.

4.3.2 Depolama

Daha önce de bahsettiğimiz gibi depolanan büyük verinin %80'i yapılandırmamış ve yarı-yapilandırılmış veriden oluşmaktadır. Şu anda büyük veri depolama belirli yöntemler kullanılarak ölçülebilir olan ve masraflı olmayan X86 sunucu küme sistemleri üzerinde yapılandırılmaktadır.

Yapilandırılmamış veri dağıtık dosya sistemlerinde depolanmaktadır. Geleneksel ağ dosya sisteminin (NFS) aksine, bir dağıtık dosya sistemi ölçeklenebilirliği geliştirmek için kontrol akışlarını veri akışlarından ayırrı. Bir dağıtık dosya sisteminde veri seti sunucuları tüm veri setlerini ve veri depolama sunucularının küme bilgilerini yönetir. Örneğin Google File System (GFS), Hadoop Distributed File System (HDFS), Lustre ve Ceph sistemlerinin kapasiteleri 10, hatta 100 petabaytlara kadar çıkabilir. Bu depolama sistemleri POSIX (UNIX için Taşınabilir İşletim Sistemi Arabirim) arayüzleri üzerinden dosya erişimini sağlar.

Eğer bir veri seti sunucusu bu tür bir depolama sisteminde başarısız olursa, tüm dosya sistemi kullanıcılarla hizmet sağlamada başarısız olur. Bir tek düğüm veri seti sunucusunun işleme ve depolama kapasitesi de oldukça sınırlıdır. Sistemdeki veri trafik miktarı artarken veri seti sunucularının işleme kapasitesi sistem ölçeklenebilirliği için bir dar boğaz halini alır.

NoSQL Data Stores

DEĞER DEPOSU	Anahtar / değer çiftlerinin toplanması Anahtar aracılığıyla veri erişimi: get (key), put (key, value)	 redis	 Amazon DynamoDB *
GENİŞ SÜTUN DEPOLARI	(Birçok) dinamik sütuna sahip kayıtlara sahip tablolar Anahtar aracılığıyla erişim, SQL benzeri sorgu dili...	 cassandra	 Apache HBase  Cloud Bigtable
METİN DEPOSU	Belgelerde yarı-yapılardırılmış veriler (örn. JSON) Anahtar veya basit API / sorgu dili aracılığıyla erişim	 mongoDB	 Couchbase  CouchDB relax
GRAFİK VERİTABANI	Özelliklerle düğümler ve kenarlar olarak veriler Veritabanı sorguları dahil grafik algoritmaları	 neo4j	 OrientDB *  TITAN

*çoklu model

Şekil 4.12 NoSQL Data Stores. (Kaynak: <https://www.slideshare.net/AnikaGro/nosql-data-stores-for-big-data>)

Veri seti sunucuları bundan dolayı aktif/uyku modunda çalıştırılmaktadır. Normal koşullarda aktif veri seti sunucusu tüm talepleri yerine getirir, tüm dağınık dosya sistemlerini yönetir ve düzenli bir şekilde senkronizasyon için uyku halindeki veri seti sunucusuna veri gönderir. Eğer aktif sunucu başarısız olursa, uyku modundaki sunucu devreye girer ve kesintiye uğramadan tüm faaliyetler devam eder. Uyku modundaki sunucuya gönderilmeyen veri kaybolabilir. Bu modu kullanmak tek nokta hataları için çözüm olabilir, fakat sistem ölçeklenemez. Bu yüzden yüksek derecede ölçeklenebilir veri sunucu kümlesi bir dağınık dosya sisteminde hayatı öneme sahiptir.

Büyük verinin çoğu yapılandırılmış düz metin veri olan yarı-yapılardırılmış veriyi içerir, fakat ilişkisel veri tabanlarından daha esnek modellere sahiptir. Yarı-yapılardırılmış veri katı veri tabanı işlemlerini gerektirmez. Çoğunlukla basit tablo sorgusundan oluşur ve bazı durumlarda düşük tutarlılık gereksinimlerine sahiptir. Bundan dolayı veri tabanı işlem yönetimi ağır yüklü veri tabanları içindeki bir yüktür. NoSQL veritabanı geniş bir alanda diğerleriyle konuşan, ilişkisel veritabanı arasında bağlantı kuran ilişkisel olmayan veri tabanıdır. NoSQL sabit bir tablo yapısı ya da bağlantılı işlemler gerektirmez, bu durum büyük veriye erişimi açısından önemli avantajlar sağlar.

NoSQL veri tabanlarında ilişkisel veri tabanı modelleri dağınık yatay genişlemeyi ve büyük veriyi destekleyen bir özgür-şema prensibi sayesinde terk edilmiştir. Dynamo, BigTable, Cassandra ve MongoDB’yi kapsayan birçok NoSQL veritabanı ürünleri ve açık kaynak projeleri vardır. NoSQL sistemleri henüz olgunlaşmamıştır: Birçoğu küçük ticari desteği sahip açık kaynak projeleridir. NoSQL sistemlerinin birleşik uygulama arayüzleri eksiktir ve SQL desteklemezler.

4.3.3 İşleme

Karışık yükler ile büyük veri setlerini işleme süreci karmaşıktır ve Şekil 4.13’te görünen belirli işleme gereksinimlerine ihtiyacı bulunmaktadır.

Yapılardırılmamış büyük veri temel olarak dağınık bilgisayar mimarisi kullanılarak işlenmektedir. Farklı hesaplama modelleri üzerine yapılandırılmış 3 tip dağınık bilgisayar mimarisi bulunmaktadır.

Büyük Veri İşleme Gereksinimleri	
Yüksek Ölçeklenebilirlik	Dağınlıklaştırma ölçeklenebilirliği, geniş ölçekli paralel veri işlemeyi destekleme
Yüksek Performans	Yığın veri sorgulamasına hızlı yanıt ve analiz gereksinimleri
Düşük Maliyet	Evrensel donanım sunucuları üzerine yapılandırma, yüksek fiyat/performans oranı
Hata Toleransı	Donanım güvenilirliğinin bağımsızlığı, bazı operasyonlar (diğerlerinden ayrı olanlar) hata durumunda çalışmalıdır
Kullanım Kolaylığı ve Açık Arayüzler	Geleneksel SQL ara yüzleri ile uyum, sorgulama ve çok boyutlu veri analizini destekleme
Geçmişe Yönelik Uyumluluk	Geleneksel BI (Business Intelligence-İş Zekâsı) araçlarını destekleme

- **MapReduce Modeli:** Bu model (giriş anahtarı, giriş değeri) → (çıkış anahtarı, çıkış değeri) olarak ifade edilmektedir. Çıkış anahtarı/değer çiftleri işlenmektedir, giriş anahtarı/değer çiftleri MAP ve Reduce fonksiyonları kullanılarak üretilmektedir. Bu model basit bir mantığa sahiptir ve yaygın olarak kullanılmaktadır.
 - **Bulk Synchronous Parallel (BSP) Modeli:** Bu basit bir bilgisayar modeline benzeyen tekrarlamalı hesaplama modelidir. Bununla beraber farkı iletişim içerisinde yatomaktadır. Bu modelde tüm düğümler her hesaplama turundan sonra senkronize edilmektedir. Bu tekrarlamalı senaryolar için uygundur. Örneğin Google Pregel BSP mimarisi üzerinde yapılandırılmıştır.
 - **Directed Acyclic Graph (DAG) Modeli:** Bu model karmaşık hesaplama süreçlerini ve aralarındaki ilişkileri açıklama için DAG kullanır. Microsoft, Dryad projesinde bu hesaplama modelini kullanmaktadır.

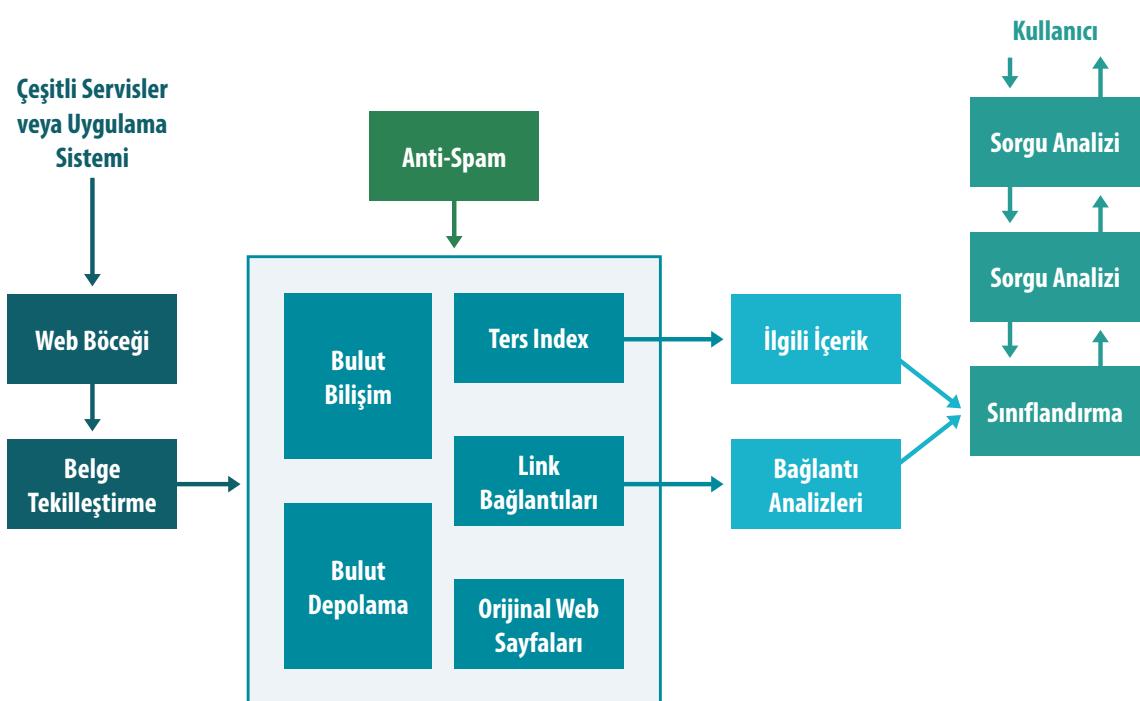
Örneğin Yahoo S4 ve Twitter Storm gibi gerçek zamanlı akış işleme platformları büyük veri işleme gereksinimlerini yerine getirmek için oluşturulmuştur. Bu platformlar hafızadaki bir veri akışını gerçek zamanlı işler ve çok fazla veriyi tutmazlar.

4.3.4 Veri Madenciliği

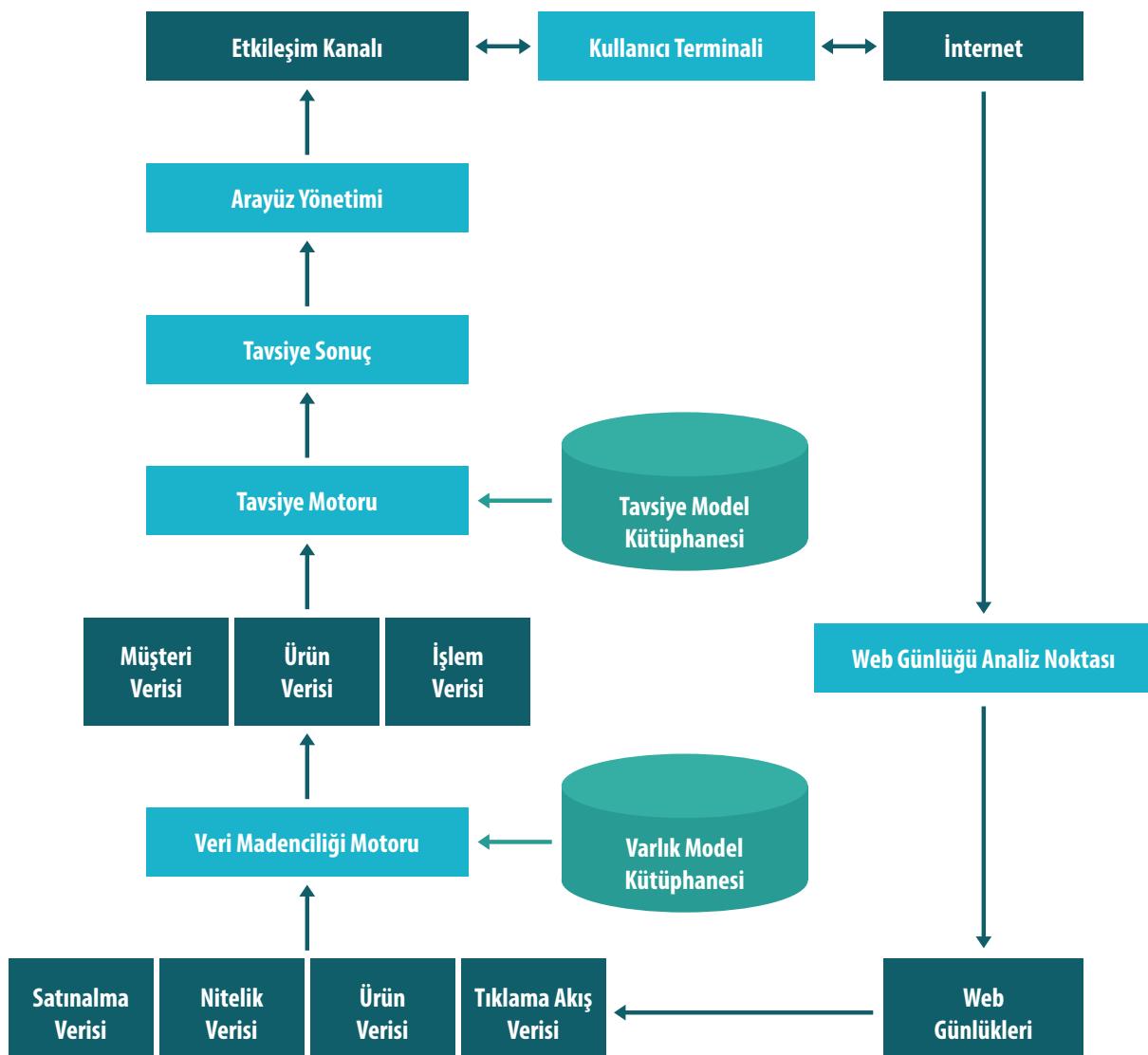
Büyük veri madenciliği teknolojisi büyük veriden etkin bir şekilde veri ayıklamak için kullanılmaktadır. Bu teknoloji paralel veri madenciliği, arama motoru, tavsiye motoru ve sosyal ağ analizlerini kapsamaktadır.

Paralel veri madenciliği büyük ölçüde veri madenciliği algoritmaları uygulananarak büyük veri madenciliği hızını artırmaktadır. Hadoop ve HDFS bu uygulamadan yararlanmaktadır. Veri madenciliği algoritmaları paralel sınıflandırma algoritması ve paralel kümeleme algoritmasını içermektedir.

Örneğin arama motoru çeşitli servislerden ya da uygulama sistemlerinden veri toplayan bir bilgi erişim sistemidir. Veriyi depolar, işler ve yeniden organize eder, sorgulama fonksiyonları ve sonuçlarıyla kullanıcılara sunar. Arama motorları, veri depolama sistemlerinden elde edilen büyük veriden sonra, veri yönetimi için önemli bir araçtır. Kullanıcıların basit sorgular girmesine izin verir ve kullanışlı bilgiler sağlar.



Şekil 4.14 Arama motoru mimarisinin şematik gösterimi. (Kaynak: http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)



Şekil 4.15 Akıllı öneri sistemi mimarisi. (Kaynak: http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

Aşağıdaki Şekil 4.14'te birçok teknolojiyi içeren evrensel arama motoru mimarisi gösterilmektedir. Anahtar teknoloji web crawler (web gezgini), belge anlama, belge indeksleme, ilişki hesaplama ve kullanıcıyı anlama öğelerini içerir. Tavsiye motoru bilgi yiğini içerisindeki kişiselleştirilmiş hizmetleri ve içeriği sağlamak için kullanıcıya yardım eder. Bu bir arama çağından keşif çağına geçişin göstergesi olmaktadır. Tavsiyelerin kalitesi sadece model ve algoritmalarla bağlı değildir, aynı zamanda ürün şekli ve hizmet modu gibi bazı teknik olmayan faktörlere de bağlıdır.

Şekil 4.15'te kullanıcı işlemi veri tabanı, veri madenciliği motoru, veri ambarı, tavsiye model kütüphanesi, tavsiye motoru ve kullanıcı etkileşimi aracını içeren tipik bir tavsiye mimarisi gösterilmektedir. Bununla birlikte bu alt sistemler, veri tabanları ve temel işlemler kullanıcılarla kişisel hizmetler sağlar. Tavsiye motoru içerisindeki temel algoritmalar içerik tabanlı filtreleme algoritması, ortak tabanlı filtreleme algoritması ve ilişki analiz algoritmasıdır.

Sosyal ağ analizi yeni analiz problemleri için yeni bir kavramdır. Bunlar üyeler arasındaki ilişkiler üzerinde temellendirilmiş ve interaktif veri madenciliği için metodlar ve araçlar sağlar. Bu çok kaynaklı bilgi ve fikirleri önüne koyar. Sosyal filtreleme, pazarlama, tavsiyeler ve aramalar için önemlidir. Sosyal ağ analizi kullanıcı ilişkileri, konular, ilgi alanları, kimlik tanımlama, etki ve duygusal analizi içerir. Ayrıca topluluğun keşfini sağlamaktadır.

Bulut tabanlı uygulamaların, mobil internetin ve internetin her türlü uygulamasının yaygınlaşması ve sosyal ağların yükselişi ile büyük veri odağı haline gelmiştir. Büyüklük verinin birçok çeşitli tipleri var ve işleme metodları çok karmaşık hal almaktadır. Bu durum çok büyük zorluklar oluşturmaktadır. İlişkisel veri tabanları ve veri ambarları gibi geleneksel veri işleme metodları büyük veri işlemek için gayret göstermektedir. Dağıtılmış depolamaya sahip büyük veriye uygun sistemler Map/Reduce paralel hesaplama ve veri madenciliğidir. Fakat bunlar da henüz gelişim aşamasındadırlar.

Büyük veri analizini geliştirmek için sistem, ölçme ve anlamlandırma üzerine araştırma yapılmalıdır. Ayrıca büyük veri depolamanın maliyeti de azaltılmalıdır ve esnek, yüksek verimli büyük veri hesaplama mimarileri ve veri madenciliği algoritmaları geliştirilmelidir.

4.4 BAZI ÖZEL BÜYÜK VERİ YAZILIMLARI VE ŞİRKETLERİ

Big data, yani büyük veri konusu üzerine son yıllarda sıkça konuşulur oldu. Hiç kuşku yok ki yakın gelecekte de büyük veriye ve analizine ihtiyacın artmasıyla birlikte büyük veri konuşulmaya devam edecek gibi görünüyor.

Kendi başına bir uzmanlık alanı haline gelen büyük veri konusıyla uğraşan çoğu proje ve şirket, belirli noktalarda hazır çözümler kullanmayı tercih etmektedir. Bunalardan bazıları şunlardır:



Şekil 4.16 2020'deki en iyi büyük veri şirketleri. (Kaynak: <https://datawider.com/top-big-data-companies-usa/>)

- **Flurry:** Twitter'ın günlük tweet trafiğinin yaklaşık 3 katı kadar veriyi kendisi ile entegre olan mobil uygulamalardan toplayan Flurry; Android, iOS, Windows Phone ve Blackberry üzerinde çalışabiliyor. 230.000 uygulamayı bünyesinde barındıran Flurry her ay 700 milyon mobil cihaza ulaşıyor. Sahip olduğu bu veriler ve pazarın geleceği ile yıldızı parlayan şirket, aldığı yatırımlar ile birlikte yıllık 100 milyon dolar kâr eden bir yapıya büründü.

- **Continuity:** Facebook (Jonathan Gray) ve Yahoo'da (Todd Papaioannou) büyük veri üzerine çalışan iki mühendisin kurucusu olduğu bir start-up olan, içerisinde Andreessen Horowitz'in de bulunduğu bir gruptan 10 milyon dolar yatırım alan California merkezli şirket, büyük veri konusunda yazılımcıların kullanabileceği bir PaaS (Platform as a Service) olmayı hedefliyor. Çoğu ürünü henüz beta aşamasında olan Continuity, Hadoop ve Hbase gibi konularda kendisi ispatlamış iki kurucu ortağı ve aldığı yatırımlarla dikkat çekiyor.

- **Infochimps:** Texas/ABD merkezli şirket kullandığı teknolojiler ve bulut çözümleriyle Continuity'ye oldukça benziyor diyebiliriz. Kısa bir süre önce Infochimps Enterprise Cloud isimli ürününü yayına alan şirketin Cloudera ve Rackspace gibi araçlarla da çeşitli ortaklıklarını bulunuyor.

- **Kontagent:** Aralarında EA Sports, BBC ve Warner Bros gibi isimler bulunan müşteri kitlesine sosyal ve mobil internet dünyalarına özel analiz araçları sunuyor. Sistem üzerinde her bir saniyede 10.000 adet sosyal tetikleme gerçekleşiyor. Ayrıca

şirket kSuit aracı ile müşterilerinin kullanıcı panellerine mobil cihazlarından (iOS) da ulaşmasına olanak veriyor.

- **Precog:** Hem bir Start-up Weekend hem de TechStars girişimidir. Kendilerini bu yukarıdaki diğer şirketlerden ayıran ise yazılımcılar için geliştirdikleri Labcoat IDE isimli platformdur. Bu platform sayesinde yazılımcılar, Quirrel sorgu dilini kullanarak özelleştirme yapabiliyorlar. Şirketin iddiasına göre söz konusu platformu yazılımcı olmayan kişilerin bile kullanabileceği yönünde.
- **StatsMix:** Oldukça sade ve net bir büyük veri analiz aracıdır. Colorado/ABD merkezli şirket aynı zamanda bir TechStars üyesidir. Servis üzerinde toplanan birçok veriye ek olarak diğer istediğiniz özellikleri de kendiniz tanımlayarak takip yapabiliyorsunuz. StatsMix'in gelişmiş kontrol paneli üzerinde bu verilerin görselleşmiş sonuçlarına ulaşmanız, dilerseniz düzenli raporlar almanız mümkün oluyor.

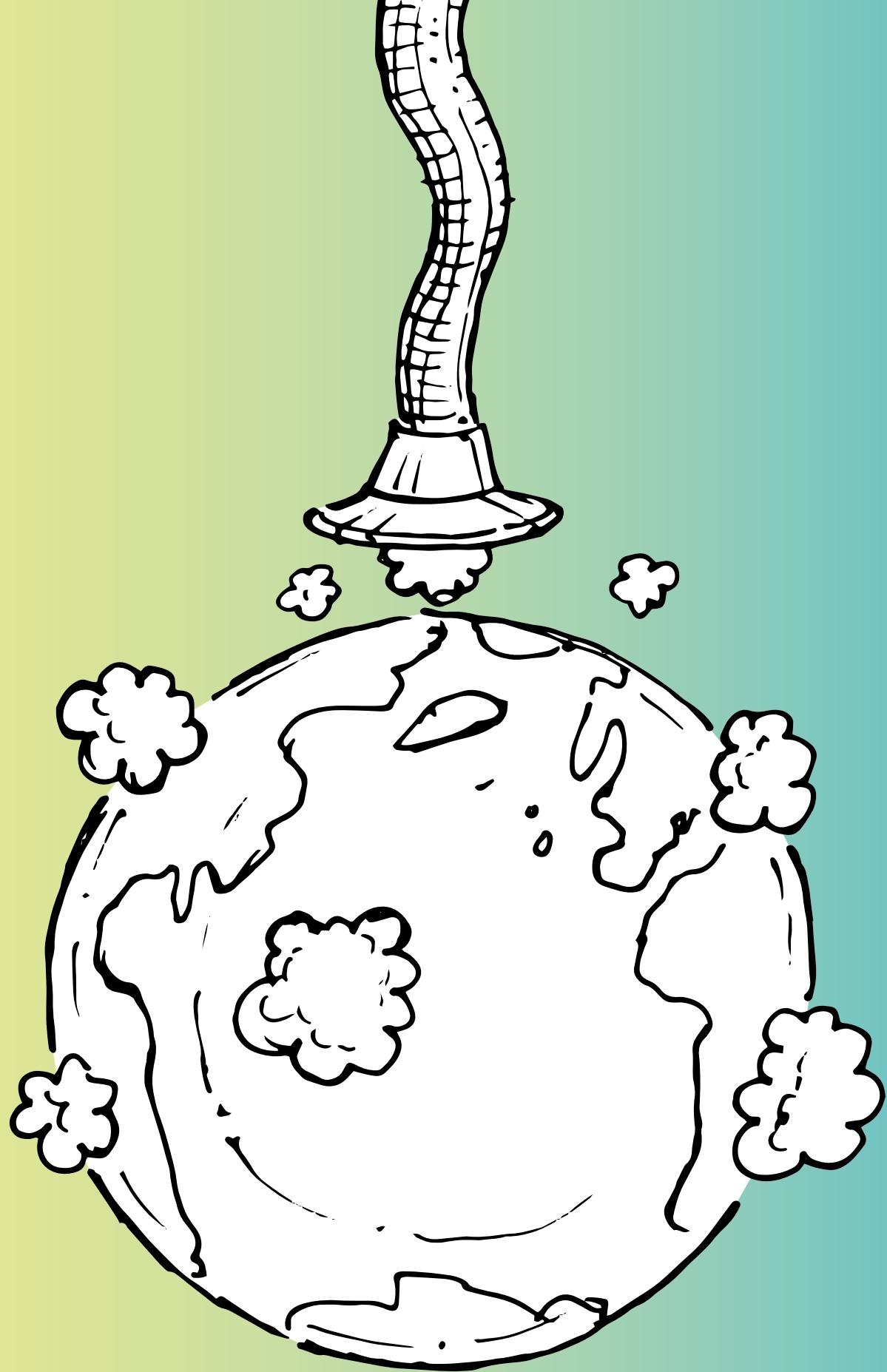
KAYNAKLAR

- Buytendijk, F., Horiuchi, H., Heudecker, N., Adrian, M., McIntyre, A. ve Ingelbrecht, N., 2014, Cool Vendors in Big Data, <http://www.clearstorydata.com/gartner-cool-vendors-report/>, [Ziyaret Tarihi: 17 Eylül 2014].
- Dumbill, Edd (2012). Planning for Big Data, O'Reilly Media.
- F.J.Provost and V.Kolluri, “A survey of methods for scaling up inductive learning algorithms”, Data Mining and Knowledge Discovery, c. 3, 1999.
- <https://techvidvan.com/tutorials/big-data-technologies/>
- <https://www.zeltisinfotech.com/products-solutions/database-management-system.html>
- <https://www.kdnuggets.com/2016/07/seven-steps-understanding-nosql-databases.html>
- <https://dashbouquet.com/blog/web-development/guide-for-freelance-skills-to-land-a-project-in-2019>
- <https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/>
- <https://reports.thomsonreuters.com/9billionbowls/>
- <https://techvidvan.com/tutorials/how-mapreduce-works/>
- <https://factspan.com/hadoop-distribution-file-system/>
- <https://datawider.com/top-big-data-companies-usa/>
- <https://www.slideshare.net/AnikaGro/nosql-data-stores-for-big-data>

- <https://www.researchgate.net/publication/334032994/figure/fig1/AS:776617282514945@1562171397916/Self-quantification-data-from-fitbit-project-Number-of-public-records-from-January-2009.png>
- http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_4_02922.html, 2013
- <http://wiki.apache.org/hadoop/PoweredBy>
- http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7293298, 2015
- https://www.bogotobogo.com/Hadoop/BigData_hadoop_Ecosystem.php
- <https://unhp.org/crg/indy-cd.html>
- Provost, Foster ve Tom Fawcett (2013). "Data Science and Its Relationship to Big Data and Data-Driven Decision Making," https://www.researchgate.net/publication/256439081_Data_Science_and_Its_Relationship_to_Big_Data_and_Data-Driven_Decision_Making
- Thomson Reuters, 2015, Filling 9 billion bowls: The answer is data, <http://blog.thomsonreuters.com/index.php/filling-9-billion-bowls-answer-data/>, [Ziyaret Tarihi: 20 Eylül 2015].
- Thomson Reuters, 2015, How Will We Fill 9 Billion Bowls By 2050?, <http://reports.thomsonreuters.com/9billionbowls/>, [Ziyaret Tarihi: 20 Eylül 2015].

5. BÖLÜM

BÜYÜKVERİNİN ZORLUKLARI VE FIRSATLARI

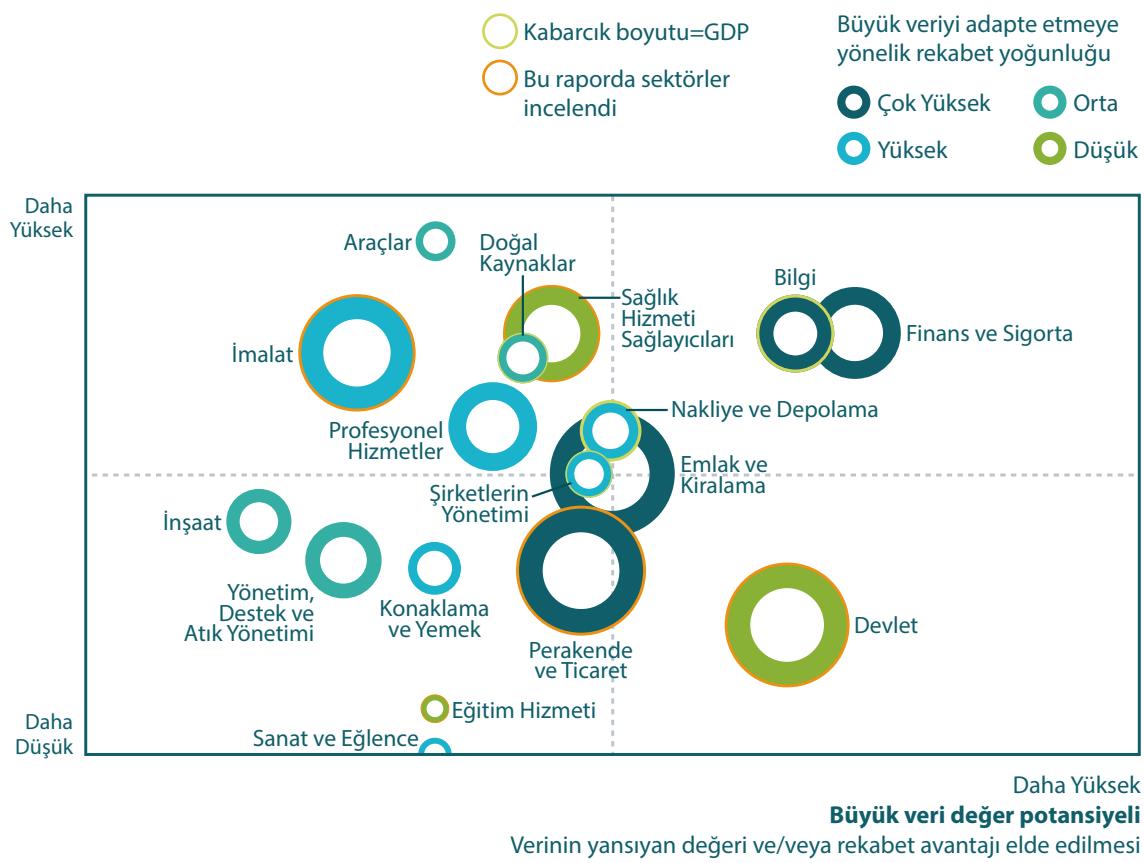


Ayrıca büyük verinin ilerleyişi birçok yeni fırsat getirebilir, fakat kurumlar ve şirketler için zorluklar da oluşturur. Yukarıda belirtilen teknik gereksinimlere ek olarak şirketlerin de ulusal kimlik numaraları dahil (ABD'de sosyal güvenlik numaraları, Türkiye'de kimlik numaraları), tıbbi kayıtlar ve diğer potansiyel sınıflandırma gibi potansiyel "zehirli veriye" dikkat etmesi gerekmektedir ki kişi bilgileri ortaya çıkarsa (kasıtlı veya başka bir şekilde) bu durum bir şirket için yasal sorun oluşturabilir (Shuster, 2012). Computing Research Association (CRA, 2011) tarafından yayınlanan bir makale büyük veri yönetiminin birkaç zorluğuna işaret etmiştir: Teknik mülahazalar, örneğin hataları minimize etme ki bu bile veri temizleme, düzeltme çabaları ve güç tüketimi zorluklarının arkasından gelir. Zamanında erişim, güç tüketimi ve hata azaltılmasına dair bu teknik endişelere ek olarak dikkate alınması gereken sosyal-politik-hukuki hususlar da vardır. Bu hususa örnek olarak etik olarak toplanan büyük verinin ve çeşitli kullanıcıların coğrafi konumlarının gizliliğine saygı gösterilebilir.

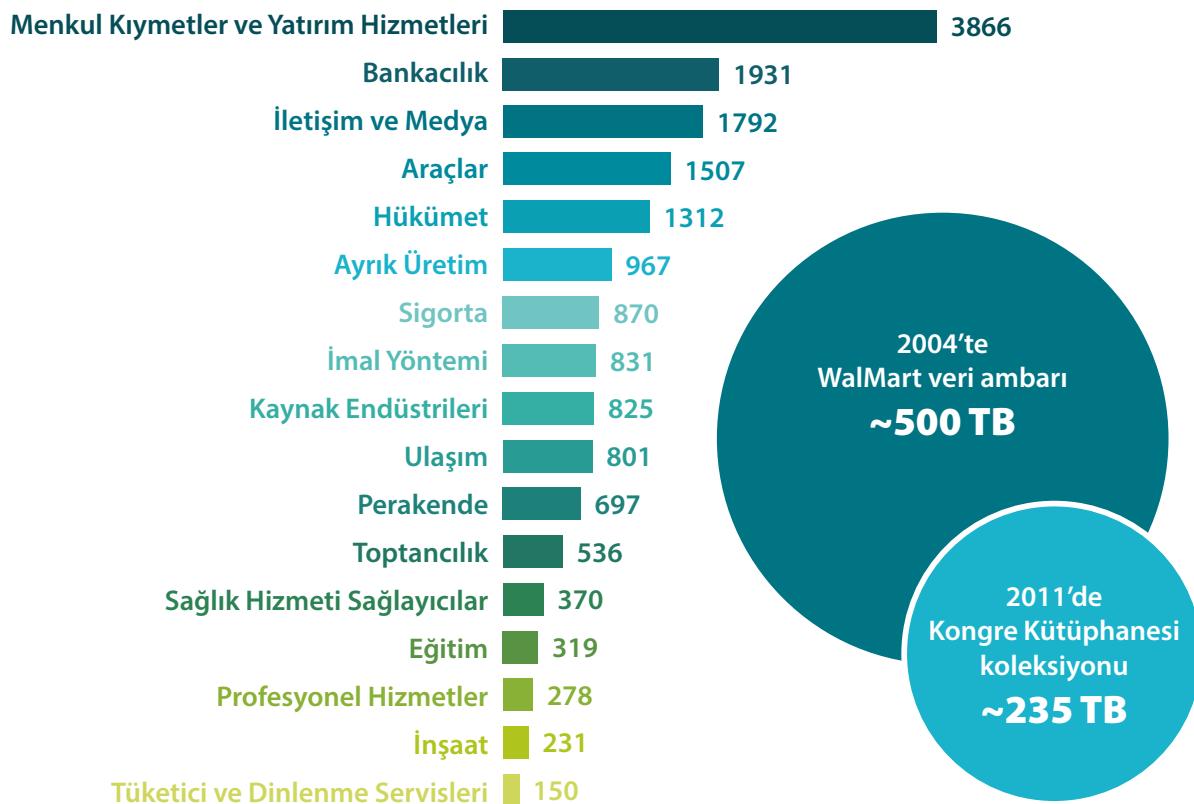
Elbette bu zorluklarla birlikte şunu da söylemeye gerek yok: Büyük veri şirketler ve işletmeler için burada hepsini adlandırmadığımız birçok potansiyel kullanımına sahiptir. Pazardaki verinin kullanım potansiyelini, standart muhasebeye göre varlıklarının toplamı 6,3 milyar dolar olmasına rağmen, pazara giriş yaptığı zaman Facebook'un ilk halka arz değerinin neden 104 milyar dolar olduğunu açıklar (Mayer-Schonberger, 2012, s. 100-120). Wikibon firmasından açık kaynak analisti, beş yıl içerisinde büyük verinin toplam pazar değerinin 50 milyar dolara ulaşacağını tahmin etmektedir.

	Başladı ve değer gördü ↓	Başladı ve değer görmedi ↓	Başlamadı ↓
Giderlerin azalması	%49.2	%23.4	%27.4
Yeni inovasyon yolları bulma	%44.3	%20.2	%35.5
Yeni ürünleri/hizmetleri başlatma	%36.1	%26.8	%37.1
Gelir ekleme	%32.8	%22.0	%45.2
Mevcut çabaların hızını artırma	%31.1	%33.4	%35.5
Gelecek için işi dönüştürme	%27.9	%23.7	%48.4
Veri tabanlı bir kültür kurma	%27.9	%41.5	%30.6

Şekil 5.1 Büyük veriyi kullanan Fortune 1000 yöneticilerinin raporu. Başlayan projeler ve değer bulduğu yerler. (Kaynak: Newvantage Partners Big Data Executive Survey, 2017)



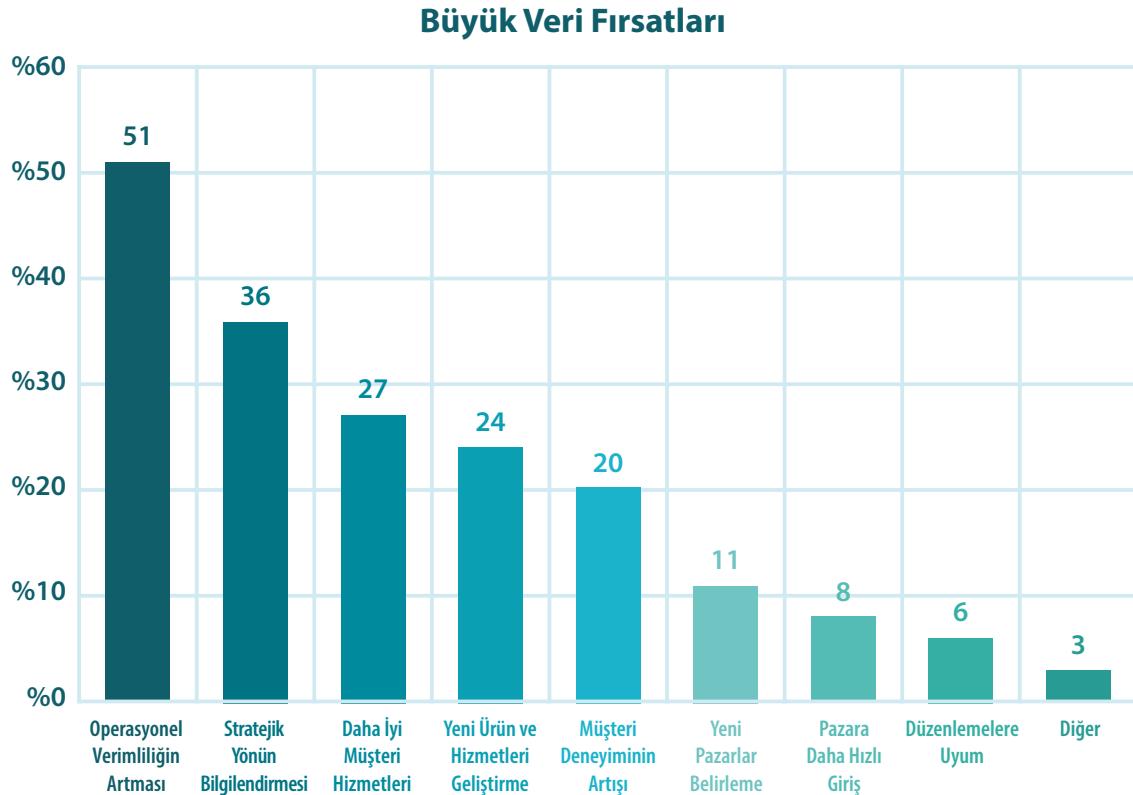
Şekil 5.2 Sektörlerin büyük veriye adaptasyonu. (Kaynak: <https://www.dasca.org/world-of-big-data/article/big-data-value-potential>)



Şekil 5.3 Sektörler arasında ABD'deki şirketler en az 100 terabayt veri depolar -çoğu 1 petabayt'tan fazla veriye sahiptir. 1.000'den fazla çalışanı olan firma başına ortalama depolanan veri (Terabayt, 2009). (Kaynak: McKinsey Global Institute)

Gelecekte aşırı spekülatyonu önlemek amacıyla şirketlerin büyük veri girişimlerinin gerçekçi, elde edilebilir amaç ve umutlar sağlaması gerektiğinden, büyük veri etkileri üzerinde artan çalışmaya daima ihtiyaç vardır.

Aynı zamanda McKinsey Enstitüsü'nün raporuna göre, büyük verinin etkin kullanımı dönüşen ekonomiler için belli başlı faydalara sahiptir ve yeni bir büyüyen üretim dalmasını doğurmaktadır. Büyük verinin ötesinde değerli bilgi avantajları sağlama bugünün işletmeleri için ana rekabetçi unsur olacak ve büyük verinin kabiliyetlerine sahip ilgi çekici işler yapabilen yeni rekabetçi firmalar oluşturacaktır. Araştırmacılar, politikacılar ve karar verenler alanlarındaki büyüyen gelecek dalgayı ortaya çıkarmak için büyük veriyi kullanma potansiyelini fark etmek zorundalar. Aşağıdaki şekilde görüldüğü gibi, operasyonel verimliliği artırma, stratejik değerlendirme, daha iyi müşteri hizmetleri gelişimi, yeni ürünler ve hizmetler tanımlama ve geliştirme, yeni müşteriler ve pazarlar bulma vb. iş bölümlerinde büyük veri kullanımıyla elde edile-



Şekil 5.4 Büyük veri fırsatları: 560 işletmenin %50'den fazlası büyük verinin operasyonel verimliliğini artırmada faydalı olabileceğini düşünüyor. (Kaynak: McKinsey Global Institute Report, <http://www.sciencedirect.com/science/article/pii/S0020025514000346>, 2011)

bilen birçok avantajlar vardır. Dikey eksen kaç tane işletmenin büyük verinin özgün amaçlarıyla ilgili onlara yardımcı olabileceğini düşündüğünü yüzde olarak gösterir.

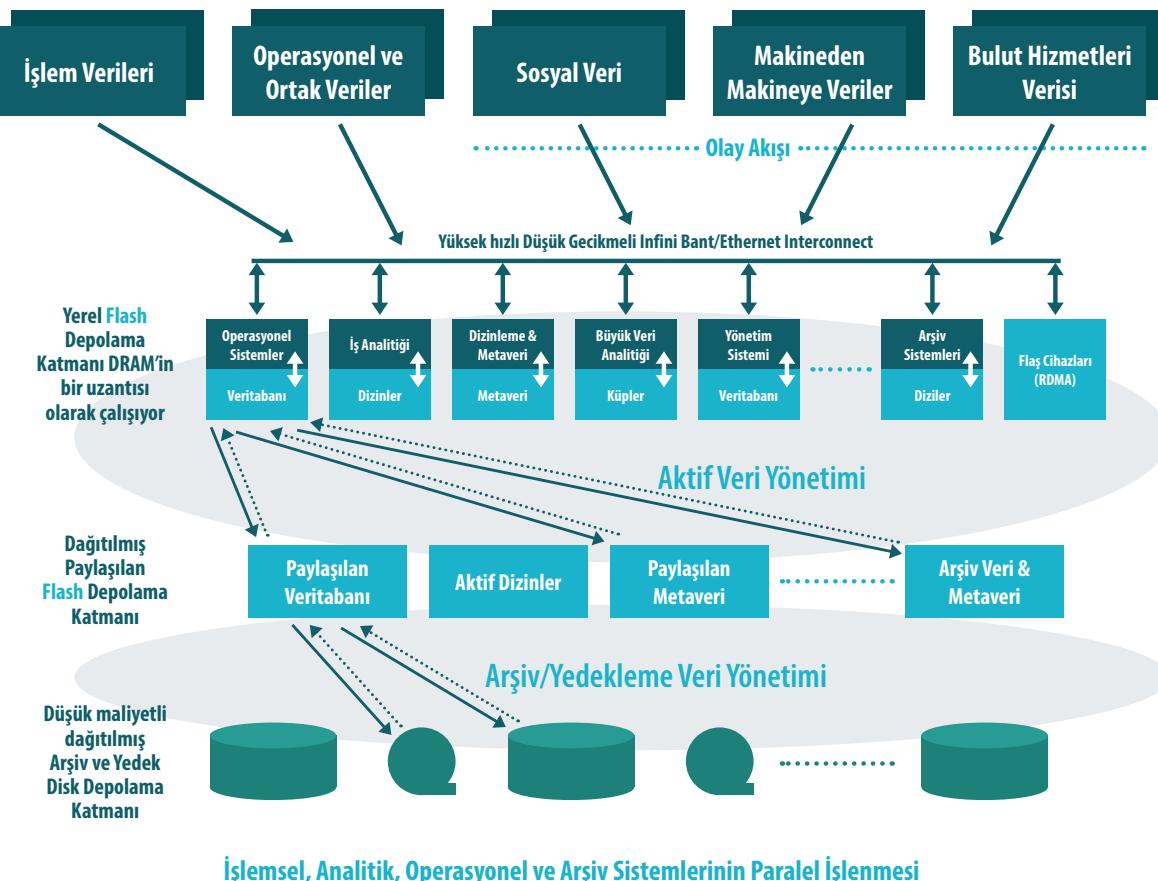
Liberal tahminlere göre, büyük veri ABD'deki sağlık bakımı için yıllık 300 milyar dolar, AB kamu yönetimi için 250 milyar euro potansiyel değer üretebilir. Dünya çapındaki kişisel lokasyon verisi kullanılarak yıllık 600 milyar dolar potansiyel müşteri fazlası olacak ve %60 oranında potansiyel bir artış sağlanacaktır. Sadece ABD'de büyük veri 140.000 ila 190.000 arasında derin analitik yetenek pozisyonları ve 1,5 milyon veri meraklısı yönetici üretmektedir. Kuşkusuz eğer doğru bir şekilde araştırılıysa büyük veri genel olarak çekici ve kârlı olmaktadır. (Bkz. Şekil 5.4)

Fırsatların ardından daima zorluklar gelmektedir. Bir yandan büyük veri çok cazip fırsatlar getirir. Diğer yandan da büyük veri problemlerini ele aldığımızda birçok zorlukla yüz yüze kalırız. Bu zorluklar veri toplama, depolama, arama, paylaşma, analiz

ve görüntüleme üzerine yayılır. Eğer bu zorlukları aşamazsak, özellikle bilgi bizim kullanım kabiliyetimizi aştığı zaman, büyük veri bir altın değerinde olacaktır, fakat onu keşfetme imkânlarına sahip olamayız. Birkaç on yıldır bilgisayar mimarisinde var olan bir engel CPU'nun güçlü fakat I/O'nun zayıf olmasıdır. Bu sistem dengesizliği büyük veriden faydalama gelişimini hâlâ kısıtlamaktadır.

CPU performansı yaklaşık her 18 ayda bir ikiye katlanıyor, aynı zamanda disk sürücülerinin performansı da aynı oranda katlanıyor, fakat disklerin rotasyonel hızları son on yılı aşkın sürede çok daha az gelişti. Bu dengesizliğin sonucu ardışık I/O hızları çok yavaş bir şekilde artarken, rastgele I/O hızları ortalama şekilde gelişti. Dahası bilgi üstel oranlarda artıyor, fakat bilgi işleme yöntemlerinin gelişimi görece

Gerçek Zamanlı Büyük Veri İşleme



Şekil 5.5 Gerçek zamanlı büyük veri işleme. (Kaynak: http://wikibon.org/wiki/v/Enabling_Real-time_Big_Data_Processing_with_Xflash_-_Very_Low_Latency_Server_SANs)

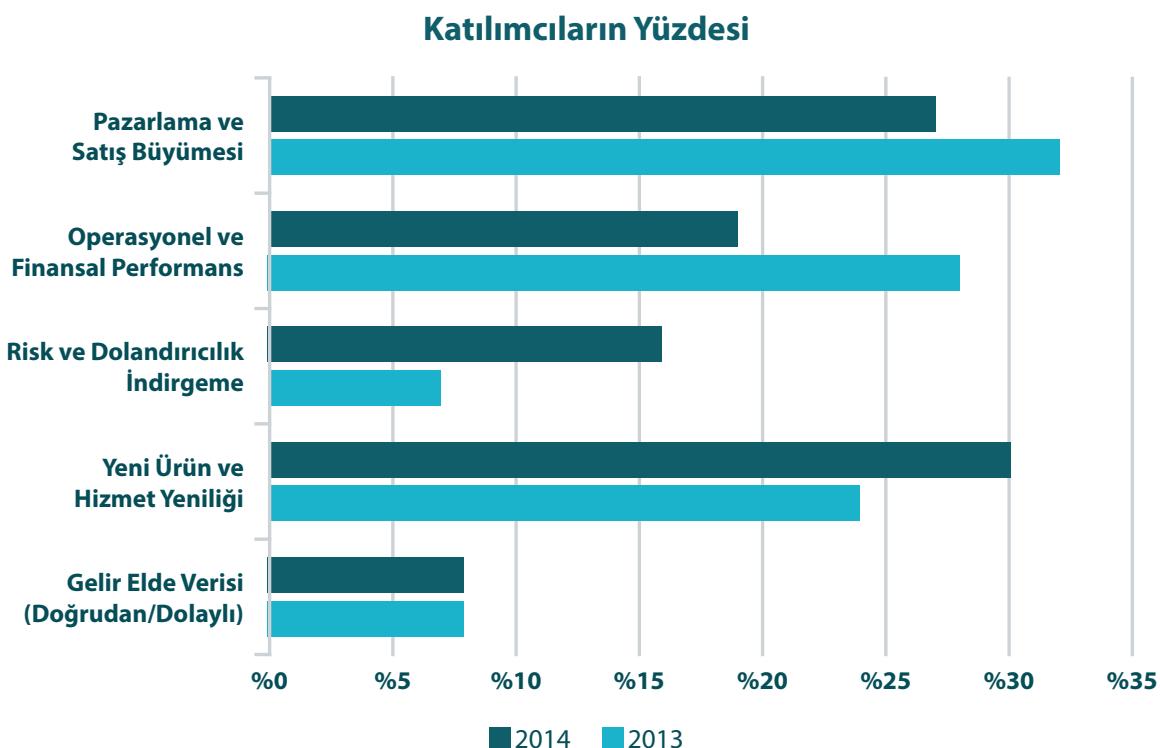


Şekil 5.6 Bilgi keşif süreci. (Kaynak: <http://www.sciencedirect.com/science/article/pii/S0020025514000346>)

daha yavaş oluyor. Birçok büyük veri uygulamasında, özellikle de gerçek zamanlı analizler için teknikler ve teknolojiler ideal bir şekilde problemleri çözemiyor. Şu ana kadar kısmen ifade edilen genel kanı şudur; altın cevherini hakkıyla işleyebileceğimiz tamamen uygun araçlara sahip değiliz.

Tipik bir genel analiz süreci Şekil 5.6'da gösterilmiştir.

Büyük veri analizindeki diğer zorluklar veri tutarsızlığı ve eksikliği, ölçeklenebilirlik, güncelilik ve veri güvenliği olarak değerlendirilebilir. Veri analizi için ilk basamak verinin iyi yapılandırılmış olmasıdır. Fakat veri setlerindeki dikkate değer çeşitlilik hedefe ulaşmak için bize büyük engeller çıkartmaktadır. Veri setleri genelde çok büyütür, bazen birkaç gigabaytlara, hatta daha fazlasına ulaşmaktadır, heterojen kaynaklardan gelmektedir, gerçek dünyanın veri tabanları süreksiz, tamamlanmamış ve gürültülü olmaktadır. Bu ve bu gibi nedenlerden dolayı veri temizleme ve veri dönüşümünü kapsayan veri işleme tekniklerine gürültü kaldırma ve tutarsızlıkları düzeltmek için başvurulabilir. Veri tabanı uygulamalarından geldiğinde, farklı engeller her bir alt işlemi artırmaktadır.



Şekil 5.7 Büyük verinin fırsat sunduğu alanlar. (Kaynak: Gartner Büyük Veri web seminerinde veri, <https://www.gartner.com/doc/2703617/customer-analytics-art-possible-big>, Mart 2013 ve 2014)

5.1 BÜYÜK VERİNİN AVANTAJLARI

Büyük verinin ana önemi farklı veri tipindeki büyük hacimli veri kullanımını bağlamında verimliliği artırmak için potansiyel oluşundandır. Eğer büyük veri düzgün bir şekilde tanımlanır ve buna uygun olarak kullanılırsa, kurumlar işlerine daha iyi bir biçimde bakabilirler. Bundan dolayı satış, ürün geliştirme ve daha başka farklı alanlarda etkin bir şekilde faydalananabilirler. (Bkz. Şekil 5.7)

McKinsey Global Enstitü'nün yayınladığı bir raporda (McKinsey Global Institute Report, 2011) büyük veri kullanımının değer artırın bazı avantajlarından bahsedilmektedir:

- Bilgiyi daha sıkıkla şeffaf ve kullanılabılır yaparak değer katıyor.
- Şirketler daha fazla işlemel veri üretip dijital şekilde sakladıkça, ürün envanterinden hastalık günlerine kadar her şey hakkında daha doğru ve ayrıntılı bilgi edinbiliyorlar. Böylece çeşitliği artırıyor ve performanslarını katlayabiliyorlar. Lider şirketler veri toplamayı ve analizini daha iyi yönetim kararları vermek kontrollü

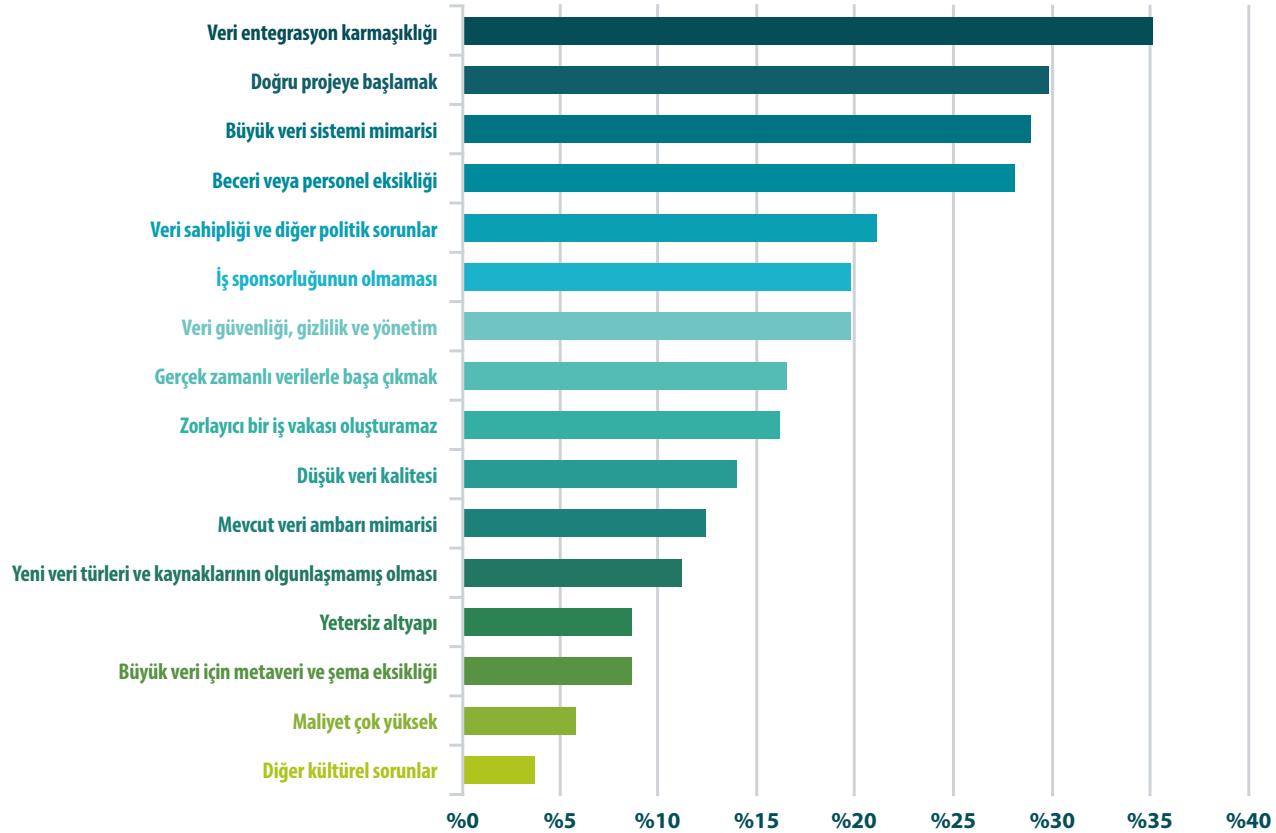


Şekil 5.8 Büyük veride endüstri trendi. (Kaynak: <https://www.openpr.com/news/1128600/big-data-in-healthcare-market-medical-insights-healthcare-analyzed-industry-trends-major-players-research-2023.html>)

deneyler yapmakta kullanıyor. Diğerleri veriyi seyrek ve uzun vadeli yerine, sık aralıklı ve mevcut-vadeli tahminler ile işlerini gerçek zamanlı yapmak için kullanıyor.

- Büyük veri müşteri segmentlerini giderek daha da daraltıyor. Böylece daha dikkatle kişiselleştirilmiş ürün ve hizmetler mümkün oluyor.
- Sofistike analizler karar vermeyi ciddi olarak kolaylaştırıyor.
- Yeni nesil ürün ve hizmetlerin geliştirilmesinde kullanılabiliriyor. Halihazırda üreticiler ürünler içine gömülü sensörlerle elde edilen veriyi proaktif bakım (hata oluşmadan, hatta fark edilmeden önce koruyucu önlemleri almak) gibi yaratıcı satış sonrası hizmetlerde kullanıyorlar.

Örneğin perakendeci sektöründe operasyonel kârin % 60 dolayında artması, ABD sağlık hizmetlerinde yılda 300 milyon dolar civarında maliyet azalması gibi vurucu rakamlar bu raporda ifade edilmiştir.

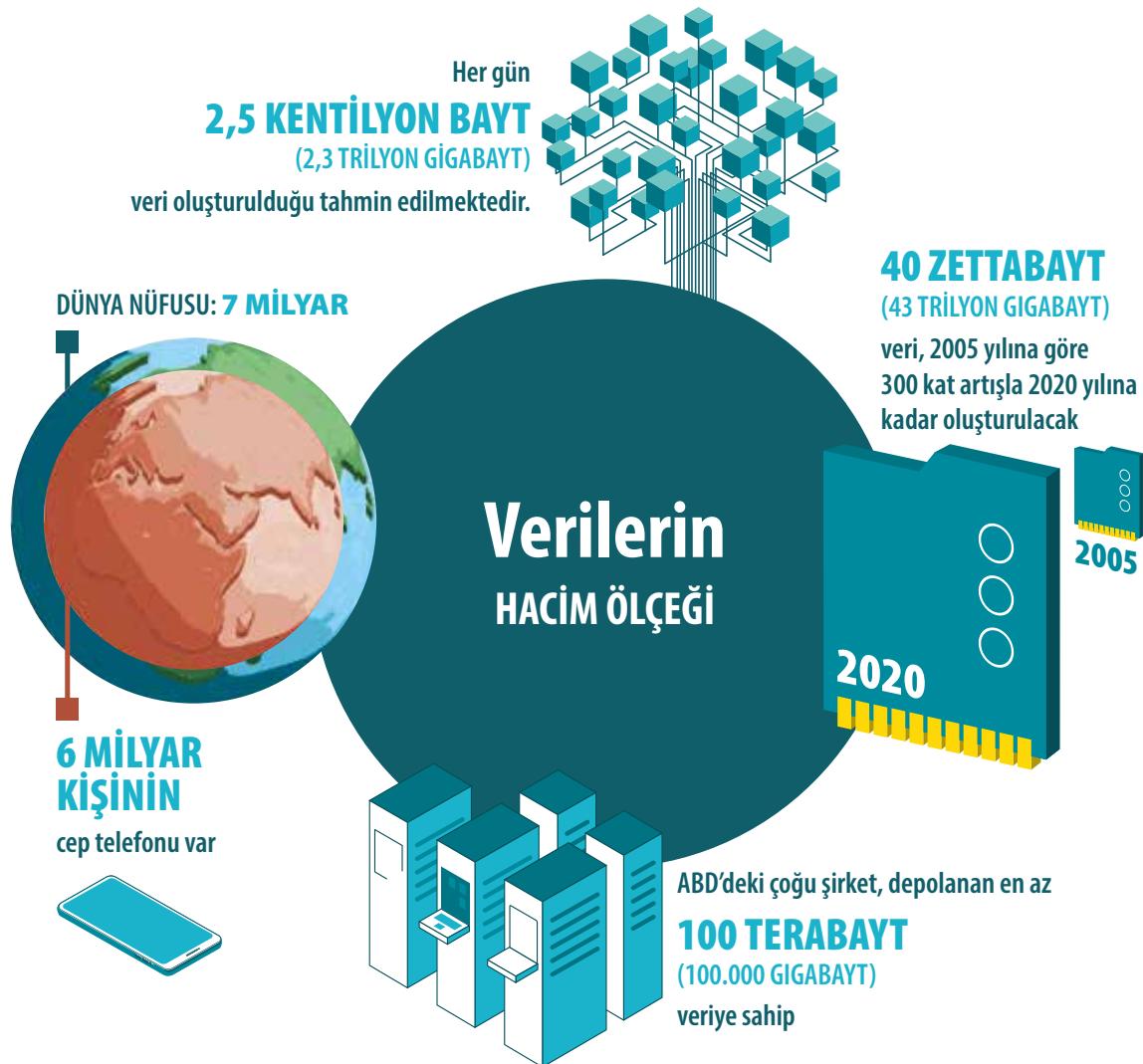


Şekil 5.9 Büyük verinin zorlukları. (Kaynak: <https://tdwi.org/Blogs/TDWI-Blog/2013/10/Four-Big-Data-Challenges.aspx?m=1>)

5.2 BÜYÜK VERİ ANALİZİNDEKİ ZORLUKLAR

Günümüzde internet şirketlerinin servisleriyle ilgili büyük veri hızlı bir şekilde büyümektedir. Örneğin her ay Google yüzlerce petabayt veri işler, Facebook her ay 10 petabaytı aşın veri üretir, Çinli bir şirket olan Baidu onlarca petabayt, Alibaba'nın bir yan kuruluşu olan Taobao çevrimiçi ticaret için her gün onlarca petabayt veri üretmektedir. Bilgi teknolojilerinin en son ilerlemeleri veri üretimini daha kolay bir hale getirmiştir. Örneğin Youtube'a ortalama 1 dakikada 72 saatlik video yüklenmektedir.

Bu sebeple veri kaynaklarının çok geniş dağılımından ve devasalığından dolayı bir entegrasyon ve toplama zorluğu ile karşı karşıyayız. İnternet araçlarının ve bulut bilişimin çok hızlı gelişmesi veri büyümesinde keskin bir artış sağlar.

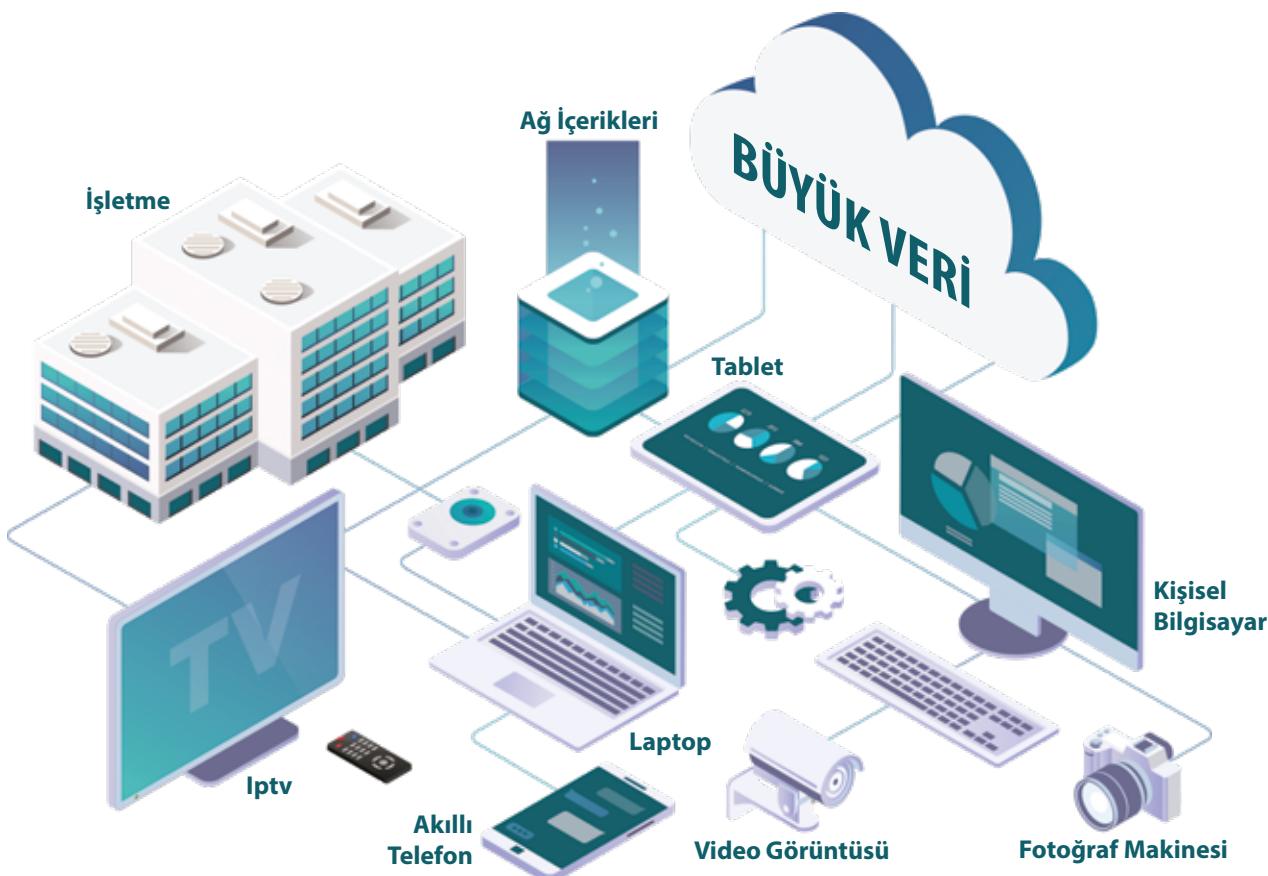


Şekil 5.10 Farklı veri kaynaklarından gelen verilerin hacim ölçüleri. (Kaynak: https://www.researchgate.net/figure/Volume-scale-of-Data-from-different-data-sources-26_fig1_324220787)

Veri işleme teknolojileri uzun bir dönem boyunca sürekli olarak gelişmiştir, fakat büyük veri anı bir şekilde iki önemli değişikliğe neden oldu. Şu anda tüm veriler kaydedilebilir. Bunun anlamı şudur: Uygulama öncesinde biriktirilmiş veri gerektiren uygulamalar daha kolay bir şekilde uygulanabilir.

Ayrıca kısıtlı veriden veri seline bir kayma meydana gelmiştir. Öyle ki veri seli uygulamalar için yeni zorluklar meydana getirmiştir. Basit bir şekilde, sadece arama motorlarından elde edilen veriler artık bugünkü uygulamalar için yeterli değildir.

Bu büyük veri kütlesinden kullanışlı veri elde etme ve işleme süreci gittikçe zorlaşmaktadır. Büyük verideki teknolojik atılımlar yapılmaya devam edecek ve bu yüzden büyük veri önmüzdeki 2 ila 5 yıl içerisinde daha yaygın bir şekilde kullanılacaktır.



Şekil 5.11 Bir büyük veri senaryosu. (Kaynak: http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013)

Geleneksel olarak veriler tek bir kaynaktan gelen ve nispeten düşük bir hacme sahiptir. Depolanması, yönetimi ve analizi büyük zorluklar getirmez ve en çok işlem süreci ilişkili veri tabanları ve veri ambarları aracılığıyla yapılmaktadır. Büyük veri alanında ise verinin hacmi o kadar büyütür ki geleneksel veri işleme sistemleri bu verinin depolanması, yönetimi ve analizi ile başa çıkamaz. Geleneksel iş zekâsı yazılımları yapılandırılmış veri analizi ve işlemesi için eksik araçlar ve metotlara sahiptir.

Büyük veriyi anlama çok önemlidir. Bir şirket ya da kurum stratejik kararlar vermek için sizin analizlerinize güveneceğinden, doğru ve düzgün bir şekilde analiz edilmesi çok önemlidir. Ayrıca bu analizlerin süresi çok önemlidir, iş dünyasındaki hızlı değişimlerden dolayı sıkılıkla ve hızlı bir şekilde uygulanması gerekmektedir.

Büyük Veri ve Küresel Veri Kanunları

KARIŞIK BİR BÖLGESEL TAHMİN



BİRLEŞİK KRALLIK
Kişisel veri hakları, tarihsel, bilimsel veya istatistiksel amaçlarla kullanılan veriler için göz ardı edilebilir.



İSRAİL
PPDS, veritabanı boyutuna ve verilerle kaç kişinin temasa geçtiğine bağlı olarak farklı güvenlik seviyeleri gerektirir.



GÜNEY KORE
Tam olmayan ancak diğer mevcut verilerle birleştirilebilen kişisel veriler de korunmaktadır.



BREZİLYA
LGPD, kişisel verileri, bir kişiyi tanımlamak için tek başına veya diğer verilerle birlikte kullanılabilen herhangi bir veri türü olarak yorumlar.

Şekil 5.12 Büyük veri ve küresel veri kanunları. Karışık bir bölgesel tahmin. (Kaynak: <https://blog.ipswitch.com/global-data-privacy-laws-us-eu-china-and-more>)

Diğer bir yönü ise her gün yeni teknolojiler geliştirilmektedir. Büyük verinin kuruluşlar için yeni olması göz önüne alındığında, yeni teknolojiler pazara girer girmez onların nasıl kullanılacağını öğrenmeleri gereklidir. Bu işletmeler için rekabet avantajı getirecek yönüdür.

Bilişim uzmanları ihtiyacı da büyük verinin diğer zorluklarından biridir. McKinsey'in büyük veri üzerine çalışmasına göre, sadece ABD'de 190.000'den fazla analitik uzman ve 1,5 milyondan fazla veri okuryazarı müdüre ihtiyaç vardır. Bu istatistikler büyük veri girişimini ele alan ve bu konuda zayıf olan şirketlerin ya uzmanlar kiralamasını ya da kendi personelini bu yeni alan üzerine eğitmek zorunda olmasını beraberinde getirmektedir.

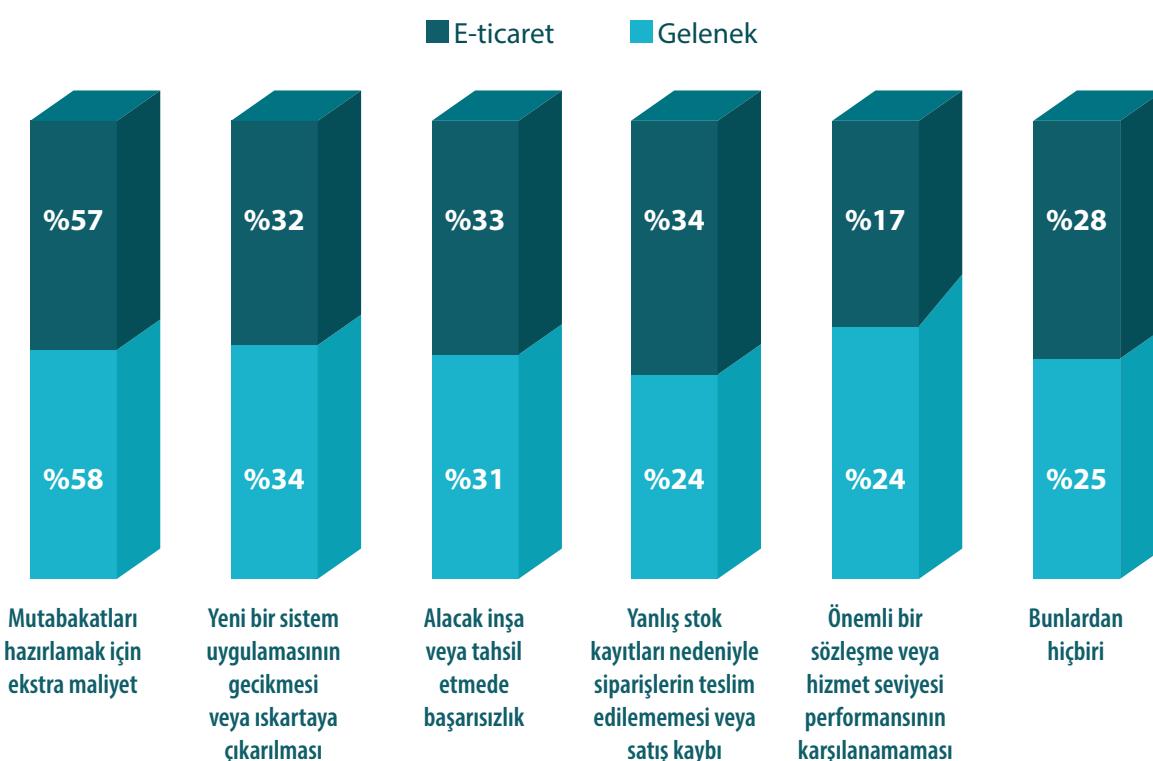
Gizlilik ve güvenlik de büyük verinin zorluklarından biridir. Büyük veri çok karmaşık verilerden oluşmaktadır, bu yüzden bir şirket için bilgileri gizlilik seviyelerine göre sıralamak ve güvenlik uygulamak çok zordur. İlaveten bugünlerde birçok şirket ülkeler ve kıtalararası çaprazlama işler yapıyor ve bunların gizlilik hukuklarındaki farklılıklar dikkat çekiyor. Bir şirket büyük veri girişimine başladığında bunları göz önünde bulundurmak zorundadır.

Bu sebeplerle büyük veri uygulaması yapılırken tüm bu faktörlerin hepsi çok dikkate alınmalıdır. Büyük verinin analizi bazı zorlukları beraberinde getirmektedir. Bu zorluklar şu şekilde sıralanabilir.

5.2.1 Kapsamlı Veri Kaynakları ve Kötü Veri Kalitesi

Büyük veri senaryoları heterojen veri kaynakları ile karakterize edilmektedir. Örneğin işlem kayıtları, metinler, resimler ve videolar. Bu veri farklı şekillerde açıklanmakta ve veri giriş oranı saniyede yüzlerce megabaytlara, hatta gigabaytlara kadar ulaşabilmektedir. Yapısal veriyi açıklamak için kullanılan geleneksel metodlar büyük veriyi açıklamak için yeterli değildir. Ek olarak büyük veri gürültüden kolayca etki-

Yetersiz Veri Kalitesinin Neden Olduğu Sorunlar

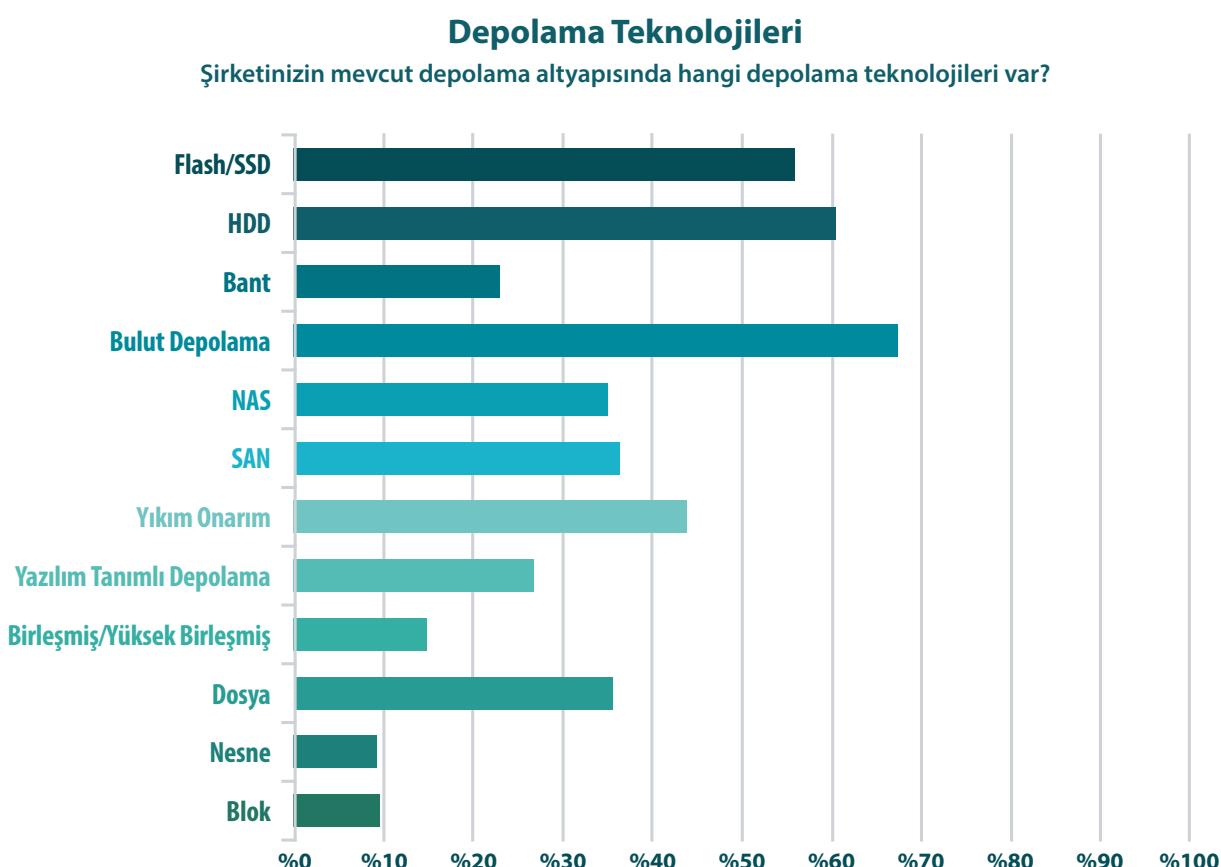


Şekil 5.13 Yetersiz veri kalitesinin neden olduğu sorunlar. (Kaynak: https://www.researchgate.net/figure/Problems-caused-by-poor-data-quality-Adapted-from-8_fig2_293015090)

lenebilir, kaybolabilir ya da tutarsızlaşabilir. Eksik, gürültülü ve tutarsız veriyi filtreleme ve bütünlendirme büyük veriyi verimli bir şekilde depolama ve işleme süreci için önkoşuldur.

5.2.2 Büyük Verinin Oldukça Verimli Depolanması

Büyük verinin depolanması yalnız maliyetleri değil, aynı zamanda analiz ve işleme verimliliğini de etkiler. Büyük veri çoğunlukla kurumsal depolama alanı ağları ve buna bağlı ağlarda ele alınmayan petabaytlarla, hatta exabaytlar ile ölçülmektedir. Büyük veri çağında hizmet ve analiz gereksinimlerini karşılamak için güvenilir,

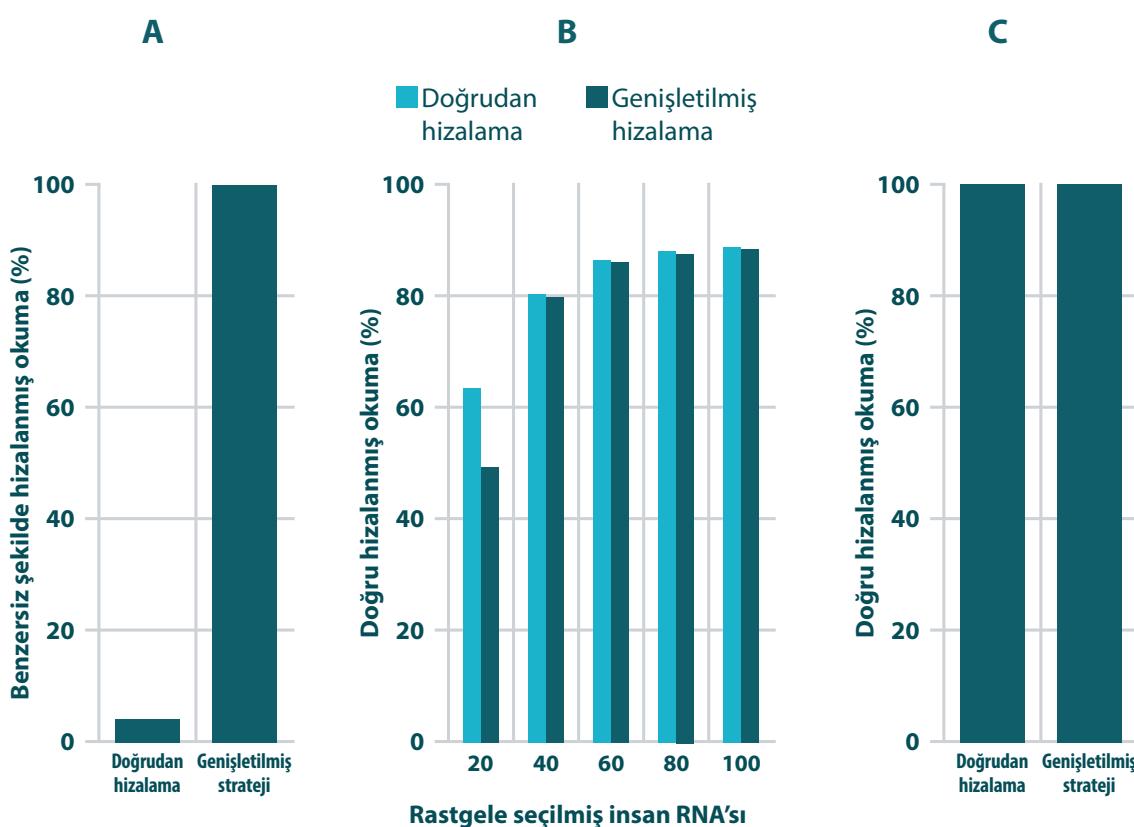


Şekil 5.14 Bu grafik şirketin mevcut depolama altyapısına dahil edilen Veri Depolama Teknolojilerinin sayısını vermektedir. (Kaynak: <https://www.softwaretestinghelp.com/data-storage-companies/>)

yüksek performanslı, yüksek kullanılabilirliğe sahip, düşük maliyetli çözümler geliştirmeye ihtiyaç vardır. Büyük verinin çeşitli kaynaklardan gelmesinden dolayı, aynı veri sistemde zaten var olabilir ve bu fazlalığa neden olabilir. Algılama ve fazlalıkları eleme depolama yerini artırır. Bu büyük veri depolama platformları için gerekli olan temel unsurdur.

5.2.3 Veriyi Verimli Bir Şekilde İşleme

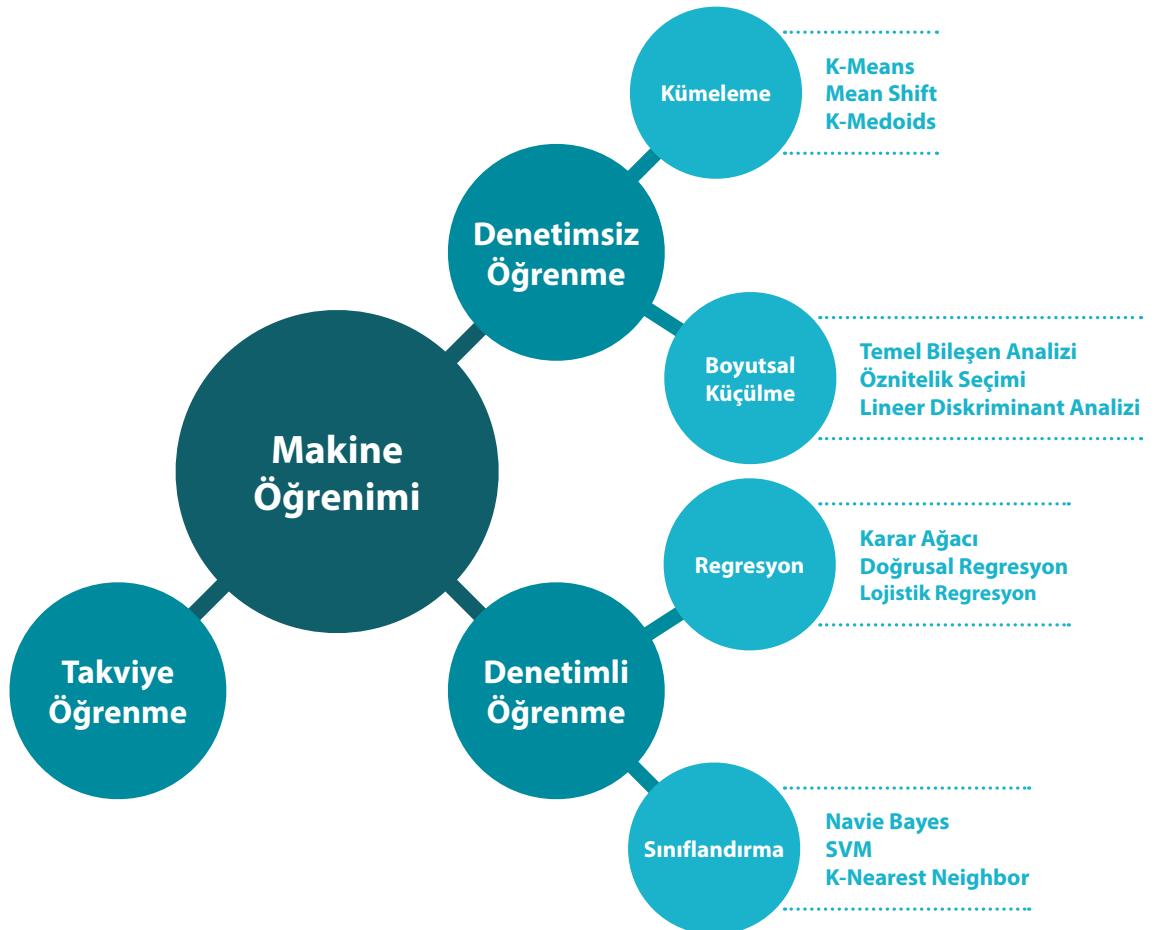
Kurumsal veri ağırlıklı olarak ilişkisel veri tabanları ve veri ambarlarında işlenmektedir. Ancak bu tür veri tabanları ve veri ambarları yapılandırılmış ve ya-



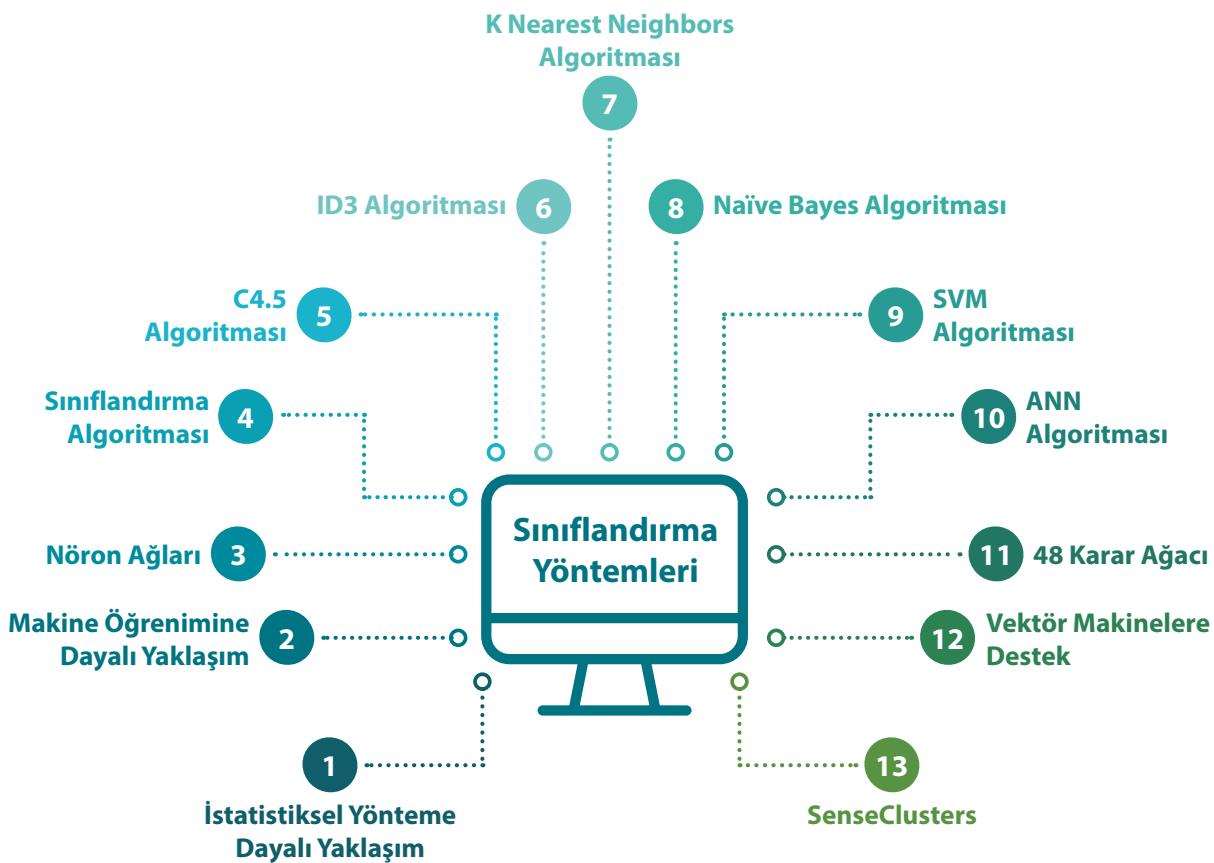
Şekil 5.15 RC-Seq'in veri işleme verimliliğinin simülasyonla değerlendirilmesi. (Kaynak: https://www.researchgate.net/profile/David_Corey3/publication/274257509/figure/fig10/AS:324996351184899@1454496576237/Evaluating-data-processing-efficiency-of-RC-Seq-by-simulation-A-The-problem-the.png)

rı-yapılmalıdır büyük veriyi işlemek için yetersizdir. Büyük veri ile yazma/okuma işlemleri çok sayıda kullanıcı için oldukça eş zamanlı olmalıdır. Büyük verinin depolanması ve veriye erişimin yüksek verimde olması gerekmektedir ve sistem son derece ölçülebilir olmak zorundadır.

Veri setlerinin büyüklükleri arttıkça, algoritmalar verimsiz hale gelebilir ve bölgelendirme, ilişkisel veri tabanlarının tutarlılık, yalıtkanlık, dayanıklılık özellikleri yoğun kaynaklardır. CAP teoremi dağınık bilgisayar sistemleri için tutarlılık, ulaşılabilirlik ve bölünebilme toleransının garanti olmadığını ifade eder. Çünkü tutarlılık paralel ilişkili veri tabanlarında gereklidir, bu veri tabanları iyi derecede ölçülebilir ve mevcut değildir. Büyük veri analizleri için yüksek sistem ölçülebilirliği en önemli gerekliliklerdir ve son derece genişletilebilir veri analiz teknikleri geliştirilmiş olması gereklidir.



Şekil 5.16 Makine öğrenme algoritmaları. (Kaynak: <https://www.geeksforgeeks.org/top-10-algorithms-every-machine-learning-engineer-should-know/>)



Şekil 5.17 Veri madenciliği algoritmaları. (Kaynak: <https://data-flair.training/blogs/data-mining-algorithms/>)

5.2.4 Yiğin Veri Madenciliği

Veri setleri boyutlarının artmasıyla büyük veri için çok daha fazla makine öğrenme algoritmaları ve veri madenciliği algoritmaları ortaya çıkmıştır. Araştırmalar göstermiştir ki daha büyük veri setleri için makine öğrenimi daha doğrudur ve makine öğrenme algoritmaları arasında çok az farklar vardır.

Ancak geniş veri setleri veri madenciliği algoritmaları için olduğu kadar geleneksel makine öğrenme algoritmaları için de problemdir. Aslında en geleneksel veri madenciliği algoritmaları büyük veri setleri tarafından geçersiz hale getirilir. Bununla birlikte veri seti boyutları petabaytlar seviyesine ulaştığında, seri algoritmalar kabul edilebilir bir süre içerisinde hesaplanmak konusunda başarısız kalabilir. Etkili makine öğrenme ve veri madenciliği algoritmaları büyük veri setleri için geliştirilmek zorundadır.

KAYNAKLAR

Computing Research Association, 2011, “Challenges and Opportunities with Big Data.”
<http://www.cra.org/ccc/files/docs/init/bigdatawhitepaper.pdf>, [Ziyaret Tarihi: 12 Ağustos 2014].

Gartner Büyük Veri web seminerinde veri, <https://www.gartner.com/doc/2703617/customer-analytics-art-possible-big>, Mart 2013 ve 2014

<https://www.dasca.org/world-of-big-data/article/big-data-value-potential>

http://wikibon.org/wiki/v/Enabling_Real-time_Big_Data_Processing_with_Xflash_-_Very_Low_Latency_Server_SANs

<http://www.sciencedirect.com/science/article/pii/S0020025514000346>

<https://www.openpr.com/news/1128600/big-data-in-healthcare-market-medical-insights-healthcare-analyzed-industry-trends-major-players-research-2023.html>

<https://tdwi.org/Blogs/TDWI-Blog/2013/10/Four-Big-Data-Challenges.aspx?m=1>

https://www.researchgate.net/figure/Volume-scale-of-Data-from-different-data-sources-26_fig1_324220787

http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, 2013

<https://blog.ipswitch.com/global-data-privacy-laws-us-eu-china-andmore>

https://www.researchgate.net/figure/Problems-caused-by-poor-data-quality-Adapted-from-8_fig2_293015090

<https://www.softwaretestinghelp.com/data-storage-companies/>

https://www.researchgate.net/profile/David_Corey3/publication/274257509/figure/fig10/AS:324996351184899@1454496576237/Evaluating-data-processing-efficiency-of-RC-Seq-by-simulation-A-The-problem-the.png

<https://www.geeksforgeeks.org/top-10-algorithms-every-machine-learning-engineer-should-know/>

<https://data-flair.training/blogs/data-mining-algorithms/>

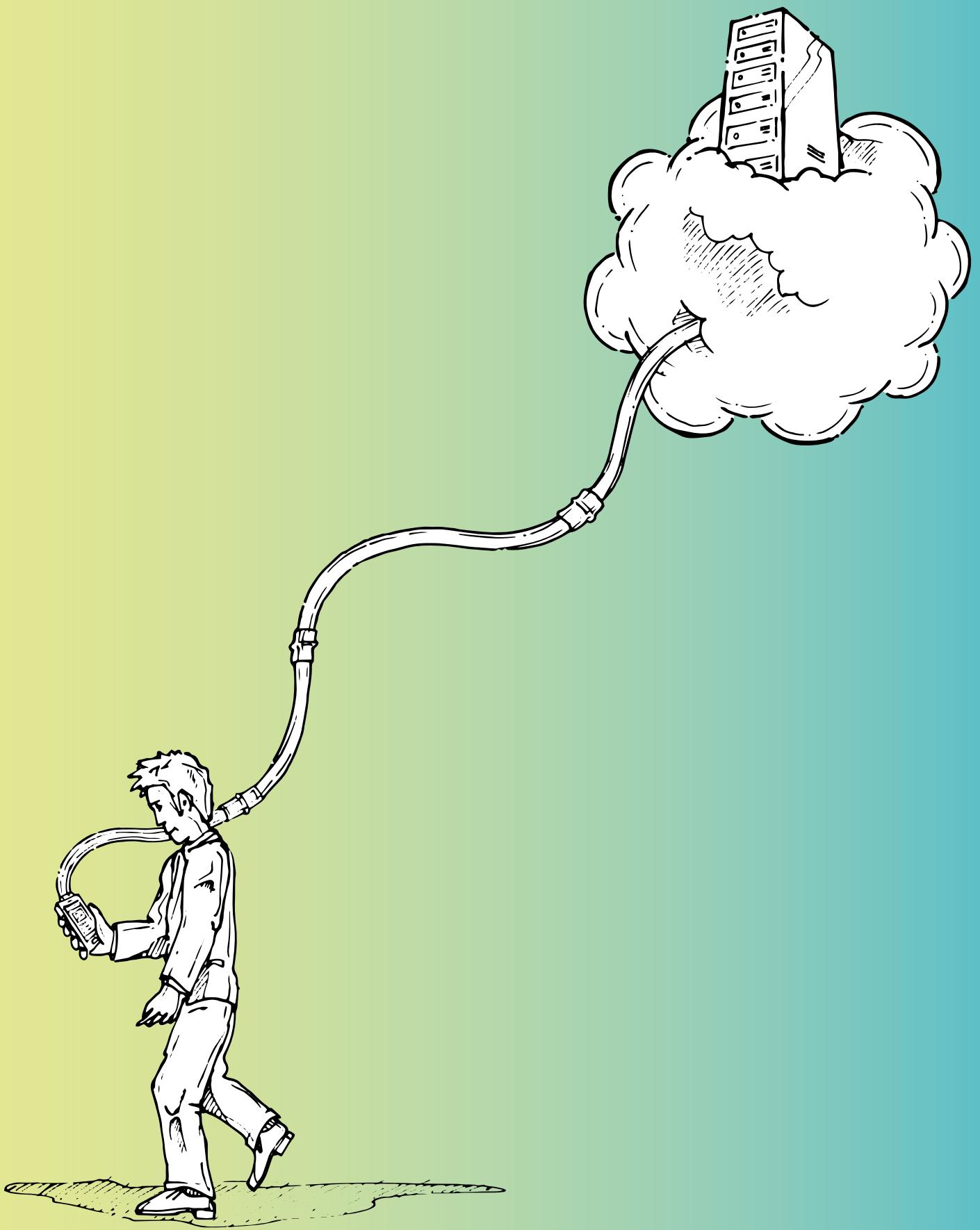
Mayer-Schonberger, V. ve Cukier, K., 2013, Big Data: A Revolution that will transform how we live, work and think, John Murray An Achette UK Company, ISBN: 978-1-84854-790-2.

McKinsey Global Institute Report, 2011, Big data: The next frontier for innovation,

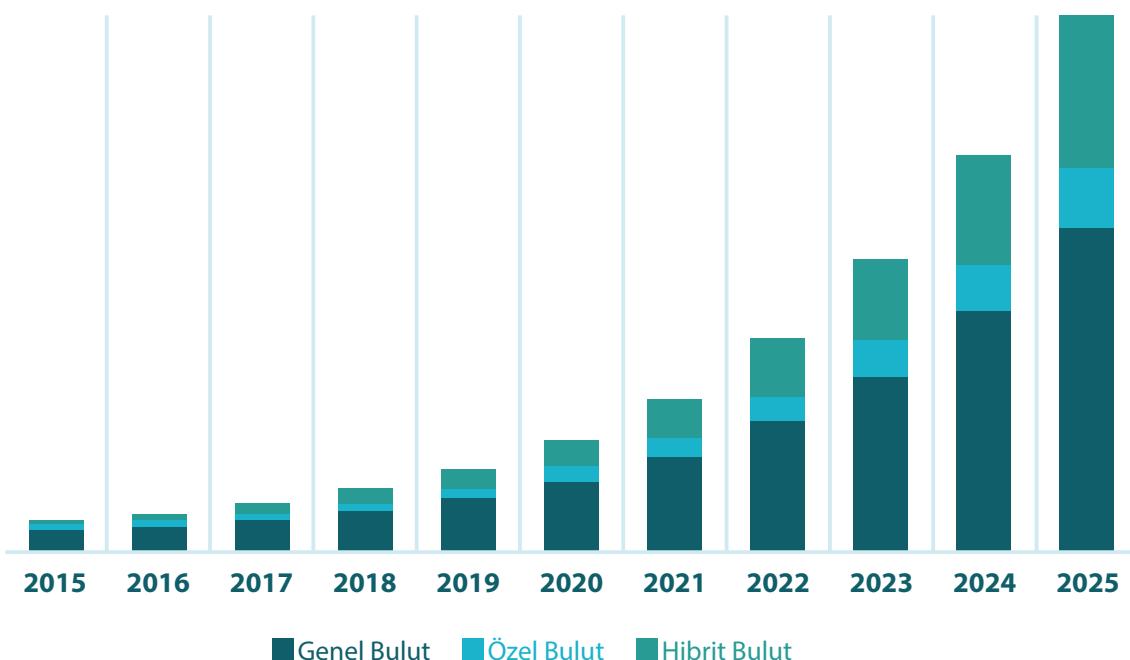
competition, and productivity, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, [Ziyaret Tarihi: 2 Mayıs 2015].
Newvantage Partners Big Data Executive Survey, 2017

6. BÖLÜM

BÜYÜK VERİ UYGULAMA ALANLARI



Internetin çok fazla kişi tarafından kullanılır olması, günlük yaşamda aldığımız hizmetlerin internet üzerine taşınan uygulama yazılımları yoluyla kolayca erişilebilir olması nedeniyle oluşan yaygın kullanım, özellikle hizmet sektöründe satış sonrası müşteri memnuniyeti sağlama gereğine yönelik olarak son kullanıcıların her türlü bilgisinin firmalarca alınıp saklanması sonucunu doğurdu. Saklanacak ayrıntılı bilgilerin, üzerinde tutulacağı sayısal ortamlarda baş gösteren “yer darlığı” yeni arayışlar için başlangıç oldu. Son birkaç yılda öne çıkan yeni yapıyı özetlersek, işletmeler müşterilerini daha iyi tanıyip onlara “bireyselleştirilmiş-kİŞİYE ÖZEL” hizmetleri sunabilmek için onlarla ilgili çok sayıda bireysel bilgiyi saklamak durumunda kalmaktadır.

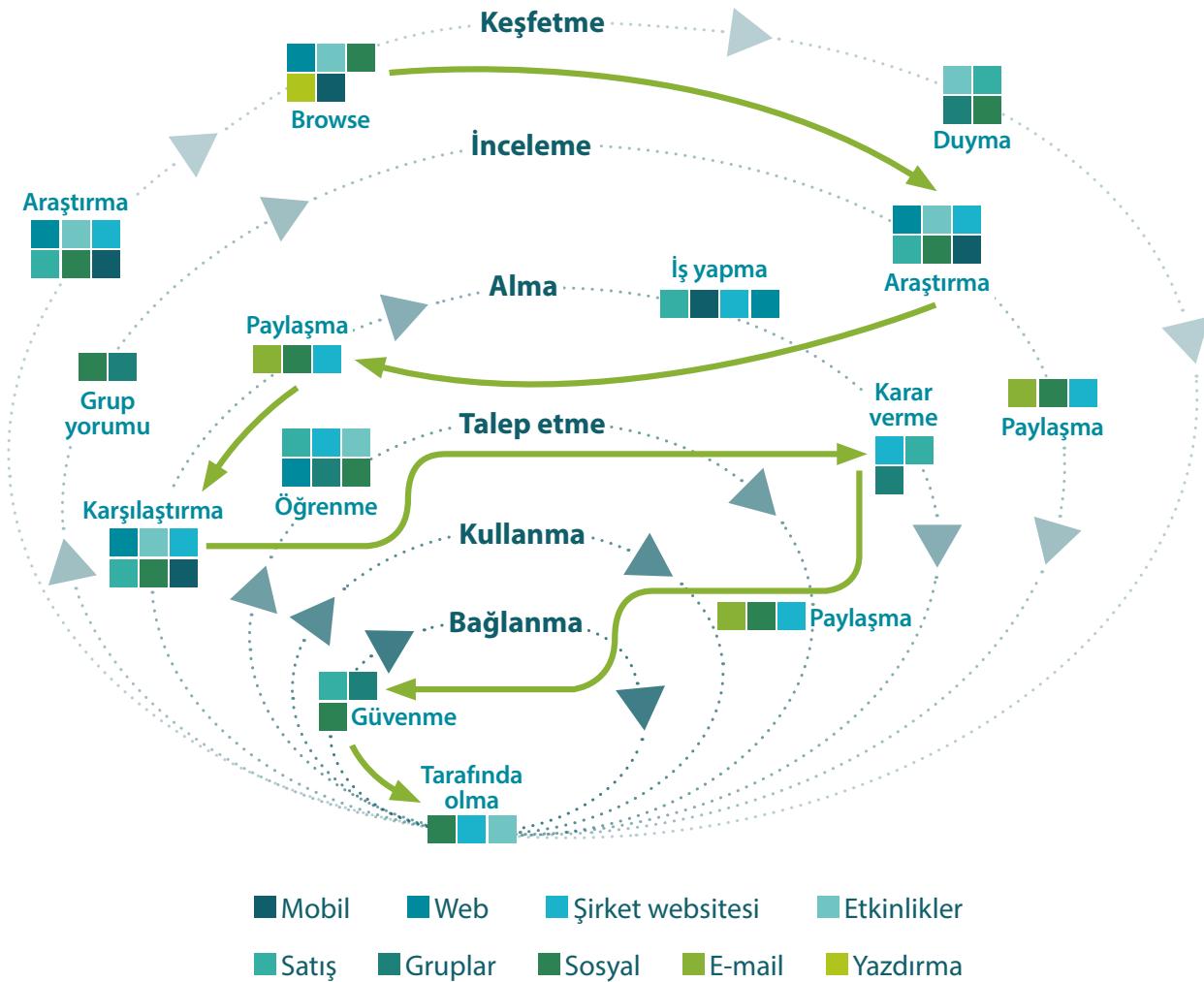


Şekil 6.1 Bir hizmet pazarı büyüklüğü olarak gelişmesine göre Kuzey Amerika'da büyük veri, 2012-2025 (Milyon ABD Doları). (Kaynak: <https://www.grandviewresearch.com/industry-analysis/big-data-as-a-service-bdaas-market>)



Şekil 6.2 1 dakikada internette neler oluyor? (Kaynak: <https://www.byteant.com/blog/7-ways-how-to-use-big-data-in-social-media/>)

Spesifik olarak bakarsak örneğin sosyal medyada süregelen reklam savaşları yüksek hacimli verinin en önemli “müşterilerinden” biridir. Facebook, Twitter, Instagram ya da müzik dinleme hizmeti sunan birbirinden farklı ya da benzer mecrə reklam verenler için durmaksızın veri toplar. Kullanıcı alışkanlıklarını belirlemek, bu alışkanlıklar gerektiğinde paraya, gerektiğinde ise kısa dönemde kâr sağlayacak stratejilere çevirmek asıl amaçtır. X bir sosyal mecrayı kullanan bir kullanıcı, Y bir sayfaya gelmeden önce neler yapmıştır, incelediği sayfada bulunma sebebi nedir, kullanıcı ne yapılrsa reklam verenin arzu ettiği alana gelir, bu noktada istenilen ürüne yönlendirilebilir mi, nasıl yönlendirilir? İşte, bunların hepsi yüksek hacimli verinin işlenmesi ile ortaya çıkan sonuçların işidir.



Sermaye açısından büyük verinin önemi, aşağı çıkardığı ticari veri havuzundan anlaşılabilecektir. Günümüzde müşteri taleplerine hızla yanıt vermek, reel ve potansiyel tüketicilere karşı şeffaf bir kurum kültürü oluşturmak, yeni müşteri deneyimleri oluşturmak amacıyla fiziki ve sanal dünya arasındaki bağlara yönelik inovasyona öncülük etmek ve veri analitiği ile ilgilenmek, sıradan şirketlerin bağımsız olarak gerçekleştirebilecekleri bir işlem değildir.

Forrester anket verilerinden elde edilen bulgulara dayalı olarak oluşturulan grafik, ticari alıcıların % 74'ünün araştırmalarının yarısından fazmasını çevrimdışı bir satın alma yapmadan önce çevrimiçi olarak gerçekleştirdiğini gösteriyor.

IBM'in 56 ülkeden 524 şirket üst düzey yöneticisinin katılımıyla gerçekleştirdiği araştırmada, yöneticilerin % 35'i çoklu veri kaynaklarını entegre etmede zorlandıklarını, % 30'u analitik yaklaşımı nasıl uygulayacaklarından emin olmadıklarını ya da bunu yapacak yeteneklerinin bulunmadığını ve % 22'si veriye hiç erişemediğini ya da yetersiz oranda eriştiğini ifade etmiştir (Bloomberg Businessweek Türkiye, 2014). Bu kapsamda "büyük verinin nasıl analiz edileceği ve yönetileceği" sorusu gündeme gelmekte, yanıtı ise "çok uluslu şirketler tarafından kontrol edilen ticari veri havuzu aracılığıyla" şeklinde ifade edilmektedir. ABD Patent Ofisi (USPTO) enformasyon yönetiminin toplumsal kontrolünün temel göstergelerinden biri olarak görülmektedir. USPTO tarafından kayıt altına alınan patent sayısına baktığımızda 2012 yılında IBM'in 6457, Samsung'un 5043, Canon'un 3173, Sony'nin 3017, Matsushita Electric'in (Panasonic) 2748, Microsoft'un 2610, Toshiba'nın 2415, General Electric'in 1650, LG Electronics'in 1617 ve Fujitsu'nun 1527 patent aldığı tespit edilmiştir (USPTO, 2014). Aynı dönemde (2012 yılında) Türkiye'de tescil edilen yerli patent sayısının 1025 ve bu rakamın % 6,8'inin elektrik-elektronik alanında olması ise büyük veri havuzundan ulusal ölçekte ne düzeyde yararlanıldığının açık göstergesidir.

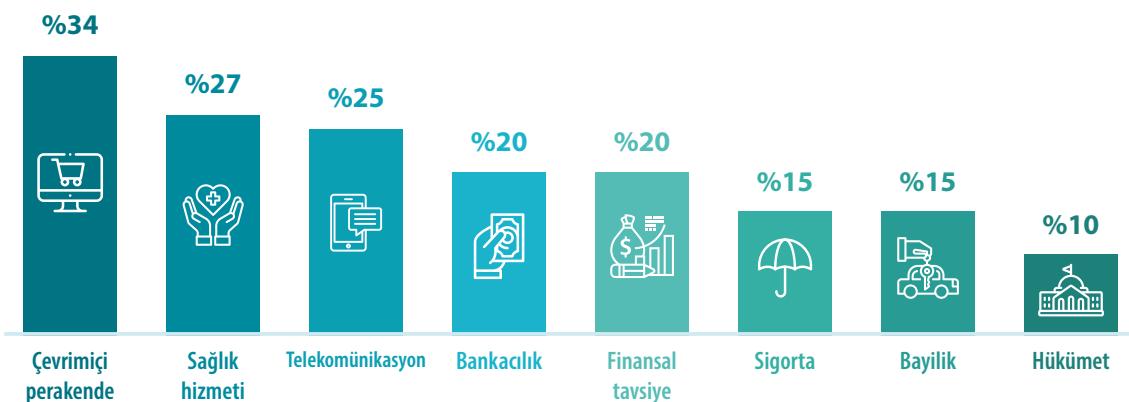
Büyük verinin sermaye açısından işlevsellliğini sağlayan bir diğer konu ise yeni iş modelleri geliştirilebilmesine imkân tanımıştır. Büyük veri müşterilere, organizasyonlara, ürünlere, pazarlara ve diğer tüm ticari unsurlara ilişkin ayrıntılı analizlerle büyük bir ekonomik değer açığa çıkarmaktadır. Büyük veri ile öngörlülebilir pazar bölümlendirmesi yapılması, tüketici karar verme süreçlerine ilişkin karmaşık analiz tekniklerinin kullanılması ve mal ya da hizmet geliştirme/iyileştirme çalışmalarında yeni yöntemlere başvurulması mümkün hale gelmiştir (Gürsakal, 2013).

Şirketler veri madenciliği yoluyla selin önüne file germekte ve fileye takılanları analiz ederek ticari kararlar almaktadırlar. Sahip olunan veriler ile sorunların nasıl çözümleneceğinin öğrenilmeye başlanması şirketleri "ürütim yönetimi," "pazarlama," "insan kaynakları" ya da "muhasebe-finans" departmanlarının elinden kurtarak "bilgi işlem" birimlerine doğru yönlendirmektedir.

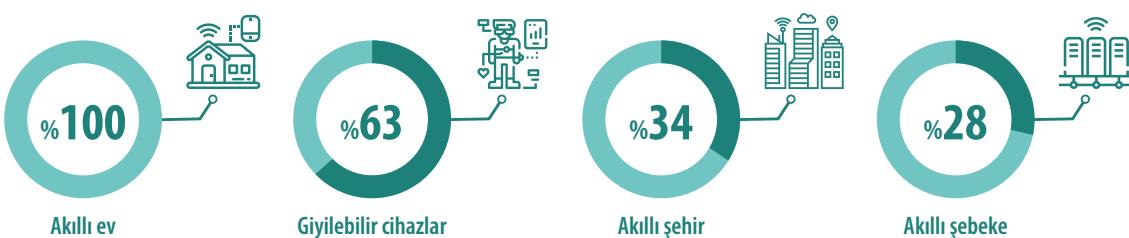
2012 yılının sonlarında ABD merkezli teknoloji araştırma şirketi Gartner, büyük verinin ilerleyişini destekleyecek biçimde 2015 yılı itibarıyla küresel ölçekte tüm sektörleri kapsayacak biçimde 4,4 milyon enformasyon teknolojisi (IT) temelli istihdam geleceği öngörüsünde bulunmuştur. Gartner, yetişmiş personel eksikliği nede-

Bilişim Teknolojisinde Bilmeniz Gereken Başlıca 3 Trend

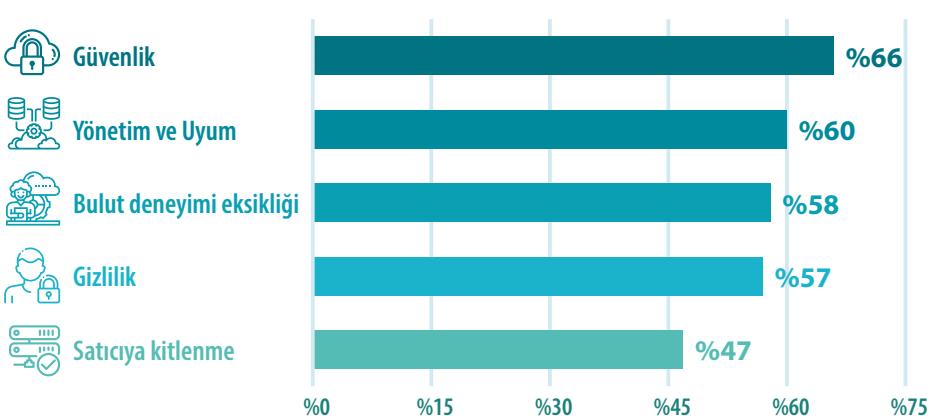
1 AI sohbet robotlarının hizmete göre artan kabulü



2 IoT cihazlarının artan uygulaması



3 Bilişim yöneticileri için onde gelen genel bulut zorlukları



Şekil 6.4 Bilişim teknolojisinde bilmeniz gereken başlıca 3 trend. (Kaynak: <https://financesonline.com/it-trends/>)

nileyle bu pozisyonların sadece 1/3'ünün dolacağını ifade etmiştir. Veri uzmanlığının, piyasa yeterli sayıda veri uzmanı yetiştirinceye ve nihai mamul üreticisi şirketler ileri teknoloji, finans ve biyoteknoloji sektörlerindeki şirketlerle rekabet edebilecek düzeye gelinceye degen, orta vadede “kıt ve değerli bir ürün” olacağını belirtmiştir (*The Economist*, 2013, s. 7-8).

Şunu da belirtmek gerekir ki büyük veri tüm ölçekteki organizasyonlar tarafından kullanılmaktadır. Büyük çaplı firmalara ek olarak küçük işletmeler, girişimciler ve devlet ajansları büyük veriye arzu duyan kullanıcılardır. FTC (Federal Ticaret Komisyonu) Başkanı Bayan Edith Ramirez bu konuyu şöyle ifade etmektedir: “Büyük veri artık birkaç dev şirketin uzmanlık alanında değildir.” Sonuç olarak daha küçük firmalar tarafından kullanımı yeni rekabet şekli ile tanıştıracak bir potansiyele sahiptir ve yerleşmiş endüstrileri rahatsız etmektedir.

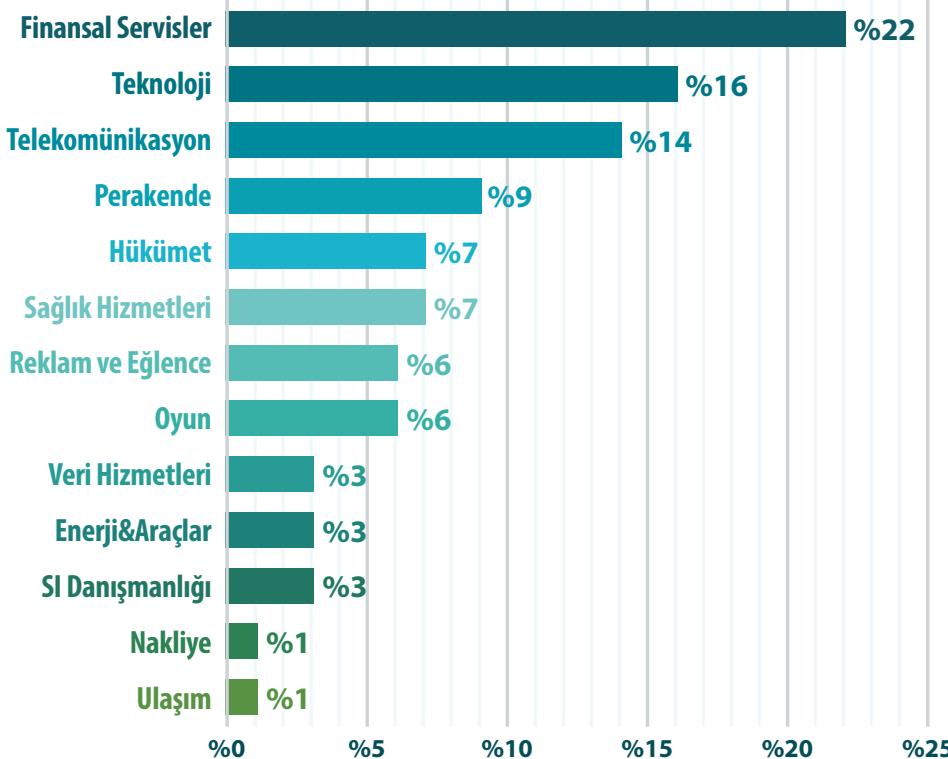
“Büyük veri” yönetiminin nerelerde kullanıldığıyla ilgili birkaç örnek olarak şunları sayabiliriz.

6.1 FİNANSAL SERVİSLER

Bu alan belki de büyük veriden faydalanan en geniş alanlardan biridir. Mobil para hizmetlerinden toplanan veri sektörler ve bölgeler arasında harcama ve tasarruf alışkanlıklarını içerisinde derin kavrayış sağlar. Dijital ödeme tarihleri, emanetler ve diğer kredi tabanlı finansal servisler için talep ettikleri kredi tarihlerini yapılandırmak için kişilere imkân verir.

Bankalar, müşterileriyle ilgili olarak toplayıp sakladıkları bilgiler yoluyla kullanıcılarını tanıyan, internet şubesine o gün ne için girdiğini bilen ve buna göre ana sayfayı, menüyü en etkin hale getiren, müşterisine hatırlatmalar yapan, özelleştirilebilir arayüzler sunan, zengin içerikli, hızlı ve kullanışlı bir 7/24 şube haline gelmiştir.

Sosyal medya/aglار her geçen gün insan hayatındaki yerini büyütmektedir. Twitter üzerinde 2008 Beijing Olimpiyatları ile ilgili 6 milyon kullanıcı tweet atarken, 2012 Londra Olimpiyatları’nda kullanıcı sayısı 83 katına çıkarak 500 milyonu bulmuş, bir günde atılan tweet sayısı ise 300 binden 400 milyona çıkararak 1333 kat artış göstermiştir. 2013 yılı itibarıyla Twitter’da günlük atılan tweet sayısı 500 milyonu aşmakta, Facebook kullanıcı sayısı 1 milyarı, LinkedIn kullanıcı sayısı 250 milyonu



Şekil 6.5 Sektöre göre büyük veri kullanımı. (Kaynak: <https://www.guires.com/blog/what-is-big-data-analytics-and-why-is-it-important-to-business/>)

geçmiş durumdadır. Bu rakamlar, bu kaynaklardan üretilebilecek yapısal olmayan verinin boyutu hakkında fikir vermektedir. Üretilen bu kadar veri içinden kendilerini ilgilendiren kısmını hızlı bir şekilde filtreleyip ayrıştırma, işleme ve saklama için bankalar büyük veri teknolojilerinden yararlanabilir.

Müşterilerin yaşadığı kötü bir deneyim hızla sosyal medya/ağlar üzerinden yayılabilmekteidir. Ernst ve Young'un yaptığı bir ankete göre, Amerika Birleşik Devletleri'nde müşterilerin % 63'ü bankacılık ürünleri ile ilgili kişisel ağlarındaki insanların fikirlerine güvenmekte ve müşterilerin % 45'i aldıkları hizmetle ilgili sosyal medya üzerinden paylaşımada bulunmaktadır.

Bankalar bu yapısal olmayan verileri içindeki yapısal olan verilerle entegre ederek müşterileri ve hizmetleri ile ilgili düzenlemeler yapabilir, gerçek zamanlı aksiyonlar alabilir. Örneğin banka hakkında kötü bir tecrübe paylaşan müşteriye bunu telafi etmek için bir teklife bulunulabilir veya kredi ihtiyacı olduğunu belirten bir müşteriye



Şekil 6.6 Finans alanında veri bilimin uygulanması. (Kaynak: <https://data-flair.training/blogs/data-science-in-finance/>)

anında uygun bir teklif sunulabilir. Yine müşteri olmayan insanların paylaşımılarına göre ihtiyaçlarına uygun ürünler önerilebilir. Bankalar aynı zamanda sosyal medya üzerinde rakipleri ve ürünleri hakkındaki düşünceleri de yakalayabilir.

Bununla birlikte müşterilerin sosyal medya/ağ erişim bilgilerini bankalar ve diğer kurumlarla paylaşmak istememesi bu noktada karşılaşabilecek bir engel olarak göz önüne alınmalıdır, bankalar bu açıdan paylaşımı teşvik çalışmaları yapabilir.

Çağrı merkezleri bankaların vazgeçilmez kanallarından biri ve müşterilere istedikleri bilginin ve bankacılık servisinin verilmesi ve müşterilerin iyi / kötü geri dönüşlerini almak için kullanılan dağıtım kanalıdır. Çağrı merkezi verimliliği için kullanılan ve yapısal verilerden elde edilen ortalama yanıt verme süresi, ortalama görüşme süresi, ortalama aktif / bekleyen çağrı sayısı birçok metrik olmasına rağmen aslında çağrı merkezinin en büyük verisini oluşturan, yapısal olmayan çağrı merkezi görüşme kayıtlarının analizi oldukça önemlidir.

Ayrıca yine bu alanda dijital ortamlarda dolandırıcılık tespit konusu çok önemlidir. Buna da en çok ihtiyaç duyan bankalar ve benzeri finans kuruluşlarıdır. Bankacılık gibi kritik bir finansal sektörde en önemli konulardan biri güvenliktir. Her banka



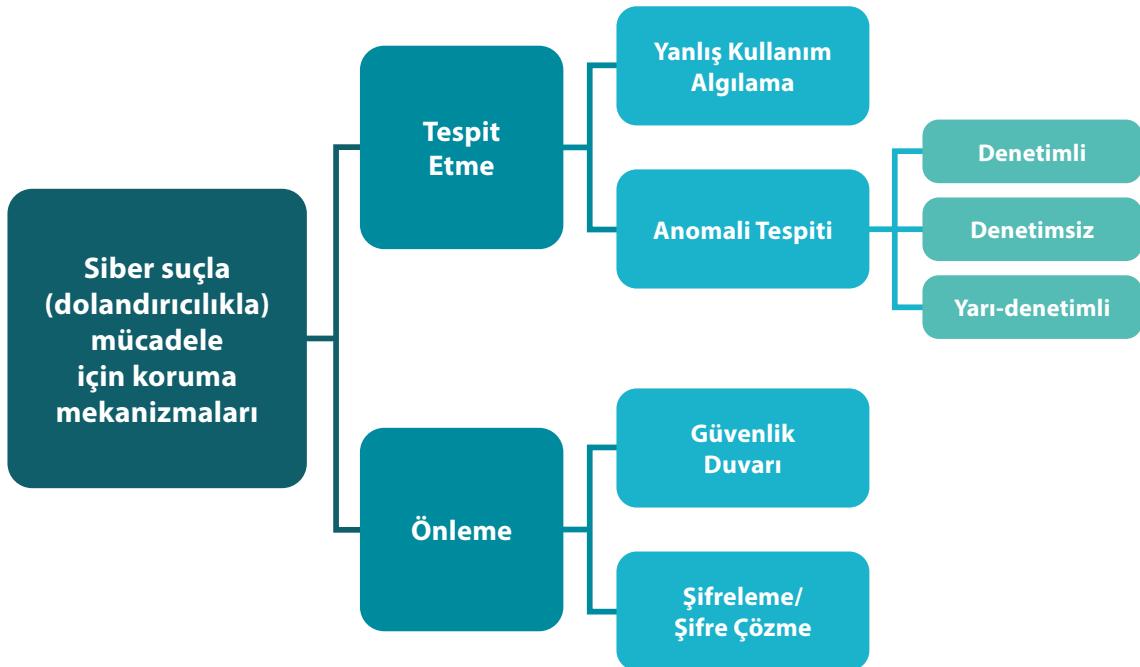
Beyaz renkli rakamlar önceki yılın toplamındaki yüzde değişimini göstermektedir.

Şekil 6.7 2004-2014 çevrimiçi bankacılık dolandırıcılık kayıpları. (Kaynak: http://www.theukcardassociation.org.uk/plastic_fraud_figures/)

İçin her an dolandırıcılık tehdidi mevcuttur. ABD Adalet Bakanlığı Tüketiciler Koruma Ağı'nın verdiği rakamlara göre, dünya çapındaki toplam kredi kartı dolandırıcılığı 5,5 milyar dolar olup her geçen gün artmaktadır. Bu yüzden bankalar dolandırıcılığı tespit etmek ve engellemek için sistemler geliştirmeye ve bunları sürekli iyileştirmeye yönelik büyük yatırımlar yapmaktadır.

Dolandırıcılık tespitinin gerçek zamanlı yapılabilmesi için çeşitli kanallardan yüksek hızda gelen büyük hacimdeki verilerin gerçek zamanlı olarak işlenerek çırakımlar yapılması gerekmektedir. Geleneksel sistemler genellikle geçmişe yönelik bu tarz durumların tespitini yapabilmektedir. Ancak büyük veri teknolojileri ile satış noktası, sosyal medya, müşteri veri tabanları ve veri sağlayıcılarından alınan yüksek hacimli veriyi kullanarak gerçek zamanlı tespit ile yanında aksiyon alınabilmesi mümkün olmaktadır.

Büyük veri teknolojilerinin dolandırıcılığı önleme için kullanımında öncelikle bankalarda bulunan verilerin analizi, müşterilerin aktivite örüntülerinin çıkarılması



Şekil 6.8 Dolandırıcılık tespit sistemi. (Kaynak: <https://www.sciencedirect.com/science/article/abs/pii/S1084804516300571>)

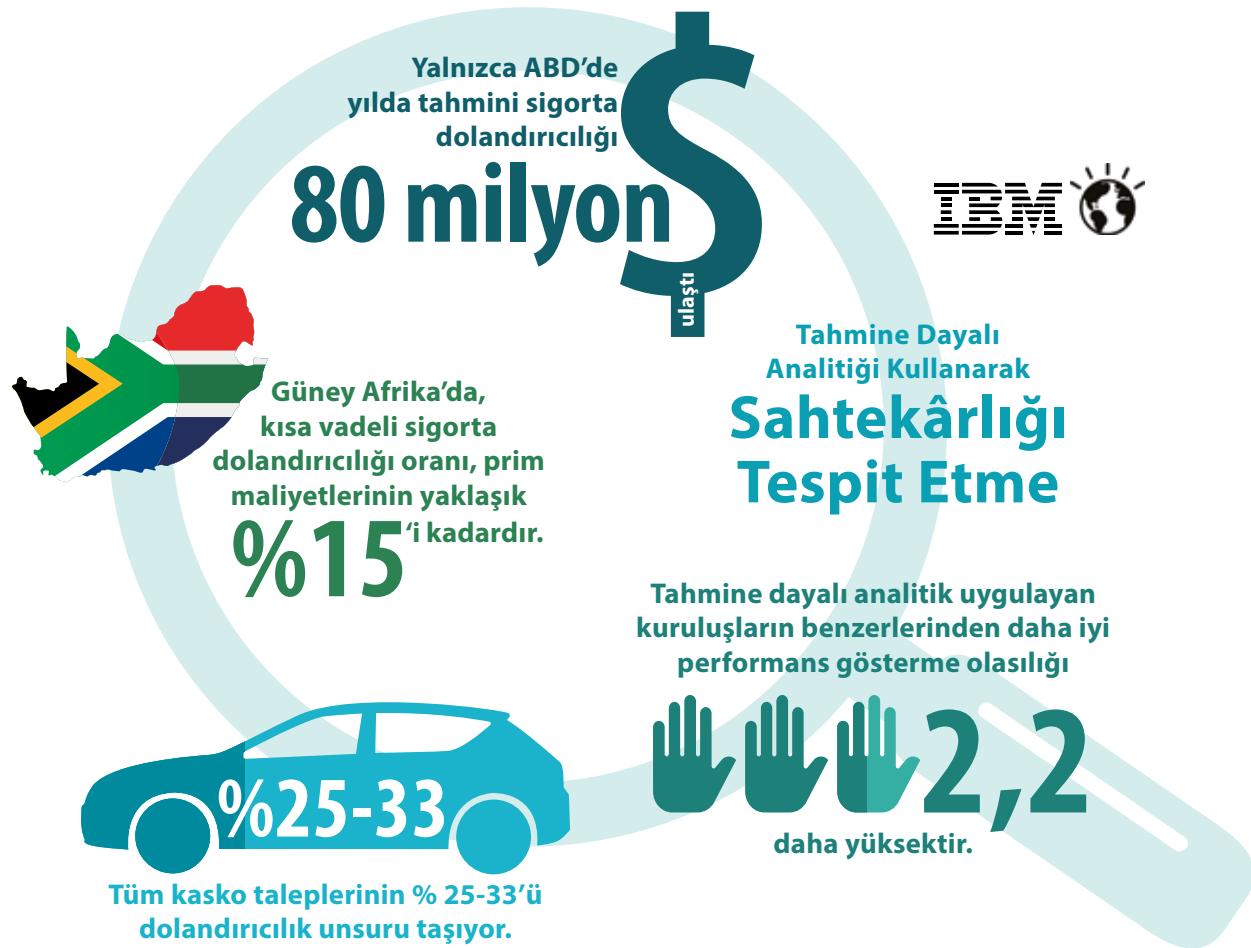
önerilmektedir. Bir kullanıcının hangi periyotta hangi kanaldan ve hangi lokasyondan bankaya giriş gerçekleştirdiği, hangi işlemi yaptığı, kullanıcının tuş dinamikleri gibi bilgiler dolandırıcılık tespitinde kullanılan başlıca yöntemlerdir. Aktivite örüntülerinin çıkarılması esnasında bankacılık kanallarının izole olarak ayrı ele alınmamaları, tespit etmesi zor çapraz kanal dolandırıcılıklarını yakalamak için bütün kanallardan yapılan işlemlerin göz önüne alınması gerekmektedir. Büyük veri ayrıca davranışsal kimlik doğrulamaya da olanak sağlamaktadır.

Dolandırıcılık tespitinde hileli aktivitelerin normal aktivitelerden ayrılmasında ayrıca mobil veriler ve sosyal ağlardaki paylaşılardan da yararlanılabilir. Örneğin bir banka müşterisinin tatil çıkışındaki harcamalarında bankanın dolandırıcılık izleme sistemi müşterinin mobil veri ve sosyal ağlarda yapmış olabileceği tatil ile ilgili paylaşımıları da göz önüne alırsa yanlış pozitif sonuçların üretilmesi önlenir. Bununla birlikte dikkate alınması gereken bir konu da müşteri gizliliği ile ilgili yasal düzenlemelere uygun bir şekilde yasal sınırlar içinde kalmaktır. Dolandırıcılık tespitinde izlenen yaklaşım aynı zamanda kredi riski için de izlenebilir.

Bankaların geleneksel veri ambarı modellerinde çok büyük miktarda veri, günlük yüksek hacimde veri aktarımı, saatlerce süren sorgular mevcuttur. Bu yüzden bankalar sorgu performanslarında ciddi kazançların yollarını aramaktadır. Ayrıca bankaların geleneksel veri ambarı yaklaşımlarında mevcut durumda cevap aradıkları sorular için bazı çıkarımlar yaparak o çıkarımlara gerek duydukları verileri saklama, şu anda cevap aramadıkları konularda kalan verileri silme eğilimi bulunmaktadır. Bu eğilime neden olarak verilerin saklanması maliyeti, korunma ve yönetim zorlukları ve veri boyutu büyündükçe sorgulama performansının kötüleşmesi örnek verilebilir. Ancak değişebilen iş ve piyasa koşulları ileride bankalar için cevabını aradıkları yeni sorular ortaya çıkmasına neden olabilmekte ve bu silinen verilerden çıkarımlar yapılmasına ihtiyaç olabilmektedir. Büyük veri teknolojilerinin sağladığı imkânlar ile minimum veri silme, büyük miktarda veri saklama ve istenilen bilgiye tüm bu verilerden erişilebilme sağlanabilmektedir.

Var olan müşterileri elinde tutmak ve yeni müşterileri cezbetmek bankaların en öncelikli hedeflerinden biridir. Banka müşteri verilerinin kapsamlı olarak öngörüllererek yapılan analizi ile sunulacak yeni ürün, servis ve hizmetler ile rekabet avantajı ve yeni gelir kapıları elde edilebilir. Bankalar her ne kadar hesap hareketleri ve bölümeleme ile müşterileri hakkında bilgi sahibi olsalar da büyük veri teknolojileri ile birlikte müşterilerin davranış ve ihtiyaçları daha derin ve öngörü ile analiz edilerek teklif optimizasyonu ve çapraz satış gibi konularda fark yaratılabilir, gerçek zamanlı kişiselleştirilmiş ürünler ve servisler sunulabilir.

Geleneksel bankacılık sistemlerinde pazarlama ve ürün geliştirme departmanlarının müşteri verilerine erişimi kısıtlıdır. Bu yüzden örneğin yeni bir kredi kartı ürünü geliştirileceği zaman hangi müşterilerin kredi kartının olduğu, dahası bu müşterilerin başka bankalardan daha çok harcama yaptıkları başka kredi kartlarının olup olmadığı bilgisini çıkarmak oldukça zordur. Fakat birçok veri kaynağından veri alan yüksek veri içeren Hadoop kümeli veri ambarı dahilinde, bunun ve hangi müşterilerin hem mobil bankacılığı kullanıp hem de çek bozdurmak için şubeye uğradığı sorusunun cevabı kolaylıkla alınabilmektedir. Ayrıca çapraz kanal aktiviteleri izlenerek müşterilerin ne kadar dijital olduğunu gösteren yaşam stili skoru modeli geliştirilebileceği bilgisi paylaşımaktadır. Bu tarz bir veri ambarında edinilebilecek bilgilere başka bir örnek olarak bankayla çalışmayı bırakılan müşterile-



Şekil 6.9 Büyük verinin finansta kullanımı. (Kaynak: <https://datasciencegeyan.com/big-data-use-cases-in-financial-services/>)

rin aktivitelerinin analizi sırasında fatura ödeme işlemi gerçekleştirmeyi bırakan müşterilerin genellikle yaklaşık 6 ay içinde bankaya çalışmayı bıraktığının tespit edilmesi verilebilir. Buna göre artık fatura ödeme işlemini yapmayan bir müşterinin bankaya çalışmayı bırakabileceğinin öngörüsü ile banka müşterisini kalmaya ikna edecek kişiye özel faaliyetlerde bulunabilir.

İş analitiğinin hızlandırılması da büyük verinin getirdiği çok büyük bir avantajdır. Fortune 20 şirketlerinden biri büyük veri teknolojilerinden yararlanarak risk raporlama süresini 45 dakikadan 45 saniyeye indirerek daha hızlı karar verebilmeyi ve iş yapabilmeyi sağlamıştır. Ayrıca tüm bu verilerin iç süreçlerdeki detaylı analizi ile operasyonel verimlilik sağlanabilir.

Bankalar, müşterilerle daha iyi etkileşim kurmak için büyük veri analizi ve gerçek zamanlı işletim üzerine oynuyor

(önümüzdeki iki yıl içinde hangi hedeflerin ve teknolojilerin önemli olacağını belirten küresel bankacılar)

%81

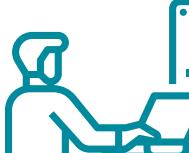
Büyük bankaların % 81'i müşteri merkezliliğine öncelik veriyor



Her 2 yöneticiden 1'i müşteri stratejilerini desteklemek için olgunlaşmış yeteneklere sahip olmadığını

her 2'nden 1'i

...ve teknolojinin müşteri odaklı bir stratejide önemli bir rol oynadığını düşünüyor:



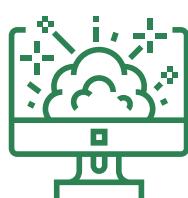
%59

Veri ve analitinin gerçek zamanlı işlenmesi



%59

Tahmine dayalı analitik



%54

Veri görüntüleme

Şekil 6.10 Önümüzdeki iki yıl içinde hangi hedeflerin ve teknolojilerin önemli olacağını belirten küresel bankacılar. (Kaynak: <https://www.ciklum.com/blog/what-are-the-big-data-trends-in-banking/>)

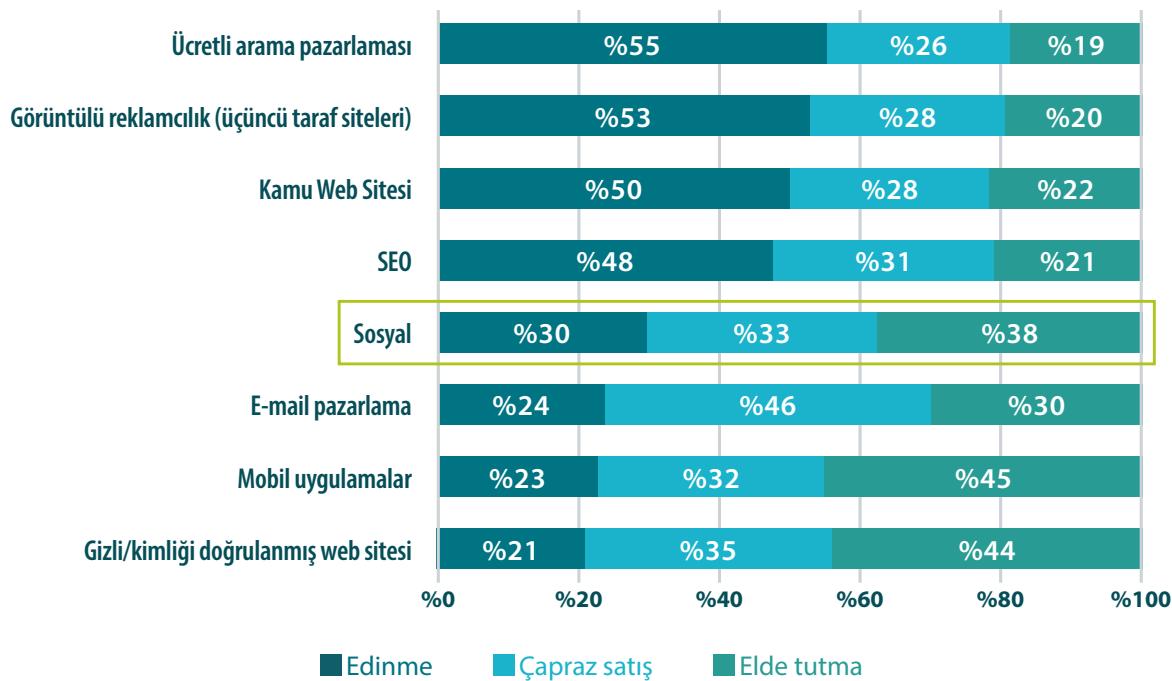
Yine bir finansal kurum içindeki sistemle ilgili tutulan birçok veri vardır. Kullanıcıların dosya sistemi erişim bilgileri, internet erişim bilgileri, veri erişim bilgileri, e-postalar, güvenlik kayıtları, uygulamaların ve donanımların ürettiği sistem günlükleri, değişiklik ve erişilebilirlik günlükleri örnek olarak verilebilir. Yasal zorunluluklardan ötürü yapılan dış denetimlerde tüm bu verilerin incelenmesi, çıkarımlarda bulunulması gerekmektedir. Geleneksel veri tabanı araçları bu noktada zorlanmakta ve günlerce çalıştırılan sorgularla karşılaşılabilir ve arşivlenmiş verilerin erişilebilir hale getirilmesi sürecinde de uzun süren birçok işlem yapılması gerekmektedir. Yine iç denetimciler tarafından sürekli kontroller gerçekleştirilmektedir. İç denetçiler tarafından veri büyülüğu sıkıntısından ötürü örneklenmiş veriler üzerinden bilgi çıkarımı gidilebilmektedir. Büyük veri teknolojilerinden yararlanarak bu büyük miktardaki verilerin saklanması, yönetimi daha kolay ve istenilen bilgiye erişimi çok daha hızlı olmaktadır, böylece tüm veri üzerinden kaliteli denetimler yapılabilmektedir.

6.2 PAZARLAMA VE PERAKENDECİLİK

Hem klasik ticaret hem de e-ticaret alanında büyük veri çok etkin bir biçimde kullanılan araçlardan biridir. Dünyanın en büyük alışveriş sitelerinden alibaba.com, amazon.com, ebay.com müşterileriyle ilgili her türlü bilgiyi satışa dönüştürmek için büyük veri analizlerinden faydalananmaktadır.

Büyük veri pazar bölümlemesi, müşteri değerlendirme ve çapraz satış analizleri, satış noktası veri analizleri, alışveriş sepeti analizleri, tedarik ve mağaza yerleşim optimizasyonları şeklinde kullanılarak satışları ve kârlılığı artırmaya yönelik çalışmalar yapılabilmektedir. Örneğin McKinsey'e göre, tipik bir şirketin gelirinin % 75'i aynı

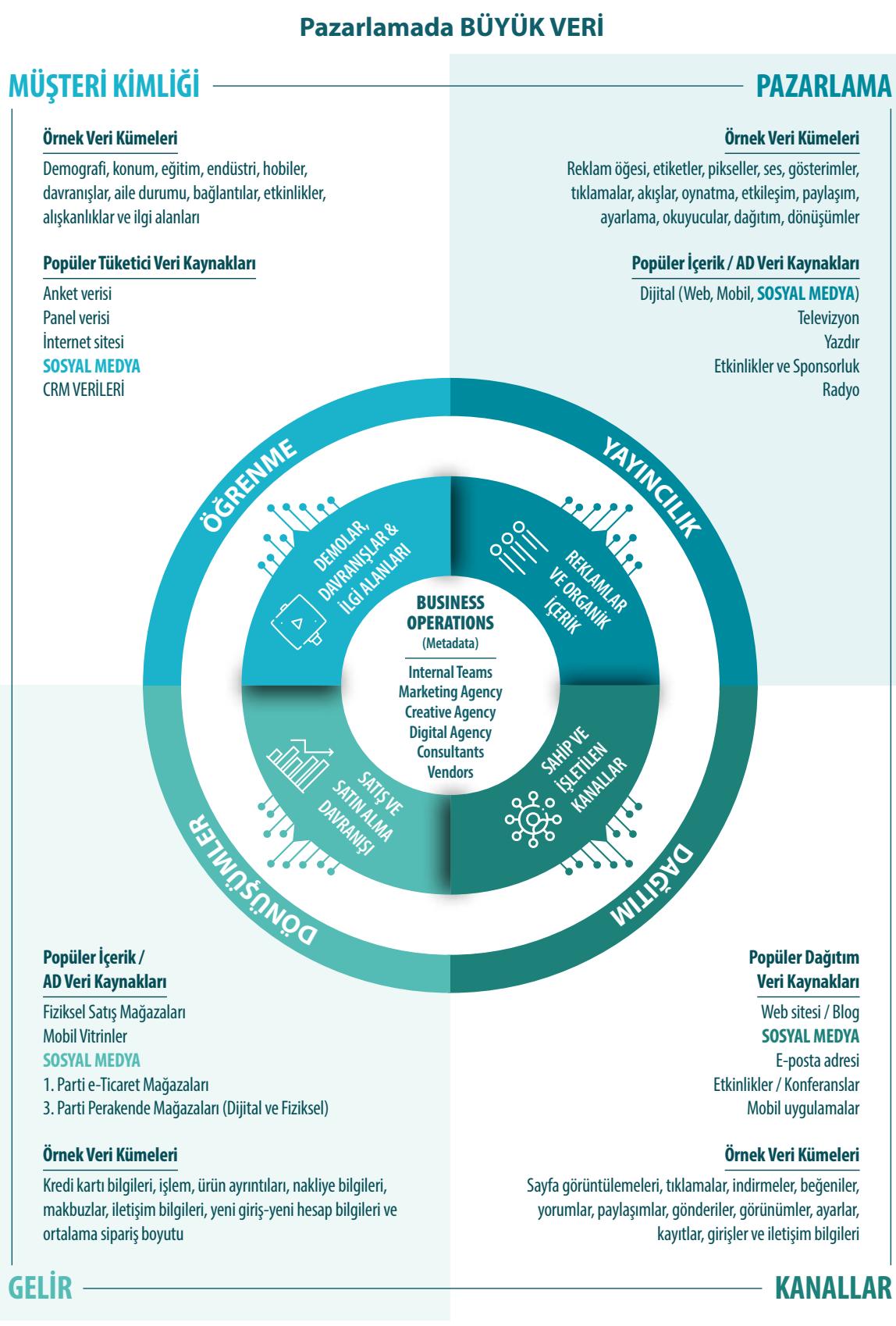
Dünyanın en büyük alışveriş siteleri müşterileriyle ilgili her türlü bilgiyi satışa dönüştürmek için büyük veri analizlerinden faydalananmaktadır.



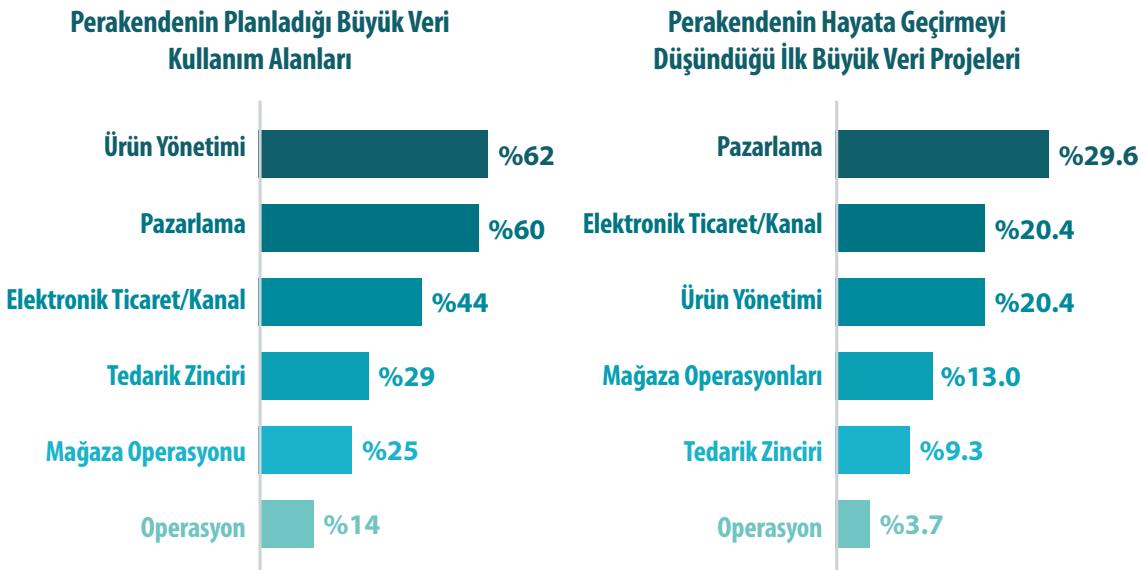
Şekil 6.11 Yukarıdaki dijital kanalların edinme, çapraz satış veya elde tutmaya yönelik olma-lığına dair şe'k. (Kaynak: <https://www.hallaminternet.com/social-media-in-the-financial-industry/>)

anda standart ürünlerinden kaynaklanıyor ve bu şirketlerin her yıl aldığı fiyatlandırma kararlarının % 30'u en iyi fiyata ulaşamıyor. Operasyon kârlarında % 8,7'lik bir artışla sonuçlanan nominal % 1'lik bir fiyat artışında hacim kaybı olmadığı varsayıldığında fiyatlandırma kârlılığı önemli ölçüde artırılabilir. Büyük veri analizi, pazar analistlerinin çok çeşitli pazar faktörlerini incelemesine ve bir ürün veya hizmetin en uygun fiyatına ulaşmak için verileri şirketin hesap beyanları ile birleştirmesine olanak tanır. Bununla birlikte McKinsey'e göre, bir perakende zincirindeki büyük veri analizi, işletme marjını yaklaşık % 60 oranında artırabilir.

Gartner ise 2025 yılına kadar ilk 10 perakende devinin gerçek zamanlı fiyatlandırmadan yararlanacağını tahmin ediyor. Büyük veri analizleri müşteriler için mağaza içi fiyatları ayarlamak için gerçek zamanlı fiyatlandırmaya ulaşmaya yardımcı olacaktır. Bu noktada perakendeciler çeşitli ürünlerin fiyatlarındaki değişimin etkisini analiz etmeye çalışıyor. Büyük veri analizi, perakendecilerin satışlarını artıracak ve böylece maksimum gelir elde edecek en uygun fiyatın tahmin etmelerine yardımcı olacaktır.



Şekil 6.12 Pazarlamada büyük veri. (Kaynak: <https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/>)



Şekil 6.13 Perakende alanında büyük veri kullanımı ve projeleri. (Kaynak: *Retailer's Guide to Big Data, Infographic, 2103*)

6.3 KAMU VE HÜKÜMETLER

Hükümetler yurttaşlarına yönelik bilgi ve hizmetleri işleyip saklama konusunda çok büyük ölçekli veri ile çalışmak zorundadırlar. Örneğin RTÜK kararları gereği, ülkemizdeki televizyon kanallarının son bir yıllık yayınlarını saklama zorunluluğu vardır. Saklanacak bilgiler “büyük veri” olarak tanımladığımız türdendir.

Kamu verisinin potansiyel değeri yüksektir. Bu verinin içinden gizli ve kişisel bilgi içeren kısımlar ayıklandıktan sonra, araştırmacılar özel şirketler ve girişimcilerle paylaşılabilirse bu veriler üzerinde çalışan katma değerli hizmetler önemli ekonomik faydalara sağlayabilir. Kamu verisinin faydaları ekonomiye katkı, çevrenin korunmasına katkı, bilimsel gelişime katkı ve katılımcı demokrasiye katkı ana başlıklarını altında incelenebilir (European Commission, Aralık 2011, Open data, an engine for innovation, growth and transparent governance).

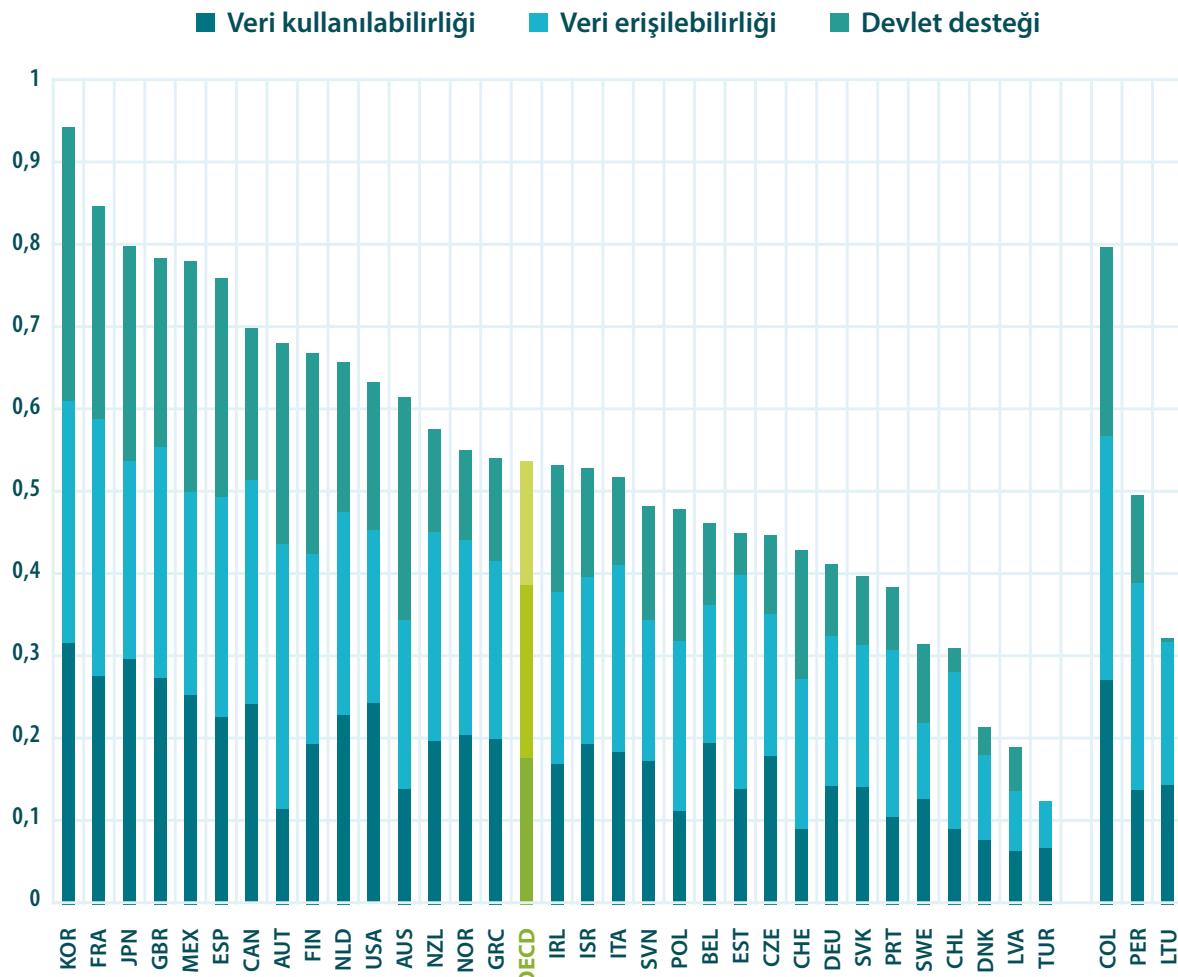
McKinsey Global Institute raporuna göre, kamu büyük verisi kullanımı ABD sağlık sektöründe yılda 300 milyar dolar, AB kamu sektöründe yılda 250 milyar avroluk değer yaratabilecek potansiyele sahiptir (McKinsey Global Institute, 2011, Big data: The next frontier for innovation, competition, and productivity).



Şekil 6.14 Dijital hükümet stratejileri hakkında OECD tavsiyesi. (Kaynak: <https://slideplayer.com/slide/17000001/>)

AB tarafından hazırlatılan başka bir çalışmaya göre ise daha iyi veri altyapısı ve veriye erişim engellerinin azaltılmasıyla kamu sektörü verisinin AB ekonomisine doğrudan ve dolaylı etkisinin toplamı 2008 itibarıyla 200 milyar avro seviyesinde tahmin edilmiştir (Graham Vickery, Information Economics, Paris, 2011, Review Of Recent Studies On Psi Re- Use and Related Market Developments).

Ayrıca günümüzde e-devlet kavramı oldukça kritiktir. E-devlet uzmanlarının en önemli hedefi bilgiye eş zamanlı olarak ulaşmak ve daha iyi hizmet vermektedir. E-devlet uygulaması gerçekleştirilen ülkelerde kamu kuruluşları ziyaretçilerin sayfalarını nasıl kullandığı, ihtiyaç duyulan formlara kolayca ulaşılıp ulaşılmadığı, web sayfa tasarımının nasıl en iyi kullanılabilir hale getirileceği, hangi sayfaların hangi sıra ile ziyaret edildiğinin anlaşılması, geçmişteki ziyaretçi davranışlarına göre kurumun web sayfasını vatandaşın ihtiyacına daha iyi yanıt verecek şekilde nasıl yeniden düzenlenmesi gereği sorularına çözüm aramaktadırlar. Bu anlamda büyük veriden faydalananabileceklerdir.



Şekil 6.15 OECD OURdata Endeksi: Açık, Yararlı ve Yeniden Kullanılabilir Veriler. (Kaynak: <https://www.slideshare.net/OECD-GOV/ou-rdata-index>)

6.4 EĞİTİM

Mobil katma değerli servislerin kullanımından elden edilen veri, kritik bilgiyi yaymak için daha fazla odaklanma ve zamanlama girişimlerine izin vererek bilgi boşluklarını ve eğitim ihtiyaçları kavranarak kamu sektörünün gelişmesi için kullanılabilir.

Eğitim sektöründe, verilerin işlenmesi ve buna ihtiyaç gün geçtikçe artmaktadır. İnternet teknolojisinin gelişimindeki artış eğitim kurumlarındaki “tebeşir ve konuşma” sistemini değiştirdi. İnternet ve bilişim teknolojilerinin etkisiyle web tabanlı uygulamalar hayatı geçirildi.

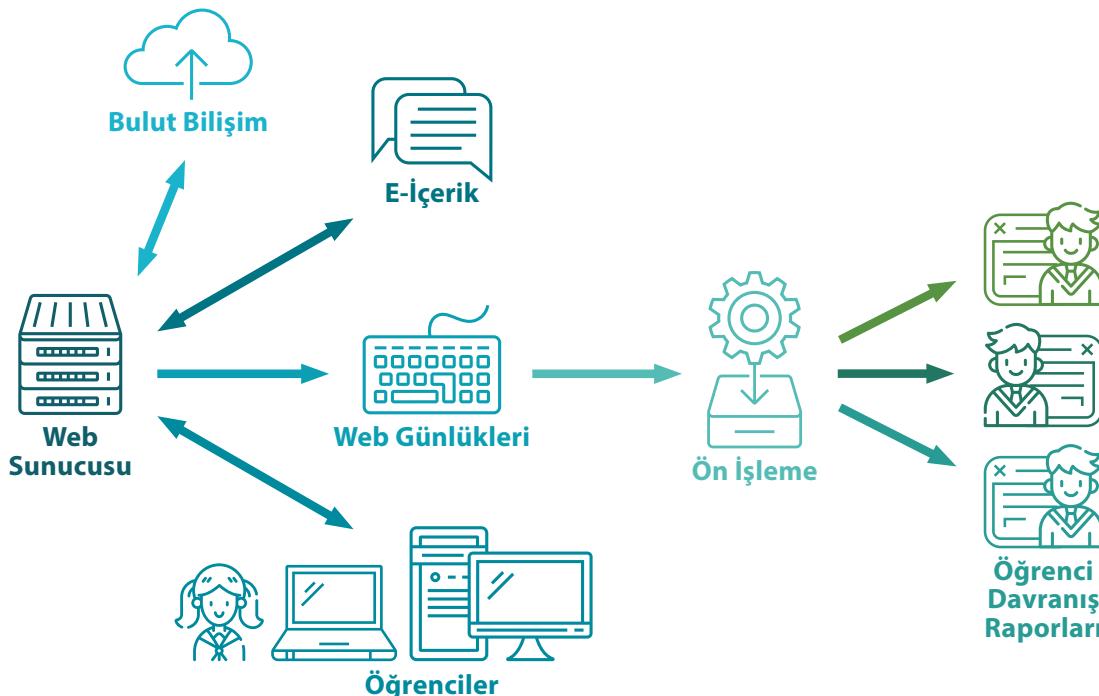


Şekil 6.16 Eğitimde büyük veri. (Kaynak: <https://ismailimail.blog/2015/09/13/aga-khan-university-east-african-institute-big-data-can-look-at-the-whole-child/>)

Modern ve saygın teknik yükseköğretim kurumlarının çoğu şu anda kâğıtsız çevrimiçi eğitim sistemi ile çalışmak için gerekli adaptasyonu sağlamış durumdadır. Bu teknolojilere adaptasyon, yeni nesil eğitim sistemini hayatı geçirmede daha iyi iletişim ve daha gelişmiş yetenekler sunarak baskıyı azaltmakta ve zaman kazandırmaktadır.

Kısaca şu faktörler yükseköğrenim kurumlarında büyük veri neslini oluşturmaktadır:

- **Akademik Trend:** Yükseköğrenim kurumlarının çoğu farklı yönetsel iş amaçları için çeşitli özelleştirilmiş araçlar uygulamaya koymaktadırlar. Örneğin öğrenci kayıt kabulü, ücret toplama, öğretmenler ve diğer çalışanlar için personel takibi, rapor sistemi vb. iç değerlendirme, devam kayıtları, not tabloları, bilgisayar laboratuvarı iş yükleri ve planlama gibi günlük öğrenci aktiviteleri büyük miktarda veri üretmektedir.
- **Performans İzleme:** Kurumların sağlıklı istatistiksel veriler elde edebilmesi için bütün mezun etkikleri öğrencilerinin verilerini de korumaları gerekmektedir.



Şekil 6.17 Öğrenci davranışları analizi yaklaşımı. (Kaynak: <http://research.ijcaonline.org/icaccth-pa2014/number4/icaccthpa6041.pdf>, 2014)

Mezun olan öğrenci raporları devam eden öğrencilerle ilgili öngöründe bulunmakta fayda sağlamaktadır. Öngörü analizlerinin çıktılarından faydalananak öğrencinin yetenekleri belirlenir ve üzerinde durulur, eksik yanları tespit edilir ve tamamlanır. Böylece öğrenciler rüyalarındaki işi elde etmek için en iyi donanıma yükseltilirler.

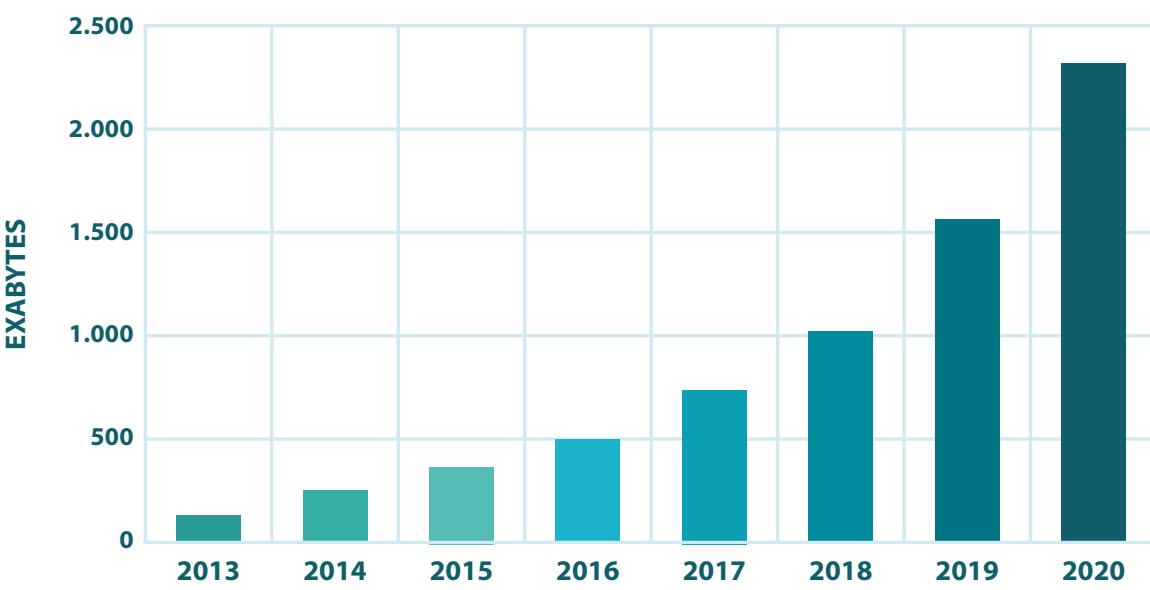
- **Sanal Sınıflar:** Akademik eğitimdeki yüksek rekabetin daha da artması ile öğrenciler yeteneklerini daha etkin bir biçimde güncelleyebilecekleri çeşitli mekanizmalara adapte oluyorlar. Gelişmiş eğitim yazılımlarıyla uzman eğitmenlerin rehberlikleri sanal toplantılar ile kolayca sağlanmaktadır. Çeşitli tanınmış kurumlar şu anda zaten bulut tabanlı sanal sınıflara adaptasyonu sağlamış durumdadır. Bu sınıflarda öğrencilerine en iyi eğitim paketleri ve çeşitli dijital içerikler sunabilmektedirler. Bu dijital içerikler genellikle PowerPoint sunumları ve ses dosyaları halindedir. Bazı durumlarda sanal sınıflar öğrencinin sistemde onaylanması ile kayda izin vererek çevrimdışı de görünebilmektedir. Bu nedenle bu tip sanal sınıflar muazzam veri üretirler ki geleneksel sunucularda barınamaz ve bulut tabanlı depolama sistemlerine ihtiyaç duyurlar.

6.5 SAĞLIK

Mobil aygıtlardan toplanan veri sağlık çalışanlarında elde edilmiş, bireyler tarafından sunulmuş ya da serbest veri şeklindeki analizle salgınları durdurma ya da sağlık eğilimlerini anlamada önemli bir araç olabilir. Bireysel elektronik sağlık kayıtları bağlamında toplandığında, bu veri sadece bireylerin bakımının sürekliliğini geliştirmez, aynı zamanda tedaviler ve çıktılar ile büyük veri setleri oluşturmak için kullanılabilir ve bunları etkin verim ve maliyetle karşılaştırabilir.

Hastaneler hastalarına yönelik etkili, bireysel, kişiselleştirilmiş tıbbi hizmetler sunabilmek için bireysel bazdaki verileri kendi sayısal ortamlarında depolamaktadır. İlaç sanayisinde örneğin “kanser araştırmaları” için oluşturulan büyük genomik veri tabanları araştırmacıların sürekli erişimine açık olmak durumundadır.

Aslında şunu söyleyebiliriz: Araştırmacılar halen sağlık alanında büyük verinin sınırlarını ve tanımlarını tartışıyorken büyük verinin sağlık alanındaki faydalari şimdije kadar üç ana alanda kendini göstermektedir. Bunlar hastalıkları önlemek, hastalıklar için risk faktörlerini değiştirebilir olarak tanımlamak ve farklı sağlık davranış de-



Şekil 6.18 Sağlık hizmeti verilerinde büyümeye. (Kaynak: https://www.researchgate.net/figure/Growth-in-healthcare-data-Source-6-Multiple-linear-regressions-are-used-to-show-the_fig2_342615272)

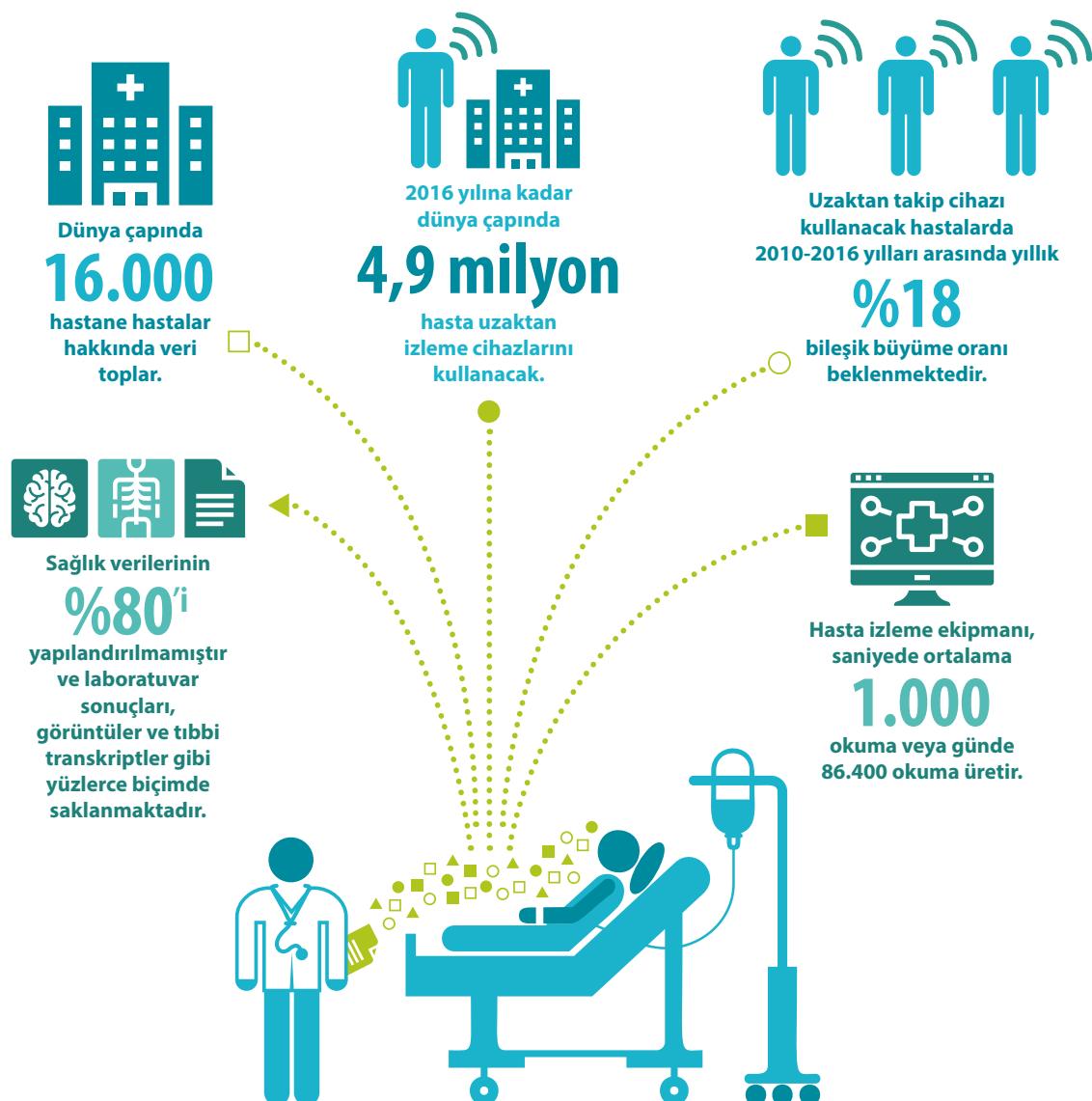
gişimleri için müdahaleler tasarlamak olarak ifade edilebilir. Dünya çapında kurumlar büyük veri hareketini fark ediyor ve veri destekli karar alma ve bilgiyi keşfetmek için yeni girişimleri uygulamaya koyuyor. Örneğin Ulusal Sağlık Enstitüsü (The National Institute of Health-NIH) geniş çaplı biyomedikal veri analizlerini kolaylaştırmak için paylaşılan sayısal bir çevre olan (veri standartları, ontolojiler, veri katalogları, sanallaştırılmış bulut bilişim vb.) Altyapı Plus Programı ve Enformasyon için büyük veri kuruyor. Özellikle NIH Tıp Kütüphanesi etkileyici bir zengin veri seti kaynağına ev sahipliği yapmaktadır. Bunlar öncelikle biyomedikal veri ve NIH'nin finanse ettiği araştırmacılarından gelen bilgi paylaşım sistemlerinden oluşmaktadır. İlaveten Birleşmiş Milletler (The United Nations-UN) Küresel Nabız Projesi'ni başlatmaktadır. Bu proje ayrıca bireylerle kurumlardan veriye katkı sağlanması isteyerek "veri hayırseverliğini" de savunmaktadır.

Sağlıkta büyük veri akışı üç kategori içerisinde özetlenebilir:

- **Geleneksel Medikal Veri:** Öncelikle sağlık kurumları orijinlidir. Örneğin EMR, kişisel ve aile sağlığı tarihçesi, ilaç tedavisi geçmişi, laboratuvar raporları, patoloji sonuçları. Bu verilerin analizinden hedeflenen hastalık çıktılarını ve kişilerin risk faktörlerini daha iyi anlamak, sağlık sistemi masraflarını düşürmek ve etkisini artırmaktır.
- **Omk Veri:** Gen bilimi, protein bilimi, mikrobiyotik, metabolomiks vb. gibi biyolojik ve moleküler alanların genel adı olarak tanımlanan "omik," büyük veri setleri oluşturmaktadır. Bu verilerin analizi ile amaçlanan da hastalıkların mekanizmalarını anlamak ve "en doğru ilaç" gibi tıbbi tedavilerin kişiselleştirilmesini hızlandırmaktır. 2013 Stanford Biyotipta Büyük Veri Konferansı'nda Alice Whittmore'un dediğinde gibi, genomik ile test ve haritalama, örneğin göğüs kanser gelişimi yüksek risk içindeki kadınları işaret edebilir. Bu bilgiler tıp insanlarına diğer düşük riskli kadınlar için önceden potansiyel tehlikelere erken müdahale, geniş çaplı ihtiyaçları azaltma ve öncelikli tedaviyi sağlamaya imkân vermektedir.
- **Kişisel Veri:** İnternet kullanımı, sosyal medya, mobil uygulamalar, sensörler, giyilebilir aygıtlar, diğer teknolojik ya da teknolojik olmayan bilgi işlem cihazları kişilerin davranışları ve sağıyla ilgili işaretleri kendi yaşamlarında ve hareket tarzlarından alarak daha iyi bilgi sağlar ve kişilerin yaşam sağlığını artırmaya katkıda bulunurlar.

Sağlık Hizmetlerinde Büyük Veri: Hayat Kurtarmak için Yeni Öngörülerden Yararlanmak

Sağlık hizmetleri, çeşitli, yapılandırılmamış ve katlanarak büyüyen hareket halindeki büyük miktarda veriyle karşı karşıya. Veriler, bir doktor veya hemşirenin yetişebileceğinden daha hızlı gerçek zamanlı olarak birbirine bağlı sensörler, monitörler ve aletler aracılığıyla sürekli olarak akışı sağlar.



Şekil 6.19 Sağlık hizmetlerinde büyük veri. (Kaynak: <https://www.ibmbigdatahub.com/infographic/big-data-healthcare-tapping-new-insight-save-lives>)

Spesifik birkaç örnek vermek gerekirse Ontario Üniversitesi her gün, erken doğan bebeklerden (prematüre) yaklaşık 100 milyon adet veri toplayarak analizini en hızlı şekilde gerçekleştiriyor. Bunun sonucunda hasta muayenesi sırasında erken teşhis edilen değişimler bir hastalık durumıyla ilişkilendirilebiliyor. Asya Sağlık Bürosu, hasta görüntüleme verilerini Hadoop üzerinde tutup analiz ederek radyoloji ve patoloji uzmanlarının hem daha hızlı hem de daha az hata yaparak teşhis koymasını sağlamıştır.

Şekil 6.20'de büyük verinin sağlık alanında nasıl kullanıldığı ve örneklerini özet olarak görmekteyiz. Bu bağlamda Türkiye'de de SGK, toplamış olduğu büyük mikardaki yapılandırılmış ve yapılandırılmamış veriler üzerinde çeşitli analizler yaparak verimliliği artırmak, kayıp-kaçak oranlarını düşürmek ve hizmet kalitesini yükseltmek için büyük veri konusunda çalışmalara başlamıştır (Özbilgin, 2015).

Veri Tipi	Sağlık Alanında Nasıl Kullanıldığı
Sayısallaştırılmış Kişisel Veri (aygıtlar, kendi raporları ve sensörler vasıtasıyla)	Kişisel ya da grupların proaktif davranışlarının izlenmesi / Biyolojik, fiziksel, davranışsal ya da çevresel potansiyel risk faktörleri üzerinde daha zengin ve detaylı bilgi sağlar / Standart sorgular kullanarak şuna da mümkün olandan daha uzun takip sürelerinde veri toplanmasına izin verir
	Örnekler Gıda tüketimi / Beslenme bilgisi / Beklenmedik kas hareketleri / Kahve tüketimi, sosyal etkileşim ve ruh hali / Fikir izleme süreci / Bir soluk alma cihazı sensörü ile astım ilaçlarının kontrolü ve yardım kullanımı / Şeker hastalarında kan glikoz seviyesini izleme / Psikolojik, mental ve bilişsel durumlar ve tedavileri / Fiziksel aktiviteleri izleme / Diyet / Uyku kalitesi / İlaç bağımlılığı
Konum tabanlı bilgiler	Global konum bilgilendirici servislerden bilgi sağlama / Çevresel ve sosyal belirleyicilerden bilgi sağlanması / Konum civarındaki salgın hastalıkların izlenmesi
	Örnekler Hava örnekleri / Kirlilik seviyeleri / Alerjenler / Trafik modelleri / Su kalitesi / Çevre yürengibilirliği / Taze meyve ve sebzeye erişim (örn. supermarkerler gibi.) / Sağlık Haritası

Twitter (Not: 2011'deki bir çalışmaya göre; İngilizce tweetlerin %8,5'i hastalıkla ve %16,6'sı da sağlıkla ilgilidir)	<p>Gerçek zamanlı hastalık yayılmasını değerlendirme / Duygu ve ruh hallerini değerlendirme / Tıbbi asistan ihtiyacı olan ve izin veren kişilere acil hizmetini kolaylaştırma / Kriz haritalamayı kolaylaştırma / Acil olmayan sağlık üzerindeki araştırmaları kolaylaştırma</p>
	Örnekler <p>Tıbbi yanlışları ölçmek (beyin sarsıntısı gibi) / Antibiyotik kullanımı gibi kötü tedavilerin yaygınlaşması / Kalp durma ve yeniden harekete geçme trendleri / Rahim ağzi ve göğüs kanseri taraması / Doğum sonrası depresyon / Grip salgını (hastalık aktivitesi ve kamu endişesi) / 2010 Haiti kolera salgını / Boston maraton patlamasındaki acil durumlar</p>
Sağlıkla ilgili sosyal ağ siteleri	<p>Hastalar ve tüketiciler arasındaki tavsiyeler ve kişisel sağlık paylaşımlarını kolaylaştırma / Kalabalık yoluyla bulaşıcı hastalıkların yaygınlaşmasının izlenmesi</p>
	Örnekler
	<p>PatientsLikeMe.com / Katılımcıların raporlarının toplandığı hastalık gözetim siteleri / Bulaşıcı salgın hastalıklar hakkında bilgi yayılması, haritalanması ve analiz edilmesini destekleyen resmi olmayan veri kaynakları (örn. Flu Near You, HealthMap, GermTracker, Sickweather)</p>
Diğer sosyal ağ siteleri (örn. Çevrimiçi tartışma, Facebook)	<p>Hastaların endişelerini tartışmak için sosyal medyayı nasıl kullandıklarının izlenmesi / Sokaktaki insanın ne söylediğini farkındalığına sağlar</p>
	Örnekler
	<p>Yan etkiler ve ilgili ilaç bağımlılık davranışları (örn. ilaç değiştirme ve kesme)</p>
Arama sorguları ve web günlükleri	<p>Nüfus seviyesi sağlık davranışlarının yaygınlaşması için daha yüksek öngöründe bulunmak / Arama kelimesi seçimi güvenilir küratör sağlık içeriğine varış için kritik olabilir / Web günlüklerinden tıklama akışlarının yönü kişisel özellikler hakkında bilgilendirici olmaktadır, örn. ruh sağlığı ve beslenme tercihleri</p>
	Örnekler
	<p>Google ve Yahoo arama sorguları hastalık salgılarının öngörülmesi için kullanılmaktadır, örneğin Grip (Google 2013) / Deng Humması / Ruh sağlığı, depresyon ve intihar mevsimselligi / Kene ısırıkları / Sigara içmenin önlenmesi ve elektronik sigara kullanımı</p>

Şekil 6.20 Sayısalştırılmış-Kişisel Hareket ve Sosyal Medyanın Büyük Veri ile Sağlık Alanında Kullanılmasına Örnekler.

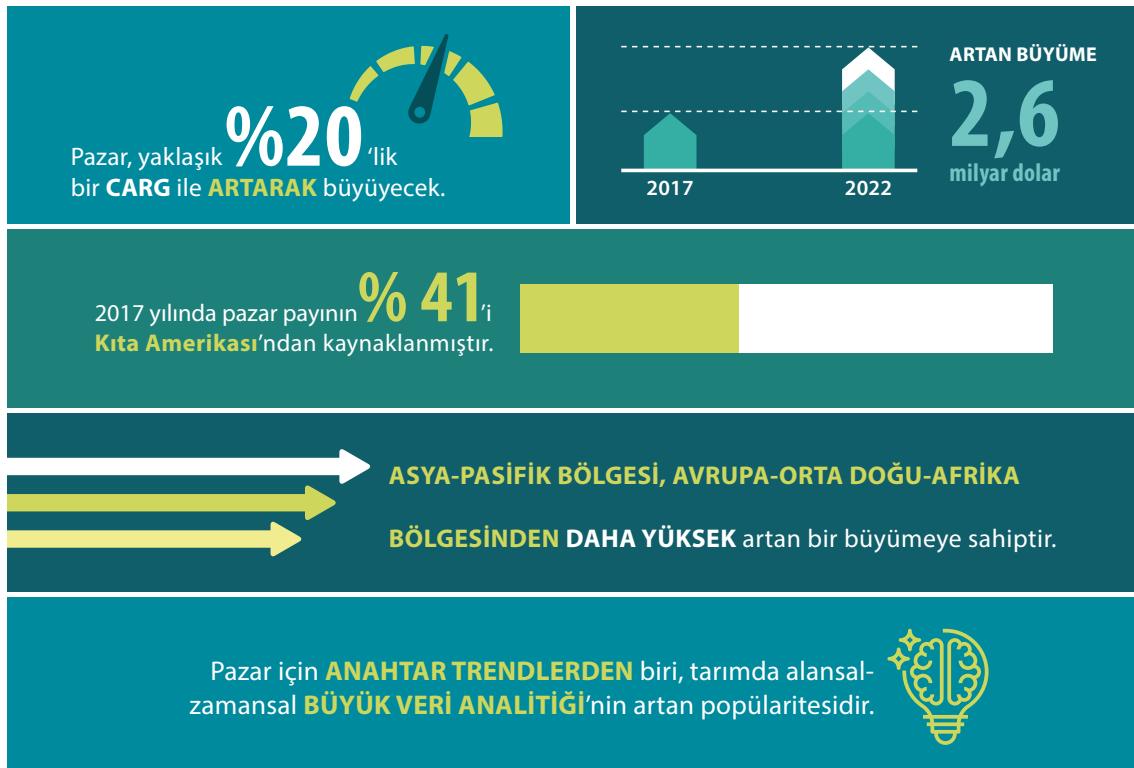
Veri Kümelerinden Sağlanan Bilgiyle Çiftçilerin Gelişimi



Şekil 6.21 Veri kümelerinden sağlanan bilgiyle çiftçilerin gelişimi. (Kaynak: www.cta.int/en/article/2016-03-03/open-data-for-agriculture-and-nutrition.html)

6.6 TARIM

Büyük veri ve analitik, modern dünyadaki çok sayıda endüstrinin iyileştirilmesine ve dönüştürülmesine yardımcı oluyor. Bu tür teknolojilerin yaptığı en etkili şey, operasyonel ve finansal faaliyetler hakkında ayrıntılı ve gerçek zamanlı içgörüler sağlamaktır. Tarımda da bu durum kendisini gösteriyor. Tarım ürünleri, girdi satın almaları ve sübvansiyonlar (devlet yardımları) için mobil ödemeler, öngörülen gıda üretim trendleri ve teşvik tedbirleri açısından hükümetlere daha fazla yardımcı olabilir. Bu bilgi uygun ürün depolama kullanılabilirliğini sağlamak, israfı ve bozulmayı azaltmak ve çiftçilerin hangi tip finansal hizmetlere ihtiyacı olduğunu bilgisini daha iyi sağlamak için kullanılabilir. Mobil kullanım şekilleri sıkıntılı bölgeleri belirlemede hükümetlere ve kalkınma örgütlerine yardımcı da olabilir. Erken teşhis tarımsal üretimin daha da azalmasını ve ailelerin topraklarından ayrılmasını engelleyebilir. Örneğin haşere ve mahsul hastalıkları, fırtınalar veya aşırı hava koşulları gibi doğal afetler tüm hasadı yok edebilir. Büyük veri ortaya çıkmadan önce, bu tür olayları tahmin etmek nere-



Şekil 6.22 Technavio, 2018-2022 yılları arasında tarım sektöründeki küresel büyük veri pazarı hakkında yeni bir pazar araştırma raporu yayınladı. (Kaynak: <https://www.businesswire.com/news/home/20180710005759/en/Global-Big-Data-Market-in-the-Agriculture-Sector-2018-2022-Increasing-Adoption-of-Smart-Farming-Techniques-to-Boost-Growth-Technavio>)

deyse imkânsızdı. Günümüzde ise büyük veri ve izleme teknolojileri bu tür olayları izleyebilir ve hatta tamamen tahmin edebilir. Veri bilimi geçmiş ve mevcut verileri bir sisteme koyarak ve geçerli algoritmalar yoluyla içgörülerini çıkararak gelecekteki getirileri etkili bir şekilde artırabilir. Bu, çiftçilerin ve tedarik zinciri paydaşlarının genel olarak büyük miktarda para tasarrufu yapmasının yanı sıra, dağıtım modellerini ve tedariki kolaylaştırmaya da yardımcı olabilir.

Büyük veri, modern teknolojinin sahaya dahil edilmesini sağlar. İnsansız hava araçları (İHA'lar), kara modellerinin üzerinden uçmak ve değerlendirmek için kullanılabilir. Toplanan haritalama verileri daha sonra analiz edilebilir ve yararlı bilgiler için araştırılabilir.

Büyük veri, tedarik zincirinde ortaya çıkan bazı sorunları hafifletebilir, çünkü her mevsim mahsuller ve hasatla ilgili daha fazla gözetim sağlar. Bu sadece bitkilerle

Büyük veri, tedarik zincirinde ortaya çıkan bazı sorunları hafifletebilir, çünkü her mevsim mahsuller ve hasatla ilgili daha fazla gözetim sağlar.

çalışan çiftçiler için değil, dağıtımcılar, paketleyiciler, perakendeciler ve daha fazlası dahil olmak üzere tedarik zincirindeki diğer herkes için de geçerlidir. Veriler aktarıldığında, beklenenden daha fazla veya daha az miktar da dahil olmak üzere herkesin mevcut ilerlemeye hazırlanmasına gerçekten yardımcı olabilir.

Büyük veri ile hemen hemen her sistem karar veya olay risk analizi planında değerlendirilebilir. Her hata veya olası engel, yalnızca uygun çözümle değil, aynı zamanda beklenen bir sonuç listesiyle birlikte hesaba katılabilir. Çiftçiler, harekete geçmenin mahsulün tamamını yok etmeyeceğinden emin olabilir. Daha da önemlisi, hasarın minimum düzeyde kalmasını sağlamak için gerçek zamanlı verileri kullanabilirler.

Çiftçiler aynı zamanda geçmişte ne kadar ürün ürettiğini, bunun müşteri etkisi açısından ne anlamına geldiğini, bunun arz ve talebi nasıl etkilediğini ve hatta operasyonlarını iyileştirme yolları için ipuçlarını tam olarak görebilirler. Örneğin alternatifleri büyütmek için hem paradan hem de yerden tasarruf etmek adına talebin daha az olduğu bir sezon için daha az ürün üreterek fazla atığı azaltabilirler.

6.7 ENERJİ

Enerji firmaları, akıllı şebeke ve sayaçlar kullanarak abonelerinin bireysel kullanımıyla ilgili oluşan verileri saklayıp işlemek durumundalar. Çağımızda bunlar gibi birçok uygulama alanında kendine yer bulan, fakat işlenmesi, saklanması ve irdelenmesi önceki nesil veri tabanlarından farklı yöntemler gerektiren büyük ölçekli verilere, kısaca “büyük veriye” odaklanmak durumundayız.

Ortak enerji ve kamu hizmetleri endüstrisi, genellikle endüstriyel veya konut komplekslerine uzun mesafeden iletilen elektrik enerjisi üretir. Bununla birlikte enerji ve kamu hizmeti kuruluşu şebekeleri daha akıllı hale getiren tahmine dayalı anali-

Enerji sektörüne büyük verinin daha büyük bir avantaj sağladığı birçok yol var. İşte birkaçı:



Şekil 6.23 Enerji sektöründe büyük verinin sağladığı avantajlar. (Kaynak: <https://www.apogaeis.com/blog/how-big-data-is-adding-a-bigger-advantage-in-the-energy-and-power-sector/>)

tik gibi teknolojilerle bir dönüşüm geçiriyor. Güç üreten kaynaklar daha temiz hale geliyor ve müşterilerin güç elde etmek için birden fazla alternatif var. Büyük veri ve analitiğin ortaya çıkışı, bu tür gelişmelerde çok önemli bir rol oynamaktadır.

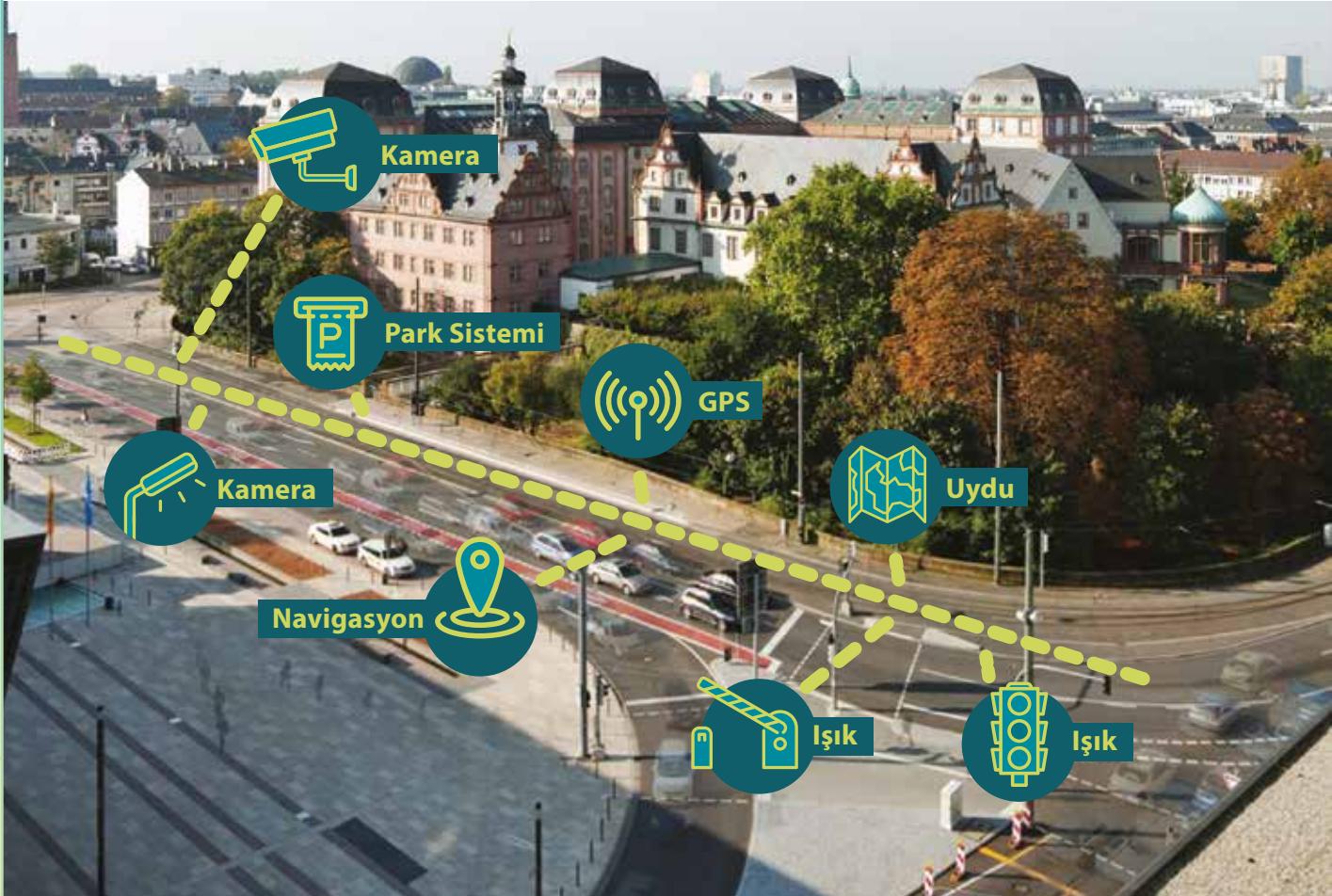
Enerji ve kamu hizmeti işletmeleri sensörleri, bulut bilişim teknolojilerini, elektrik planlamasını ve topluluk iletişimini kullanır. Bu teknolojilerin karışımı milyonlarca evden her saat petabaytlarca bilgi üretir. Sensörler ve termostatlar gibi akıllı cihazların hızlandırılmış kullanımı ile trafo merkezlerinin yardımıyla güç teknolojisinden müşteri tüketimine kadar üretilen dev bir kayıt hacmi alınır.

Enerji ve kamu hizmetleri şirketleri akıllı sayaçlardan, şebeke ekipmanlarından, iklim verilerinden, GIS verilerinden, fırtına verilerinden ve daha fazlasından istatistikler üretir. Bu verilere dayanarak kamu hizmeti kuruluşları güç planlamasını gerçekleştirmek için birden çok model çalıştırır. Ayrıca gruplar maliyetleri, karbon emisyonlarını düşürmek ve müşterilerden ayrılmak için güç talebini yönetmek için içgörülerini kullanır. Büyük veri analitiği, güç çağını ve nihayetinde fiyatlandırmayı etkileyen güç tüketiminin doğru tahmin edilmesine yardımcı olur. Güç tahmini, yenilenebilir elektrik kaynaklarından enerji üretimini etkiler, çünkü bunlar esas olarak değişen iklim koşullarına dayalıdır. Tahmine dayalı analiz bunu hava durumu sistemlerinden alınan gerçeklerden ele alır.

Akıllı şebekeler, tüketiciler ve tedarikçiler arasında iki yönlü gerçeklerin ve gücün akışını sağlar ve büyük veri ve analizler akıllı şebekelerde dinamik enerji yönetimi'ne izin verir. Bu, enerji verimliliği, elektrik sürdürülebilirliği ve güvenilirlik açısından enerjiyi optimize eder. Yük tahmini ve yenilenebilir enerji üretimi ince dinamik elektrik yönetimini gerektirir. Bu nedenle enerji ortamı, akıllı sayaçlar aracılığıyla toplanan büyük miktarda istatistiğin analizi için hesaplama cihazı algoritmaları gibi akıllı stratejiler ve seçenekler gerektirir. Dolayısıyla optimize edilmiş akıllı şebeke operasyonu için sağlam kayıt analizi, çevre dostu istatistik topluluk yönetimi, bulut bilgi işlem ve yüksek performanslı bilgi işlem çok önemlidir. Enerji ve hizmet sektörü gelişiyor ve zamanla büyük veriler ve analitik bunun gerekli bir bölüm olacaktır.



Şekil 6.24 Ulaşımda büyük veri. (Kaynak: <https://medium.com/@mobility4eu/big-data-in-transport-43748d618a1f>)



Şekil 6.25 Taşıma sistemleri akıllı servisler entegrasyonu. (Kaynak: <http://www.sciencedirect.com/science/article/pii/S0167739X13001593>)

6.8 HABERLEŞME VE ULAŞIM

Telekom sektöründe en önemli sorun müşteri kaybıdır. Kuruluşlar hangi müşterilerini kaybedebileceklerini önceden belirleyebildikleri takdirde bu müşterilerini elde tutma amaçlı stratejiler geliştirebilir, düşük maliyetli ve etkili kampanyalar düzenleyebilirler. Kaybetme olasılığı olmayan bir müşteriye kalıcılığını sağlamak amacıyla bir mesaj göndermek hem müşterinin kendisine verilmek istenen mesajın ne olduğunu algılamasını zorlaştıracak hem de maliyetleri artıracaktır. Örneğin Amerika'nın en büyük kablosuz iletişim sağlayıcısı olan Verizon kaybetme olasılığı yüksek olan müşterilerini ve müşteri kaybına neden olan faktörleri belirleme amaçlı bir çalışma yapmıştır.

Öte yandan bugün navigasyon cihazlarının geniş kullanımı, trafik sistemlerinin çevrimiçi merkezlere bağlanabilmesi, insanların konum hareketlerinin değerlendirilmesi ile ulaşım konusu da büyük verinin etkin kullanılabileceği alanlardandır. Şekil 6.25'te en ayrıntılı uygulamalardan olan trafik ve taşımacılıkla ilgili bir çalışma görülmektedir, bunun gibi diğer tüm alanlarda da benzer uygulamalar yapılmaktadır.

6.9 BÜYÜK VERİ KAYNAKLARI

Büyük veri pürzülü verinin farklı tiplerinden oluşmaktadır. Bunları şu şekilde inceleyebiliriz.



Şekil 6.26 Büyük veri kaynakları. (Kaynak: https://www.researchgate.net/figure/Data-sources-of-Big-Data_fig1_336028940)



Şekil 6.27 Açık kamusal verinin faydalari. (Kaynak: <https://www.sdsntrends.org/research/2019/3/4/maximizing-access-public-data>)

6.9.1 Kamusal Veri (Public Data)

Geniş çaplı iş ve yönetim uygulamaları için potansiyel olarak kullanılabilen ve devletler, devlet kurumları ve yerel topluluklar tarafından tutulan veridir. Erişim veya kullanıma ilişkin mevcut yerel, ulusal veya uluslararası yasal kısıtlamaları olmayan herkes tarafından serbestçe ve sonrasında yeniden kullanılabilen ve yeniden dağıtılabilen bilgilerdir. İşletmedeki kamuya açık verilere örnekler arasında basın bültenleri, iş tanımları ve genel halka yönelik pazarlama materyalleri yer alır. Bireysel gizliliğimizi korumak için belirli kısıtlar altında erişilen bu verilere örnek ulaşım, enerji kullanımı ve sağlık alanı sayılabilir.

ÖZEL VERİ NEDİR?



İsim



Adres



Yerini belirtme



Gelir



Çevrimiçi tanımlayıcı



Sağlıkla ilgili bilgiler



Kültürel profil



••• ve daha fazlası...



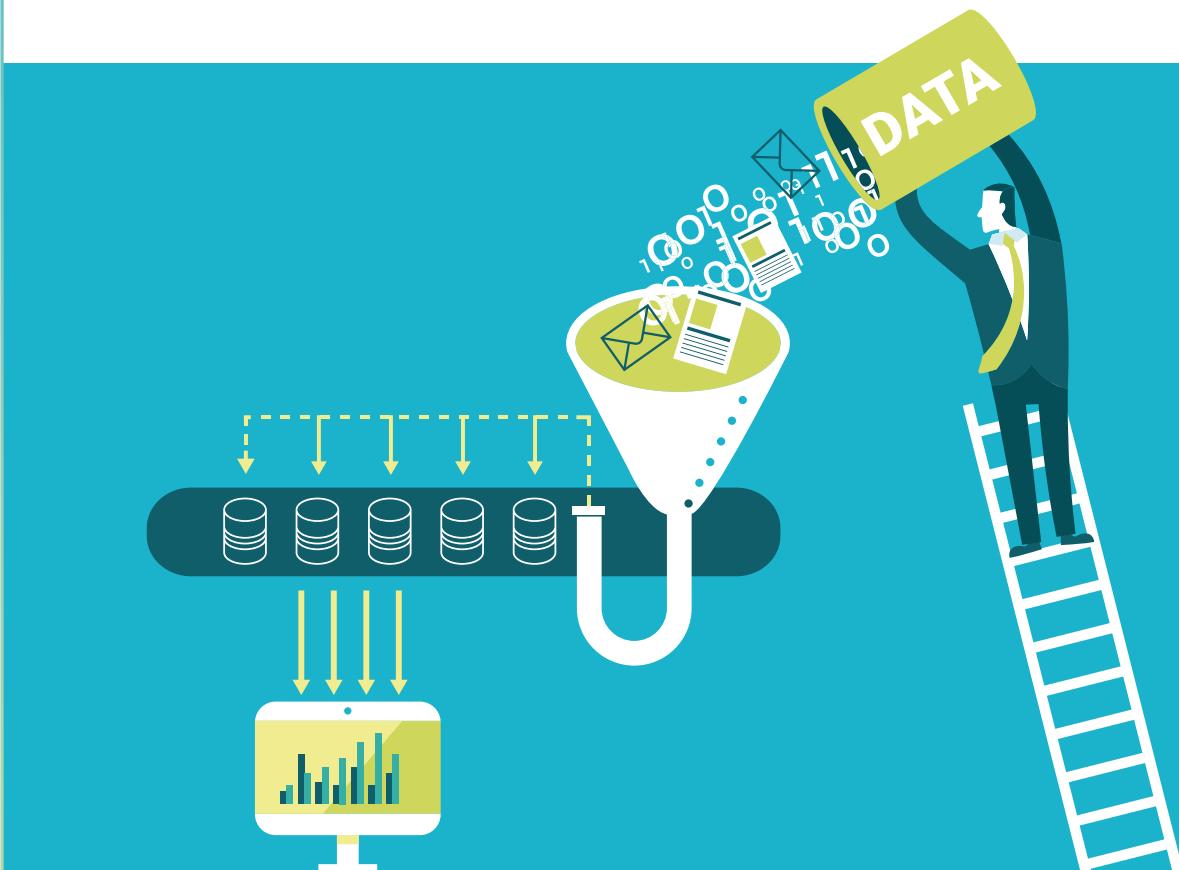
**Toplanan
veri
kullanımı**



**Kurallara
uymalısınız.**

6.9.2 Özel Veri (Private Data)

Özel şirketler, kâr amacı gütmeyen kuruluşlar ve bireyler tarafından tutulan verilerdir. Özel veri yapılandırılmış, yapılandırmamış, sözlü, yazılı, kâğıt biçiminde veya elektronik, video, fotoğraf biçiminde herhangi bir bilgidir. Özel veriler demografik bilgileriniz, kişisel tanımlayıcılarınız, finansal, eğitim, sağlık verileriniz, genetik bilgileriniz, kişisel geçmişiniz, sosyal medya, satın alma geçmişiniz, kullanım eğilimleriniz, bakım kalıplarınız, yasal veya cezai veya ailevi verileriniz vb. olabilir. Sizin hakkınızdaki bilgilerden belgelenen, toplanan, elde edilen, saklanan, aktarılan veya kullanılan her türlü veri kişisel veridir ve bu nedenle özel veri olarak kabul edilebilir. Örneğin müşteri işlemleri, örgütsel tedarik zincirleri tarafından kullanılan radyo frekans tanımlama etiketleri, şirket mal ve kaynaklarının hareketi, web sitesi tarama hareketleri, mobil telefon kullanımı bunlardan birkaçıdır.

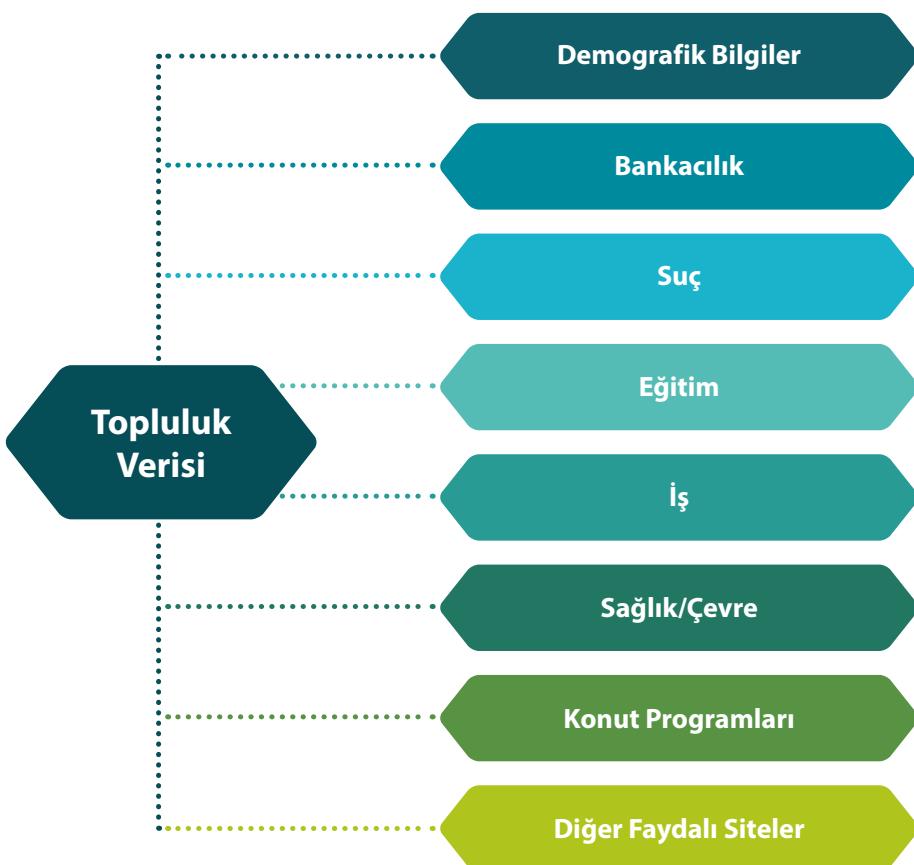


6.9.3 Çevresel Veri (Data Exhaust)

Bu veri tipi pasif bir şekilde toplanan ortama ait çevresel veriyi anlatır. Veri toplanan için sıfır değerinde ya da bir merkezi olmayan veridir. Bu veriler farklı amaç için toplanmış fakat yeni değerli kaynak oluşturmak için başka veri kaynaklarıyla yeniden kombine edilebilirler. Bireyler yeni teknolojiler benimsediği ve kullandığı zaman (örneğin mobil telefonlar) bu cihazların günlük faaliyetleri ile ortamsal veri üretirler. İnsanlar günlük yaşamlarına devam ederken pasif bir şekilde farkında olmadan veri yayıyor da olabilirler (mesela alışveriş yaparken, sağlık hizmetine erişirken, başkalarıyla etkileşim halindeyken). Bu veri türü için bir başka kaynak insanların ihtiyaçlarını, arzularını ve niyetlerini anlamaya yarayan bilgi arama davranışıdır. İnternet aramaları, telefon görüşmeleri ya da özel çağrı merkezleri bu türden veri üretmektedir.

6.9.4 Topluluk Verisi (Community Data)

Sosyal eğilimleri yakalamak için dinamik ağların içerisindeki yapılandırılmamış verinin -özellikle metin- özüdür. Tipik bir topluluk verisi müşterilerin ürün inceleme-

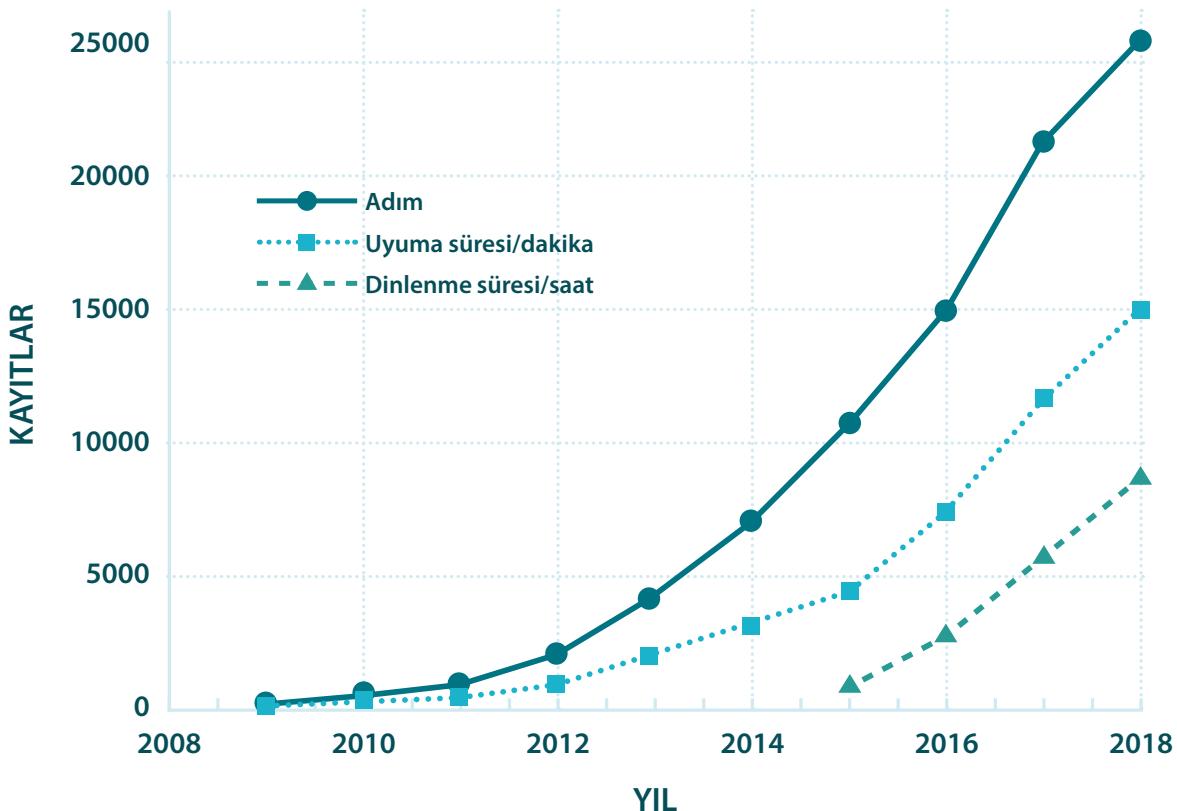


Şekil 6.28 Topluluk verisi. (Kaynak: Kaynak: <https://unhp.org/crg/indy-cd.html>)

rini, oylama butonlarını, Twitter bildirimlerini ve bunun gibi başka şekilleri içermektedir. Bu veriler sosyal yapıdaki kalıplardan damitilmiş bir sonuç çıkarmak anlamına gelebilir (Kennedy, 2008).

6.9.5 Kişisel Veri (Self-Quantification Data)

Kişi eylem ve davranışlarının belirlenmesi yoluyla kişiler tarafından ortaya çıkan veri türüdür. Örneğin bu verinin bir şekli egzersiz ve hareket izleyen bilekliklerle sağlanan veridir, bu veri bir mobil uygulamaya yüklenir ve kişi takip edilmiş olur. Psikolojide kişinin durumuyla ilgili bilgi edinilmesi o kişinin tercihlerini ifade eder. Örneğin ışıkları uzun süre açık tutmak yerine tasarruf amacıyla kapatılan bir kişi enerji tasarruflu ampul alabilecektir. Bu öz miktar verisi psikoloji ve davranış arasındaki bağlantıya köprü oluşturmaya yardım eder. Farklı alanlardaki sosyal bilimlerin uzmanları psikoloji, pazarlama ya da kamu politikası gibi bu alanlardaki araştırmalar için bu veri türünden yararlanırlar.



Şekil 6.29 Fitbit projesinden kişisel veriler. Ocak 2009'dan Ekim 2018'e kadar herkese açık kayıt sayısı (küümülatif toplam). (Kaynak: <https://www.researchgate.net/publication/334032994/figure/fig1/AS:776617282514945@1562171397916/Self-quantification-data-from-Fitbit-project-Number-of-public-records-from-January-2009.png>)

KAYNAKLAR

Bloomberg Businessweek Türkiye, 2014

European Commission, Aralık 2011, Open data, An engine for innovation, growth and transparent governance.

Graham Vickery, Information Economics, Paris, 2011, Review Of Recent Studies On Psi Re-Use and Related Market Developments.

Gürsakal, Necmi, İstatığın ABC'si, Dora Yayıncılık, 2013.

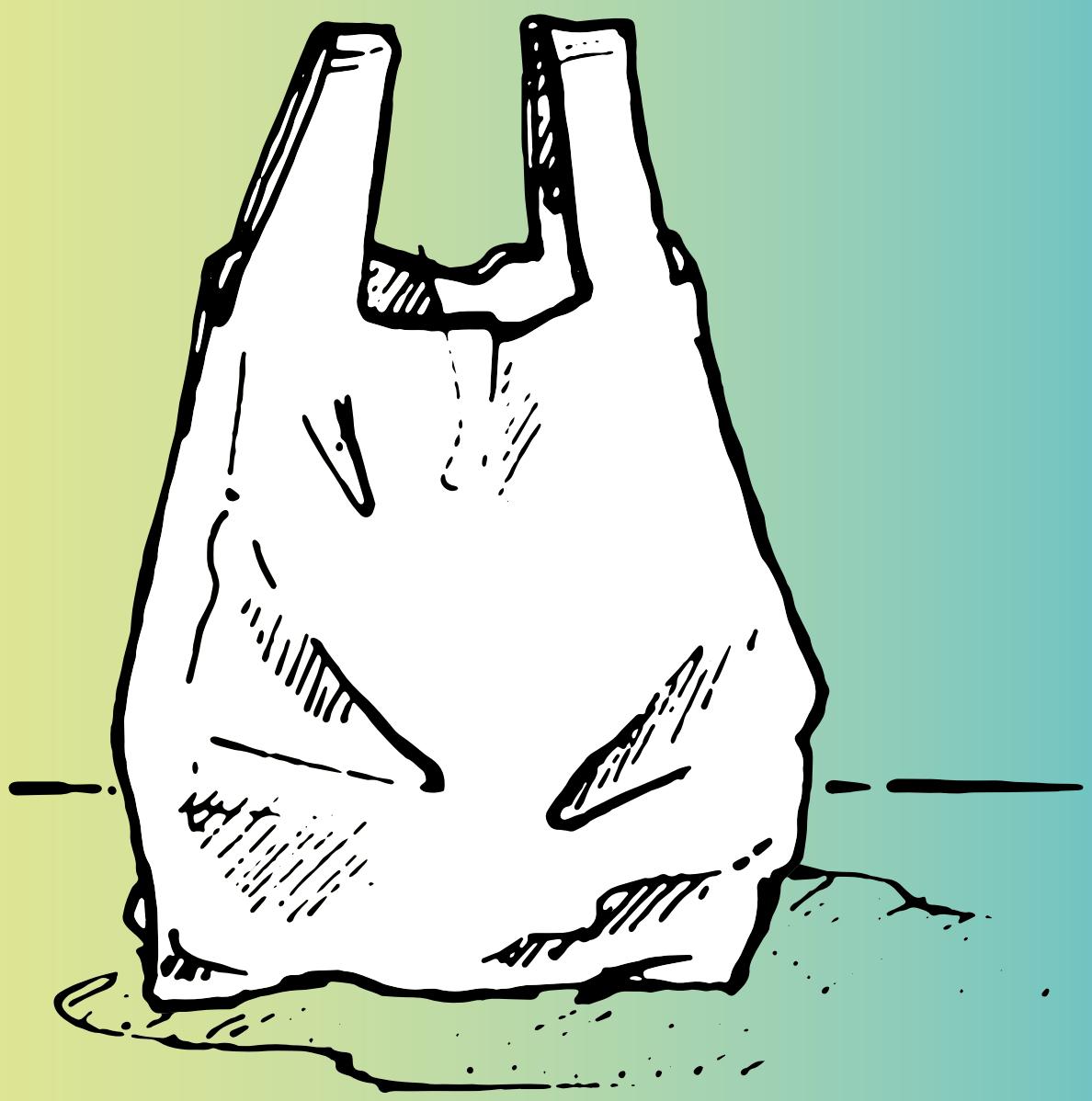
<https://www.grandviewresearch.com/industry-analysis/big-data-as-a-service-bdaas-market>

<https://ismailimail.blog/2015/09/13/aga-khan-university-east-african-institute-big-data-can-look-at-the-whole-child/>

- <https://www.byteant.com/blog/7-ways-how-to-use-big-data-in-socialmedia/>
- <https://customerthink.com/14-visualizations-mapping-the-b2b-buyer-journey/>
- <https://financesonline.com/it-trends/>
- <https://www.guires.com/blog/what-is-bigdata-analytics-and-why-is-it-important-to-business/>
- <https://data-flair.training/blogs/data-science-in-finance/>
- <https://www.hallaminternet.com/social-media-in-the-financial-industry/>
- http://www.theukcardsassociation.org.uk/plastic_fraud_figures/
- <https://www.sciencedirect.com/science/article/abs/pii/S1084804516300571>
- https://www.researchgate.net/figure/Growth-in-healthcare-data-source-6-Multiple-linear-regressions-are-used-to-show-the_fig2_342615272
- <http://research.ijcaonline.org/icaccthpa2014/number4/icaccthpa6041.pdf, 2014>
- <https://www.ibmbigdatahub.com/infographic/big-data-healthcare-tapping-new-insight-save-lives>
- <https://datasciencegyan.com/big-data-use-cases-in-financial-services/>
- <https://www.ciklum.com/blog/what-are-the-big-data-trends-in-banking/>
- <https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/>
- <https://medium.com/@mobility4eu/big-data-in-transport-43748d618a1f>
- <https://www.apogaeis.com/blog/how-big-data-is-adding-a-bigger-advantage-in-the-energy-and-power-sector/>
- <http://www.sciencedirect.com/science/article/pii/S0167739X13001593>
- <https://www.businesswire.com/news/home/20180710005759/en/Global-Big-Data-Market-in-the-Agriculture-Sector-2018-2022-Increasing-Adoption-of-Smart-Farming-Techniques-to-Boost-Growth-Technavio>
- <https://www.cta.int/en/article/2016-03-03/open-data-for-agriculture-and-nutrition.html>
- McKinsey Global Institute, 2011, Big data: The next frontier for innovation, competition, and productivity.
- Özbilgin, İ. G., 2015, Kamuda Büyük Veri ve Uygulamaları, <http://ab.org.tr/ab15/bildiri/483.pdf>, [Ziyaret Tarihi: 29 Eylül 2015].
- Retailer's Guide to Big Data, Infographic, 2103
- The Economist, 2013, s. 7-8.

7. BÖLÜM

BÜYÜK VERİ YÖNETİMİ İÇİN ÖRNEK VAKALAR



Büyük verinin zorlukları her ölçekteki kurumu etkiler, ama bunun yanında yeni fırsatlar da oluşturur. Temel bir varlık gibi görünen bilgi aracılığıyla tüm ölçekteki kurumlar en önemli satış, pazarlama ve müşteri hizmetleri girişimini ve iş büyütme hareketini destekler.

Var olan müşteri verisinin bütünüyle hacim, çeşitlilik ve hızı, pazarlama, satış ve müşteri desteginin yeni anımlarını araştırırken, bilgi stratejisi ve inovasyon açısından yöneticiler için çok büyük bir umuttur.

Artık bundan sonra yeni dönemde işletme analistleri ve veri bilimcileri müşteriler için benzersiz deneyimler oluşturduğu zaman fırsatlara kavuşmak için hızlı analitik modelleme ve simülasyon kullanmalıdır. Pazarlamacılar yeni teklifler ve pazarlar geliştirirken tekliflerin rasyonel sınıflandırılması ile müşterilerini eşleştirmek için yeni ve gizli bilgi kaynakları kullanmalıdır.

Değerli bir şirket varlığı olarak görünen bilgi aracılığıyla hemen hemen her kurum büyük veri kullanımı için yeni yaklaşımlar, CRM ve müşteri deneyim girişimleriyle yeni rekabet avantajları bulmak için analizler keşfedebilir.

Tüm coğrafyalarda ve tüm endüstrilerde kurumların en önemli satış, pazarlama ve müşteri hizmeti girişimlerini desteklemek için analizler ve büyük veri kullanımı için yenilikçi yaklaşımlar kullanan, bütün ölçeklerdeki şirketlerden birçok örnek vardır. Bu yeniliklerin birçoğu geleneksel veri deposundan ve iş zekâsı girişimlerinden gelmez, bunun yerine büyük veri ile ilgili fırsatçı deneyimlerden doğmaktadır.

Aşağıdaki şekil işletme ve bilişim uzmanlarının büyük verinin hangi alanlar için en büyük fırsatları sunduğunu algıladıklarını göstermektedir. Doğrudan müşteri odaklı aktiviteler için büyük veri kullanımı, örneğin satış ve pazarlama alanında yüksek kalmasına rağmen, yeni ürün ve hizmet gelişimi ve dolaylı müşteri odaklı aktiviteler için artan bir şekilde tercih edilmektedir.

Burada “büyük veri” analizi yaparak problemlerine çözümler üretmiş üç örnek kayayı inceleyeceğiz. Bu üç vaka çalışmasında müşterilerin bakış açılarını sezebilmek için analistik çalışmaları destekleyen ve bunu yapmaya çalışırken karşılaşılan zorlukların üstesinden gelen kurumlara göz atılacaktır. Bu örnekler farklı kurum türlerinden gelmesine rağmen her biri değerli genel dersler sunmaktadır.



Story mağazasının bir reyonundan bir kesit. (Kaynak: <https://echochamber.com/article/story-new-york-city/>)

Story, büyük veri ve analizinin büyük ya da küçük geleneksel perakende dünyası şirketlerine nasıl yardım ettiğini gösterdi.



Story mağazasından bir kesit. (Kaynak: <https://www.lauraperuchi.com/en/story-an-amazing-decor-stationery-and-gift-store-in-new-york/>)

7.1 VAKA 1: HIZLI ANALİTİK MODELLEME VE SİMÜLASYON KULLANIMI

New York'taki Story adlı bir perakende mağazası, sıradan bir teknoloji parçası kullanarak müşteri deneyimlerine benzersiz bir bakış açısı kazandı. Prism Skylabs adlı bir teknoloji ortağı ile çalışan Story, kendisinin satış, pazarlama ve müşteri destek deneyimlerini geliştirmek için zaten mağazada bulunan güvenlik kameralarından birçok perakendecinin bir masraf olarak gördüğü teknoloji ile elde ettiği bilgiyi faydalı hale getirmenin bir şeklini buldu.



Story mağazasından ürün teşhirine dönük bir kesit. (Kaynak: <https://www.homeaccentstoday.com/industry-news/macys-acquires-new-york-retail-concept-shop-story/>)

Problem: Story üç ila sekiz hafta içerisinde mağaza düzenini tamamen değiştirebildiği yeni bir perakende konsepti oluşturmak için yola çıktı. Bu fikir gelişen trendler, konular ve modalara adaptasyon için her birkaç haftada bir değişen bir bakış açısıyla yeni bir mağaza modeli olmalydı. Malların düzeni için yer planındaki her şey yeni bir “Story”i yansıtacak şekilde tasarlanmış olacaktı.

Cözüm: Prism ile çalışan Story mağazasında müşteri deneyiminin içerisindeki bakışı algılamak için ısı haritaları oluşturan güvenlik kameraları kullanıldı. Bu haritalar mağazanın düzenine bağlı olarak mağazayı ve ürünlerini incelemek için gelen müşterilerin nasıl hareket ettiklerini analiz etmeye başladı. Bu müşterilerin neleri beğendiğine bağlı olarak personelin ve ürünlerin hareket etmesine izin vermektedir.

Sonuçlar: Analitik şekilde optimize edilen bu yaklaşım Story'ye müşterilerine cazip bir ortam kurabilmeyi mümkün kıladı. Müşteriler tipik olarak Story'de 40 dakika harcamakta ve üç haftalık tatil süresi boyunca müşterilerinin % 30'unun alım yapmak için mağazaya geri döndüğü görüldü.

Çok aşılmaz engelleri olmasa bile, Story'nin yaptığı gibi, birçok perakendeci mağaza katlarını henüz bu şekilde analiz etmemiştir.

Story, büyük veri ve analizinin büyük ya da küçük geleneksel perakende dünyası şirketlerine nasıl yardım ettiğini gösterdi.

7.2 VAKA 2: MÜŞTERİLERİ EŞLEŞTİRME

Satış ve pazarlama için büyük veri ve analizlerin en ortak kullanımlarından biri şirketin teklifleri ile ilgili müşterileri eşleştirmektir. Bunu akıllıca yapmak için kurumlar ürün ve hizmet tekliflerinin karakteristikleri ve müşterilerle ilgili karmaşık bilgiye ihtiyaç duyarlar. Trendler, müşteri lezzetleri, ihtiyaçlar ve durumlar zamanla değişebilir, bu yüzden veri taze olmayı ve trendler tatmin edici olmayı gerektirir. Analizler çok fazla varsayımlar yapmamalıdır ve sadece iyi bir eşleştirme olup olmadığını değil, tercihen kabul edilen tekliflerin olasılığını öngörmelidir.

Şirketler tekliflerini rasyonelleştirmelidir. Bugün birçok şirket üretim, parça evvanteri, kullanılabilirlik, uzun süreli satış öngörüleri, teslimat ve hizmet ile ilişkisindeki konulara neden olan sonsuz varyasyonlar, konfigürasyonlar ve herhangi belirli bir ürün için seçenekleri kolayca üretebilir. Bazı şirketler bu konuları adreslemek için ürün konfigürasyonlarını rasyonelleştirecek/optimize edecek çözümlere dönmüştür.

Bugün birçok şirket üretim, parça evvanteri, kullanılabilirlik, uzun süreli satış öngörüleri, teslimat ve hizmet ile ilişkisindeki konulara neden olan sonsuz varyasyonlar, konfigürasyonlar ve herhangi belirli bir ürün için seçenekleri kolayca üretebilir.



PDZA. (Kaynak: <https://www.pdza.org/>)

ABD Washington Tacoma'daki 29 akrelilik (~117.363 m²) bir hayvanat bahçesi olan Point Defiance Zoo&Aquarium (PDZA) satış ve pazarlama girişimlerini desteklemek için bilgisayar analizleri kullanmıştır.

Problem: Tam zamanlı seksen çalışanıyla küçük bir kurum olan PDZA gelirlerini artırmak istedti.

Çözüm: PDZA, büyük veri analizleri için IBM'in bir ortağı olan BrightStar Partners ile çalışmayı seçmiştir. Bu teknolojiyle önceki ve gelecekteki üyeleri ve ziyaretçiler için daha akıllıca hedeflenmiş kampanyalar yapma olanağı sağlamasına yönelik hayvanat bahçesinin yıllık 600.000 ziyaretçisinin örnek kalıplarını ve davranışlarını inceledi. Örneğin PDZA'nın üyelikleri sona eren insanlar için e-posta tabanlı bir tanıtıcı kampanya amaçladığını tanıttı. Kişiselleştirilmiş e-postaların hızlı oluşum ve yürütülmesi sayesinde, tanıtımı hayvanat bahçesinin geleneksel doğrudan e-posta kampanyalarının iki katına çıkarmayı başararak satışa %6 katkı sağladı. PDZA'nın müşteri analizleri doğru zamanda doğru tanıtım türleriyle müşteri sağlanması imkân veren anlayış geliştirdi.

Sonuçlar: Yeni analitik teknolojileri kabul ettiğinden beri, PDZA çevrimiçi bilet satışlarında 2012'de % 700'ü aşan ve 2013'te de ilaveten % 50 daha büydü.

 Let's build a smarter planet Media & Entertainment



Point Defiance Zoo & Aquarium

Making the right decisions – whatever the weather.

Smart is...

Combining ticket sales and visitor data with real-time weather forecasting optimizes operational efficiency and enhances the customer experience.

What if your business needed to serve ten times as many customers tomorrow as it did today? Would you be able to adjust quickly enough to meet their needs? And what if the fluctuation in demand was caused by something as arbitrary and unpredictable as the weather?

For an organization like Point Defiance Zoo & Aquarium (PDZA), these questions aren't a joke; they are issues that need addressing on a daily basis.

Working with BrightStar Partners, an Arnet Services company, and architecting the PDZA team has developed a new and efficient solution that combines real-time weather data with ticket sales and attendance information to help forecast visitor numbers much more accurately than ever before.

Like all open-air attractions, zoos live and die by the weather. When sunny days alternate with rainy ones and daily attendance can fluctuate wildly, how can zoos avoid the slings of over- or under-staffing – and the potentially serious consequences for cost-efficiency and customer service?

Point Defiance Zoo & Aquarium (PDZA) – part of Point Defiance Park, the largest attraction in Pierce County, Washington – wanted to counteract the uncertainties of the weather by gaining a better understanding of its effects on ticket sales and visitor numbers. With this new insight, the Zoo hoped to be able to manage all aspects of its business more efficiently.

The first step on this journey was to find a way to unlock the value of the data that PDZA already possessed. In partnership with BrightStar Partners, the Zoo embarked on a journey that now promises to transform its entire operations marketing to staffing and facilities management.

At the starting point

Donna Powell, Business & Administrative Services Manager, Zoological & Environmental Education Division, Tacoma, which owns PDZA, explains: "We already had a point-of-sale (POS) system in place, which was great for us to assess key metrics on ticket sales. However, without the tools to extract that data in a form that was easy to use, it was limited."

Powell's team met with BrightStar Partners, a company with extensive experience of helping organizations in the industry develop innovative business analytics. BrightStar's groundbreaking work at Cincinnati Zoo helped create a new analytics landscape.

PDZA tanıtımı. (Kaynak: <https://www.slideshare.net/ibmsoftwareindia/point-defiance-zoo-aquarium-optimizes-operational-efficiency-and-enhances-the-customer-experience>)

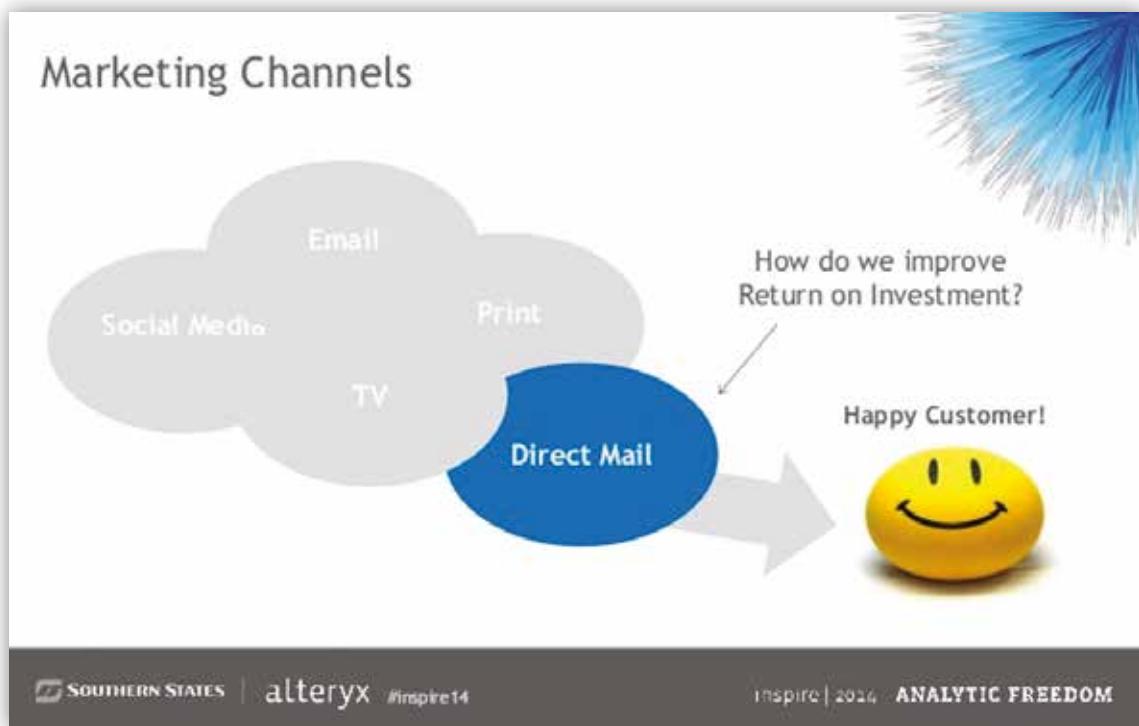


PDZA bilet. (Kaynak: <https://www.pdza.org/>)

7.3 VAKA 3: MÜŞTERİ DAVRANIŞLARININ DEĞERLENDİRİLMESİ

Kurumlar yalnızca bir müşteri iletişim listesine ihtiyaç duymaz, aynı zamanda tüm kullanılabilir müşteri verisinden yararlanabilmeye ihtiyaç duyar.

Büyük veride Gartner'in "3V"si (Volume, Velocity, Variety: Hacim, Hız, Çeşitlilik) ile veri kaynaklarındaki çeşitlilik teknoloji şirketlerinin müşterileri tarafından hem en büyük zorluk hem de en büyük fırsat olarak görülmüştür. Firmaların kendi ERP, CRM ve diğer iç uygulamalarında kullanılan veri, mobil, sosyal ve web kanalları gibi dış kaynaklarla birlikte müşterilere daha iyi bir anlayış sağlar ve daha etkili bir şekilde müşteri ilişkileri yapılandırılmalarına yardım edebilir.



Southern States e-posta çözümü. (Kaynak: <https://www2.slideshare.net/Alteryx/inspire-2014-southern-states-cooperative-improving-marketing-roi-with-predictive-analytics>)

ABD'deki en büyük çiftlik işletmecisi kooperatiflerinden biri olan Southern States Kooperatifi, doğrudan e-posta için bir öngörü modeli yaklaşımını benimserek sonuçlarını aldı. Karşılık oranı % 34'e kadar artarken katalogların sayısının % 63'e kadar azalabildiğini keşfetti.



Southern States. (Kaynak: <https://www.southernstates.com/farm-store/store-locations/16243>)

Problem: Southern States 1200 parekende satış noktası ve 23 eyalet üzerinde 300.000 çiftlik olmak üzere binlerce tarım ürününe sahiptir. Çoklu sistemler, karmaşık ve rekabetçi analitik platformlarda veri ile mücadele etmekte ve bunu yapabilecek birçok personele ihtiyaç duymaktadır.

2012 yılında Southern States'in doğrudan e-posta kampanyası önceki yıllarda gibi aynı şekilde yürütüldü. Sonuçlar % 3 oranında borç ödemesi ve % 39 pazarlama ROI (Return On Investment)



*Southern States'in reklamının yapıldığı bir çubuk.
(Kaynak: <https://www.ebay.com/itm/Vintage-Southern-States-Cooperative-Coop-Paper-Hand-Fan-Advertising-SSC-/223926719191>)*

yatırım getirişiyle önceki yıllarda kampanyalar çizgisindeydi. Sonuca Southern States'e bu durumu artan kârlar üretiminden daha fazlaya mal oldu ve kooperatif bunun yeterince iyi olmadığını farkındaydı. Southern States müşterilerinin ve onlara nasıl hizmet edeceğini tam bir resmine ihtiyaç duyduğuna karar verdi.

Cözüm: Kooperatifin katalog ve promosyonlarının kime postalanması gerektiğini belirleyen gelişmiş bir yaklaşımı yoktu, fakat bir analitik tedarikçi olan Alteryx ile ortaklık yoluyla pazarlama aktivitelerini sürekli bir şekilde optimize edebildi. Alteryx herhangi bir kod yazmadan veya veri verme ya da alma olmadan gelişmiş optimizasyon ve mekânsal analiz sağladı.

Sonuçlar: 2013 yılında Southern States bu kez öngörücü modelleme yaklaşımını kullanarak fakat aynı doğrudan posta kampanyası yürüttü. Sonuçlar % 10 oranında borç ödemesi ve % 59 ROI (Return on Investment) yatırım getirişiyle dramatik bir şekilde farklıydı. Yalnız geçen çeyrek boyunca Southern States % 189 ortalama bir pazarlama ROI (Return on Investment) yatırım getirişiyle sekiz doğrudan posta kampanyası (hepsi modellenmiş) gerçekleştirdi.

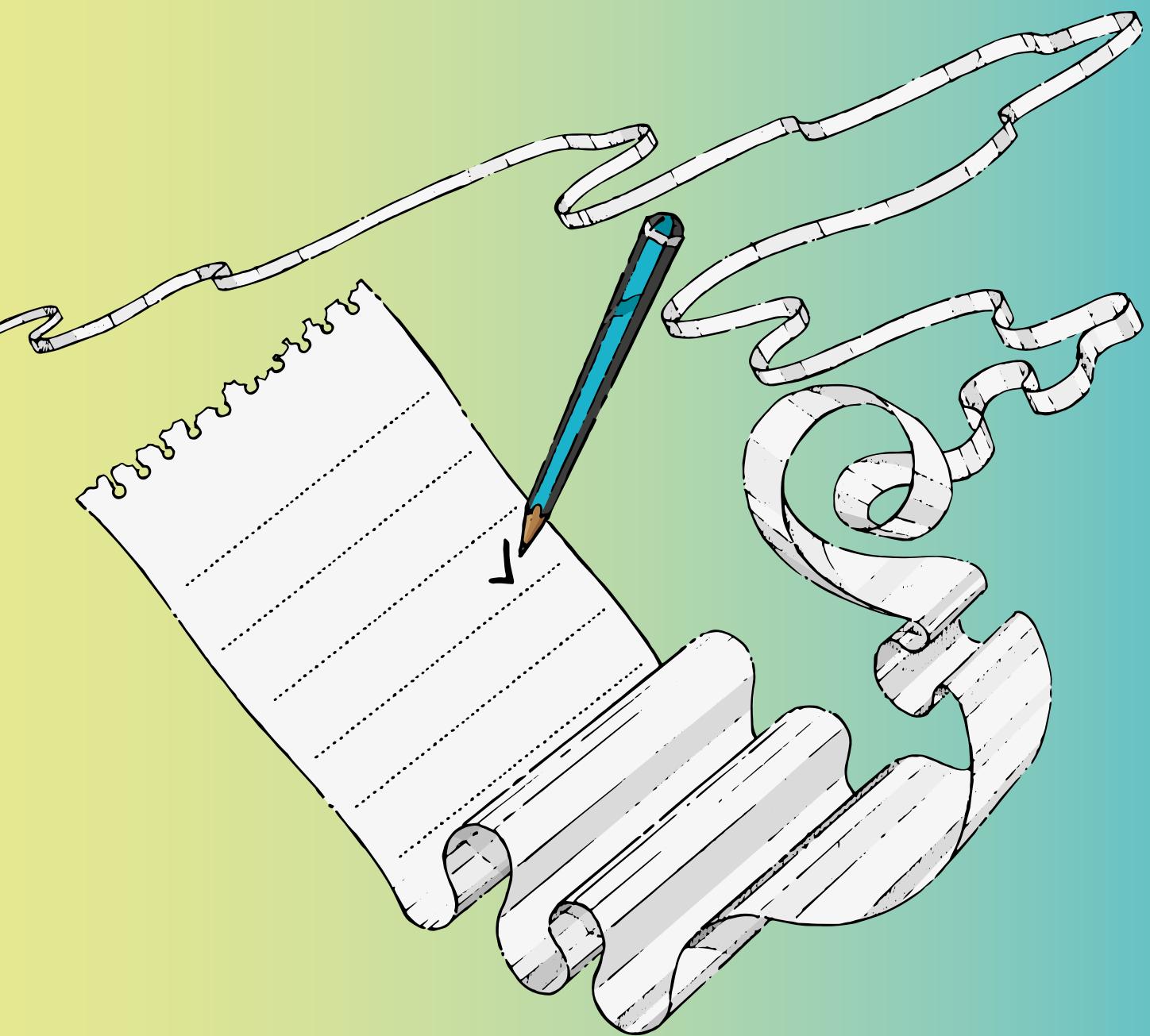
Doğrudan posta kampanyalarının başarısını takiben Southern States daha iyi işletme kararlarımasına yardımcı olması için kurum çapında analiz uygulamalarını uyguladı.

KAYNAKLAR

- <https://echochamber.com/article/story-new-york-city/>
- <https://www.lauraperuchi.com/en/story-an-amazing-decor-stationery-and-gift-store-in-new-york/>
- <https://www.homeaccentstoday.com/industry-news/macys-acquires-new-york-retail-concept-shop-story/>
- <https://www.pdza.org/>
- <https://www.slideshare.net/ibmsoftwareindia/point-defiance-zoo-aquarium-optimizes-operational-efficiency-and-enhances-the-customer-experience>
- <https://www.southernstates.com/farm-store/store-locations/16243>
- <https://www2.slideshare.net/Alteryx/inspire-2014-southern-states-cooperative-improving-marketing-roi-with-predictive-analytics>
- <https://www.ebay.comitm/Vintage-Southern-States-Cooperative-Coop-Paper-Hand-Fan-Advertising-SSC-/223926719191>

8. BÖLÜM

BÜYÜK VERİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ İÇİN SAHA ARAŞTIRMASI



8.1 ARAŞTIRMANIN YÖNTEMİ VE KISITLARI

Veri analizi için Collis ve Hussey'in (2009, s.21) prosedürü şu ana özellikler ile takip edildi: Verinin azaltılması, yeniden yapılandırılması ve kavramsalştırılması. Kodlayıcı kullanım sayesinde, araştırmancın amacı açısından daha az önemli konulara atıf yapan kişilerden alınan uygun veriler seçilmiştir. Daha sonra veri faktörler grubu ve katılımcılar üzerinde yapılandırılan kategoriler içeresine uygun hale getirildi. Bunun tamamlanması üzerine, veri yakından incelenerek, karşılaştırılarak, etiketlenerek ve organize edilerek farklı yollar ile incelenmiştir.

Tüm sonuçların derlenmesinden sonra, farklı cevaplar alanlarının üzerinde yapılandırılarak kümelendirilmiştir. Eğer büyük veri uzmanları tarafından örnek cevaplar için belirgin farklılıklar olduysa, farklı değişkenler ve cevaplar arasındaki potansiyel ilişkiler incelenmiştir.

Araştırmancın nitel yaklaşımı ve röportajların kısıtlı örneklerinden dolayı kusursuz sonuçlara sahip olmak mümkün değildir, bu bir istatistiksel temsildir ve bu çalışma değerlendirilirken dikkate alınmalıdır. Bu kısıt her alandaki en yaygın ve en güncel bibliyografik kaynaklar kullanılarak kısmen aşılmıştır. Bu bulgunun sonuçları geniş bir şekilde bütün kurumlara veya şirketlere mutlaka anlamlanmış olmayabilir. Bununla birlikte büyük veri girişimleri bir iş operasyonunun artan şekilde daha geniş

Büyük veri girişimleri bir iş operasyonunun artan şekilde daha geniş bir parçası olmasından dolayı, bu çalışmanın ölçüğünü sayısal verilerin de takviye için kullanılması nedeniyle mümkün olduğunda artırılmasına çabalamak gereklidir.

bir parçası olmasından dolayı, bu çalışmanın ölçüğünü sayısal verilerin de takviye için kullanılması nedeniyle mümkün olduğunda artırılmasına çabalamak gereklidir: Diğer veri türleriyle birleştirilen röportaj verisi için çalışan birçok nitel araştırma içerisinde kesinlikte ortaktır (Flick, 2009).

Diğer kısıtlar, varsayımlar ve önerilen araçlar test edileceği için, bu çalışmanın içeriğindeki sınırlı imkânları kapsar, diğer taraftan önerilerin geniş çaplı bir uygulamasını yapmak için yeterli güveni verir. Diğer bileşenlerdeki benzer araçlar çalışan başarılı bir analiz bize kısmen de olsa önerinin fizibilitesini test etmeye izin verir.

Son olarak nitel ve nicel araştırma yaklaşımının kombinasyonu ideal olacaktır. Bu yöntem ankete katılanların sayısının artmasıyla birlikte konunun her yönüyle idrak edilmesine imkân verecek (Buhalis ve dig., 2004) ve bütün sektörler için sonuçları genelleştirmeyi mümkün kılacaktır. Sonuç olarak araştırma ana amacı olan nitel yaklaşımıyla başarılı bir şekilde elde edilen, konunun derinlemesine anlaşılmasına odaklanıp devam etmiştir.

8.2 ARAŞTIRMA EVRENİ VE VERİ TOPLAMA

Veri ya da büyük veri ile ilgilenen kurumların yöneticileri ya da operasyonel birim personeli gibi çeşitli aktörler röportaj için seçildi, işini iyi bilen uzmanlar ve belli bir ürün üretimi, müşteri memnuniyetini dikkate alan kurumlar olmasına özen gösterildi (Bkz. Şekil 8.1)

Görüşülen kişilerin seçimi, aktörlerin, bilgi alanlarının hem de içeriğin çeşitliliğini yakalamak niyetiyle tasarlanmıştır. Çeşitli insanlar grubu içerisindeki yönetici ve

operatörlerin bakış açılarını getirmek için örnek olarak tercih edildi. Bu kişiler dışarıdan uzmanların ve müşterilerin bakış açılarıyla birlikte kendi dönemlerindeki büyük veri kullanımının nasıl olduğunu yönetmeye etkin bir şekilde gayret gösterirler. Bu kişiler yenilikçi uygulamaları ve net bir şekilde prosedürleri önererek yeni ürün ve servis türleri için alışılmışın dışında, orijinal fikirler gelişmesine yardım edebilirler, var olan problemlerin çözümü için yeni yollar ortaya koyarlar ve iki tarafa (kurum ve müşterileri) karşılıklı fayda gelişmesi için girişimde bulunurlar.

Ayrıca görüşülen kişilerin hem pozisyonları itibarıyla hem de tecrübeleriyle kurumlardaki ilgili oldukları projelere etki edebilecek kişilerden seçilmeleri de önemli bir husustur. Bundan dolayı yeni fikirleri ortaya koyma, takibini yapma, geliştirme süreçlerinde aktif rol oynamaktadırlar. Projelerin saha uygulama safhalarına da bizzat dahil olmaları, araştırma konusu ile ilgili olarak pozitif ya da negatif yanları daha iyi ortaya koyabilmelerini sağlamıştır.

Yine kendi pozisyonlarındaki lokal sorumlulukları ve bilgileri dışında konu ile ilgili genel literatür bilgisine ve konunun sürekli gelişmesinin gereği olarak güncel bilgilere de sahiptirler. Bu görüşme esnasında kişilerle güncel olarak aynı hususları paylaşabilme açısından avantajlar sağlamıştır.

Bu iç-dış yaklaşımlar ve bakış açıları arasında güçlü ve sağlam kıyaslama yapmak için fırsat sağlayacaktır. Bu da bu araştırma sorusu için en uygunu olan yorumlayıcı-olgusal araştırmamanın bir başka Özelliğidir.

Vurgulanması gereken şudur ki aktörlerin farklılığı büyük veri kullanımındaki geniş kapsamlı bir perspektif kazanmak ve ana aktörlerin görüşlerindeki farklılıklar araştırmak için önemliydi. Farklı aktörler, özgün ve farklı avantajlı noktalarının tümü, algılarını çalışmak, fikirlerin görünen çatışmalarını gözlemlemek ve ilginç bulguları sonuçlandırmak için girişimde bulunmuştur.

Bu çalışmanın tamamlayıcı bir parçası yöneticiler, operatörler ve uzmanlar ile olan röportajlar ve konuşmalar üzerine dayandırılmıştır. Yarı yapılandırılmış röportajı yürütmek için bir standart görüşme rehberi (EK.1) yapıldı. Soruların amacı çalışma tarafından kapsanan ana konular hakkında daha fazla şey öğrenmek, karşı karşıya kaldıkları yönetimsel ve teknik zorluklar ile ilgili yöneticilerin/operatörlerin düşünelerini öğrenmeye çalışmaktır. Bu çalışmadan gelen geribildirimler, kullanabilecekleri ya da bilecekleri (kullanılan büyük verinin içeriğindeki) araçlar ve tekniklerdir.

Düzen taraftan uzmanlar ile gerçekleştirilen konuşmalar, belirli alanlar içerisindeki anlayışlara, bu araştırmmanın uygunluğu ve oryantasyonu hakkında geribildirime, kurumların büyük veriden nasıl yararlanmış olabileceği üzerine tavsiyeler hakkında daha başka bilgilere erişmek için tasarlanmıştır.

Aktörlerin kimlik belirleme süreci ve son röportajların ayarlanması 4-5 hafta devam etti. Her biri yaklaşık 60-90 dakika süren röportajlarla görüşme yapılan kişilerin işyerlerinde yürütüldü.

Kurumların çalışanları ile yürütülen tüm görüşmeler “Büyük veri sektörünüzde kullanılıyor mu?” sorusuyla başladı. Bu konudaki derinlemesine tartışmadan sonra, kurumdan iş operasyonlarının geliştirilmesi için büyük verinin nasıl kullanılacağı ile

Kişi	E/K	Pozisyon	Sektör	Kurum Genel Özellikleri
1	E	Takım Lideri	Çağrı Merkezi, Telekomünikasyon	~10.000 Kişilik. 2007'den beri faaliyette. Müşteri odaklı çözümler sunan bir firma.
2	E	Proje Yöneticisi	Ar-Ge, Bilişim	~3.500 Kişilik. Proje odaklı çalışan kamu kurumu.
3	E	IT Müdürü, Süreç ve Sistem Geliştirme	Otomotiv	~25.000 Kişilik. Uluslararası bir otomotiv firmasını Türkiye grubu.
4	E	Büyük Veri Teknik Takım Lideri	Büyük Veri, Telekomünikasyon	2011'de kurulmuş ve büyük veri üzerine çalışan firma.
5	E	İş Geliştirme Uzmanı	Ar-Ge, Bilişim	~3.500 Kişilik. Proje odaklı çalışan kamu kurumu.
6	E	Yönetici	Bilişim	Mühendislik danışmanlık hizmeti veren Kobi.
7	E	Satın Alma Süreç Sorumlusu	Ar-Ge, Bilişim	~3.500 Kişilik. Proje odaklı çalışan kamu kurumu.
8	E	Müdür	Bilişim	Yazılım ve e-imza konularında Kobi.

Şekil 8.1 Örneklem tablosu.

ilgili anahtar yönetim tekniklerinin verilmesi istendi. Ondan sonra, kurumlarındaki yenilikçi yetenekler için zorluklar ve kolaylıklar olarak hareket alanı tartışıldı. Son bölümde, görüşülen kişiye konuya ilgili tamamlayıcı görüş ve potansiyel çözüm soruldu.

Görüşme yapılan kişiler özgürce konuşmaka, düşüncelerini ve deneyimlerini ifade etmekte teşvik edilmiştir. Ayrıca araştırma (izleme), konunun daha iyi anlaşılmasını sağlamak, farklı faktörleri bağlamak, bir faktörün önemini sınmak, diğer görüşlerle karşılaştırma yapmak ve cevabı etkileyebilen potansiyel önyargıyı tanımlamak için kullanılmıştır.

Bütün röportajlar görüşülen kişi ile yakın ilişki kurulmak suretiyle kaydedildi. Görüşme esnasında notlar tutuldu, karşılıklı tartışmalar yapıldı ve bunlardan anlaşılan sonuçlar soru listesinde cevaplar olarak işaretlendi.

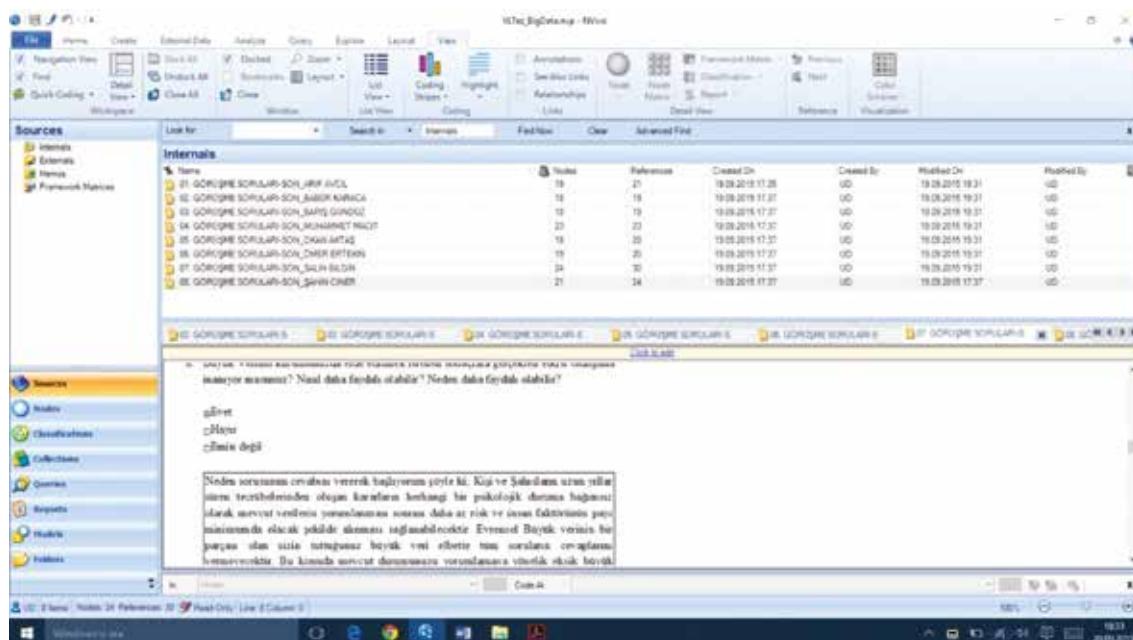
8.3 ARAŞTIRMA ETİĞİ

Katılımcılar araştırmaya katılmayı isteyerek kabul etmiştir. Ayrıca katılımcılar kendi kişisel önyargılarının ve bakış açılarının araştırma üzerinde hiçbir önceliği olmayacağına ve de görüşlerinin adil ve eşit bir tarzda dikkate alındığına ve değerlendirilmiş olduğuna ikna edilmiştir.

8.4 ARAŞTIRMA SONUÇLARI VE DEĞERLENDİRME

Bu bölümde yaptığımız görüşmelerde verilen cevaplar yer alacak ve değerlendirilmesi yapılacaktır. Görüşme analiz tablosunda katılımcıların cevaplarından oransal olarak nicel bir veri elde edilecektir. Elde edilen oranlar ile yapılan görüşmelerdeki yaklaşımalar birlikte değerlendirilerek analiz edilecektir.

Görüşmeler için her bir görüşmeci ile yapılacak röportaj 16 sorudan oluşmaktadır. Bu sorularla birlikte konu ile ilgili 8 kişi ile görüşülmüş, 128 cevap alınmış, görüşmelerin Word düzenlemeleri 36 sayfadan oluşmuştur. Öncelikle yapılan görüşmeler Word dosyası formatında NVivo ver.10 programının Sources/Internals bölümüne Import edilmiştir (aktarılmıştır). Sources kısmı yaptığımız görüşmelerin dökümlerinin yer aldığı bölümlereidir.



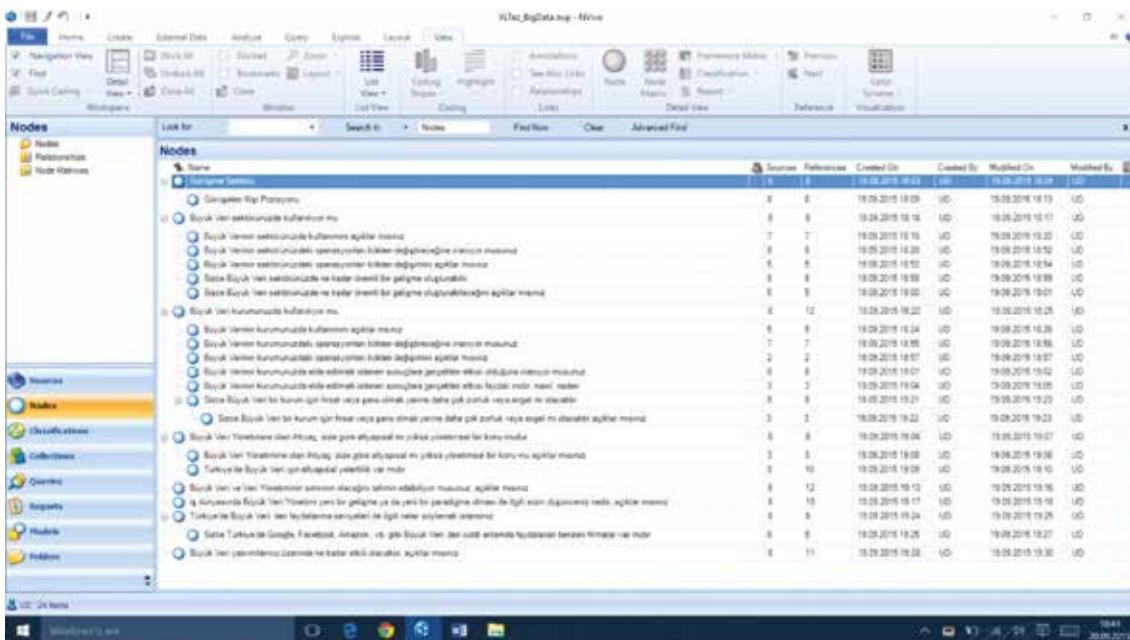
Şekil 8.2 NVivo programına görüşmelerin yüklenmesi. (Kaynak: NVivo çıktısı, 2015)

Görüşme kaynaklarını programa yükledikten sonraki ikinci adım, kaynakların Nodes olarak ifade edilen kodlamalarının yapılması sürecidir. Nodes kavramı analiz sürecinin en temel noktası olarak ifade edilebilir. Nodes'lar verileri kodlarken kullanacağınız case, kod, kategori, temalar vb. yani aynı şeyi anlattığını düşündüğünüz cümleleri toplayacağımız kelime veya kelime gruplarıdır. İki şekilde node'ları düzenlemekteyiz: Free Nodes'lar herhangi bir kategori, kod vs. ile ilişkisi olmadığını düşündüğünüz nodes'lardır. Tree Nodes'lar ise, Nodes'ların alt-üst gibi ilişkileri olduğunu düşündüğümüzde kullandığımız Nodes'lardır. Her alt veya üst kategori eklemenizde ağaç şeklinde ilişki ağını görebilirsiniz.

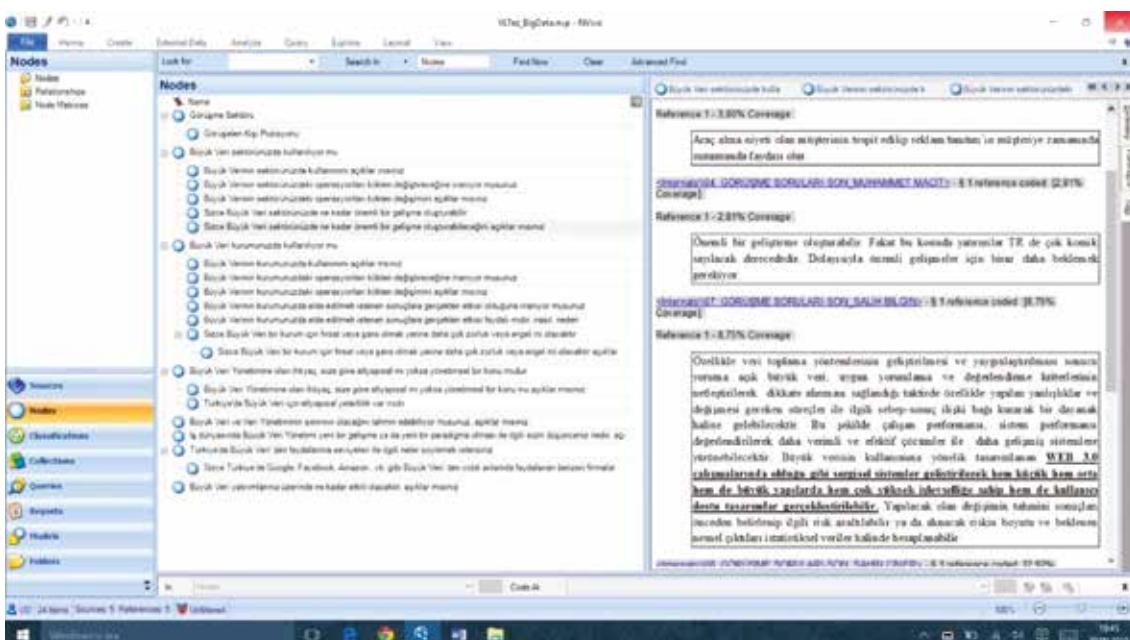
Bu şekilde Sources bölümüne yüklediğimiz görüşmelerin Nodes'lar ile kodlanması yapıldı. Aşağıdaki ekran görüntüsünde oluşturulan Nodes ve Tree Nodes'lar görülmektedir.

Her bir ilgili Nodes'lara diğer ilgili Nodes'lar ilintileştirilerek Tree Nodes'lar oluşturulmuştur. Böylece birbirine bağlı konuların ve noktaların incelenmesi daha kolay hale gelmiştir.

NVivo her bir Node ve ona bağlı Tree Nodes'ların hangi kaynaklardan geldiği ve kaç referans noktası olduğu ve bu kaynaklara ya da referanslara kolayca ulaşımı sağlamaktadır.



Şekil 8.3 Nodes ve Tree Nodes kodlama görünümü. (Kaynak: NVivo çıktısı, 2015)

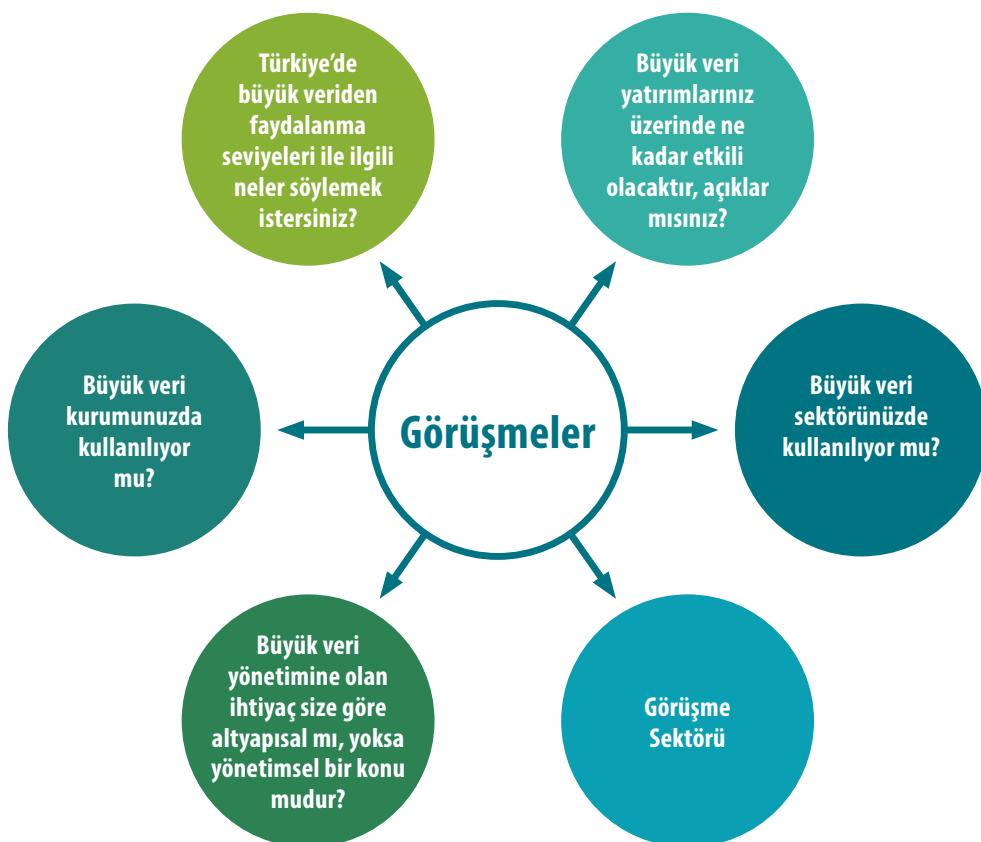


Şekil 8.4 Nodes ve Tree Nodes kodlamalarının incelenmesi. (Kaynak: NVivo çıktısı, 2015)

Bunların yanı sıra çalışmalarında NVivo'nun bize sunduğu diğer olanaklardan yararlanılmıştır. NVivo ile birbirile ilişkili metinler ve konular kolayca bize rehberlik edecek şekilde görüntülenebilmektedir. Böylece elimizdeki yarı yapılandırılmış veriden daha fazla olgusal yorumlar çıkarma imkânını elde etmiş olmaktadır.

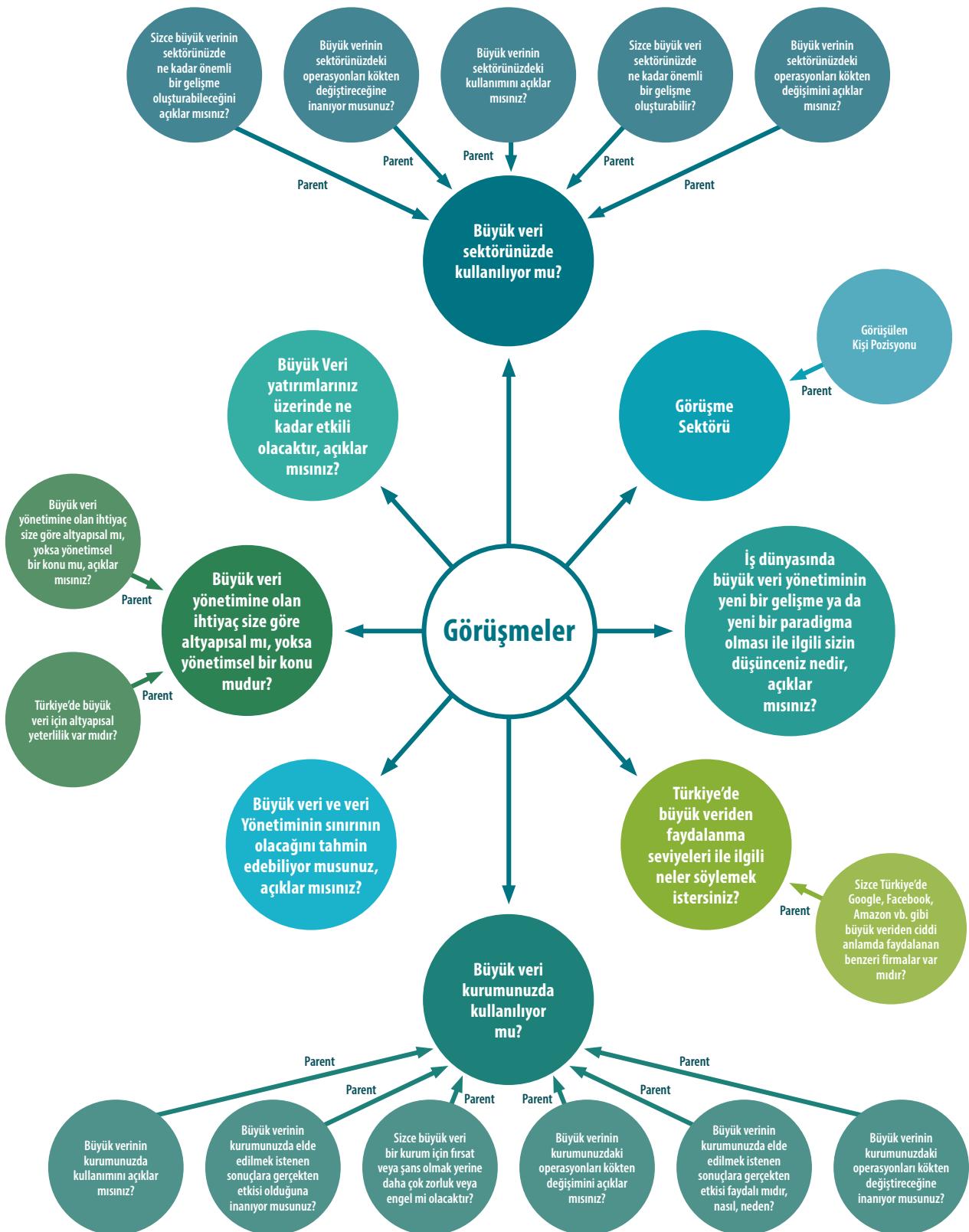
Aynı sorulara farklı kişilerin verdiği cevaplar ya da kişilerin tüm sorulara verdiği cevapları istediğimiz biçimde görme imkânını sağlamaktadır. Yine elimizdeki veriler üzerinde özel olarak ilgilendiğimiz herhangi bir kelimenin ya da kelime grubunun sorgulamasını yaparak bu kelimenin geçtiği tüm paragrafları bir arada görme ve değerlendirmeyi daha sağlıklı bir şekilde yapabilmekteyiz.

Kodlamaların yapılmasıyla çalışma için “model” ortaya konulmuş olmaktadır. Bu modeli NVivo yardımıyla aşağıdaki şekillerde görebilmekteyiz.



Şekil 8.5 Analizin genel çerçeve modellemesi. (Kaynak: NVivo çıktısı, 2015)

NVivo yaptığımız görüşmelerde öne çıkan kelimelerin kaynaklarını ve miktarlarını da raporlayabilmektedir. Herhangi bir şekilde sorgulama (query) yapabilmek olanağı vermektedir. Böylece en çok geçen kelimelerin kaynaklarına giderek analizimizde odaklanmayı sağlayabilmeye yardımcı olmaktadır.



Şekil 8.6 Analizin detay modellemesi. (Kaynak: NVivo çıktısı, 2015)



Şekil 8.7 NVivo kelime sorgulamaları. (Kaynak: NVivo çıktısı, 2015)

Yaptığımız çalışmada en çok geçen kelimelerin NVivo görsel hali şu şekilde olmuştur:

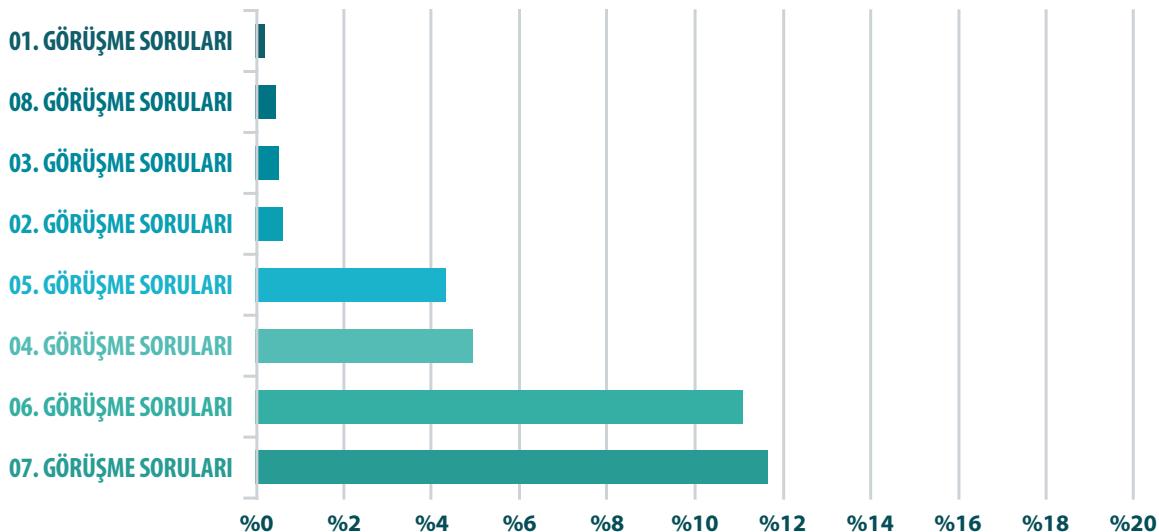
İlaveten, NVivo ile hangi kaynakta, hangi Nodes ile ilgili daha fazla bilgi ve değerlendirme yapılmış olduğunu da göstermekte ve incelemelerimize kaynaklara yönelik açısından destek sağlamaktadır. Örneğin Nodes'lardan biri ile ilgili grafik şu şekilde verilmiştir:

Bu düzenlemeler yapıldıktan sonra yapılan kodlamalar incelenerek sonuçlar, analizler ve yorumlar yapılarak değerlendirilmeye geçilmiştir.

Bu analizleri yaparken şu hususlar göz ardı edilmemiştir:

Programdaki kodlama işlevini kullanabilmek ve bu sürecin teknik kısmına karşılık gelen adımları eksiksiz yerine getirmek bizim geçerli, tutarlı ve teorik olarak an-

Büyük veri yatırımlarınız üzerinde ne kadar etkili olacaktır, açıklar misiniz?



Şekil 8.8 NVivo'da grafik desteği örnek gösterimi. (Kaynak: NVivo çıktısı, 2015)

lamlı kategoriler oluşturduğumuzun garantisini değildir. Başka bir ifadeyle, bir bilgisayar programı kullanıyor olmak tek başına analizimizi güçlendiren bir unsur değildir.

İyi bir analiz, eldeki verilerin niteliği ve kapsamına da bağlı olarak bilgisayar programının kullanımı ile daha da güçlendirilebilir. Ancak gerçekte analizimizi güçlü kıalan şey, bir bilgisayar programının kullanılması değil, metodolojik duruşumuzun netliği ile bunu inşa ve ifşa etme becerimizdir. Yani nasıl ki nicel veri analizi programları iyi bir analizi garantilemez, aynı biçimde nitel veri analizi programlarının kullanımını da tek başına iyi bir analizin garantisini değildir. Bu nedenle araştırmacılar, metodolojik bilgilerini güçlendirmeli ve metodolojik refleksivite -neden ve sonucun karşılıklı birbirini etkilemesi- geliştirmelidirler (Kuş, 2006).

Şekil 8.9'da verilen görüşme analiz tablosu katılımcıların büyük veri hakkındaki görüşlerini sayısallaştırma gayreti olup elde edilen veri yaptığımız olgusal nitel değerlendirmeye niceł olarak katkı sağlayacaktır. Açıklama soruları yüzdelik değerlendirilmeye alınmamıştır. Görüşme rehberinin soru listesinde olduğundan dolayı tabloda gösterilmiştir. Nitel değerlendirmeler ile analiz edilmiştir.

Görüşme tablosundan da ilk bakışta anlaşılacağı üzere, büyük veriye uzmanlar tarafından sınırının olmadığı düşünülen yeni bir paradigma gözüyle bakılmakta ve

SORU	CEVAP	KATILIMCILARIN CEVAPLARI								Yüzde
		1	2	3	4	5	6	7	8	
1. Büyük veri sektörünüzde kullanılıyor mu?	A. Evet B. Hayır C. Yeni yeni	A	A	B	C	A	A	C	C	A. 50 B. 12,5 C. 37,5
2. Nasıl? Açıklar misiniz?										
3. Büyük veri kurumunuzda kullanılıyor mu?	A. Evet B. Hayır C. Yeni yeni	A	B	B	A	A	B	C	C	A. 37,5 B. 37,5 C. 25
4. Nasıl? Açıklar misiniz?										
5. Büyük verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz? Nasıl? Açıklar misiniz?	A. Evet B. Hayır C. Emin değil	C	A	B	B	A	B	A	C	A. 37,5 B. 37,5 C. 25
6. Büyük verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz?	A. Evet B. Hayır C. Emin değil	C	C	B	B	A	B	B	B	A. 12,5 B. 62,5 C. 25
7. Sizce büyük veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir? Açıklar misiniz?	A. Çok önemli B. Önemli C. Kısmen D. Önemsiz	B	C	C	B	B	C	A	C	A. 12,5 B. 37,5 C. 50 D. 0
8. Büyük verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? Nasıl daha faydalı olabilir? Neden daha faydalı olabilir?	A. Evet B. Hayır C. Emin değil	A	C	B	A	A	C	A	B	A. 50 B. 25 C. 25
9. Büyük veri yönetimime olan ihtiyaç, size göre altyapısal mı, yoksa yönetimsel bir konu mudur? Açıklar misiniz?	A. Altyapısal B. Yönetimsel C. Her ikisi	C	C	C	C	C	C	C	C	A. 0 B. 0 C. 100
10. Türkiye'de büyük veri için altyapısal yeterlilik var mıdır?	A. Var B. Yok C. Kısmen var	C	B	C	B	B	B	C	C	A. 0 B. 50 C. 50

11. Büyük veri ve veri yönetiminin sınırlarının olacağını tahmin edebiliyor musunuz? Açıklar misiniz?	A. Evet B. Hayır	B	B	A	B	B	B	B	B	A. 12,5 B. 87,5
12. İş dünyasında büyük veri yönetiminin yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir? Açıklar misiniz?	A. Evet, yeni bir paradigma veya olgu B. Hayır, verinin genişleyen devamı	A	A	A	A	A	B	B	B	A. 62,5 B. 37,5
13. Sizce büyük veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır? Açıklar misiniz?	A. Fırsat-Şans B. Zorluk-Engel C. Her ikisi	C	A	C	C	A	A	A	A	A. 62,5 B. 0 C. 37,5
14. Türkiye'de büyük veriden faydalama seviyeleri ile ilgili neler söylemek istersiniz?	A. Yeterli B. Değil C. Kısmen	B	B	C	B	B	C	B	B	A. 0 B. 75 C. 25
15. Sizce Türkiye'de Google, Facebook, Amazon vb. gibi büyük veriden ciddi anlamda faydalanan benzeri firmalar var mıdır?	A. Evet, var B. Hayır, yok C. Önümüzdeki 10 yıl içinde öngörüyorum D. Önümüzdeki 20 yıl içinde öngörüyorum	A	C	B	B	C	D	C	C	A. 12,5 B. 25 C. 50 D. 12,5
16. Nihai olarak, yaptığınız açıklamalar ve değerlendirmeler doğrultusunda, büyük veri yatırımlarınız üzerinde ne kadar etkili olacaktır? Bu etkilerin nedenlerini açıklayabilir misiniz? Eğer yanınız olumlu ise, büyük verin yatırımlar üzerinde ne tür etkileri olacaktır? Açıklar misiniz?	A. Çok etkili B. Etkili C. Kısmen D. Etkisiz	B	C	C	A	B	D	C	C	A. 12,5 B. 25 C. 50 D. 12,5

Tablo 8.9 Görüşme analiz tablosu.

gerçek anlamda fayda sağlanması için kurumların hem altyapısal hem yönetimsel bir konu olarak ele alması ve bu şekilde yaklaşım geliştirmesi gerekmekte olduğu görülmektedir.

Teknolojik gelişmelere ve yazılım gelişmelerine doğrudan bağlı bir olgu olan büyük veri yönetiminin Türkiye'deki durumunun -bu altyapıların kısmen yeterli olmasından dolayı- ağırlıklı olarak önumüzdeki yıllarda gelişimi öngörmüştür.

Genel olarak büyük veri yönetimi kurtarıcı bir konsept olarak görünmemesine rağmen, kurumların planlamaları ve stratejik yatırımları açısından artık “kaçınılması zor bir olgu” olarak karşımızda durmaktadır.

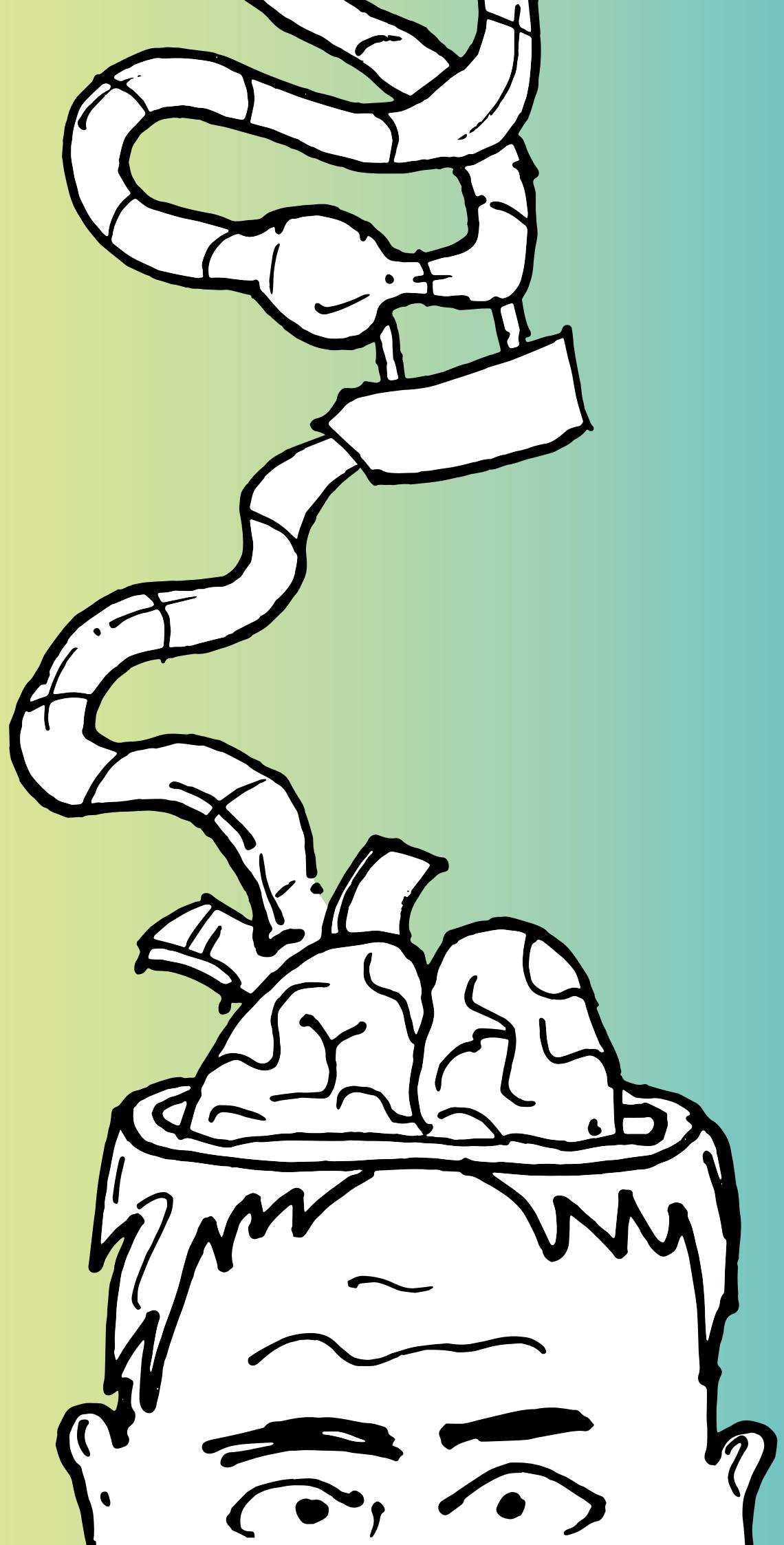
Yapılan görüşmeler, analizleri ve literatür çalışması ile elde edilen veriler “Tartışma ve Sonuç” bölümünde detaylı bir şekilde incelenerek konu ile ilgili değerlendirilmeler ortaya konulmuştur.

KAYNAKLAR

- Buhalis, D. ve Deimezi, O., 2004, E-tourism developments in Greece: Information communication technologies adoption for the strategic management of the Greek tourism industry, *Tourism and Hospitality Research*, 52, 103-130.
- Collis, J. ve Hussey, R., 2009, *Business research: A practical guide for undergraduate and postgraduate students*, Palgrave Macmillan, New York, ISBN: 978-0-230-30183-2.
- Flick, U., 2009, *An introduction to qualitative research*, Sage Publications Ltd., Hamburg, ISBN: 978-1-84787-1.
- Kuş, E. S., 2006, *Sosyal Bilimlerde Bilgisayar Destekli Nitel Veri Analizi: Örnek Program NVivo ile Gösterimler*, Anı Yayıncılık, Ankara.
- NVivo çıktısı, 2015

9. BÖLÜM

TARTIŞMA

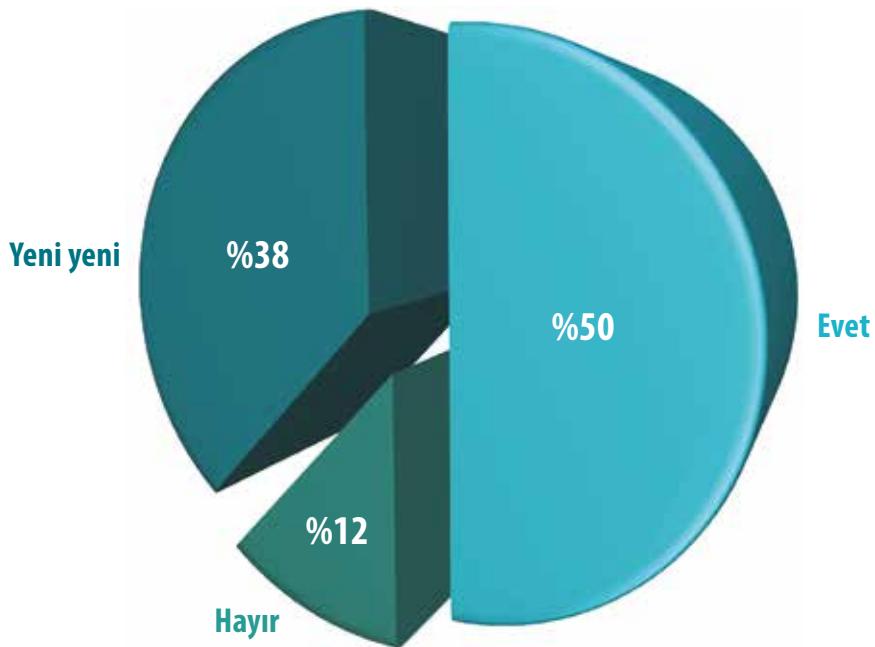


9.1 BÜYÜK VERİNİN KULLANIMI

Genel olarak görüşmeler kurumların hem dahili hem harici odaklı kavramlarda “büyük veri” kullandığını ya da kullanılmaya çalışıldığını gösterdi. Şekil 9.1’deki grafikten de anlaşılacağı üzere % 50 oranında büyük veri kullanımına yeni yeni başlayanlar da eklendiğinde ortalama % 78 oranında büyük veri ile ilgilenme söz konusudur.

Örneğin bir kurumun Büyük Veri Teknik Takım Lideri olan bir kişi sorulara şu şekilde yanıt verdi; “Artan log miktarıyla beraber sektör daha hesaplı bir şekilde bu datayı incelemek istiyor, bu sebeple firmalar Big Data teknolojilerini kullanmayı düşünmeye başladılar,” diğer bir deyişle şirketler artık büyük veri konusunu değerlendirmeyi arzulamaktadır. Bununla birlikte bir otomotiv firması yöneticisi şuna deheimiştir; “Araç alma niyeti olan müşterinin tespit edilip reklam/tanıtımın müşteriye zamanında sunumunda faydası olmaktadır” ve bu projeler “kurumun iş modelleri üzerinde doğrudan etkiye sahip olan müşterilere ulaşmada ve onları tanımlama/karakterize etmede” hedeflenmiştir, diğer tabirle harice odaklanmıştır.

Birçok katılımcının cevapları gösterdi ki kurumlar Suther'in “çok boyutluluk” olarak tanımladığı, hem kurumun var olan müşteri tabanı hem de yeni müşterilere ulaşmanın anlaşılmasını güçlendirmek için büyük veriyi kullanıyordu.



Şekil 9.1 Büyük veri sektörünüzde kullanılıyor mu? (Kaynak: Analiz çıktısı, 2015)

Kurumlar ayrıca rakipler karşısında rekabet gücü kazanmak için bir çaba içerisinde büyük veri girişimlerini kullandı. Bir yönetici “Rekabet artacaktır,” dedi ve şu konuya değindi, yapılandırılmış veri aynı zamanda “kurum tarafından rakip firmalarının müşterilerini kazanmak ve kaybetme ihtimali yüksek olan müşterileri yeniden kazanmak” amacıyla kullanılıyordu. Sınıflandırılmamış sosyal medya verisi rakip firmaların müşteri profiline ve firmaların eylemlerini analiz etmek için kullanılabilir. Örneğin Facebook üzerindeki rakip kurumun sayfasını beğenmiş insanların profil ve sayısı sınıflandırılmamış veridir ve veri analizleri bu müşterileri -hatta onlar kurumun kendi müşterisi olmasa bile- derinlemesine anlamak için kullanılabilir.

Bazı katılımcılar sosyal medya sitelerinden veri toplamanın [Hurtwitz ve diğerlerinin “insanlar tarafından oluşturulan yapılandırılmış veri olarak tanımladığı”] kurumların farklı yönlerdeki operasyonlarında da önemli bir rol oynayabileceğini düşünmektedirler. Bir yönetici bu durumu şöyle açıkladı: “Kurumu arama ve şikayetlerini seslendirme yerine, müşteriler doğrudan doğruya sosyal medya siteleri üzerinde onlardan (kuruma ya da birbirilerine) bahsederler ve böylece şikayetlerini ve önerilerini daha etkili bir şekilde yapmayı denerler. Kurum bu şikayetleri ayırtırır,

Yapılandırmamış veri aynı zamanda “kurum tarafından rakip firmalarının müşterilerini kazanmak ve kaybetme ihtimali yüksek olan müşterileri yeniden kazanmak” amacıyla kullanılıyordu.

önemlerine ve derecelerine göre sıralar, dolayısıyla müşteri şikayetlerini tam olarak saptamaktadır.” Bu, tam olarak aynı olmamasına rağmen, Hurwitz ve arkadaşlarının daha fazla deneyimi olan fiziksel şubeleri tam olarak tespit etmek ve şikayetleri daha verimli bir şekilde analiz etmek için analiz verisi kullanmaya dayalı bir araç kiralama şirketine verdikleri örnek gibidir.

Ek olarak bir katılımcı şuna işaret etmiştir: Kurumlarındaki yapılandırmamış büyük veri formlarının müşteri tarafından üretilen yaratıcı şekildeki analizlerine -doğrudan adreslenmiş olmasa bile- önem veriyor. Örneğin herhangi bir kullanıcı açıkça kurumdan bahseden herhangi bir Tweet’i görebilir ve kurumdan bahseden kullanıcı tarafından üretilen diğer formlar, mesela internet forum postları da halka açık olabilir. Bu mesajlar derlenebilir, analiz edilebilir ve günlük olarak kurumların müşteri hizmetinin gelişmesinde kullanılabilir.

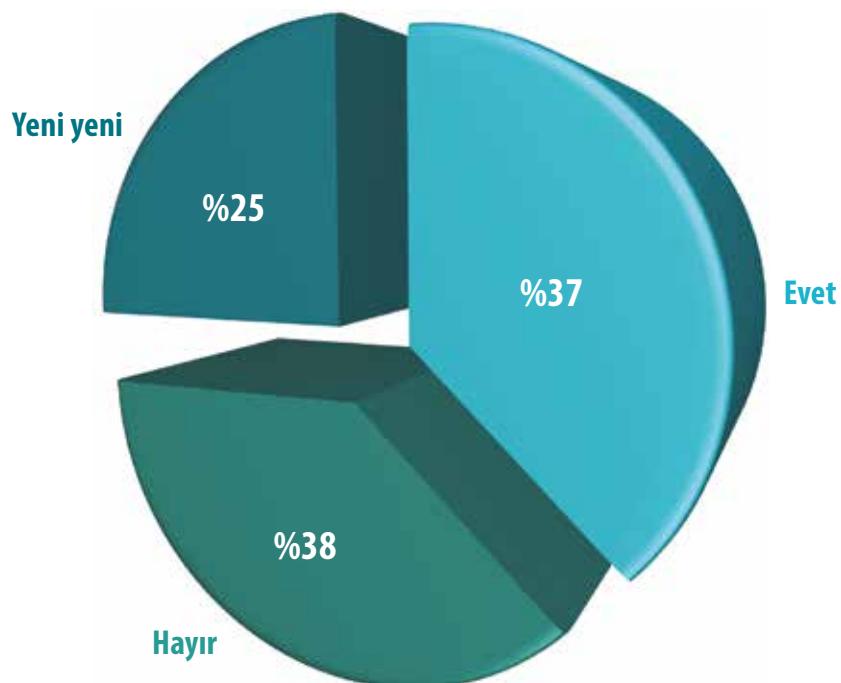
Suther ve diğerleri şu şekilde ifade etmiştir: Firma tanıtımının büyük bir yüzdesi istedikleri hedef kitleye ulaşmaz ya da sıklıkla aynı yolla benzer reklamın yanlış yönlendirmesi oluşur. Kurumun hizmet ve ürünlerinin pazarlama ve dağıtımında büyük veri ve analizlerinin nasıl kullanıldığı sorulduğunda, bir yönetici şu şekilde yanıldırı: Bir sektörde büyük verinin kullanımı ile tüm sektörlerdeki kullanımı tam olarak aynı değildir.

Diğer bir katılımcı da “Büyük verinin imkânları hakkında heyecanlıyız fakat aynı zamanda bu fırsatlarla gelen zorlukların da üzerine eğilmek zorundayız,” dedi. Katılımcılar ayrıca birçok yönden bu “zorlukları” açıkladı. Bir teknik müdür şu şekilde ifade etmiştir, “Türkiye’de bulut sistemlerin kurulması mümkün olsa da bu sistemlerin yurtdışı ile fiyatta rekabet etmesi pek mümkün görünmemektedir. Bu konudaki 5651 Sayılı Yasa ile verilerin Türkiye dışına çıkarılması da mümkün olmadığı için bulut üz-

rinde büyük veri çalışması yapabilmek olanaksız hale gelmektedir. Bu sebeple geçici süreli ve esnek kabiliyetli sunucu kümeleri kurulamamaktadır. Statik, eski sistem ile oluşturulan sunucu kümelerinin barındırmadan tutun elektriğe kadar birçok masrafi olmaktadır.”

Gizlilik yasaları ve haklar ile ilgili olarak bir katılımcı “büyük verinin yasalar ve hükümet tarafından desteklenmesine ek olarak Türkiye’deki firmaların çok ciddi yatırımlar yapması lazım ki ciddi anlamda faydalanan firmalar olsun. Ama azımsanmayacak miktarda faydalanan firmalar var ve daha da artacaktır. Kişisel bilgilerin güvenliği konusunu da kapsayan bu mesele çok ortada kalmıştır,” şeklinde ifade etti.

Büyük veri ayrıca şirketin bazı limitli (fakat yaygın değil) iç yeniden yapılanmasında neticeleniyordu: Bir katılımcı “Veriye sahip olan zaten yönetmektedir. Bunların hiçbirini ciddi anlamda veri biriktirme ile çalışmamakta, kayıtlarını bir yerlerde tutup bunları bir sonraki nesle aktarmamaktadır. Beynimizde gerçekleştirilen tüm faaliyetler büyük verinin yorumlanması olarak düşünüldüğünde asıl amaç olan deneyimlerin ya da deneyimlemenin bireye, şirkete, topluma kattığı değer birikmediği takdirde ilgili veri



Şekil 9.2 Büyük veri kurumunuzda kullanılıyor mu? (Kaynak: Analiz çıktısı, 2015)

kişi ile yaşamakta ve ölmektedir. Bir sonraki birey teslim alamadığı veriyi yorumlama hususunda hata yapmakta, bu da zarara yol açmaktadır. Önümüzdeki dönemde günümüz konuları arasında yer alan kalite yönetimi, kurumsal hafiza, ortak akıl, kurumsallık gibi kavramların subjektif kriterlerinin büyük veri ile yorumlanabilir verilere dönüştürülmesinin önemini anlaşılacığını düşünüyorum,” dedi. Konu biraz daha derinlemesine irdeledikten sonra kendisinden “kurumlarda büyük veri ile ilgili kişilerin yetişirilmesi, yöneticilerin bu konulara hâkim olması ve yönetebilir hale gelmesi, bilginin diğer personele aktarılabilmesi” hususunun çok önemli olduğu fikri edindi.

Genel olarak katılımcıların cevapları gösterdi ki kurumlar istenilen seviyede yüksek oranda ve etkin olmasa da çeşitli şekillerde büyük veriyi kullanıyordu.

9.2 BÜYÜK VERİ HAKKINDAKİ GÖRÜŞLER VE TAHMİNLER

Daha önce belirtildiği gibi, olgusal yaklaşım araştırmacıya ayrıca bir çalışma içerisinde karışan katılımcıların subjektif deneyimlerini ve duygularını saptamaya izin verir. Bundan dolayı büyük veri üzerine, uzmanlar ve yöneticilerin düşünceleri, önyargıları ve genel görüşleriyle ilgili bazı sorular da sorduk.

Genel olarak katılımcılar büyük veriyle ilgili pozitif fakat gerçekçi bir genel görüşe sahipti. Onlar büyük veriyi “her derde deva bir ilaç” gibi görmedi ve birkaç Seigel's yanıkları ya da Kent ve diğerlerinin iddiaları olan büyük verinin bir anlayışını ifade ettiler.

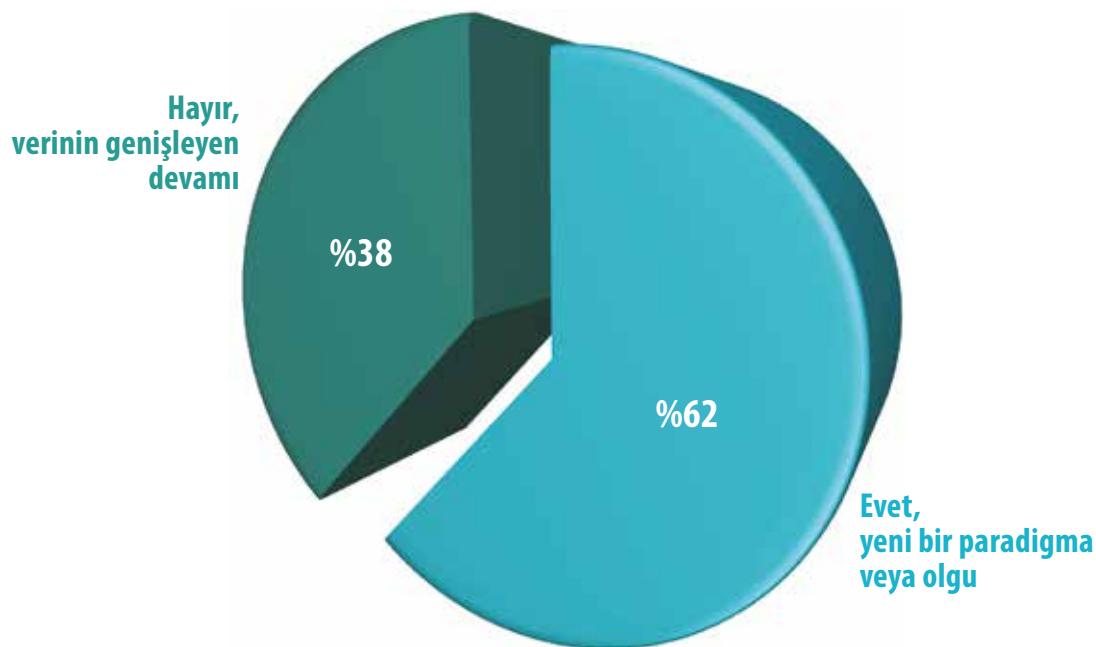
Sekiz katılımcıdan 5'i büyük verinin bir şekilde özellikle “yeni bir paradigma veya vaka” olduğunu ifade ederken, 3 katılımcı tamamen yeni bir paradigma olmasından ziyade, basit bir şekilde, ne olursa olsun durmadan büyüyen veri miktarının genişlemesi olarak ifade etmiştir.

Bununla birlikte hiçbir katılımcı kurumlar için herhangi bir şekilde “zararlı” ya da “istenmeyen” şeklinde tarif etmemiştir; derinlemesine görüşlerini almak yerine, potansiyel faydalar ile potansiyel zorlukları denelemek için yeni yaratıcı düşünce şekilleri gereklidi. Her halükârda, hiçbir katılımcı (büyük veriyi özellikle farklı bir olgu olarak görmeyen birisi bile) “geçici” bir gelişme olarak nitelmedi ve genel olarak hepsi şunu kabul etti; büyük veri “kalmak için burada” idi ve bundan dolayı kurumların uzun dönemli odağını ve konsantrasyonunu gerektirdiğini düşünmektedirler.

Katılımcılardan bir diğeri olan bir telekomünikasyon firmasının Big Data Takım Lideri yerel teknik durumla ilgili olarak da şunu ilave etti: Büyük veri sektörü sadece yeni soluk getirecektir. Yapısı itibarıyla klasik veri yönetimi sistemleriyle yer değiştirebilecek teknoloji değildir. Daha ziyade beraber kullanılması daha doğru olacaktır. Operasyonlar hemen hemen aynı kalmakla beraber elde edilen işletme değerleri artacaktır, çünkü eskiden yapılamayan raporlama ve veri işleme süreçleri çok daha kolay bir şekilde yapılabilecektir. Özellikle Türkiye'de firmaların onlarca-yüzlerce sunucu kullanarak büyük veri işleme gibi girişimleri bulunmamaktadır. Halbuki büyük veri gücünü paralel işleyebilme özelliğinden almaktadır.

Şu yanıtlar da önemliydi, katılımcılardan 3'ü büyük veri bu sektörde "devrim yapamaz" (belki de onlar onu "yeni" bir paradigma olarak görmediği için ya da altyapısal-yönetimsel sorunlar nedeniyle), diğer 3'ü olabilir dedi ve kalan 2'si emin olmadı.

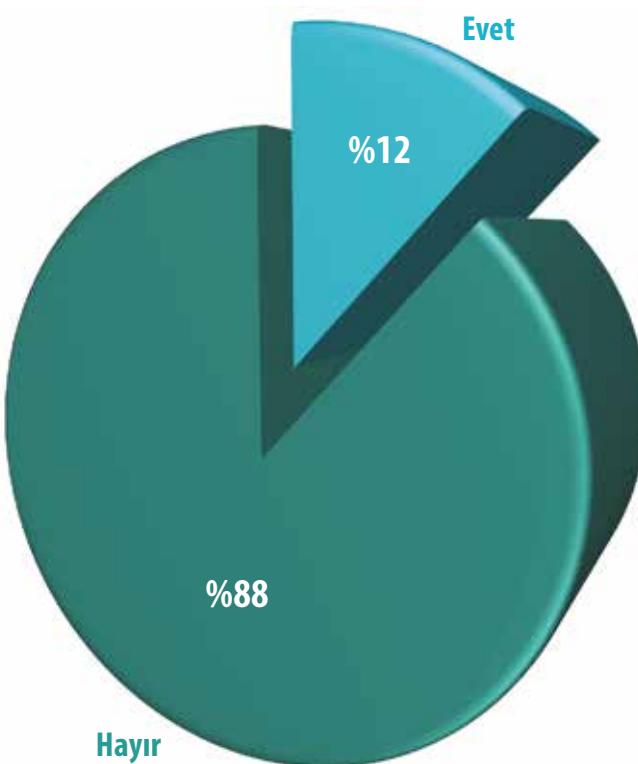
Bir katılımcı, o bir paradigmadır fakat "Yeni bir paradigma olmasına rağmen ne yazık ki iş dünyası bu değişimi henüz anlayabilmiş değildir. Hâlâ eski geleneksel kıstaslara göre değerlendirmeler yapılmaktadır" şeklinde ifade etmiştir.



Şekil 9.3 İş dünyasında büyük veri yönetiminin yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir? (Kaynak: Analiz çıktıları, 2015)

Başa bir katılımcı ise buna karşı olarak şunu ifade etmiştir: Şirketler kendi resmi ya da gayri resmi raporlama ihtiyaçları için tutmuş oldukları verilerin zamanla genişleyen ihtiyaçlara yeterli gelmediğini, büyüyen işletme için gereken faaliyetlerinin yönetilmesine yönelik ihtiyacın ortaya çıkması ile yeni kayıtlar tutmanın gerekliliğini anladıkları anda bu amaca yönelik yapmış oldukları tüm girişimler, firmaların veri tutma ve analiz etme kabiliyetlerinin gelişmesine sebep olur. Durumdan da görüldüğü üzere her kararın bir sebebi, her sebebin belirli bir dönemlik farkındalığı, dolayısı ile her geliştirmenin bir ihtiyacın ortaya çıkması ile yapıldığı görülmektedir.

Büyük veri bu örnekte de olduğu gibi bir ihtiyacın giderilmesine yönelik verinin bulunmayışı sonucu iradi olarak genişletilen veri bütününden oluşur ve genişletildikçe büyük veriyi yöneten kişiye yeni karar alanları ve yöntemleri hakkında farklı ufuklar açan bir araç haline gelir.



Şekil 9.4 Büyük veri ve veri yönetiminin sınırının olacağını tahmin edebiliyor musunuz?
(Kaynak: Analiz çıktıları, 2015)

Büyük Verinin kurumlar için “gelecekteki büyük şey” olup olmadığı sorulduğunda, katılımcıların çoğu Büyük Veri için “gelecek sınırın” ne olabileceğini kesin olarak öngöremedi.

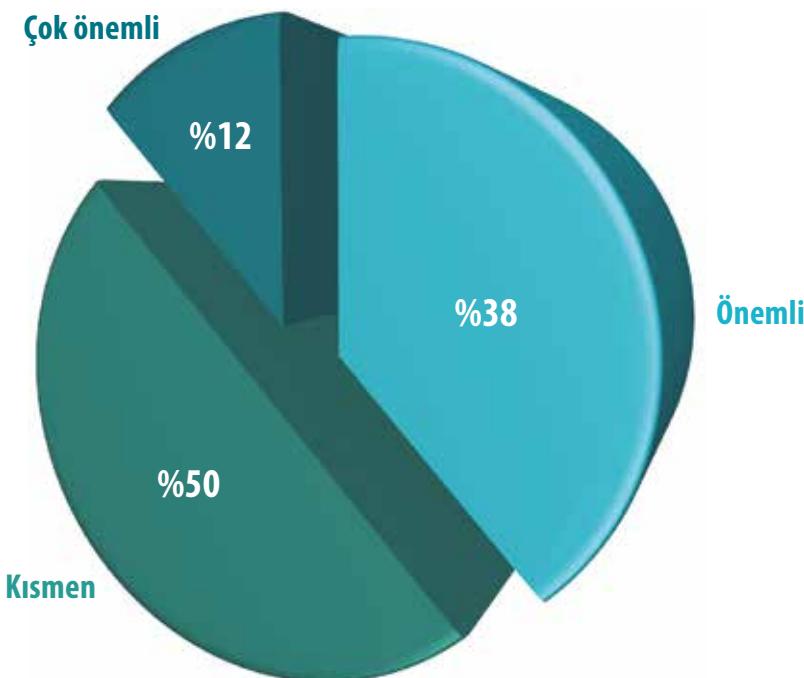
Büyük verinin kurumlar için “gelecekteki büyük şey” olup olmadığı sorulduğunda, katılımcıların çeşitli cevapları vardı. Katılımcıların çoğu büyük veri için “gelecek sınırın” ne olabileceğini kesin olarak öngöremedi.

Bir yönetici şuna deyindi, “Tüm şirketler aynı şekilde -özellikle bir grup yerine bizim müşterilerimize ürünlerimizi uygun hale getirmek yoluyla- büyük veriden faydalananmayı umar. Bundan dolayı [büyük veri ile] pazarın geleceğinde ne olacağını tam olarak öngörmek çok zordur.”

Başka bir katılımcı ise “Büyük veri sürekli gelişen sistemler tasarılanmasına yol açar. Tasarlanarak değişen sistem üzerinden de veriler birikmeye devam eder. Bu da yeni yöntem ve sistemlere sebep olacağından sürekli yaşayan, yaşadıkça gelişen, gelişikçe değişerek büyüyen bir sistem ve bu sistemin kararlarını oluşturan etkenlerin içinde bulunulan değişken duruma göre, bağlı fonksiyonların izlendiği sürecin çıktısı sadece değişim, gelişim ve yeni kararlar ile bir sonraki dönemdeki faaliyetlere karar vermeye yarayan yeni verilerin büyük veriye eklenmesi olacaktır. Kanımcı tek sınır kişisel hak, hürriyet ve kanunla belirlenmiş kurallar olabilir” şeklinde bir yaklaşımda bulundu.

Bu görüşmecinin veri ile birlikte yönetsel pozisyonundan ve bu ifadelerinden dolayı büyük verinin, yönetim anlayışına etki etmesi bekłentisi zihinlerde bulunmaktadır.

Kurumlarda elde edilmek istenen sonuçlara etkisi ve neden/nasıl sorusunu ise çoğu katılımcı ihtiyatlı bir şekilde yanıtladı. Bir katılımcı, “Evet, faydalı olabilir, neden sorusunun cevabını vererek başlıyorum, söyle ki: Şahısların uzun yıllar süren tecrübelerinden oluşan kararların herhangi bir psikolojik duruma bağımsız olarak mevcut verilerin yorumlanması sonrası daha az risk ve insan faktörünün payı mini-



Şekil 9.5 Sizce büyük veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir?
(Kaynak: Analiz çıktısı, 2015)

mumda olacak şekilde alınması sağlanabilecektir. Evrensel büyük verinin bir parçası olan sizin tuttuğunuz büyük veri elbette tüm soruların cevaplarını vermeyecektir. Bu konuda mevcut durumunuzu yorumlamaya yönelik eksik büyük verinin diğer büyük veri kullanıcılarından temini ve/veya büyük veri paylaşımı durumunda çok daha tutarlı, entegre ve yönetilebilir faaliyetler tasarlanarak tüm girdileri ve çıktıları belli olan bir matematiksel model yerine belirli koşullara bağlı olarak değişen, sizin yeri-nize düşünen, karar vermede sizlere seçeneklerle yardımcı olan bir Stokastik Model tasarımlı yapılabilir. Nasıl sorusuna vereceğim yanıt ise basit bir şekilde her alınan karar veya riskin, yapılacak kâr odaklı çalışmaların öznel sebeplerinin bulunduğu matematiksel dayanağı olacağının şeklindedir” dedi.

Katılımcılardan elde edilen genel düşünce şuydu: Büyük veri girişimlerimizin kritik bir parçası olmaya devam edecek, fakat kendimizde fazla abartmadan ve gerçek bir fikir çerçevesini korumalıyız. Sadece pazardaki bütün diğer önemli gelişmeler gibi davranışmalıyız. Şirketler için son derece önemlidir, çünkü “Şirket için yapılandırılmış

veri kullanımı artık ihtiyaç duydukları avantajı sağlamada yeterli değildir. Daha hızlı bir şekilde internet ortamındaki yapılandırılmamış veriyi toplama ve işlemek için, büyük veriyi anlamak ve veri yönetimindeki en son gelişmelerle ilgili güncel kalmak kurumlar için son derece önemlidir.”

Yine bir katılımcı büyük veri fırsat-şans mı, yoksa zorluk-engel mi sorusuna, büyük veri kurumlar için bir fırsat-şans olarak karşımıza çıkar diyerek şu şekilde yorum getirmiştir; “Doğru yorum ve değerlendirme sonucu iyi tasarlanmış ve işletilen kararlar doğru yönetildiği takdirde başarıya ulaşacaktır. Bu sürecin başarısı süreç sahibinin genel kabiliyeti ve iradesi ile doğru orantılıdır. Büyük veri ilgili süreç sahibine yukarıda saymış olduğum; yorum, değerlendirme, tasarım, işletim, kararın niteliği ve bağlı unsurları, yönetim fonksiyonları gibi daha birçok başarı odaklı çalışmalar süresince gerekli ya da gerekebilecek nice veriyi nitel fonksiyonlar halinde önüne serecektir. Burada en önemli unsur yorumlanacak büyük verinin yorumlanması için gerekli bilginin doğru kaynaklardan doğru şekilde elde edilmesi olacaktır. Bu konudaki başarısızlık; gelişimdeki zorluk-engel ve sonrasında zarar olarak karşımıza çıkacaktır” şeklinde yanıt verdi.

KAYNAKLAR

- Hurwitz, J., Alan N., Fern H., ve Marcia K., 2013, Big Data for Dummies, John Wiley & Sons, ISBN: 978-1-118-50422-2.
- Kent, Paul, Radhika Kulkarni, Udo Sglavo (2013). Big Data and Business Analytics, Auerbach Publications.
- Siegel, E., 2013, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, Wiley, ISBN: 978-1-118-35685-2.
- Suther, Tim, Bill Burkart, Jie Cheng (2013). Big Data and Business Analytics. Auerbach Publications.

10. BÖLÜM

**VERİ BİLİMİ,
YAPAY ZEKA VE
MESLEKLERİN
GELECEĞİ**



Birçok insan çalışma dünyasının geleceğinde kaçınılmaz teknolojik değişikliklere işaret ediyor. Bununla birlikte bağlam, seçimler ve bireyler, firmalar, sivil toplum kuruluşları, hükümetler ve uluslararası kuruluşlar için fırsatları şekillendiren yeni koşullara uyarlanabilirlik arasındaki karmaşık dinamiklere bağlı çok sayıda olası geleceğin var olduğu öngörülmektedir. Konuya ilgili son araştırmalarda ülkeler arasında ve içinde artan eşitsizlik, iş kaybı, genç nesillerin karşılaştığı zorluklar ve işyerinde haklardan mahrumiyet gibi zorluklara vurgu yapılmakta. Ayrıca işlerin değişen doğası ve organizasyonu ve ekonomi genelinde yeni teknolojilerin benimsenmesi hem teknolojiye dayalı hem de geleneksel endüstrilerde yeni işler gibi yeni fırsatlar doğurmaktadır. Pazar erişimini genişletmek, işin değişen doğasını ve organizasyonunu yönetmek ve düzenlemek için daha uygun modeller ve geliştirilmiş işyeri teknolojisi değişiklikleri olumlu sonuçlara yol açabileceği diğer alanlardır. Bu değişiklikler dünyanın dört bir yanındaki ekonomileri ve toplumları farklı şekilde etkilerken, zorluklar ve fırsatlar araştırma olanakları doğurmaktak ve işbirliğiyle öğrenme için de yollar açılmasını sağlamaktadır.

İşin değişen doğası ve organizasyonu ve bunun toplumlar üzerindeki çeşitli etkileri yalnızca yeni konulara uygulanan en iyi uygulamaları değil, aynı zamanda makine öğrenimi, robotik, ağ bilimi ve veri bilimi gibi önemli alanları da hayatımıza dahil etmiştir.

10.1 VERİ BİLİMÇİ KİMDİR VE ROLÜ NEDİR?

Hızla gelişen teknoloji, yapay zeka ve nesnelerin internetinin hızla büyümesi sonucu dünya engellenemez bir değişim süreci içine girmiştir. Örneğin 2025 yılında nesnelerin internetinden elde edilecek kazancın 1.1 trilyon dolar olacağı öngörülmektedir. Bu noktada insan kaynağına olan ihtiyaç da yeni gelişen sektörler ve alanlarda artma eğilimi içindedir. Büyük veri alanında şirketlerin artan yatırımları da bu noktada yeni iş alanları ve iş tanımlarını ortaya çıkarmıştır. Yeni yeni duymaya başladığımız bu mesleklerden biri de “veri bilimci”dir.

Veri bilimciler belirli iş ihtiyaçlarını ve hedeflerini şekillendirmeye veya karşılamaya yönelik ihtiyaçları gidermek için büyük miktarda yapılandırılmış ve yapılandırmamış veriden içgörü elde etmekle sorumlu olan kişilerdir. İşletmeler karar verme sürecinde daha etkin ve verimli olmayı amaçladıklarından veri analitiğine daha fazla güvenmektedirler ve bu noktada veri bilimciler bilişim teknoloji stratejilerinin temel bileşenleri olarak otomasyon ve makine öğrenimi alanında uzmanlaşlıklarını için veri bilimcinin rolü giderek daha da önemli bir hale gelmektedir.

Bir veri bilimcinin temel amacı ve rolü genellikle görev için özel olarak tasarlanmış yazılımları kullanarak büyük miktarda veriyi düzenlemek ve analiz etmektir. Bir veri bilimcinin analizinin nihai sonuçları yatırım yapılan tüm paydaşların, özellikle de bilişim teknolojisi dışında çalışanların anlayabileceği kadar kolay ve basit olmalıdır.

Bir veri bilimcinin veri analizine yaklaşımı, sektörüne ve çalıştığı işletmenin veya departmanın özel ihtiyaçlarına göre değişkenlik göstermektedir. Bir veri bilimci yapılandırılmış veya yapılandırılmamış verilerden bir anlam çıkarmadan önce, işletme liderleri veya departman yöneticileri ne aradıklarını ya da istediklerini doğru ve net bir şekilde belirterek iletmelidir. Bu nedenden dolayı bir veri bilimci, şirket veya departman hedeflerini tahmin motorları, model tespit analizi, optimizasyon algoritmaları ve benzeri veriye dayalı çıktıya dönüştürmek için yeterli iş alanı uzmanlığına sahip olmalıdır.

2025'E KADAR NESNELERİN İNTERNETİ

IoT'YI ÖLÇEKLENDİRMEK

IoT, mobil operatörler ve daha geniş endüstri için muazzam bir büyümeye fırsatı sunuyor, ancak bu büyümeye bağlantılılıkta daha fazlasına dayanıyor.

GSMA Nesnelerin İnterneti Programı, ölçüği artırmak ve bağlantılı bir dünya vizyonunu gerçeğe dönüştürmek için stratejik araçlar kullanıyor.

1.1 TRİLYON DOLAR

IoT GELİR FIRSATI

25 Milyar Bağlı Cihaz



KAPSAM

küresel ve evrensel faydalara sağlayan makine dostu, uygun maliyetli ağlar



YETENEKLER

bağlantının ötesinde daha yüksek değerli hizmetleri geniş ölçekte yakalamak



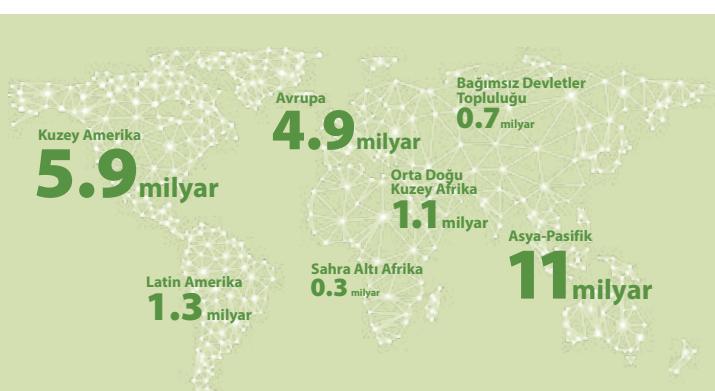
SİBERGÜVENLİK

IoT değer zincirinin her aşamasında güvenliğin yerleşik olduğu güvenilir bir IoT'yi etkinleştirmek



GELECEK YILLARDA DÜŞEN BAĞLANTILAR

TOPLAM IoT BAĞLANTILARI			
2018: 9.1 milyar		2025: 25.2 milyar	
HÜCRESEL BAĞLANTILAR			
2018: 760 milyon		2025: 3.1 milyar	
2G, 3G, 4G	Licensed Spectrum LPWA	2G, 3G, 4G, 5G	Licensed Spectrum LPWA
694 milyon	66 milyon	1.3 milyar	1.8 milyar



Şekil 10.1 Nesnelerin İnternetinin ölçeklendirilmesi. (Kaynak: <https://www.gsma.com/iot/the-gsma-iot-infographic/>)



Bir veri bilimcinin başlıca sorumluluğu, veri toplama ile başlayan ve veri bilimcinin nihai veri analitiği sonuçlarına göre iş kararlarıyla biten bir süreç olan veri analizidir.

■ Temel (kavramsal okuryazarlık) ■ Orta (pratik uygulama) ■ Gelişmiş (uygulamalı)

Beceriler	Analitik özellikli iş			Veri bilimi işleri	
	Veriye dayalı karar vericiler	Fonksiyonel analistler	Veri analistleri	Veri mühendisleri (donanım ve yazılım)	Veri bilimci ve ileri düzey analistler
Alan bilgisi Araştırma veya işletme	■	■	■	■	■
Görselleştirme Verilerdeki hikâye	■	■	■	■	■
Veri yönetimi Etik ve güvenlik dahil	■	■	■	■	■
Mühendislik Donanım, yazılım, depolama		■	■	■	■
Yönetim/İyileştirme Kaynak bulma, temizleme, kullanma		■	■	■	■
Analitik yaklaşımlar Kesinlik seviyesi	■	■	■	■	■
Makine öğrenimi Bilgisayarlarla kalıpları tanımayı öğretme			■	■	■

Şekil 10.2 İhtiyaç duyulan becerilere dair tablo. (Kaynak: <https://www.pwc.com/us/en/library/data-science-and-analytics.html>)

Veri biliminin diğer meslek grupları arasında hızla büyüyen bir alan olduğu ve veri bilimci ihtiyacının gün geçtikçe artması su götürmez bir gerçekdir. Bu bağlamda Indeed'den alınan verilere göre, veri bilimcileri için iş ilanları Ocak 2015'ten Ocak 2018'e kadar yüzde 75 oranında artarken, veri bilimci iş ilanları için yapılan aramalar da aynı zaman sürecinde yüzde 65 oranında artış göstermiştir.



İşletmelerin yapılandırılmış verileri gelir hedeflerine hizmet etmesi için giderek daha fazla kullanmak istemesi veri bilimcinin rolünü daha da eşsiz ve değerli bir hale getirmektedir.

Bir veri bilimcinin başlıca sorumluluğu, veri toplama ile başlayan ve veri bilimcinin nihai veri analitiği sonuçlarına göre iş kararlarıyla biten bir süreç olan veri analizidir. Veri bilimcilerin analiz ettiği, genellikle büyük veri olarak adlandırılan veriler bir dizi kaynaktan alınır. Büyük veri çatısı altında yer alan iki tür veri vardır: Yapılandırılmış veriler ve yapılandırılmamış veriler. Yapılandırılmış veriler, genellikle bir bilgisayarın otomatik olarak sıralaması, okuması ve organize etmesi-

ni kolaylaştırın kategorilere göre düzenlenir. Bu; hizmetler, ürünler ve elektronik cihazlar tarafından toplanan verilerdir, ancak nadiren de olsa insan girdisinden toplanan verileri de içerebilir. Akıllı telefonunuz tarafından toplanan web sitesi trafik verileri, satış rakamları, banka hesapları veya GPS koordinatları, bunların tümü yapılandırılmış veri biçimleridir.

Diger taraftan büyük verinin en hızlı büyüyen biçimini olan yapılandırmamış verilerin insan girdilerinden (müşteri incelemeleri, e-postalar, videolar, sosyal medya gönderileri vb.) gelme olasılığı daha yüksektir. Bu verilerin sınıflandırılması genellikle daha zordur ve teknolojiyle yönetilmesi daha az verim sağlamaktadır, bu nedenle yapılandırmamış verilerin yönetilmesi için büyük bir yatırım gerekebilir. İşletmeler, aranabilir terimleri kullanarak ilgili verileri çıkarmanın bir yolu olarak yapılandırmamış verileri anlamlandırmak için genellikle anahtar kelimelere yönelir.

Tipik olarak işletmeler bu yapılandırmamış verileri işlemek için veri bilimcilerine ihtiyaç duyarlar, bilişim teknolojisi alanındaki diğer personel ise yapılandırmış verileri yönetmek ve korumaktan sorumludur. Veri bilimcilerinin kariyerlerinde büyük olasılıkla birçok yapılandırılmış veriyle ilgileneceği yadsınamaz bir geçektir, ancak işletmelerin yapılandırmamış verileri gelir hedeflerine hizmet etmesi için giderek daha fazla kullanmak istemesi ve yapılandırmamış verilere yaklaşımı ise veri bilimcinin rolünü daha da eşsiz ve değerli bir hale getirmektedir.

10.2 BİR VERİ BİLİMCİNİN GELİRİ NE KADARDIR?

ABD Bureau of Labor Statistics'e (Çalışma İstatistikleri Bürosu) göre, veri bilimcisinin ortalama olarak geliri 100.560 dolardır. Veri bilimindeki yüksek gelirlerin ardından itici güç, kuruluşların büyük verinin gücünü fark etmesi ve bunu işlerinde zekice kararlar almak için kullanmak istemesidir. Bunun yanı sıra veri uzmanının arzı henüz talebi karşılayamadığından bu pozisyonlar için başlangıç maaşları, özellikle veri biliminde veya ilgili bir alanda ileri dereceye sahip olanlar için çok daha yüksek olmaya devam etmektedir.

Veri bilimcilerinin çalışıkları kurum ya da iş kolunda muazzam miktarda bir değer ürettiği yadsınamaz bir gerçekir. Diğer taraftan bir veri bilimciden beklenen



Veri bilimindeki yüksek gelirlerin ardındaki itici güç, kuruluşların büyük verinin gücünü fark etmesi ve bunu işlerinde zekice kararlar almak için kullanmak istemesidir.

tüm görevleri yerine getirebilecek birini bulmak zordur ve bu yüzden bu profesyonelleri işe alma konusunda ciddi bir rekabet söz konusudur. Bunun sonucu olarak da işverenler yetenekli veri bilimcilere en yüksek maaşları ödeme bakımından oldukça istekli davranmaktadır.

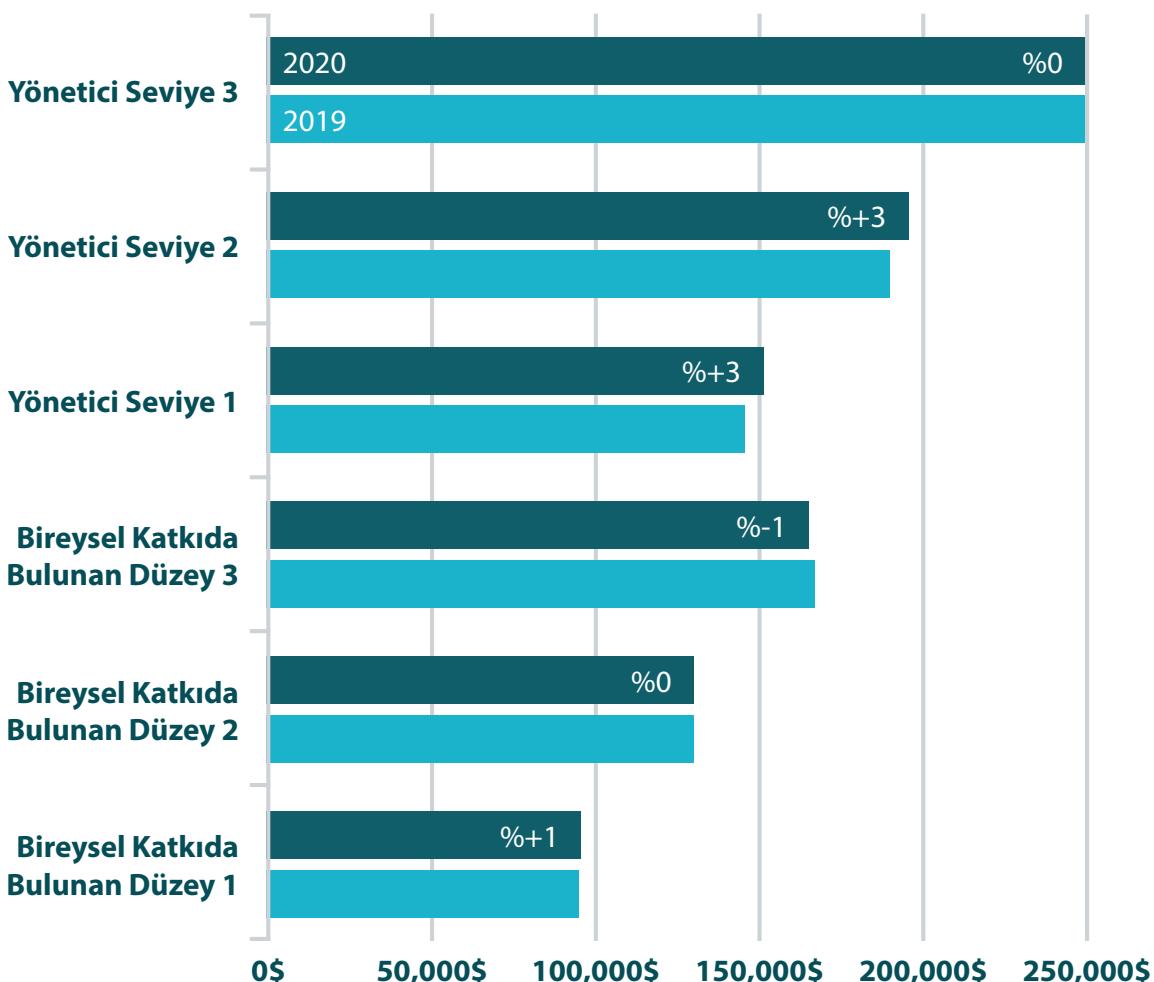
Bir veri bilimcinin çok rağbet görmesine rağmen kazancını etkileyen unsurlar söz konusudur. Bu faktörler şunlardır:

A. Deneyime Göre Bir Veri Bilimcinin Geliri

O'Reilly'nin 2016 Veri Bilimi Gelir Anketi'ne göre, deneyim bir veri bilimcinin maaşındaki en önemli faktörlerden biridir. Veri bilimi uzmanları her yıl artan deneyimleri için ortalama 2.000 ila 2.500 ABD doları daha fazla kazanıyor.

Veri bilimi gelirleriyle ilgili 2020 Burtch-Works çalışması deneyime dayalı en son gelir eğilimini ortaya koydu:

Giriş düzeyi veri bilimcisi geliri: Kariyerinin başındaki profesyonellerin son zamanda ciddi oranda sektörle girmesine rağmen, bir veri bilimcisinin ortalama başlangıç maaşı 95.000 dolarla oldukça yüksek bir miktarla tekabül etmektedir.



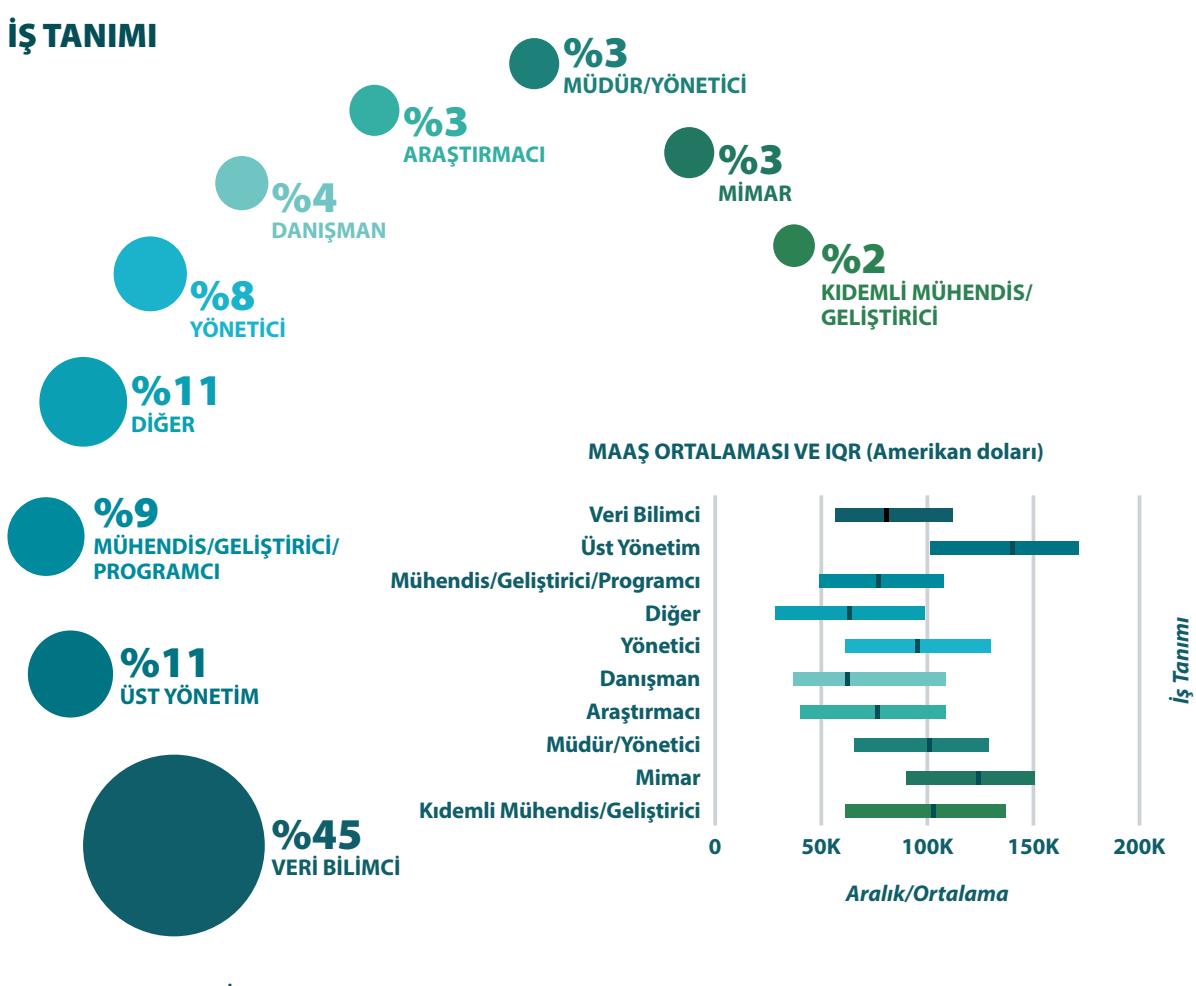
Şekil 10.3 Veri bilimcilerinin iş kategorisine göre ortalama taban maaşlarının karşılaştırılması. (Kaynak: <https://datasciencedegree.wisconsin.edu/data-science/data-scientist-salary/screen-shot-2020-09-25-at-2-43-57-pm/>)

Orta düzey veri bilimcisi geliri: Orta düzey bir veri bilimcinin ortalama maaşı 130.000 ABD dolarıdır. Bu veri bilimcinin aynı zamanda yönetimde bir rolü varsa ortalama maaşı 195.000 dolara kadar çıkmaktadır.

Deneyimli veri bilimcisi geliri: Deneyimli veri bilimi profesyonellerinin ortalama maaşı 165.000 dolarken, yönetici düzeyindeki deneyimli profesyoneller için ortalama maaş 250.000 dolar gibi önemli ölçüde yüksek bir rakam karşımıza çıkmaktadır.

B. İş Tanımına Göre Bir Veri Bilimcisinin Geliri

O'Reilly'nin Veri Bilimi Gelir Anketi'ne göre, ankete katılanların yüzde 45'i "veri bilimci" ünvanına sahip olduğunu ifade etti. Yüzde 31'i üst düzey yönetici olduklarını belirtti; bir mühendis, geliştirici veya programcı ya da diğeri. Genel olarak veri bilimi alanında bir profesyonel ne kadar çok yönetici pozisyonunda görev alırsa (ekip projeleri yönetme, analitikle çözülecek sorunları belirleme veya taraflarla iletişim kurma) maaşı da o kadar yüksek olmaktadır.



Şekil 10.4 O'Reilly Maaş Veri Bilimi Maaş Raporu. (Kaynak: O'Reilly Salary Data Science Salary Report, 2016, <https://www.oreilly.com/library/view/2016-data-science/9781492049029/>)

Veri Bilimci: Datajobs.com'a göre, "veri bilimci" ünvanına sahip olan kişiler genellikle veriye dayalı organizasyonlarda deneyimli, uzman seviyesinde profesyonellerdir. Bir veri bilimcinin maaş aralığı da 85.000-170.000 ABD doları civarındadır.

Veri Analisti: Veri analistleri verilerle uygulamalı çalışır ve veri bilimi araçları ve beceri setleri oluşturmaya odaklandıklarında kariyerlerinde bir noktaya gelirler. İşe giriş seviyesinde maaşları 50.000-75.000 ABD doları arasındadır. Deneyimli bir veri analistinin maaşı ise 65.000-110.000 ABD doları civarındadır.

Veri Bilimi/Analiz Yöneticisi: Bu alanda uzman kişiler bir ila üç kişinin doğrudan rapor verdiği, ince teknik ve nicel becerilerle güçlü liderlik ve iş becerisine sahiptir. Analiz yöneticisinin maaşı aralığı ise 90.000-140.000 ABD doları arasında değişmektedir.

Büyük Veri Mühendisi: Mühendisler, veri bilimcilerin veri analitiğini gerçekleştirmek için kullandığı platformları ve uygulamaları oluşturup sorunları çözerek iş değerini artırabilmektedir. Yeni işe başlayanların maaşı genel olarak 70.000-115.000 ABD doları civarındadır. Alanında uzman olan mühendisin maaşı ise 100.000-165.000 ABD doları arasında değişmektedir.

C. Sektörlerle Şirket Büyüklüğüne Göre Veri Bilimecinin Geliri

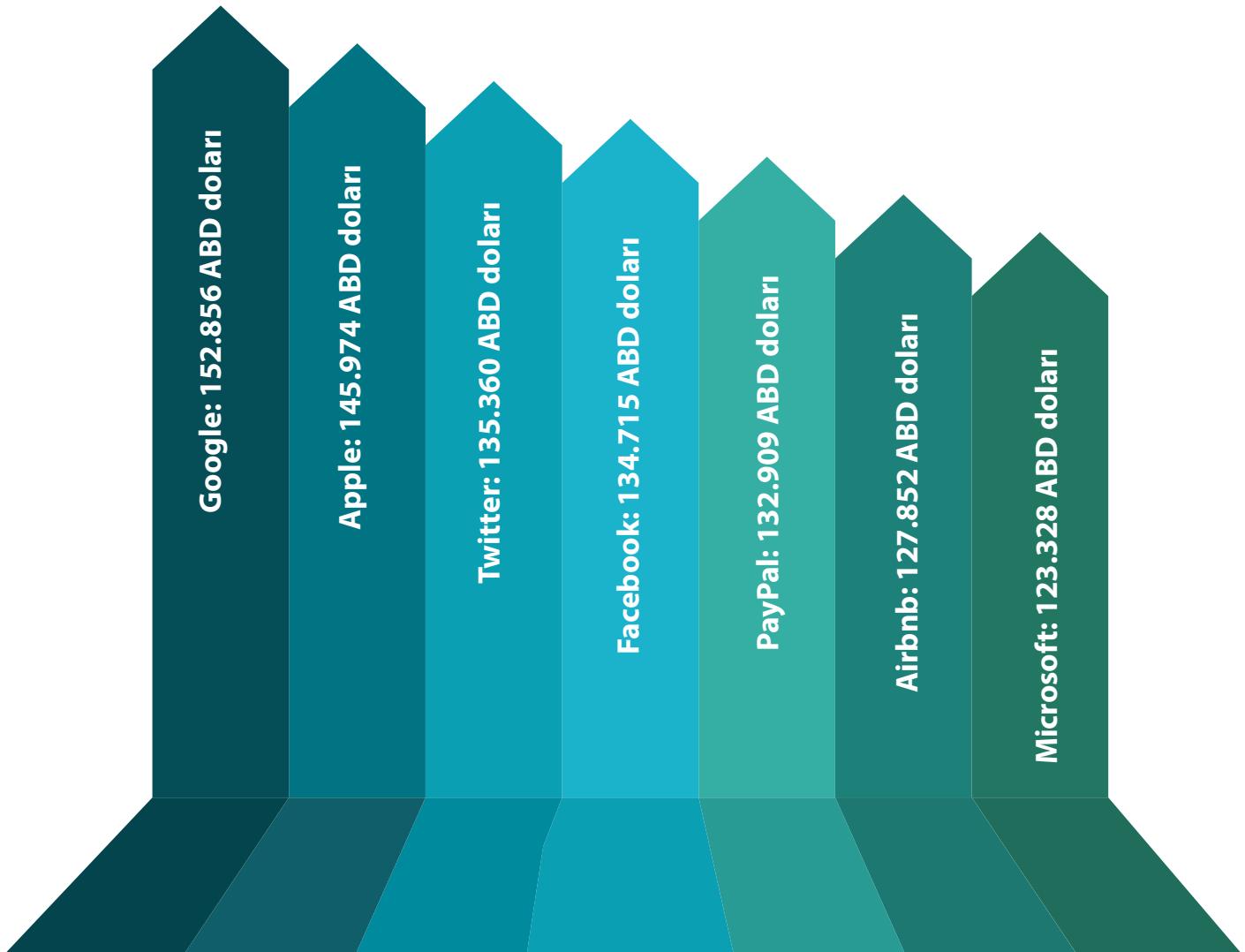
Veri bilimi alanında ortalama olarak en yüksek geliri sağlayan endüstriler şunlardır:



MAAŞ ORTALAMASI VE IQR (Amerikan doları)



Şekil 10.5 Sektörler ve ortalama gelir. (Kaynak: O'Reilly Salary Data Science Salary Report, 2016, <https://www.oreilly.com/library/view/2016-data-science/9781492049029/>)

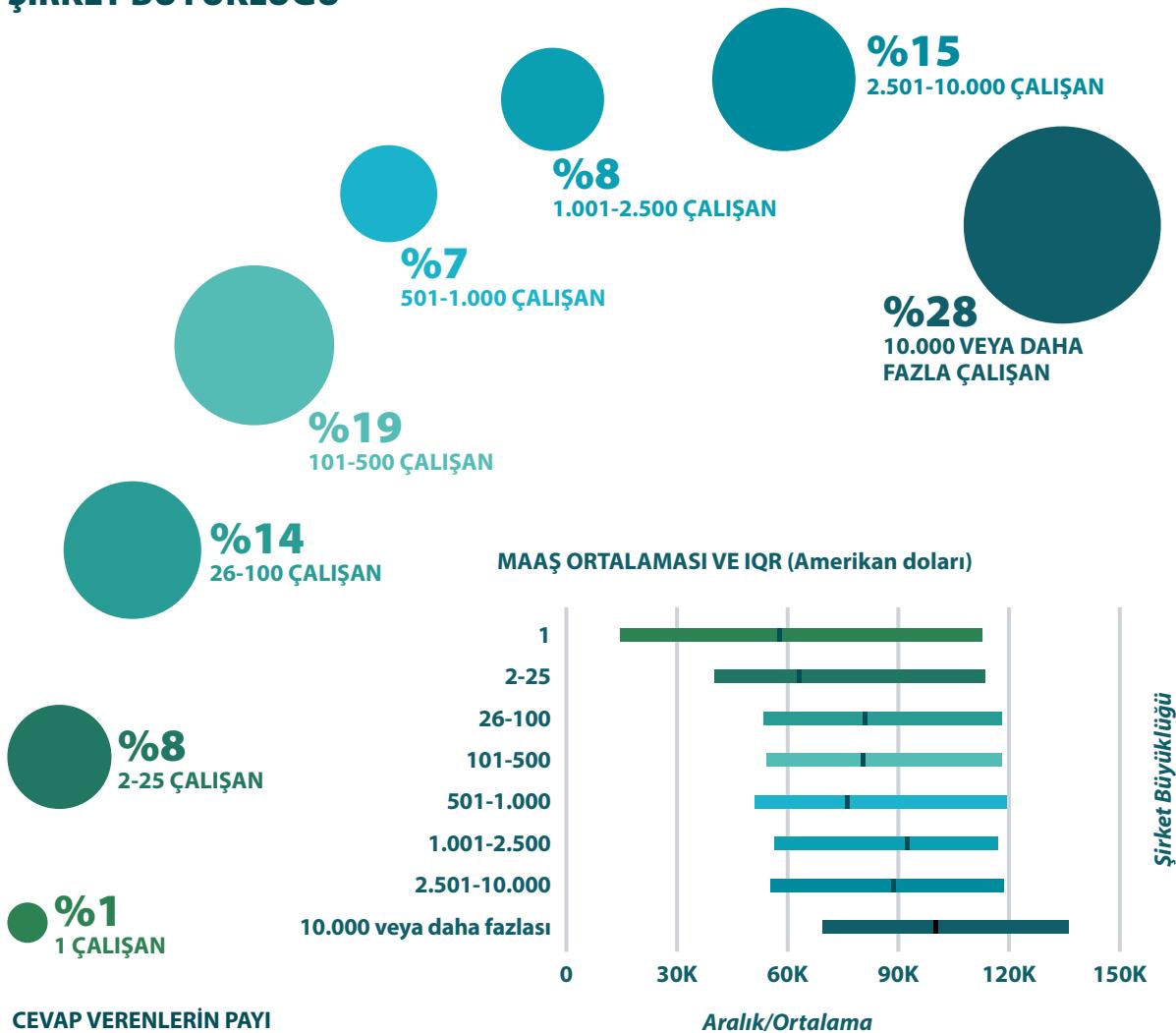


O'Reilly tarafından 2016 yılında yapılan ankete katılan veri bilimcilerin yalnızca yüzde 12'si bu üç sektörde çalışmaktadır. Yaklaşık yüzde 30'u ise danışmanlık veya yazılım alanında (SAAS, web ve mobil) çalışmaktadır.

Bunun yanı sıra veri bilimcileri arasında en yüksek geliri sağlayanlar ise önde gelen teknoloji şirketlerinde çalışmaktadır. Bu alanda yüksek profili olan yedi kuruluşun ortalama maaşı:

Bir diğer unsur olan şirket büyüklüğünde ise, kuruluş ne kadar büyükse maaşı da o kadar artmaktadır. Örneğin en az 10.000 çalışanı olan bir şirketteki bir veri bilimci aynı görevi yerine getirerek 1000'den az çalışanı olan bir şirkettekine göre daha yüksek bir gelir elde edebilir.

ŞİRKET BÜYÜKLÜĞÜ

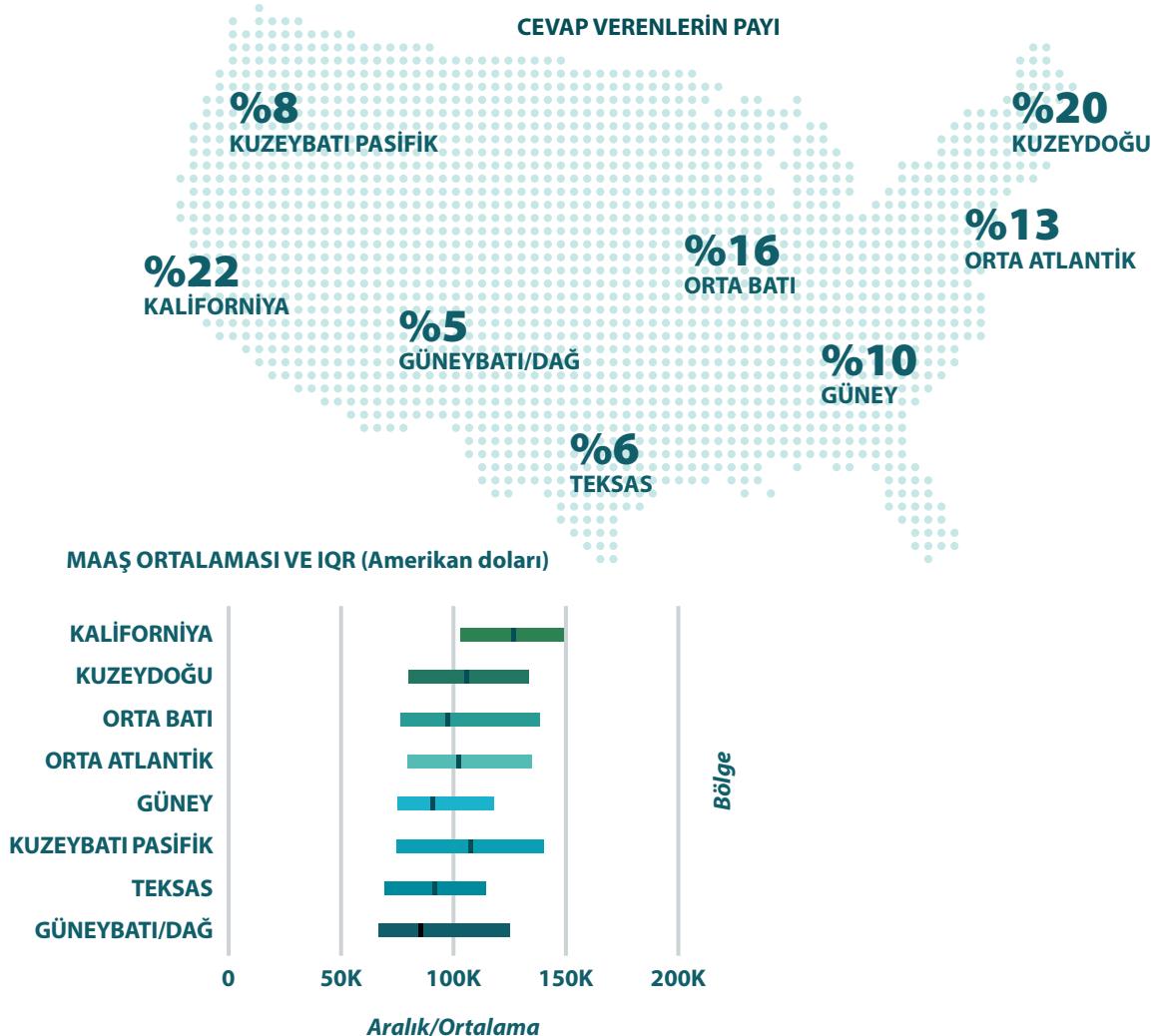


Şekil 10.6 Şirket büyüğününe göre ortalama gelir. (Kaynak: O'Reilly Salary Data Science Salary Report, 2016, O'Reilly Salary Data Science Salary Report, 2016)

D. Bölgelere Göre Veri Bilimcisinin Geliri

Veri bilimcilerin maaşları büyük ölçüde yaşadıkları ve çalışıkları bölgeye bağlıdır. Amerika'da en yüksek maaşı O'Reilly'nin araştırdığı veri bilimcilerin çoğunun çalıştığı yer olan Kaliforniya'daki kuruluşlar sağlamaktadır. Pacific Northwest'te ise ülkenin çoğu bögesinden daha az veri bilimci olmasına rağmen verilen maaşlarla bölgeler kıyaslandığında en yüksek maaşların sağlandığı bölgedir.

ABD BÖLGESİ



Şekil 10.7 ABD'deki bölgelere göre maaş ortalaması. (Kaynak: O'Reilly Salary Data Science Salary Report, 2016, O'Reilly Salary Data Science Salary Report, 2016)

E. Eğitime Göre Veri Bilimcinin Geliri

Veri bilimciler arasında özellikle doğru eğitim ve beceri kombinasyonuna sahip olan kişi sayısının mevcudiyeti oldukça nadir görülür. Veri bilimcilerin en temel düzeyde kod yazmayı bilmesi gerekmektedir. Aslında en yüksek veri bilimci maaşını haftada dört ila sekiz saat kod yazanlar almaktadır, en düşük maaş alanları ise hiç kod yazmayanlar almaktadır.

VERİ BİLİMÇİ VE IT'İN MAAŞINI ETKİLEYEN BECERİLER	
Apache Hadoop	▲ %8
Hadoop	▲ %7
Python	▲ %5
Machine Learning	▲ %4
Big Data Analytics	▲ %3
Data Mining/Data Warehouse	▲ %2
SAS	▲ %2
Data Analysis	▲ %1
Data Modeling	▲ %1
SQL	▲ %0
Ulusal Ortalama	92.000\$

Şekil 10.8 Veri biliminde ihtiyaç duyulan beceriler. (Kaynak: Payscale)

Veri bilimcilerin ayrıca büyük veriyi işleyen teknolojileri kullanmaları ve yeni ortaya çıkanları öğrenebilmelerinin yanı sıra açık kaynaklı araçlar, bulut bilgi işlem ve veri görselleştirmesi de gerekiyor. Ancak bu araçların nasıl kullanılacağını bilmek yeterli değildir; veri bilimcilerin kuruluşlarını iyileştirecek eyleme geçirilebilir iç-görüler elde etmek için bunları hangi şekilde ve nerede kullanacaklarını bilmeleri de çok önemlidir.

Ayrıca veri bilimi pozisyonları keskin iş zekası, bilimsel merak ve güçlü liderlik ve iletişim becerileri gerektirir. Bu özelliklere sahip olmak elde edilen geliri büyük ölçüde etkileyecektir.

10.3 BİR VERİ BİLİMÇİYE İHTİYAÇ DUYULAN ALANLAR VE İŞ TANIMLARI

Her sektörde her kuruluş ya da şirketin bir veri bilimcinin analiz etmesi için kendi büyük veri profili vardır. Genel eğilime göre her sektörde kullanılan daha yaygın büyük veri biçimleri vardır ve buna göre bir veri bilimcinin yapması gereken analiz türleri mevcuttur. Bu alanlar ve gereksinim duyulan analiz türleri şu şekilde sınıflandırılabilir:

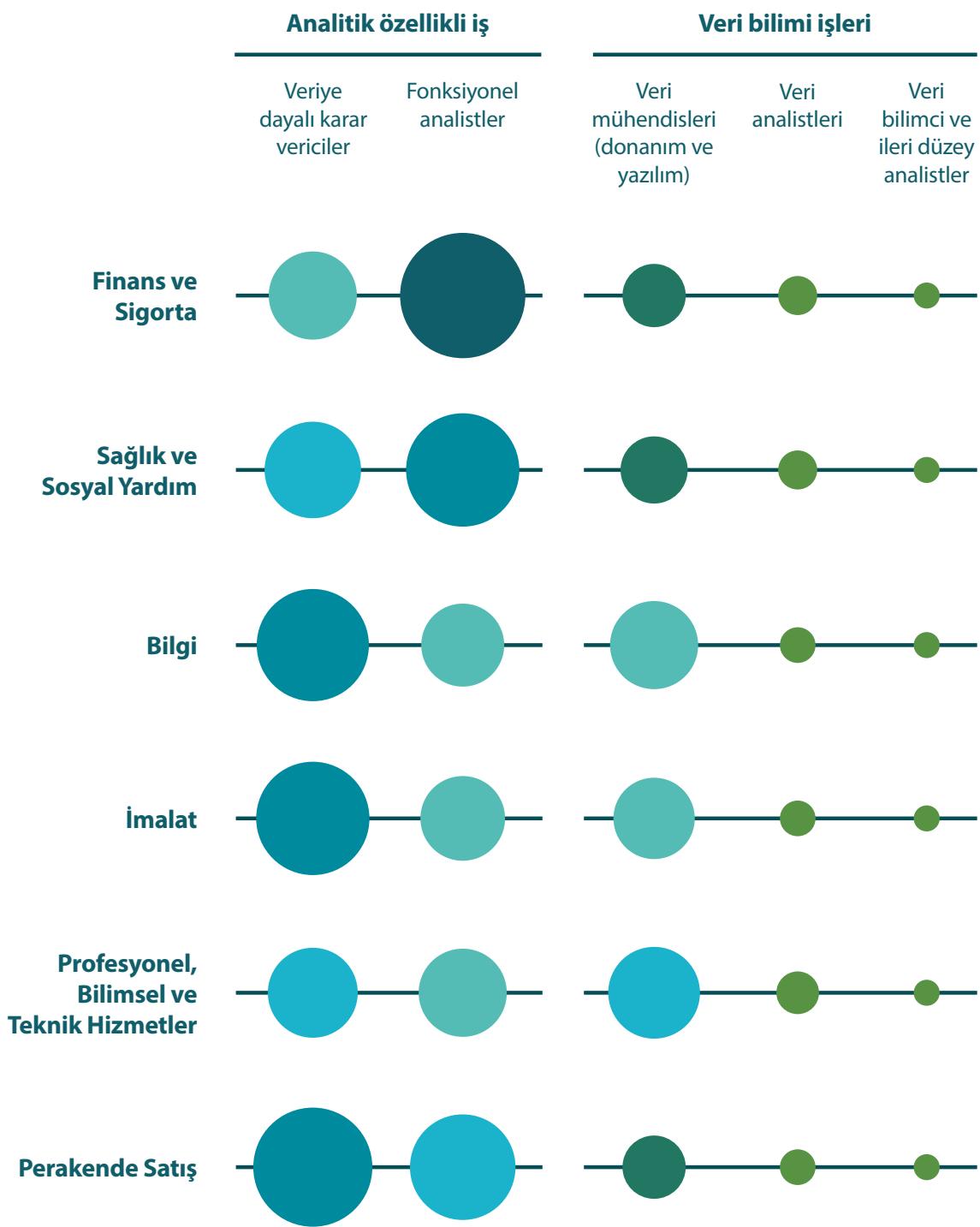
İş Dünyası: Günümüzde veriler neredeyse her şirketin iş stratejisini şekillendirmeye başlamıştır, ancak işletmelerin bilgileri anlamlandırmak için veri bilimcilere ihtiyacı bulunmaktadır. İş dünyasına ait verilerin veri analizi verimlilik, envanter,

Büyük veri, hükümetlerin karar vermesine, seçmenleri desteklemesine ve genel memnuniyeti izlemesine yardımcı olur. Finans sektöründe olduğu gibi, bu alanda da güvenlik ve uyumluluk veri bilimcileri için en önemli konulardandır.



STRATEJİK BÜYÜK VERİ YÖNETİMİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ

250



Not: Analiz yöneticilerinin iş kategorisi gösterilmemiştir. Toplamlar %100'e eşit olamaz.

Şekil 10.9 Sektörlerde veri beceri setlerine duyulan ihtiyaç. (Kaynak: <https://www.pwc.com/us/en/library/data-science-and-analytics.html>)

üretim hataları, müşteri sadakatı ve daha fazlasıyla ilgili kararların alınmasında bilgi edinilerek fayda sağlayabilir.

E-ticaret: Artık web siteleri satın alma verilerinden fazlasını toplamaktadır. Bu doğrultuda veri bilimcileri e-ticaret işletmelerinin müşteri hizmetlerini iyileştirmesine, trendleri bulmasına ve hizmet veya ürün geliştirmesine yardımcı olabilir.

Finans: Finans sektöründe hesaplar, kredi ve borç işlemleri ile ilgili veriler ve benzer finansal veriler bir işletme için hayatı önem taşır. Ancak bu alandaki veri bilimcileri için dolandırıcılık tespiti de dahil olmak üzere güvenlik ve uyumluluk da üzerinde durulması gereken önemli konulardır.

Devlet: Büyük veri, hükümetlerin karar vermesine, seçmenleri desteklemesine ve genel memnuniyeti izlemesine yardımcı olur. Finans sektöründe olduğu gibi, bu alanda da güvenlik ve uyumluluk veri bilimcileri için en önemli konulardandır.

Bilim: Bilim insanları tarih boyunca her zaman verileri bir şekilde işlemiştir, ancak şimdi teknoloji sayesinde deneylerden elde edilen verileri daha iyi toplayabiliyor, paylaşabiliyor ve analiz edebiliyorlar. Veri bilimcileri bu süreçte onlara yardımcı olabilir.

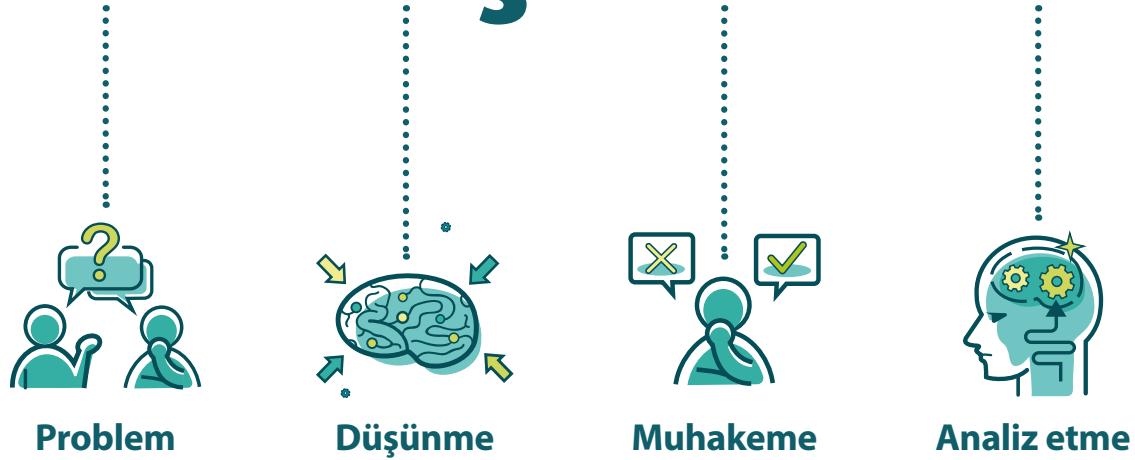
Sosyal Ağ Oluşturma: Sosyal ağ verileri, hedeflenen reklamcılığa bilgi vermeye, müşteri memnuniyetini artırmaya, konum verilerinde trendler oluşturmaya ve özeliklerle hizmetleri geliştirmeye yardımcı olur. Gönderilerin, tweetlerin, blogların ve diğer sosyal medyanın sürekli veri analizi, işletmelerin hizmetlerini süreklilik arz edecek şekilde iyileştirmesine yardımcı olabilir.

Sağlık: Büyük veriyi gerektiren ve güvenlikle uyumluluğa dayanan elektronik tıbbi kayıtlar artık sağlık tesisleri için standart ve elzemdir. Burada veri bilimciler, sağlık hizmetlerini iyileştirmeye ve diğer taraftan fark edilemeyecek eğilimleri ortaya çıkarmaya yardımcı olabilir.

Telekomünikasyon: Tüm elektronik cihazlar veri toplar ve tüm bu verilerin depolanması, yönetilmesi, bakımı ve analiz edilmesi gereklidir. Veri bilimciler şirketlerin hataları gidermelerine, ürünleri iyileştirmelerine ve istedikleri özellikler sunarak müşterilerini daha mutlu etmelerine yardımcı olur.

Diger: Büyük veri aktarımına karşı bağıskılık kazanan bir endüstri yoktur ve politika, kamu hizmetleri, akıllı cihazlar ve daha fazla benzer diğer alanda da veri bilimciye gereksinim duyulmaktadır.

ELEŞTİREL



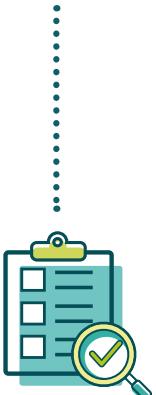
10.4 VERİ BİLİMCİLERİN TEMEL BECERİLERİ VE ÖZELLİKLERİ

Veri bilimciler, şirketlerin gelişen bilgi kaynaklarından en yüksek değeri elde etmek için neredeyse her sektörden yüksek talep görmeye devam etmektedir. Fakat veri bilimcilerin diğer alanlardaki meslek gruplarına göre bazı önemli özellikler ve becerilerinin olması gerekmektedir.

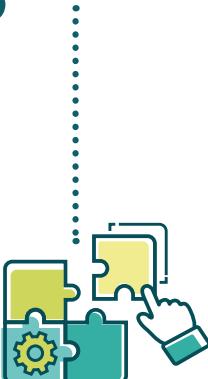
Eleştirel Düşünme: Veri bilimcilerin görüşlerini formüle etmeden veya yargıda bulunmadan önce belirli bir konu veya problemle ilgili gerçeklerin objektif analizini uygulayabilmek için eleştirel düşünme becerilerinin gelişmiş olması gerekmektedir. Bu beceri her şeyden fazla bir veri bilimcinin başarısını etkilemektedir. Gerektiğinde bir veri bilimci kendi inancını arka sıralara almalıdır. Sorun ya da konudan uzaklaşarak ve belki de bir adım geri durarak büyük resme bakıp sorun ya da konuya birçok açıdan yaklaşabilme yeteneği bulunmalıdır.

Makine Öğrenimi ve Yapay Zeka: Endüstrilerin artan bilgi işlem gücü bağlanabilirlik ve toplanan devasa veri hacimleri nedeniyle bu alanlarda son derece hızlı ilerleme kaydetmektedir. Bir veri bilimcinin bir araştırmada hangi teknolojinin ne zaman uygulanacağını anlaması gerekmektedir. İstatistiksel beceriler önemli olmasına

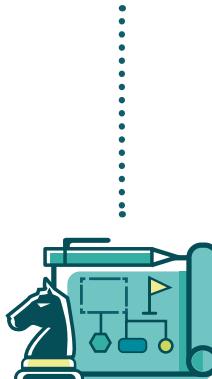
DÜŞÜNMЕ



Değerlendirme



Problem Çözme



Karar Verme



Çözüm

rağmen çoğu işveren bu becerileri dikkate almaz, çünkü bugünün otomatik araçları ve açık kaynak yazılımı çok kolay erişilebilir durumdadır. Diğer taraftan makine öğrenimi algoritmaları işlevsel arayüzleri anlamak kendi başına yeterli değildir. İş için uygun algoritmayı seçmek adına başarılı bir veri bilimcinin herhangi bir modelin genel performansını en üst düzeye çıkarmak için yöntemler içindeki istatistikleri ve uygun veri hazırlama tekniklerini anaması gerekmektedir. Bunun yanı sıra veri bilimi esas olarak klavye başında gerçekleştiğinden, yazılım mühendisliğindeki güçlü temeller de veri bilimciye yardımcı olur.

Kodlama: Birinci sınıf veri bilimciler kod yazmayı bilirler ve çeşitli programlama görevlerini rahatlıkla yerine getirebilmedirler. Örneğin veri biliminde tercih edilen dillere hâkim olmalıdırlar. Bir veri bilimci olarak gerçekten başarılı olmak için programlama becerilerinin her iki hesaplama yönüne de vakıf olması gerekmektedir; büyük hacimli verilerle ilgilenme, gerçek zamanlı verilerle çalışma, bulut bilgi işlem, yapılandırılmamış veriler ve istatistiksel yönler ve regresyon, optimizasyon, kümeleme, karar ağaçları vb. gibi istatistiksel modeller. 1990'ların sonlarında başlayan büyük verinin etkisiyle giderek daha fazla veri bilimcinin Python, C ++ veya Java gibi dilleri anlaması ve kodlayabilmesi gerekmektedir.

Matematik: Veri bilimi, matematiği sevmeyen veya yetkin olmayan insanlar için muhtemelen iyi bir kariyer seçimi değildir. Bir veri bilimci yapılan çalışmalarla karmaşık finansal veya operasyonel modeller geliştirmek isteyen müşterilerle ilişki kurulduğunda, bu modellerin istatistiksel olarak anlamlı olabilmesi için büyük miktarlarda veri gerektiğini bilir ve veri bilimcinin rolü, temel iş stratejilerini geliştirmek veya değiştirmek için kullanılabilecek istatistiksel modeller geliştirmek için matematikteki derin uzmanlıklarından yararlanmaktadır. Aynı zamanda karmaşık denklemlerin “kara kutusunda” gerçekten neler olup bittiğini yeniden güvence sağlayacak şekilde iletmek için iş kolundaki yöneticileriyle yakın işbirliği yapma yeteneğine sahip olmalıdır.

İletişim: İletişim becerilerinin önemi tekrarlanmayı gerektirir. Bugün teknolojide neredeyse hiçbir şey bir boşlukta gerçekleştirilmiyor; sistemler, uygulamalar ve veriler örnek olarak verilebilir. Ayrıca insanlar arasında her zaman bir miktar entegrasyon vardır. Veri bilimi de farklı değildir ve verileri kullanarak birden çok paydaşa iletişim kurabilmek temel bir özelliktir. Veriler aracılığıyla hikâye anlatma yeteneği, matematiksel bir sonucu eyleme geçirilebilir bir kavrayışa veya müdahaleye dönüştürebilir. İşletme, teknoloji ve verilerin kesişme noktasında olan veri bilimcilerin, paydaşların her birine bir hikâye anlatmakta ustalaşması gerekmektedir. Bu, verilerin işletme yararları için işletme yöneticilerine iletmesini içerir. Verilerin teknoloji ve hesaplama kaynakları hakkında; veri kalitesi, mahremiyet ve mahremiyetle ilgili zorluklar hakkında ve kuruluşun diğer ilgi alanları hakkında işletme yöneticilerine iletmesini içerir. İyi bir iletişimci olmak zorlu teknik bilgileri eksiksiz, doğru ve sunması kolay bir biçimde dönüştürme yeteneğini de içerir. Bir algoritmanın bir tahmine nasıl ulaştığını iletmek, liderlerin iş süreçlerinin bir parçası olan tahmine dayalı modellere güvenini sağlamak için kritik bir beceridir.

Veri Mimarisi: Veri bilimcinin başlangıçtan modele ve iş kararına kadar verilere ne olduğunu anlaması elzemdir. Mimariyi anlamamak, örneklem büyülüüğündeki çıkarımlar ve varsayımlar üzerinde ciddi bir etkiye sahip olabilir ve bu da genellikle yanlış sonuçlara ve kararlara yol açabilir. Daha da kötüsü, mimari içinde işler değişimdir.

Risk Analizi, Süreç İyileştirme, Sistem Mühendisliği: Keskin bir veri bilimcinin iş riskini analiz etme, süreçlerde iyileştirme yapma ve sistem mühendisliğinin nasıl



Keskin bir veri bilimcinin iş riskini analiz etme, süreçlerde iyileştirme yapma ve sistem mühendisliğinin nasıl çalıştığını anlaması gerekmektedir.

çalıştığını anlaması gerekmektedir. Model geliştirmenin başlangıcında risk analizleri oluşturmak riskleri azaltabilir. Harcamaları süreci iyileştirmeye dahil etmek, şirkete özgü riskleri ve verileri veya bir modelin sonucunu etkileyebilecek diğer sistemleri anlamak veri bilimcinin çabalarında daha fazla müşteri memnuniyetine yol açabilir.

Problem Çözme ve Sezgi: Başarılı veri bilimcilerin sergilediği özellikler herhangi bir iyi problem çözücü tarafından sergilenen özelliklerin aynısıdır. Dünyaya birçok açıdan bakarlar, alet kutularından tüm aletleri çıkarmadan önce ne yapmaları gerektiğini anlamaya çalışırlar, titiz ve eksiksiz bir şekilde çalışırlar ve sonuçlarını sorunsuz bir şekilde açıklayabilirler. Diğer taraftan entelektüel altyapılarının yanı sıra verilerin yapısı ve nüanslarıyla modellerin nasıl çalıştığı sezgisel olarak daha iyi anlaşılabilir. Dolayısıyla yaptığı veri analizi hakkında daha derin bir anlayışa ve sezgiye sahip olan veri uzmanına gerçek bir veri bilimi uzmanı diyebiliriz.

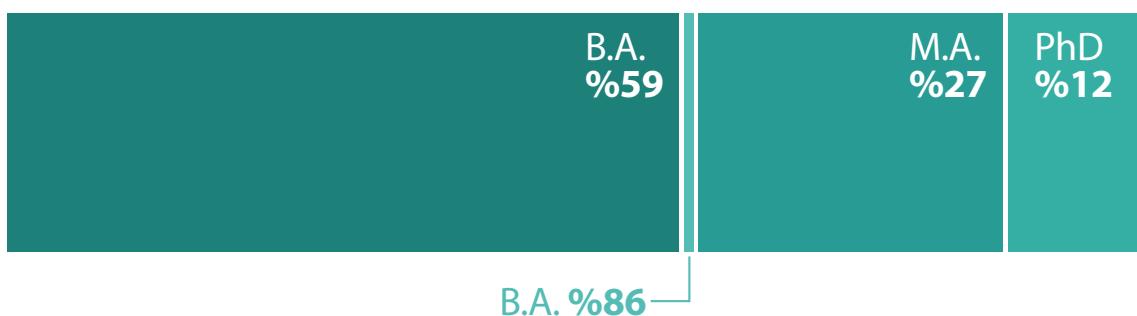
10.5 VERİ BİLİMÇİ OLMAK İÇİN GEREKENLER VE EĞİTİM

Bilgisayar bilimi veya veri analizi konusunda bir geçmişiniz yoksa, ön hazırlık kampları, üniversite programları veya sertifikalar veri bilimci olmak yolunda başlangıçtaki gerekli becerileri sağlayabilir.

ABD'deki işverenlerin analitik özellikli işler için istediği eğitim düzeyi



Veri bilimciler ve ileri düzey analistler için eğitim düzeyi



Şekil 10.10 ABD'li işverenlerin veri bilimci istihdamında eğitime bakışı. (Kaynak: <https://www.pwc.com/us/en/library/data-science-and-analytics.html>)

Veri bilimci olmanın pek çok yolu vardır, ancak en geleneksel yol bir lisans derecesine sahip olmaktır. Çoğu veri bilimci yüksek lisans derecesine sahiptir, ancak bu her veri bilimci için geçerli değildir ve veri bilimi becerilerini geliştirmenin başka yolları da vardır. Veri bilimi biraz iş alanı uzmanlığı gerektirdiğinden, bir veri bilimcinin rolü sektörle bağlı olarak değişecektir ve oldukça teknik bir sektörde çalışırsanız daha fazla eğitime ihtiyacınız olabilir.



Şekil 10.11 Dijital öğrenme yaklaşımı kavramlarının altı anahtarı. (Kaynak: <http://sb.stratbz.to/upskillscale>)

Örneğin sağlık, devlet veya bilim alanında çalışmak istiyorsanız pazarlama, işletme veya eğitim alanlarında çalışmanızdan farklı bir beceri setine ihtiyacınız olacaktır. Belirli sektör ihtiyaçlarını karşılamak için belirli beceri setleri geliştirmek istiyorsanız, becerilerinizi geliştirmenize yardımcı olabilecek çevrimiçi sınıflar, eğitim kampları ve profesyonel gelişim kurslarına da katılabilirsiniz.

Türkiye'de ise veri bilimi her ne kadar yeni yeni gelişmeye başlasa da günümüzde hızlı bir ivme kazanarak şirketlerin yatırım yapmaya başladığı alanlardan biri haline gelmiştir. Bu bağlamda da veri bilimi konusunda eğitim alanında da üniversitelerin yüksek lisans programları ve laboratuvarların sayısında artış görülmeye başlanmıştır. Buna göre aşağıdaki tabloda yüksek lisans programları ve laboratuvarlar yer almaktadır.

No	Üniversite Adı	Program Adı	Bilgi
1	Bahçeşehir Üniversitesi	Büyük Veri Analitiği ve Yönetimi	http://www.bahcesehir.edu.tr/icerik/7502-veri-analitigi-ve-yonetimi
2	İstanbul Teknik Üniversitesi	Büyük Veri ve İş Analitiği (Big Data and Business Analytics)	http://bigdatamaster.itu.edu.tr/
3	Koç Üniversitesi	Veri Bilimleri(Data Science)	https://gsse.ku.edu.tr/en/graduate-programs/data-science/program-overview/
4	MEF Üniversitesi	Büyük Veri Analitiği (Big Data Analytics)	http://bda.mef.edu.tr
5	Sabancı Üniversitesi	Veri Analitiği (Data Analytics)	http://da.sabanciuniv.edu
6	Şehir Üniversitesi	Veri Mühendisliği	http://datascience.sehir.edu.tr
7	Yeditepe Üniversitesi	Veri Bilimi	http://fbe.yeditepe.edu.tr/tr/
8	TED Üniversitesi	Uygulamalı Veri Bilimi	https://uvb.tedu.edu.tr
9	Akdeniz Üniversitesi	Veri Analitiği ve Yönetimi	http://sosyalbilim.akdeniz.edu.tr/
10	Abdullah Gül Üniversitesi	Veri Bilimi (Tezli)	http://sbe.agu.edu.tr/
11	İstanbul Medeniyet Üniversitesi	Biyolojik Veri Bilimi	https://enstitu.medeniyet.edu.tr/tr/anabilim-dallari-ve-programlar/tezli-yuksek-lisans/biyolojik-veri-bilimi
12	Dokuz Eylül Üniversitesi	Veri Bilimi	http://debis.deu.edu.tr/ders-katalog/2019-2020/tr/bolum_10075_tr.html
13	Özyegin Üniversitesi	Veri Bilimi	https://www.ozyegin.edu.tr/en/graduate-school-engineering-and-science/academic-programs/msc-data-science
14	Galatasaray Üniversitesi	Veri Bilimi	http://fbe.gsu.edu.tr/tr/vb-ylisans
15	Bakırçay Üniversitesi	İş Zekası ve Veri Analitiği Yüksek Lisans Programı	https://ybs.bakircay.edu.tr/Sayfalar/2078/is-zekasi-ve-veri-analitigi-yuksek-lisans-programi

Şekil 10.12 Türkiye'de veri bilimi alanında yüksek lisans eğitimi veren üniversiteler. (Kaynak: <http://veribilimci.org/veri-bilimi-yuksek-lisans-programlari/>)

No	Üniversite Adı	Program Adı	Bilgi
1	Gazi Üniversitesi	Big Data and Information Security Center	http://bigdatacenter.gazi.edu.tr
2	Süleyman Demirel Üniversitesi	Veri Bilimi Laboratuvarı (Data Science Lab)	https://www.facebook.com/sdudatascience/
3	Fırat Üniversitesi	Büyük Veri ve Yapay Zeka Laboratuvarı	http://buyukveri.firat.edu.tr/
4	BTK - BAB	Büyük Veri ve Yapay Zeka Laboratuvarı	https://bab.btk.gov.tr/BTKBAB/
5	İ.T.Ü	Yapay Zeka ve Veri Bilimi Uygulama ve Araştırma Merkezi	...

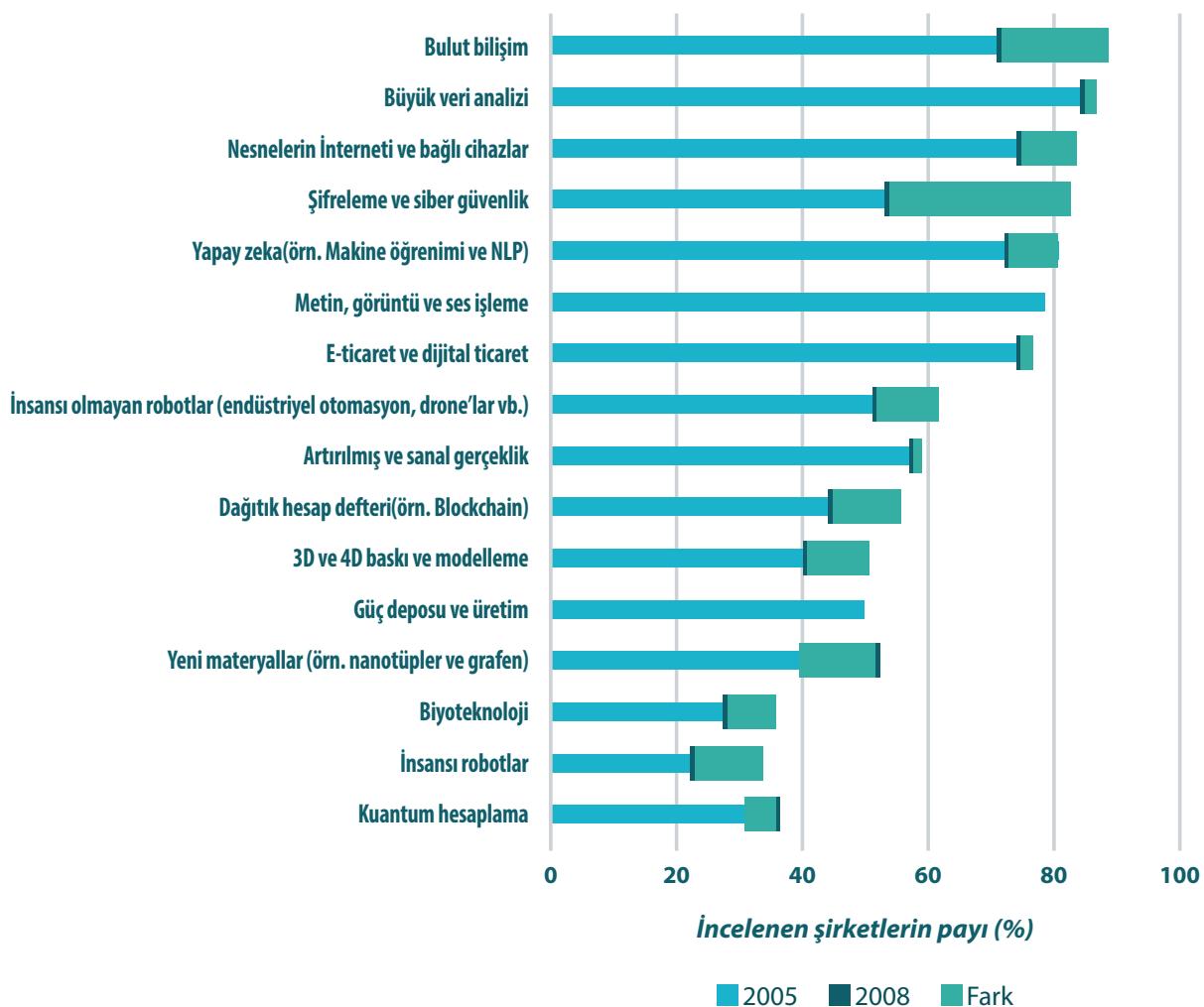
Şekil 10.13 Üniversite kapsamında yer alan merkezler ve laboratuvarlar. (Kaynak: <http://veribilimci.org/veri-bilimi-yuksek-lisans-programlari/>)

10.6 2020 DÜNYA EKONOMİK FORUMU'NUN MESLEKLERİN GELECEĞİ RAPORU VE TEKNOLOJİNİN MESLEKLERE ETKİSİ

Geçtiğimiz beş yıl içinde teknolojik otomasyonun artması nedeniyle meslek gereksinimleri ve beceri talepleri büyük bir değişim içine girmiştir. Bu rapor çalışma yaşamının geleceği için planlama ve strateji oluşturma fırsatının yanı sıra, gelişen işgücü piyasası eğilimine de kapsamlı bir genel bakış sağlamaktadır.

Raporda yeni teknolojilerin benimsenmesinde belirgin bir hızlanma görüldüğü belirtilmiştir. Şekil 10.14, şirketlerin 2025'e kadar bu teknolojileri benimseme olasılığına göre düzenlenmiş bir dizi teknolojiyi göstermektedir. Bulut bilgi işlem, büyük veri ve e-ticaret önceki yıllardaki eğilimini koruyarak öncelikler arasında yer almaya devam etmektedir. Bununla birlikte dijital çağın yeni güvenlik açıklarını yansıtan şifrelemeye olan eğilimde de önemli bir artış ortaya çıkmış ve insansız olmayan robotlarla yapay zekayı benimsemek için hevesli olan firmaların sayısı da önemli ölçüde artmıştır.

Bu teknolojik gelişmelerin benimsenmesi sektörde göre değişse de en geniş ölçekte adaptasyonun yapay zeka, dijital bilgi ve iletişim, finansal hizmetler, sağlık ve ulaşım

2025 yılına kadar benimsenmesi muhtemel teknolojiler (İncelenen şirketlerin payına göre)**Şekil 10.14** 2025 yılına kadar benimsenmesi muhtemel teknolojiler (İncelenen şirketlerin payına göre).

sektörleri alanında olacağı öngörmektedir. Büyük veri, nesnelerin interneti ve insansı olmayan robotiklerin madencilik ve metallerde güçlü bir şekilde benimsenirken hükümet ve kamu sektörü endüstrisinin şifrelemeye belirgin bir şekilde odaklanacağı belirtiliyor (Şekil 10.15).

Bu yeni teknolojiler, sektörlerde gelecekteki büyümeyi yönlendirmenin yanı sıra, yeni iş rolleri ve becerilere olan talebi artırmıştır. Bu bağlamda giderek daha stratejik ve giderek artan iş rollerine bakıldığından, veri analistleri ve veri bilimciler, yapay zeka ve makine öğrenimi uzmanları, robotik mühendisleri, yazılım ve uygulama geliştiricilerinin yanı sıra dijital dönüşüm uzmanları önemli yer tutmaktadır. Ayrıca sü-

Teknoloji/Sektör	TARIM (%)	OTO (%)	TÜK (%)	DİBT (%)	EĞİT (%)	ETT (%)	FH (%)	HUK (%)	SA (%)	İMA (%)	MM (%)	PEG (%)	PH (%)	ULAŞ (%)
3D ve 4D baskı ve modelleme	54	67	39	39	69	69	27	45	65	69	48	79	40	60
Yapay zeka(örn. Makine öğrenimi, nöral ağlar, NLP)	62	76	73	95	76	81	90	65	89	71	76	71	76	88
Artırılmış ve sanal gerçeklik	17	53	58	73	70	75	62	56	67	54	57	71	57	62
Büyük veri analizi	86	88	91	95	95	76	91	85	89	81	90	86	86	94
Biyoteknoloji	50	18	48	40	46	47	46	38	65	31	16	36	28	23
Bulut bilişim	75	80	82	95	95	88	98	95	84	92	87	86	88	94
Dağıtık hesap defteri(örn. Blockchain)	31	40	41	72	61	50	73	40	72	41	50	46	53	38
E-ticaret ve dijital ticaret	80	75	85	82	72	71	90	67	78	82	62	62	70	87
Şifreleme ve siber güvenlik	47	88	85	95	86	88	95	95	84	72	83	71	78	75
Nesnelerin İnterneti ve bağlı cihazlar	88	82	94	92	62	94	88	79	95	84	90	93	74	76
Yeni materyallar (örn. nanotüpeler ve grafen)	15	46	22	36	67	65	36	33	47	51	37	36	27	27
Güç deposu ve üretim	75	64	59	38	27	88	55	33	31	62	57	69	45	46
Kuantum hesaplama	18	21	17	51	25	41	44	36	38	21	29	25	19	39
İnsansı robotlar	42	50	38	44	47	24	47	31	47	41	15	17	25	21
İnsansı olmayan robotlar (endüstriyel otomasyon, drone'lar vb.)	54	60	52	61	59	65	53	50	56	79	90	79	35	69
Metin, görüntü ve ses işleme	50	59	82	90	89	88	88	89	88	64	76	87	79	65

NOT: TARIM=Tarım, Yiyecek ve İçecek; OTO= Otomotiv, TÜK=Tüketiciler; DİBT= Dijital İletişim ve Bilgi Teknolojileri; EĞİT=EĞİTİM; ETT=Enerji Tesisleri ve Teknolojileri; FH=Finansal Hizmetler; HUK=Hükümet ve Kamu Sektörü; SA=Sağlık ve Sağlık Hizmeti; İMA= İmalat; MM=Madencilik ve Metaller; PEG=Petrol ve Gaz; PH=Profesyonel Hizmetler; ULAŞ=Ulaşım ve Depolama

Şekil 10.15 İncelenen şirketlerin payına ve seçilen sektörlerde göre 2025 yılına kadar benimsenmesi muhtemel teknolojiler ve endüstriler.

Geçtiğimiz beş yıl içinde teknolojik otomasyonun artırması nedeniyle meslek gereksinimleri ve beceri talepleri büyük bir değişim içine girmiştir.

Sektörler Arasında Talebi Artan ve Azalan İlk Yirmi İş Kolu

↗ Talep olarak artan

- 1 Veri Analistleri ve Veri Bilimciler
- 2 Yapay Zeka ve Makine Öğrenimi Uzmanı
- 3 Büyük Veri Uzmanı
- 4 Dijital Pazarlama ve Strateji Uzmanı
- 5 Proses Otomasyon Uzmanı
- 6 İş Geliştirme Uzmanı
- 7 Dijital Dönüşüm Uzmanı
- 8 Bilgi Güvenliği Analisti
- 9 Yazılım ve Uygulama Geliştiricileri
- 10 Nesnelerin İnterneti (IoT) Uzmanı
- 11 Proje Yöneticisi
- 12 İşletme Hizmetleri ve Yönetim Yöneticileri
- 13 Veritabanı ve Ağ Uzmanları
- 14 Robotik Mühendisi
- 15 Stratejik Danışman
- 16 Yönetim ve Organizasyon Analistleri
- 17 FinTech Mühendisi
- 18 Mekanik ve Makine Tamircileri
- 19 Organizasyonel Gelişim Uzmanları
- 20 Risk Yönetimi Uzmanı

↘ Talep olarak düşen

- 1 Veri Giriş Görevlisi
- 2 İdari Sekreter
- 3 Muhasebe, Defter Tutma ve Bordro Memuru
- 4 Muhasebeci ve Denetçiler
- 5 Montaj ve Fabrika Çalışanı
- 6 İşletme Hizmetleri ve İdare Yöneticileri
- 7 Müşteri Bilgileri ve Müşteri Hizmetleri Çalışanları
- 8 Genel ve Operasyon Müdürü
- 9 Mekanik ve Makine Tamircileri
- 10 Malzeme Kaydı ve Stok Tutma Görevlileri
- 11 Finansal Analist
- 12 Posta Hizmeti Memuru
- 13 Satış Temsilcisi ve Toptancı
- 14 İlişkiler Müdürü
- 15 Banka Vezneleri ve İlgili Memurlar
- 16 Pazarlamacı ve Sokak Satıcısı
- 17 Elektronik ve Telekom Montajcısı ve Tamircisi
- 18 İnsan Kaynakları Uzmanı
- 19 Eğitim ve Geliştirme Uzmanı
- 20 İnşaat İşçileri

Şekil 10.16 Sektörler arasında talebi artan ve azalan ilk yirmi iş kolu.

reç otomasyon uzmanları, bilgi güvenliği analistleri ve nesnelerin interneti uzmanları gibi meslekler de işverenlerden artan talep gören meslekler olarak ortaya çıkmaktadır (Şekil 10.16).

Dünya Ekonomik Forumu'nun LinkedIn ve Coursera'daki ortak şirketlerdeki veri bilimcilerle ortaklaşa yazılan Geleceğin Meslekleri Raporu, gerçek zamanlı işgücü piyasası verisini kullanarak ekonomi genelinde bir dizi yeni işin ortaya çıkışını ölçmek ve izlemek için ilk kez bir yol sunmuştur. Bu işbirliğinden elde edilen verilere göre 20 ekonomide talepte sürekli olarak artan 99 iş tanımlanmış olup alt gruplara ayrılarak sınıflandırılmıştır. Şekil 10.17, her bir fırsatın ölçüne göre düzenlenmiş, her bir profesyonel kümeye karşılık gelen meslekler dizisini göstermektedir.

Bakım Ekonomisi



Sıra

Niş

Kitle

Yeşil Ekonomi



Şekil 10.17 Yarının işlerinde kümelenen yeni roller.

2025 için En İyi 15 Beceri

- | | |
|----|--|
| 1 | Analitik Düşünme ve İnovasyon |
| 2 | Aktif öğrenme ve öğrenme stratejileri |
| 3 | Karmaşık problem çözme |
| 4 | Eleştirel düşünme ve analiz |
| 5 | Yaratıcılık, özgünlük ve girişimcilik |
| 6 | Liderlik ve sosyal etki |
| 7 | Teknoloji kullanımı, izleme ve kontrol |
| 8 | Teknoloji tasarıımı ve programlama |
| 9 | Direnç, stres toleransı ve esneklik |
| 10 | Muhakeme, problem çözme ve kavrayış yeteneği |
| 11 | Duygusal zekâ |
| 12 | Sorun giderme ve kullanıcı deneyimi |
| 13 | Hizmete yönelme |
| 14 | Sistemlerin analizi ve değerlendirilmesi |
| 15 | İkna ve müzakere |

Şekil 10.18 2025 için en iyi 15 beceri.

Bu bağlamda gelişen meslekler açısından geleceğin insanının da yeni beceri ve donanımlara sahip olması gerekmektedir. Dünya Ekonomik Forumu'nun raporuna göre gerekli beceriler aşağıdaki şekilde gösterilmiştir. Şekilde de görüldüğü üzere bir veri bilimcinin en önemli özellikleri arasında yer alan analitik düşünme becerisi ve eleştirel düşünme becerisi ilk sıralarda yer almaktadır.

10.7 SONUÇ

Özellikle Covid-19 pandemisinden sonra artık hayatımızın ayrılmaz bir parçası olan teknoloji ve artan veri üretimi özünde hem gelecek için çok ciddi anlamda iş gücü oluştururken hem de bazı zorlukları da beraberinde getirmekte ve bu zorluklar karşısında da yine insan gücüne ihtiyaç duyulmaktadır.

Günümüzde veri bilimcileri, veri mühendisleri ve iş analistleri teknoloji alanında büyük ilerleme gösteren Amerika gibi ülkelerde en çok aranan pozisyonlar arasındadır. Buna rağmen birçok mevcut ve gelişmekte olan iş kolu işverenlerin ihtiyaç duyduğu tüm becerilere sahip değil. Yine de veri bilimcileri, uzun yıllardır büyük şirketlerin ele geçirmek için savaştığı nadir meslek gruplarından biridir.

Gelecekte ise veri bilimcilerin işletmelerin devrim niteliğinde atılımlar yapmasına yardımcı olacak iş açısından en kritik ve karmaşık zorlukların üstesinden gelmesi beklenmektedir. Günlük olarak toplanan büyük miktarda verinin anlamlanmasına ve bunların işle ilgili sorunları çözmek, eğilimleri tespit etmek ve yeni fikirleri desteklemek için kararlar almak için kullanılmaya çalışılırken istatistik, veritabanı, veri görselleştirme ve makine öğrenimi, kodlama, veri hazırlama, meraklı ve öğrenmeye hazır veri bilimcilere ihtiyaç artacaktır.

Bu becerilerin kombinasyonuna sahip olmak tam da veri bilimcilerin nadir olmasının nedenidir ve onlara olan talep ortadan kalkacaksız gibi de görünmeyecek tam aksine gelecekte daha büyük yer tutacağı benzemektedir.

KAYNAKLAR

- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4, 1043-1171.
- Ajwad, M. I., Hut, S., Abdulloev, I., Audy, R., de Laat, J., Kataoka, S., ... Torracchi, F. (2014). The skills road: skills for employability in Tajikistan. Washington, D.C., United States: The World Bank.
- Arntz, M., Gregory, T., & Zierahn, U. (2016). The Risk of Automation for Jobs in OECD Countries. <https://doi.org/10.1787/5jlz9h56dvq7-en>
- Autor, D. (2014). Polanyi's Paradox and the Shape of Employment Growth (Working Paper No. 20485). National Bureau of Economic Research. <https://doi.org/10.3386/w20485>
- Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3), 3-30. <https://doi.org/10.1257/jep.29.3.3>

- Autor, D. H., & Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5), 1553-1597. <https://doi.org/10.1257/aer.103.5.1553>
- Dünya Ekonomik Forumu Mesleklerin Geleceği Raporu, 2020. <https://www.gsma.com/iot/the-gsma-iot-infographic/>
- <https://www.pwc.com/us/en/library/data-science-and-analytics.html>
- <https://datasciencedegree.wisconsin.edu/data-science/data-scientist-salary/screenshot-2020-09-25-at-2-43-57-pm/>
- <https://www.pwc.com/us/en/library/data-science-and-analytics.html>
- <https://www.pwc.com/us/en/library/data-science-and-analytics.html>
- <http://sb.stratbz.to/upskillscale>
- <http://veribilimci.org/veri-bilimi-yuksek-lisans-programlari>
- International Labour Office. (2017). Future of Work. Inception Report for the Global Commission on the Future of Work. Geneva, Switzerland: International Labour Organisation.
- International Labour Organization. (2017). The Future of Work We Want: A Global Dialogue. Geneva, Switzerland: International Labour Organization.
- McKinsey Global Institute. (2017). Where Machines Could Replace Humans and Where They Can't (Yet). London, United Kingdom: McKinsey Global Institute.
- Michaels, G., Ashwini, N., & Van Reenen, J. (2014). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 years. *Review of Economics and Statistics*, 96(1), 60-77.
- Mishel, L., Schmitt, J., & Shierholz, H. (2013). Assessing the Job Polarization Explanation of Growing Wage Inequality. EPI-CEPR Working Paper.
- Moore, P. (2018). The Threat of Physical and Psychosocial Violence and Harassment in Digitalized Work. Geneva, Switzerland: International Labour Office.
- Nedelkoska, L., & Quintini, G. (2018). Automation, skills use and training (OECD Social, Employment and Migration Working Papers No. 202). Retrieved from https://www.oecd-ilibrary.org/employment/automation-skills-use-and-training_2e2f4eea-en
- Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3), 297-316. <https://doi.org/10.1002/smj.2014>
- Neffke, F., Henning, M., & Boschma, R. (2011). How Do Regions Diversify over Time?

- Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography*, 87(3), 237-265. <https://doi.org/10.1111/j.1944-8287.2011.01121.x>
- Nelson, R. R. (1982). An evolutionary theory of economic change. Cambridge, Mass; London: Belknap Press of Harvard University Press.
- O'Reilly Salary Data Science Salary Report, 2016, <https://www.oreilly.com/library/view/2016-data-science/9781492049029/>
- UNCTAD. (2017). Information Economy Report 2017: Digitalization, Trade, and Development. UNCTAD. Retrieved from http://unctad.org/en/PublicationsLibrary/ier2017_en.pdf
- Valerio, A., Sanchez Puerta, M. L., Tognatta, N., & Monroy-Taborda, S. (2016). Are There Skills Payoffs in Low- and Middle-Income Countries? Empirical Evidence Using Step Data (World Bank Policy Research Working Paper No. 7879). Washington, D.C., United States: The World Bank.
- Vandaele, K. (2018). Will trade unions survive in the platform economy? Emerging patterns of platform workers' collective voice and representation in Europe (No. 2018.05). Brussels: European Trade Union Initiative.
- Wiesmann, B., Snoei, J. R., Hilletoft, P., & Eriksson, D. (2017). Drivers and barriers to reshoring: a literature review on offshoring in reverse. *European Business Review*, 29(1), 15-42. <https://doi.org/10.1108/EBR-03-2016-0050>
- Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2018). Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society*, 0950017018785616. <https://doi.org/10.1177/0950017018785616>
- Wood, A. J., Lehdonvirta, V., & Graham, M. (2018). Workers of the Internet unite? Online freelancer organisation among remote gig economy workers in six Asian and African countries. *New Technology, Work and Employment*, 33(2), 95-112. <https://doi.org/10.1111/ntwe.12112>
- Yeung, Y., & Howes, S. (2015). Resources-to-Cash: A Cautionary Tale from Mongolia (Development Policy Centre Discussion Paper No. 42). Canberra: Australian National University. Retrieved from <https://im4dc.org/wp-content/uploads/2015/09/Combined-Yeung.pdf>

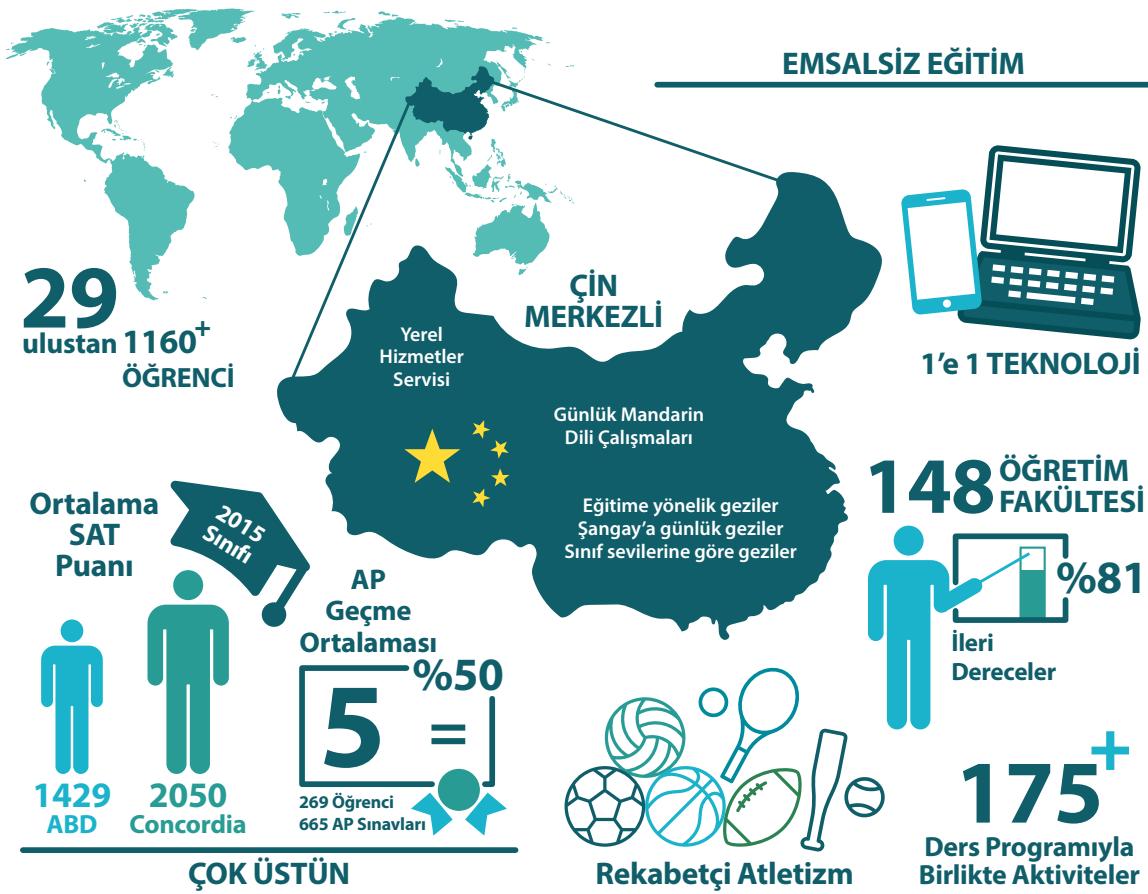
11. BÖLÜM

BÜYÜK VERİNİN YÜKSEKÖĞRETİM ÖNCESİNDEN EĞİTİME ENTEGRE EDİLMESİ



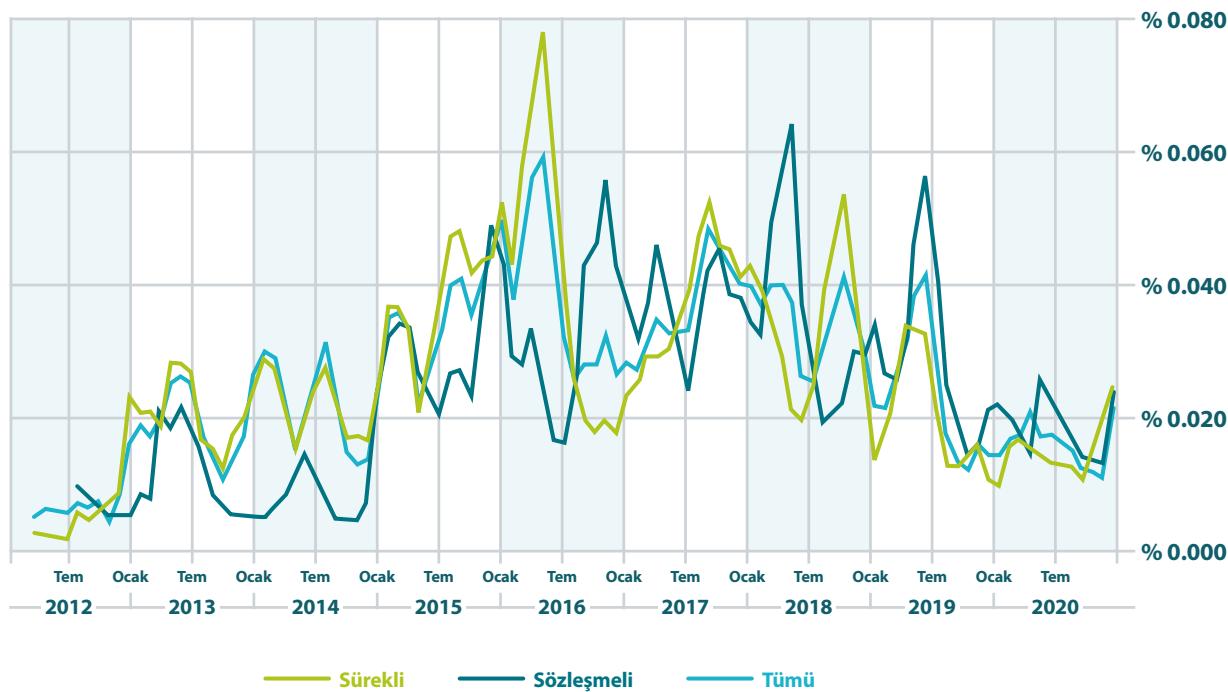
Büyük veri sürekli olarak gelişmektedir; şimdi “büyük veri” olarak adlandırılan şey, daha önceleri ortaöğretim sonrası eğitimde “veri madenciliği” veya “analitik ve veri bilimi” gibi terimlerle adlandırılıyordu. Bu alandaki ilerleme birçok yönden işlevsel olan büyük verinin avantajlarından dolayı son birkaç yılda patlama yaptı. Buna rağmen sektörlerde yetişmiş, uzman veri bilimci ihtiyacına rağmen ne yazık ki veri bilimci sayısında her şeye rağmen ciddi bir eksiklik söz konusu olmaktadır. Bu bağlamda yüksek öğrenim öncesinde de gençlere büyük veriyi tanıtmak ve anlatmak ve ayrıca bu yönde becerilerini geliştirmelerini sağlamak gelecek için büyük önem taşımaktadır.

Bu noktada eğitim sistemine büyük veriyi entegre etmek ve gençleri geleceğin dünyasına ve teknolojinin gelişimi ile önemi artan meslekler hazırlamak gerekmektedir. Dolayısıyla büyük veriyi anlamak ve işlemek için gerekli olan uygun eleştirel düşünme, tümevarımsal akıl yürütme ve analitik becerilerin kazandırılabileceği yapılandırılmış bir büyük veri programının yüksek öğrenim öncesi dönemde eğitim sistemine entegre edilmesi gençler için büyük avantajlar sağlayabilir.



Dünya ülkeleri genç nesilleri şimdiden büyük veri, nesnelerin interneti, yapay zeka gibi kavramlarla tanıtırarak ve bu teknolojileri uygulamalarını sağlayarak geleceğin teknolojilerine ve mesleklerine hazırlamakta ve doğabilecek ihtiyaç karşısında olası senaryolar üzerinden hazırlıklar yapmaktadır. Örneğin Çin'in Şangay şehrindeki Concordia International School'daki öğrencilerin talebiyle okul 2014-2015 akademik yılında yeni bir seçmeli derse öncülük etti: "Büyük Veri Analitiği." Bu pilot kurs lise öğrencilerinin üniversite eğitimlerine daha iyi hazırlanmaları için ihtiyaçlarını karşılamak üzere oluşturuldu. Ayrıca bu ders Asya'daki uluslararası okullarda verilen ilk büyük veri kursu olarak tarihe geçti.

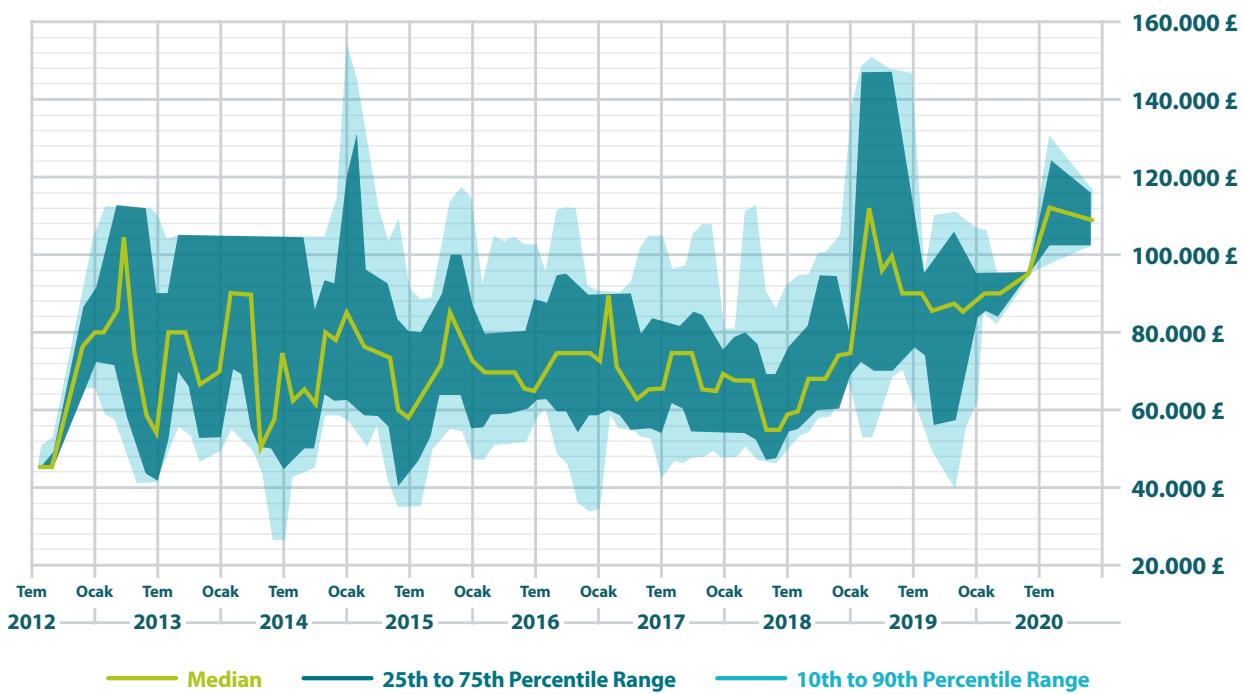
Ham haldeki veriler bir mucevher gibidir: Analiz edilene, tablo haline getirilene ve anlaşılan bir biçimde sunulana kadar hiçbir değeri yoktur. Matematikçiler, ista-



Şekil 11.2 Büyük veri yöneticiği açık pozisyon eğilimi. İlan verilen tüm BT işlerinin bir oranı olarak iş başlığında büyük veri yöneticisini gösteren iş ilanları. (Kaynak: <https://www.itjobswatch.co.uk/jobs/uk/big%20data.do>)

tistikçiler, bilim insanları ve bilgisayar bilimcileri ham verileri anlaşılır ve kullanılabilebilir bilgilere dönüştürmektedir. Büyük veri analitiği ham verilerin toplanması ve verilerin anlaşılabilecek değerli bilgilere dönüştürülmesi sürecini tanımlamak için kullanılan güncel terimdir. İşletme, ekonomi, sosyal bilimler, sanat ve beşeri bilimlerden bilime kadar çeşitli alanlardaki insanlar büyük verinin ne olduğunu ve ticari, endüstriyel, akademik alanlar ile araştırma ve geliştirme konusunda fayda sağlamak için nasıl uygulanabileceğinden doğan anlama ihtiyacının farkında olarak bu alana yönelmektedir.

Dünyada son on beş, yirmi yıl içinde üniversiteler lisansüstü düzeyde büyük veri dersleri sunarken, özellikle büyük veri ile ilgili lisans düzeyinde derslerin verilmesinin üzerinden çok fazla zaman geçmemiştir. Üniversiteler mevcut iş piyasasının talebini karşılayacak kadar hızlı veri bilimciler mezun etmediğinden, bu tür programlardaki yüksek lisans öğrencileri büyük veride lisansüstü eğitimlerini tamamlarken çok dolgun bir ücret ile zaten kapılmaktadırlar. Nitekim Birleşik Krallık'ta-



Şekil 11.3 Büyük veri yönetici maaş eğilimi. Büyük veri yöneticisine atıfta bulunulan işlerde verilen 3 aylık ortalama maaş.



Şekil 11.4 Büyük veri yönetici maaş histogramı. 6 aydan 10 Aralık 2020'ye kadar büyük veri yöneticisine atıfta bulunulan işler için maaş dağılımı.

ki maaş ortalamasının ve artan iş fırsatının gösterildiği <https://www.itjobswatch.co.uk/jobs/uk/big%20data%20manager.do> sitesinden alınan Şekil 11.2, Şekil 11.3 ve Şekil 11.4'te bu durum kendini açıkça göstermektedir.

Günümüzde artış gösteren ve artmaya da devam edecekçe benzeyen büyük veri alanındaki insan kaynağı eksikliğine rağmen veri bilimcilerin talebini karşılamaya yönelik hazırlık aşaması şu şekildedir:



Şekil 11.5 Üniversiteden iş dünyasına giderkenki doğal süreç.

Ancak bu doğal ilerleme şu anki mevcut programlarda kendisine yer bulamamaktadır. Büyük veri eğitimi göreceli olarak lisansüstü düzeyde kurulmuştur ve dünyada lisans dersleri de henüz emekleme aşamasındadır. Bunun yanı sıra ülkemizde üniversitelerde yüksek lisans programları mevcutken lisans düzeyinde programlar bulunmamaktadır. Lise düzeyinde ise ne dünya ülkelerinde ne de ülkemizde büyük veri analitiği konusunda uzmanlaşmış herhangi bir program yok denenecek kadar azdır. Oysa günümüzün dünyasındaki nesnelerin interneti, yapay zeka ve robotik alanındaki gelişmeler ve bu alandaki doğacak istihdam için öğrencileri bu yeni eğitim alanına hazırlamak adına zorunlu bir ihtiyacı vurgulamaktadır. Diğer tarafından bu tür program geliştirme için mevcut parametreler yoktur. Buna bağlı olarak da anlaşılabilir ve uygulanabilir müfredatın olmaması ve bu alandaki temel ihtiyaç ve dinamiklerin entegre edildiği yeterli ve nitelikli kaynak ya da çalışma kitapları, syllabus, öğretim materyalleri ve pedagojik formasyonun olmaması hem eğitimciler hem de danışmanlar için büyük bir zorluğu da beraberinde getirmektedir.

Büyük veri analitiğini liselerde yeni bir konu olarak tanıtmak için müfredatlar ve pedagojiler geliştirilmeli ve oluşturulmalıdır, böylece öğrenciler ortaöğretim

sonrasında bu alana girmeye hazırlanabilir. Liseler için programların hazırlanması ve uygulanmasının zorluğunun yanında ayrıca bu konuya öğretmede de zorluklar mevcuttur. Büyük veri analitiği bilgi teknolojisi, programlama, matematik, tümevarımsal akıl yürütme ve yüksek düzeyde analitik beceriler gibi ileri düzeyde bilgi birikimi gerektirmektedir. Pedagojik olarak ise müfredatta büyük veri analitiğini uygulamak çok ciddi bir görevdir çünkü bu alan aynı zamanda çeşitli konularda çok fazla arka plan bilgisi gerektirmektedir. Orta ve ilkokul seviyesi için bu alandaki bilgileri ne kadar basitleştirmeye çalışırsak çalışalım, genç öğrenciler büyük veriyi anlamak için gereken bilgiyi hiç edinemediklerinden o kadar zor olacaktır.

Bütün bu koşullar ve güçlükler göz önüne alındığında, bir lise müfredatının her seviyesine yönelik ön beceriler ve arka plan bilgisi nasıl anlatılabilir ve öğretilebilir?

11.1 ÖĞRETİMDE DEĞERLER DEĞİŞİMİ

Ortaokul ve ilkokul öğrencileri için daha da geliştirileceği anlayışıyla lise öğrencilerine yönelik büyük veri analitiği geliştirme çalışmaları yürütmektedir. Bu çalışmaların içerik olarak ana odağı farkındalık, sergileme, uygulamalar ve büyük veri hakkında genel bir kavramsal anlayış oluşturmak şeklindedir. Concordia International School'da gerçekleştirilen bu çalışmada büyük verinin erişilebilir ve sürekli gelişen doğası nedeniyle lisede öğrenciler arasında büyük veri analitiği gelişiminin ilk aşamaları iki girişime yol açmıştır: Öğretimde değerler değişimine ve büyük veri analitiğinin bir ders olarak müfredatta yer bulmasına. Bu, öğretmenin öğrencilere doğrudan bilgi verdiği öğretim yönteminin aksine, öğrencilerin öğretmen tarafından “dolaylı olarak yönlendirildiği” çok güçlü bir öğretim yöntemidir.

Yeni ders konularının ve bilgilerinin hızla geliştiği günümüzün öğrenme ortamında, öğrencilerin başvurabileceği güncel materyalleri bir araya getirmek çok zordur. Bu nedenle öğrencilerin kendi kendilerine nasıl öğreneceğini, öğrenme materyalleri ve tekniklerini nasıl geliştireceklerini, nasıl bilgi arayacaklarını ve nasıl etkili iletişim kuracaklarını öğrenmeleri önemlidir. Bu öğrenme becerileri ise öğrencilerin yalnızca ders konularını sınıfta öğrenmelerini sağlamakla kalmayıp aynı zamanda onlara gelecekteki yeni alanlarında kendilerinin nasıl öğreneceğine dair araçları tanımlarını da sağlayacaktır.



Bir eğitmenin amacı, sadece bilgiyi iletmek ve saklanarak uygulanmasını sağlamaktan çok daha fazlasıdır. Bu açıdan bakıldığından öğrenciler kendi kendilerine öğrenme eylemini gerçekleştirdiklerini hissettiklerinde öğrenme konusunda kendilerine güvenirler ve bu da bir öğrenme ortamında daha güvenli olmalarını ve güvende hissetmelerini sağlar. Bu yöntemle öğrenciler bağımsız öğrenme yetenekleriyle güçlendirilir ve sonuç olarak hem eğitimlerinde hem de kendi içlerinde daha büyük bir başarı duygusu hissedecektir.

Concordia International School'dan King, büyük veri analitiği dersi için öğretmenin rolünü “yanındaki rehber” olarak tanımlar. Öğretmenin bu ders için açık yönnergeleri, hedefleri ve hedeflerden oluşan yapılandırılmış bir çerçevesi olması gereklidir. Bu öğretim yöntemi aynı zamanda öğretmenin öğrencilerinin güçlü yönlerini ve farklı becerileriyle özelliklerini ortaya çıkarmasına ve öğrencilerin konuyu araştırması yoluyla öğrencilerin zayıf olduğu alanlarda geliştirme göstermesine de olanak tanır. Aşağıda Concordia International School tarafından hazırlanmış pilot ders ile syllabus örneği yer almaktadır.

11.2 PİLOT DERS: BÜYÜK VERİ ANALİTİĞİ

Bir örnek ya da model olması açısından Concordia International Scool'da büyük veri analitiği dersinin ortaya çıkışına ve oluşturulma sürecine göz atmakta fayda var. Concordia International School Shanghai, Şangay'daki göçmen öğrencilere yönelik öğrenci merkezli uluslararası bir okuldur. Başlangıçta büyük veri analitiği dersi bir “eXploration” kursu olması önerilmiş, ancak gördüğü ilgi ve öğrenci talebine bağlı olarak iki haftalık “eXploration” deneme kursundan hızlı bir şekilde Asya'daki uluslararası okullarda türünün ilk örneği olan bir dönemlik seçmeli derse dönüştürülmüştür.

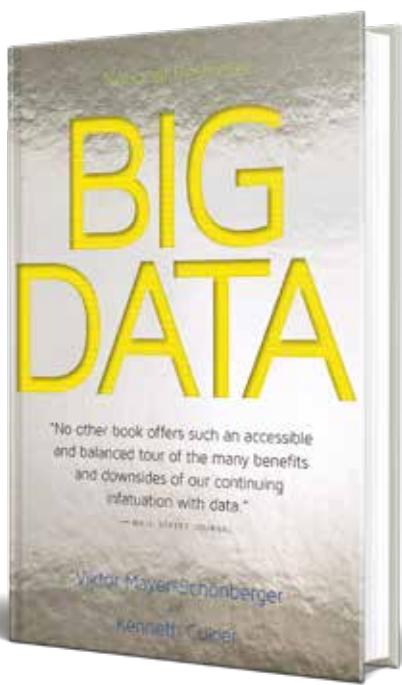
Bu yeni ders için müfredatı oluşturmanın gerekliliklerinden birinin içerik önerileri ve temel sorular geliştirmek olduğunu belirtmektedir Peter Tong ve Felicia Young. Bunu bir örnekle açıklarlar:

- **Büyük Veri nedir?**
 - Tek bir bilgi parçası nasıl büyük bir bilgi parçasına dönüşür?
 - 5 V: Hacim, Hız, Çeşitlilik, Doğruluk ve Değer
 - Büyük Verinin istatistiklerden farkı nedir?
 - Örnekler: Moneyball, H1N1 Grip, Oren Etzioni'nin Farecast'i
- **Büyük Veri ne kadar değerlidir?**
 - Veri bilgisi neden bu kadar önemli?
 - Kimler için önemli?
 - Büyük veri neden oyunun kurallarını değiştiriyor?
- **Büyük Veri nerede uygulanabilir?**
 - Finans dünyası, eğitim, sosyal bilimler, yargı sistemi.
 - Kurumlar problemlerini çözmek için büyük veriyi nasıl kullanıyor?
- **Büyük veri kullanımı için araçlar nedir?**
 - Bulut bilişim.
 - Yapılandırılmış ve yapılandırılmamış veriler.
 - Donanım/Yazılım gereksinimi.

- Yönetim sistemleri
- Analistik

Bu ders için önerilen kitaplar:

- *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Viktor Mayer-Schönberger ve Kenneth Cukier
- *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*, Eric Siegel ve Thomas H. Davenport
- *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*, Steven Finlay



Bu yeni ders öğretmenin “yanındaki rehber” rolünde olduğu bir ders olarak tasarlanmıştır. Bu pilot derste bir öğrenci, öğretmenin rehberliğinde PowerPoint ders stilini kullanarak günlük olarak dersler vermiştir. PowerPoint’ın kullanıldığı derslerde ise Viktor Mayer-Schönberger ve Kenneth Cukier’ın *Big Data: A Revolution That Will Transform How We Live, Work, and Think* adlı kitabı kullanılmıştır.

Büyük verinin genel kavramları akran tartışmaları yoluyla öğretilirken, bu derste pratik öğrenim uygulamaları aracılığıyla gerçekleştirilmiştir. Öğretmen, derste

eğitim hedefleriyle koordineli olarak hem grup hem de bireysel projeleri bir araya getirmiştir: Bilgi, anlama, uygulama, analiz, sentez ve değerlendirme. Öğrencilere, akran tartışmaları ve her dersten sonra öğretilen materyalleri daha fazla araştırmaları için sınıfta fazladan zaman verilmiş, her alt bölümün sonunda öğrenciler konuya ilgili bilgileri ve kavrayışlarını, içeriği dersten farklı örneklerle sunmanın farklı bir perspektifini içeren PowerPoint özetleri oluşturarak sözlü olarak sunmuştur.

- | | |
|---|-----------------|
| 1 Introduction: introduce the class by asking questions. Let the students make hypothesis about how big data works.
(EX: facebook suggests friends). | Videos?/Tedtalk |
| 2 Introduction: Read the first chapter of the book "Big data: a revolution that will change how we live, work, and think".
Read Pg. (Google flu trend) | Book |
| 3 Introduction: Continue to read the book Pg.(What is big data) | Book |
| 4 Introduction: | book |
| 5 Introduction | book |
| 6 Introduction: work on the presentation (show their own standing of big data , can present in any form) | book |
| 7 Introduction: presentation | Presentation |

Öğrencilerden ayrıca proje çalışmasını sınıf olarak ve bireysel olarak tamamlamaları, böylece uygulama, analiz ve sentez konusundaki hedefleri de kapsaması sağlanmıştır. Örneğin bir sınıf projesinde öğrencilerin bir CISS ortaokul beden eğitimi dersinde verileri toplamak ve analiz etmek için kalp atışı monitörlerini kullanmaları gerekmıştır. Bu projeyeyle öğrencilerin büyük verinin pratik uygulamalarını tanıtmakla kalmayarak aynı zamanda veri toplama sistemindeki kusurları belirlemek için büyük veri araçlarını ve kavramlarını kullanmalarını sağlanmıştır. Projenin yürütülmesinde öğrenciler daha sonra bir beden eğitimi öğrencisinin öğrenme konusundaki ilerleyi-



Deslerde bir YouTube videosundan yararlanabilirsiniz.



şini daha doğru bir şekilde çizelge ile göstermek ve derecelendirmek için teknikler geliştirmeye çalışmıştır. Beden eğitimi ve koçluk yöntemlerini geliştirmek için ortaokul beden eğitimi bölümüne öğretimin nasıl optimize edileceği ve ne tür verilerin toplanacağı konusunda yeni bilgiler sağladıkları için bu öğrenciler için başarılı bir pratik öğrenme fırsatı sağlanmıştır.

Bu dersle birlikte öğrenciler sadece büyük veri kavramıyla tanışmamış, aynı zamanda akademik topluluğun yanı sıra öğrenciler okulları, yerel ve küresel topluluklarla iletişim kurmuşlardır. Böylece kendilerine güvenleri gelişerek kendilerine gelecekte bir yer bulacaklarına inanan öğrencilerin ortaya çıkışını da beraberinde getirmiştir.

Teknoloji ilerledikçe, daha fazla şirket günlük operasyonlarının bir yan ürünü olarak büyük miktarda veri üretiyor. Bir veri denizinde çoğu işletme filen boğuluyor olmasa da bir gelecek isteyen şirketlerin büyük veriyi ciddiye almaya başlaması gerekmektedir. Bunun yanı sıra üstelik büyük veri parçalarını değerlendirme, analiz etme ve kullanma becerisine sahip yetenekli insanlara yönelik artan talebe rağmen bu insanların eksikliği genişlemeye devam ediyor. Inc.com'a göre, şirketlerin yaklaşık yüzde 73'ü büyük veri setlerini ihmal ediyor. Ancak mesele topladıkları verileri anlamak istememek değil, topladıktan sonra ne yapacaklarına dair hiçbir fikirlerinin

Z Kuşağının En Çok Kullandığı Sosyal Medya Uygulamaları

	Japonya	Güney Kore	Birleşik Krallık	ABD
1	Snapchat	Everytime	Snapchat	Snapchat
2	pixiv	Discord - Chat for Gamers	Discord	TikTok
3	Discord	NEXON Play	Tik Tok	GroupMe
4	KakaoTalk	Snapchat	Tumblr	Kik Messenger
5	Instagram	디시인사이드 - dcinside	Instagram	Tumblr
6	Skype	Puffin Web Browser	Kik Messenger	WeChat
7	WhatsApp Messenger	Facebook Messenger	POF Online Dating	Discord
8	Pinterest	Between app for couples	Pinterest	Telegram
9	LINE	Tumblr	Facebook	Instagram
10	WeChat	Pinterest	Telegram	KakaoTalk

Not: Mua tarafından seçilmiş en çok kullanılan 25 uygulamanın yer aldığı Z Kuşağı indeksindeki uygulamalar için sadece Android telefonlar kullanılmıştır.

Şekil 11.6 Z kuşağının en çok kullandığı sosyal medya uygulamaları. (Kaynak: https://s3.amazonaws.com/files.appannie.com/reports/1019_How_to_Win_Gen_Z.pdf)

olmamasıdır. Sonuç olarak şirketler veri kümelerini gözden geçirme ve yönetme yeteneğine sahip birini işe almazsa, rakiplerinin gerisinde kalmaya devam edecekler.

NYSCI'dan Kate Donnelly'ye göre, "ABD'li çocukların çoğu artık veri analizinin etkileriyle her gün karşılaştıkları bir dünyada büyüyor. Ancak büyük veri ve veri bilimi konuları tipik olarak trafik kalıpları, borsa eğilimleri ve dijital bilgiler etrafında dolaşıyor, dolayısıyla çocukların ilgisini çekemiyor." Veri biliminin çoğu insan için keşfedilmemiş bir konu olduğu düşünüldüğünde, yenilikçi müfredatımız ebeveynlerin ve çocukların birlikte öğrenmeleri ve bu karmaşık, zor olan konuyu birlikte koyalayıstırmaları için erişilebilir ve ulaşılabilir bir yol bulmak zorunlu hale gelmektedir.

Günümüzde batı dünyası ilkokul ve lise eğitimcileri, öğrencilerinden veri bilimini öğrenmeye başlamalarını şart koşmaya başladı bile. Örneğin Avustralya'da okullarda geleneksel okuma, yazma ve aritmetiğe ek bir konu eklenmiş olup bu Kodlamadır. Türkiye'de de kodlama ve STEM dersleri kendine yer bulsa da yine de büyük veri



Şekil 11.7 Veri bilimci ile ilgili infografik.

kavramına dönük sağlam bir müfredattan söz etmek olanaksızdır. Bu bağlamda Z kuşağı ve sonraki nesil çocuk ve gençleri düşündüğümüzde Z kuşağı zaten gelişen ve büyüyen teknolojinin içine doğmuştur. Bu noktada bu ilgilerinden de faydalananarak çeşitli yöntemlerle öğrencilere büyük veri kavramı ile ilgili farklı projelerin verildiği bir müfredatla öğrenciler geleceğe daha iyi hazırlanabilir.

Her ne kadar şu an günümüzde lise ve ortaöğretim seviyesinde büyük veriye yönelik bu başlık altında dersler olmasa da dersinizde mesleklerle dair bir konu anlattığınızda büyük veri analisti mesleğine dair bir proje yapmalarını isteyebilir ve bu noktada yukarıdaki şekle benzer görselde olduğu gibi kendi çalışmalarını hazırlamalarını

isteyebilirsiniz, böylece verileri görselleştirmenin bir yolunu da öğrenmiş olurlar. Öğrencilerinizin kendilerini bir veri mühendisi ya da bir büyük veri analisti olarak hayal etmelerini söyleyebilir ve buna göre bir sunum hazırlamalarını isteyebilirsiniz.

Sonuç olarak büyük veri yaşama, düşünme ve çalışma şeklimizi dönüştürmektedir. Bu noktada büyük veriyi eğitim sistemimize entegre ettiğimizde çocukların teknolojiye olan ilgi alanlarını keşfederek tutku duymalarını sağlayabilir ve onlara daha fazla kariyer seçeneği sunabiliriz. Uygulamaya konulacak derslerle çocukların teknolojiye olan ilgisinden de yararlanarak araştırmayı, eleştirel düşünmeyi ve analiz etmeyi öğrenen, takım oyuncusu olan, yaratıcı, detaylara odaklanan, resmin tamamına bakabilen ve toplumun aktif bir üyesi olduğunun bilincinde olan çocukların gençler yetiştirek onları yarının dünyasına hazır hale getirebiliriz.

KAYNAKLAR

Big Data Analytics for Middle School 4th International Big Data and Analytics Educational Conference 9 to 11 January 2017, University of Derby, UK Peter Tong, Ph.D. and Grace Wei, Çin.

Finlay, Steven, Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods, 2014.

<https://www.researchgate.net/publication/315780891>

<https://www.itjobswatch.co.uk/jobs/uk/big%20data.do>

<https://nysci.org/learninglab/young-learners/>

https://s3.amazonaws.com/files.appannie.com/reports/1019_How_to_Win_Gen_Z.pdf

Mayer-Schönberger, Viktor ve Kenneth Cukier. Big Data: A Revolution That Will Transform How We Live, Work, and Think, 2013.

Siegel, Eric ve Thomas H . Davenport, Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die, 2013.

12. BÖLÜM

YÜKSEK- ÖĞRETİMDE VERİ BİLİMLİ



12.1 GİRİŞ

Büyük veri ve veri bilimindeki gelişmelerin ve bu alanın önemini artmasıyla birlikte eğitimde de veri bilimin entegresi üzerine çeşitli tartışmalar gerçekleştirilmektedir. Veri biliminin nasıl tanımlanması gerektiği veya bir veri bilimi derecesinin ne gerektirdiği konusunda bir fikir birliği olmasa da çoğu araştırmacı ve akademisyen önemini giderek arttırdığı bu disiplinler arası alanın istatistik, bilgisayar bilimi, mühendislik ve matematikle çok yakından ilgili olduğu, ancak bunlarla sınırlı olmamak üzere birçok disiplinin temel unsurlarına bağlı olduğu konusunda hemfikirdir. Bu noktadan hareketle veri bilimindeki yeterliliğin ayrılmaz bir parçası olan çok sayıda temel beceri, teknik ve kavramın - bir disiplinden veya birçok disiplinde ortak olan- olduğuna dikkat çekilmektedir. Veri bilimi kursları, programları ve dereceleri gelişme gösterdikçe bu alanda nasıl hareket edilmesi gerektiğine dair ortak bir kanıya varılabilir ancak bu da kurumsal bağamlara ve fırsatlara bağlı olacaktır.

Birçok insana göre fakülteler veri bilimini öğretmek için yeni beceriler öğrenmek, yöntemleri uyarmak ve yeni (genellikle çok disiplinli) yaklaşımalar bulmakta zorlanmaktadır. Buna karşılık öğrencilerin veri becerilerine olan talebi artmaya devam etmekte ve veri, hesaplama ve yazılım araçları yaygınlık göstermeye başlamaktadır. Veri bilimi perspektiflerini mevcut verinin yoğun olduğu derslere entegre etmek gibi, orta öğretim sonrası veri bilimi eğitimine yönelik umut verici birkaç yaklaşım orta-



ya çıkmıştır. Bunlar çoklu bakış açılarını, becerileri ve alanları bütünlüğe taşıyan yeni kurslar geliştirmek ve işbirliği içinde öğretmek. Bu ve diğer birkaç öneri de bu yolda ilerleme sağlayabilir: a) Araştırma Görevlilerini çeşitli becerilerde eğitmek; b) Yeni yaklaşımı denemek ve değerlendirmek isteyen öğretim üyelerini belirlemek ve daha çok desteklemek; c) Veri bilimi için hangi disiplinlerine gereksinim duyulduğunun kavranmasını sağlamak ve bu yönde gelişim göstermek; d) Nicel eğitim olmadan öğrencilere veri bilimini tanıtmak için yöntemler geliştirmek; e) Veri bilimi eğitiminin teşvik etmek üzere standart disiplin veri setlerini entegre etmek; f) Geleneksel seminer öğretiminin ortak hedefi, bir alanda veya veri biliminde derin uzmanlığı olan ve diğeriyile etkili bir şekilde işbirliği yapmak üzere yeterli bilgiye sahip araştırmacılar yetiştirmektir.

Üniversitelerdeki Lisansüstü Veri Bilimi Programları Lisans Veri Bilimi Programlarından daha fazla gelişmiştir. Veri biliminde doktora eğitimi yaklaşımı genellikle doktora programının bir akademik kurum içinde yer olması ve diğer bölümlerle nasıl etkileşim içinde olduğuna bağlıdır. Her yaklaşım benzersiz bir amaca hizmet eder ve toplu olarak bu yaklaşımlar gelişmiş veri bilimi eğitimi için hızla seçenekler sunar. Bu yaklaşım, idari mekanizmaları ve uygulama süreçleri bakımından farklılık gösterebilir. Örneğin doktora öğrencileri doğrudan bir veri bilimi programına veya bir

ana bölüme kabul edilebilir; bazı durumlarda bir veri bilimi programına kabul yalnızca bir öğrenci kampüse geldikten sonra gerçekleşir. Bazı programlar ise öğrencileri veri bilimi için ek gereksinimleri tamamlamadan önce kendi bölümlerindeki tüm gereksinimleri tamamlamaya zorlar. Genellikle birden fazla bilimsel alanla etkileşim içinde olmak daha geniş bir tez ve araştırma yürütme yeteneği bir veri bilimi doktora programında avantaj sağlayabilir. Fakülte esnekliği ve bir kurumun kısıtlamaları dahilinde çalışma isteği, özellikle yeni programlara başlamayla ilgili zorluklar göz önüne alındığında faydalı olabilir. Değerlendirmeye dönük ölçümler ise hangi programların geliştiğini, farklı programlarda gereksinimlerin nasıl değiştigini, ne tür tezlerin üretildiğini ve mezunların nerede işe alındığını belirlemek için de yarar sağlayabilir.

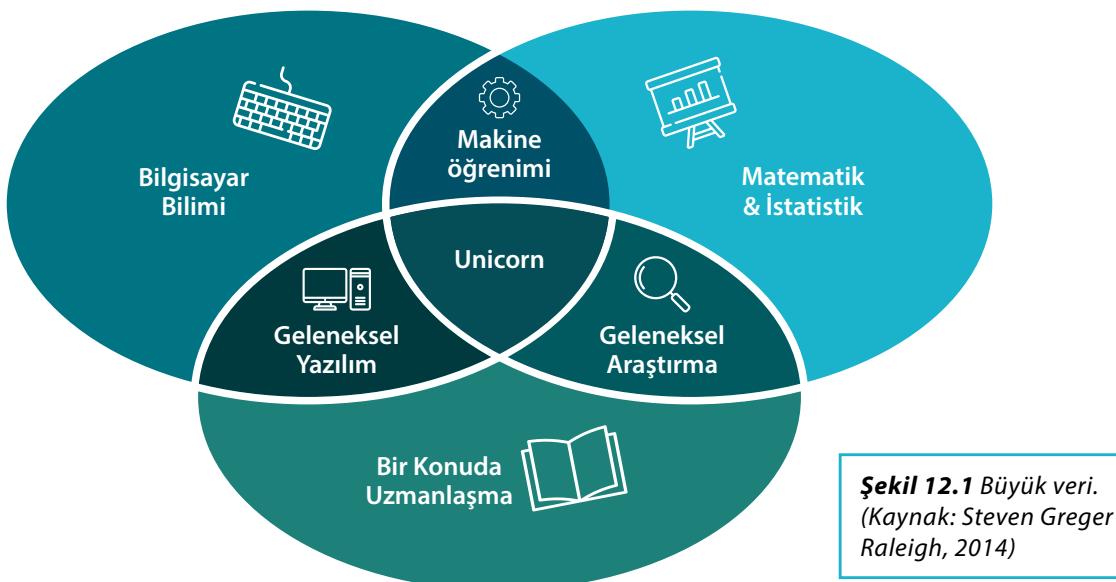
Veri bilimi becerilerine sahip çalışanlara olan talep farklı sektörlerde artmaya devam etmekte ve günümüzün veri bilimi işlerinden bazıları iki yıllık (önlisans) derecesine sahip kişiler tarafından doldurulabilmektedir. Veri konusunda kariyer yollarının ve uzman çalışan profillerinin geliştirilmesi gibi işverenlerin ihtiyaçlarını daha iyi anlamak ve bunlarla uyum sağlamak için çok çaba gösterilmektedir. Kurumlar vasıflı işgörünün değişen taleplerini karşılayacak şekilde müfredatlarını revize etmektedir. Birkaç kurum veri bilimi, veri analizi ve veri yönetimi alanında kurslar, sertifikalar ve





önlisans dereceleri geliştirmektedir. Fakülte eğitimini teşvik etmek, bu programların başarısının anahtarıdır. Finansman ve kaynak kısıtlamaları iki yıllık Meslek Yüksek Okullarında yeni programları uygulamayı zorlaştırdığından, yakındaki dört yıllık ve yüksek lisans veren kurumlarla resmi ortaklıklar kurmak oldukça değerlidir. İki yıllık Meslek Yüksek Okullarının Programlarını tasarlarken transfer potansiyelini de değerlendirmeleri ve ileri eğitim arayan öğrenciler için daha yumuşak geçişler yaratabilecek ekleme anlaşmaları geliştirmeye başlamaları da ayrıca önem arz etmektedir.

Son yıllarda veri bilimi eğitiminin hem sınıf içinde hem de sınıf dışında nasıl olacağı konusunda önemli deneyimler kazanıldı. Bunun yanı sıra diğer eğitim fırsatları arasında hackathonlar, eğitim kampları ve müzeler gibi gayri resmi ortamlardaki etkinlikler yer alabilir. Eğitim kampları veri bilimi yaşam döngüsüyle yeterlilik oluşturan ve odaklı, probleme dayalı, ekip odaklı programlarla, akademik dünyadaki derece temelli programlar ile endüstrideki iş başında öğrenme arasındaki artan boşlukları doldurmayı sağlayabilir.



12.2 VERİ BİLİMİNİN TEMELLERİ: İSTATİSTİK, BİLGİSAYAR BİLİMİ, MATEMATİK VE MÜHENDİSLİK

A. İstatistik

Hızlanan teknolojik gelişmeler, daha büyük mevcut veri kaynakları ve modellemeyle nicelendirmeye artan ilginin bir sonucu olarak istatistikler bugün 1990'larda olduğundan çok daha farklı anlamakta ve öğretilmektedir. Amerikan İstatistik Derneği'ne (ASA) göre, temel veri bilimi, veritabanı yönetimi, istatistik, makine öğrenimi ve dağıtılmış paralel sistemler alanlarını içermeli ve yalnızca lisans düzeyinde değil, aynı zamanda lise seviyesinde de tanıtılmalıdır. İstatistik, veri biliminde önemli bir rol oynar çünkü soruların verilerin daha iyi kullanımını teşvik edecek şekilde çerçevelenmesini, belirsizliğin ölçülmesine yardımcı olacak çıkarımları, nedensellik ile korelasyon arasında ayırım yaparak belirlenecek müdahaleleri, tahmin için kullanılacak yöntemlerle tekrarlanabilir olması gereken bulguları sağlar.

İstatistiksel araştırma yapmak için kullanılan döngü problemi, planı, verileri, analizi ve sonuçları içerir (genellikle PPDAC döngüsü olarak kısaltılır). ASA, verilerle çalışmak ve PPDAC döngüsünü sürdürmek için bilgi işlem, yazılım, programlama, veri tartışması, algoritmik problem çözme ve iletişim becerilerinin gerekli olduğunu ve dolayısıyla resmi müfredatın bir parçası olması gerektiğini belirtiyor. Doğru

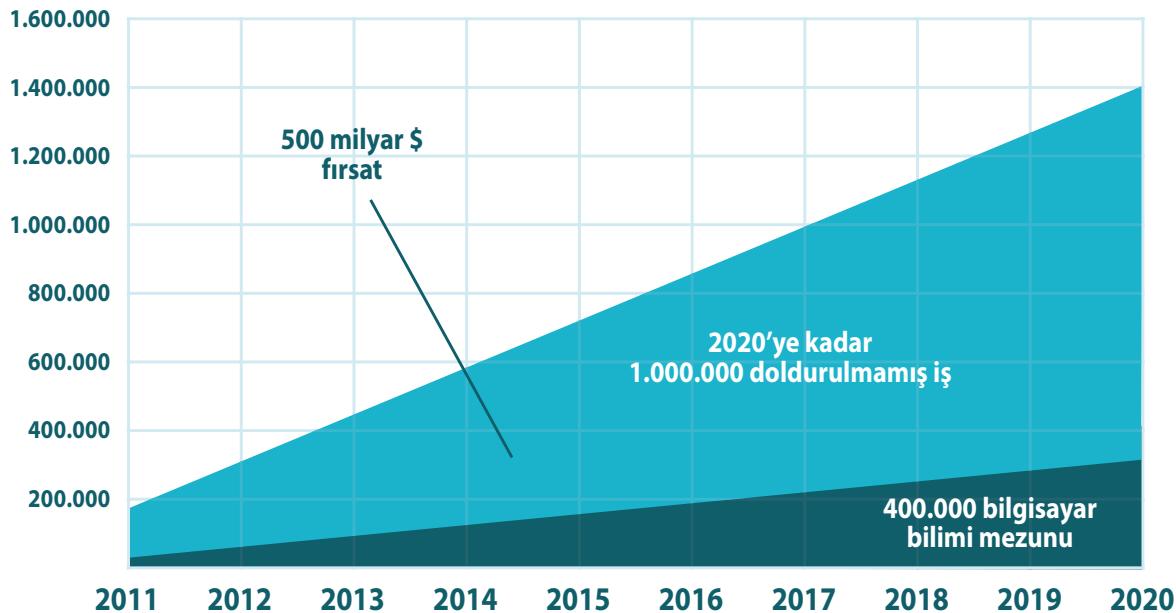


eğitimle istatistikçiler, bağlamı anlayabildikleri, değişkenliği hesaba katabildikleri, verileri tasarlayıp analiz edebildikleri, çıkarımı anlayabildikleri, tekrarlanabilirliği geliştirebildikleri, çok disiplinli ekiplerde çalışabildikleri ve veriye dayalı kararlar alabildikleri için veri bilimine değerli bir katkı sunarlar.

B. Bilgisayar Bilimi

Bilgi işlemin üç eğitim ayağı vardır. Bunlar:

1. Başlıca temeller (örneğin algoritmalar, makine öğrenimi, kürasyon, görselleştirme/modelleme ve hesaplama sistemleri aracılığıyla verileri anlama).
2. Gelişmiş temeller (örneğin yüksek performanslı bilgi işlem ve gelişmiş makine öğrenimi yoluyla büyük ölçekli verileri anlama).
3. Uygulamalar (örneğin bilgiyi veri mühendisliği yoluyla gerçek dünyadaki problemlere uygulama).

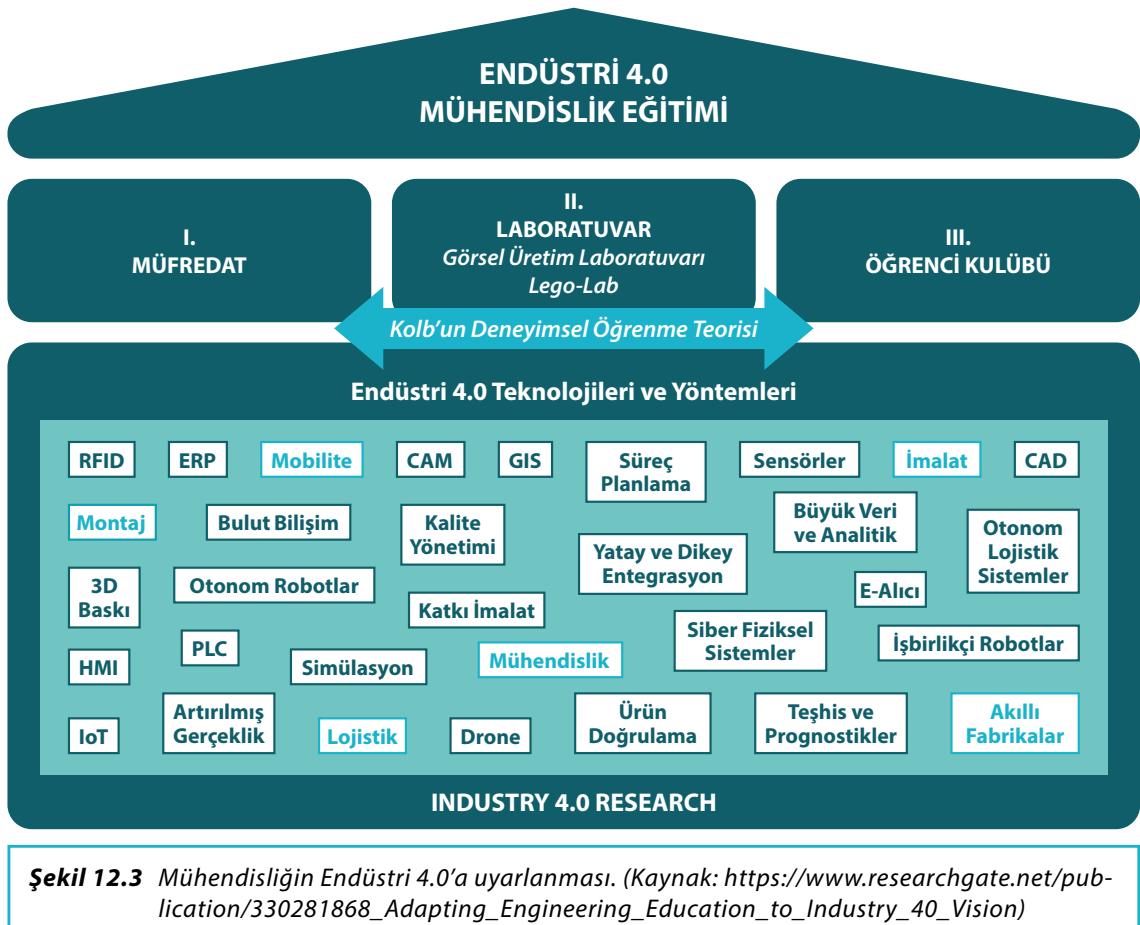


Şekil 12.2 2020'ye kadar 1.000.000 Doldurulmamış İş (Kaynak: BLS, NSF, Bay Area Council Economic Institute)

Modeller (veri içeren), diller ve makineler eşit derecede önemlidir ve bu da veri biliminin disiplinler arasılığını güçlendirir. Ayrıca algoritmaları geliştirirken yapılan seçimler politika kararlarını veya önyargıları içerebileceğinden, etik de herhangi bir veri bilimi müfredatında merkezi bir rol oynamalıdır.

C. Mühendislik

Mühendisler, öğrencileri güvenilir sistemler oluşturmaları için eğitmek ister ancak daha iyi kararlar almak için veri madenciliği ardışık dönemin yeniden tasarlanması gereklidir. Standartlar, analiz yazılımlarına artan sayıda atıf ve yazılım paketlerinin çoğalması ile başa çıkmak için standartlar da dahil olmak üzere bunun önemli bir parçasıdır. Mühendisler, veri bilimini veri toplamanın (örneğin algılama araçları ve veri havuzları aracılığıyla), verileri yönetmenin (örneğin esnek ve korumalı veritabanları aracılığıyla) ve verileri analiz etmenin (örneğin entegre hesaplama algoritmalarıyla) bir yolu olarak görürler. Mesela verinin etkin olduğu mühendislik malzeme genelinde, hassas tipta ve siber-fiziksel ağlarda kullanılır.



Veri bilimi doğal olarak multidisiplinerdir ve birçok disiplin matematikten (örneğin topolojik nesne olarak veri) gelen veri bilimi araçlarına ve ilkelerine dayanır; bilgisayar bilimi (örneğin liste/grafik olarak veriler), istatistikler (örneğin rastgele örmeklem olarak veriler), bilişim (örneğin arayüz olarak veri), fizik (örneğin doğal fenomen olarak veriler) ve mühendislik (örneğin karar verilecek veriler).

D. Matematik

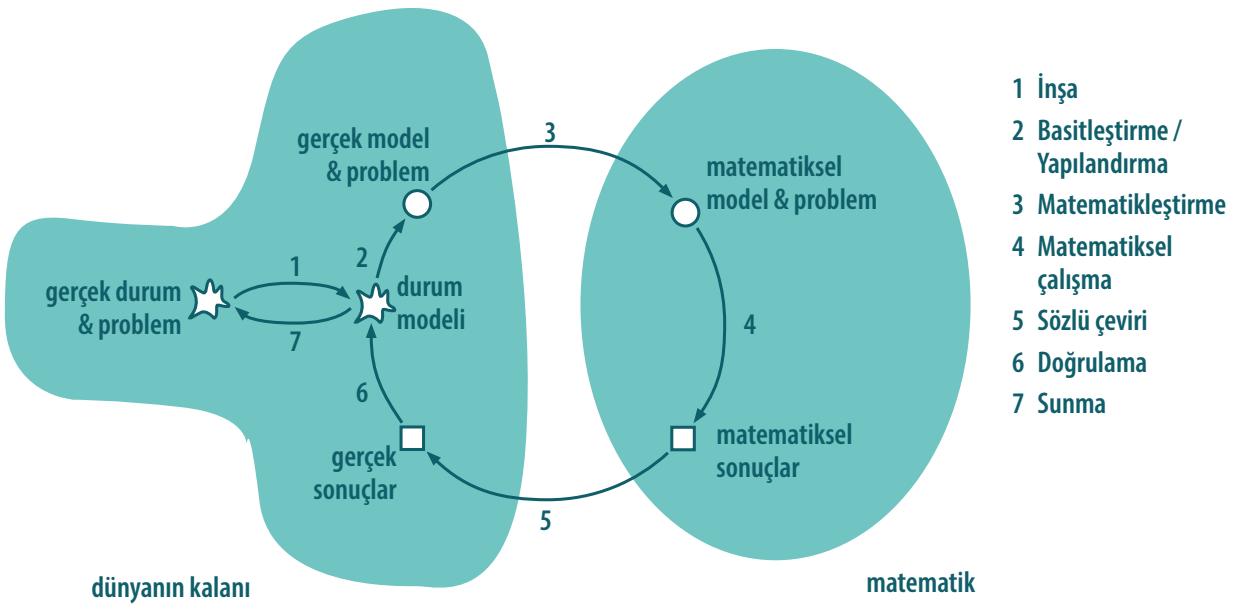
Veri bilimi tipik olarak iki kategoriden birine ayrılır: Hesaplamlı bilimler (örneğin bilgisayar bilimi, mühendislik ve istatistik) veya alan bilimleri (örneğin genomik, sinirbilim, metin analizi). Her iki alanda da matematiksel bir altyapı vardır: Hesaplama bilimleri doğrusal cebir, sayısal analiz ve grafik teorisi ile desteklenir; alan bilimleri probleme özgüdür, fiziksel ve yaşam bilimlerini kullanır ve fiziksel modellerle

Veri bilimi tipik olarak iki kategoriden birine ayrılır: Hesaplamalı bilimler (örneğin bilgisayar bilimi, mühendislik ve istatistik) veya alan bilimleri (örneğin genomik, sinirbilim, metin analizi).

matematiksel analiz araçlarına dayanır. Doğrusal cebir, analiz, geometri ve optimizasyon her zaman dünyamızı modellemek için kullanılan temel araçlar olmuştur ve bir miktar uyarlama ile öyle olmaya da devam edecktir. Matematik veri bilimi için teorik modeller, kavramsal bir çerçeve, dil ve ilgili bir “hesap” sağlayabilir. Modern veri biliminin matematiksel bir kavramsallaştırması, çeşitli matematiksel araçların açıklanlığı ve ortaklaşa motive edildiği bütünlüyici bir müfredatta alt alanların bir karışımını içerir. Eğitimciler veri bilimi ihtiyaçlarını karşılamak için matematik müfredatını nasıl geliştireceklerini ve bütünlüsterici öğretme ve öğrenmeyi nasıl daha iyi teşvik edecekleri üzerinde düşünmelidir.

Bu durumda işletmeler için bir paradoks söz konusu olur: Tüm ilgili becerilere sahip bir kişiyi mi işe almalılar, yoksa her biri benzersiz becerilere sahip bireylerden oluşan bir ekip mi kurmalıdır? Dolayısıyla “mükemmel” çalışانı bulmakta zorluklarla karşılaşabiliriz. Bir veri bilimcinin disiplinler arası iletişim kurma ve sorunları çözme becerisinin önemi artmaktadır. Diğer taraftan özel sektörün yanı sıra işe alımlar tüm devlet kurumları için de bir endişe nedenidir ve veri bilimcileri işe almak için daha iyi stratejiler geliştiren ekiplerin olması ayrıca önem arz etmektedir.

İşe alımlarda veri biliminde nitelikli insan gücünü bulmak ve gençleri geleceğe hazırlamak için farklı sertifika programları sunularak öğrenciler daha çeşitli disiplinlerden çekilebilir. Ancak günümüzde öğrencilerin veri bilimi derslerine olan talebinin tam olarak karşılanması söz konusu değildir. Buradaki en büyük engel mevcut fakülte ve kursların sınırlı sayıda olmasıdır. Bunun yanı sıra örneğin istatistik öğrencilerinin diğer disiplinlerdeki insanlarla beraber gerçek problemlerle ilişkiler geliştirmesi gibi multidisipliner çalışmalar ve işbirlikler geliştirilebilir. Bir müfredat belirlerken göz önünde olması gereken bir husus da cinsiyet değişkenliği ve dağılımı olmalıdır.



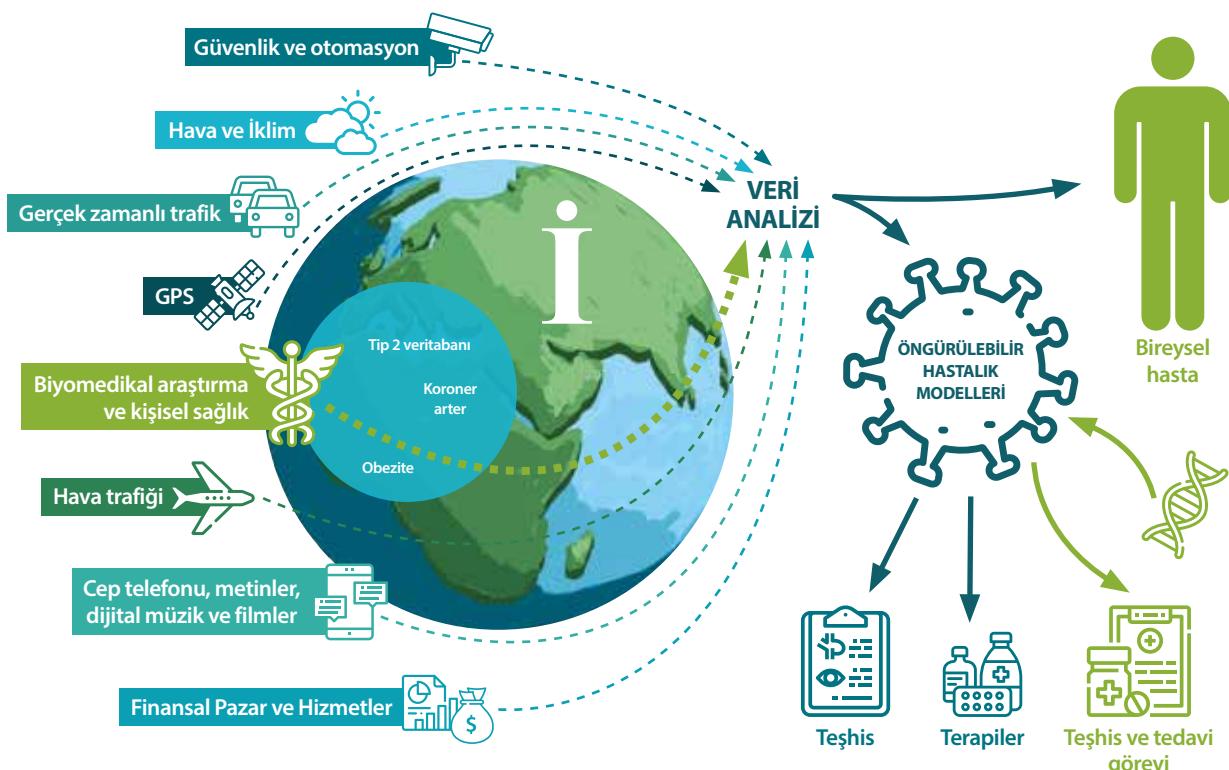
Şekil 12.4 Matematik ve büyük veri arasındaki ilişki. (Kaynak: <https://www.semanticscholar.org/paper/The-relationship-between-Big-Data-and-Mathematical-Vecchia/8a-015d40858925a36e39996ef22f9214b250e1f5>)

12.3 VERİ BİLİMİ TOPLULUKLARININ GEREKSİNİMLERİ

12.3.1 Biyomedikal Araştırma

Veri bilimi biyomedikal araştırma için çok önemli hale geldikçe, biyomedikal bilimde beş eğilim ve güçlük kendini göstermiştir:

1. Biyomedikal veri bilimi bir çalışma alanı olarak kabul edilmiş ve Türkiye'de olmasa da yurtdışındaki kurumlarda bölümler oluşturulmuştur, ancak bir netlik yoktur.
2. Biyomedikal veri bilimi eğitim programları bazen farklı finansmanlar yardımıyla oluşturulmuştur, ancak bu programların temel yeterlilikleri hakkında bir tartışmaya ihtiyaç vardır (örneğin hemen hemen tüm programların olasılık ve istatistik dersleri varken, çok azında tekrarlanabilirlik kursları bulunmaktadır).
3. Veri bilimi biyomedikal araştırmancının ayrılmaz bir parçası olarak görülmüştür, bu nedenle bir sonraki adım en iyi uygulamaları belirlemek ve benimsemek olmalıdır.



Şekil 12.5 Büyük veri ile Biyomedikal İllerlemeler. (Kaynak: <https://www.semanticscholar.org/paper/Surve%3A-Big-Data-Application-in-Biomedical-Research-Bachiller-Busch/4e-782caa179ba92d3f53a2841d2aa27f72b2113b>)

4. Biyomedikal bilimciler arasında veri bilimi eğitimi olan talep artmaya devam etmektedir. Dolayısıyla daha kitleSEL açık çevrimiçi kurslar (MOOC'lar) ve kısa kurslar eğitim programlarına entegre edilmelidir.
5. Bu alanda veri biliminde liderlik ve entegrasyona ihtiyaç vardır.

Bu alanda işe alımlarda ortalama yaşın kırk iki olduğu görülmüştür. Bu nedenle yeni nesil araştırmacılar için eğitim programlarının yüksek oranda aktarılabilir beceriler içermesini sağlamak çok önemlidir. Ancak aynı zamanda akademide kalmak isteyenler bunu gerçekleştirilebilecektir ve bu durum akademik kariyere giriş yaşını

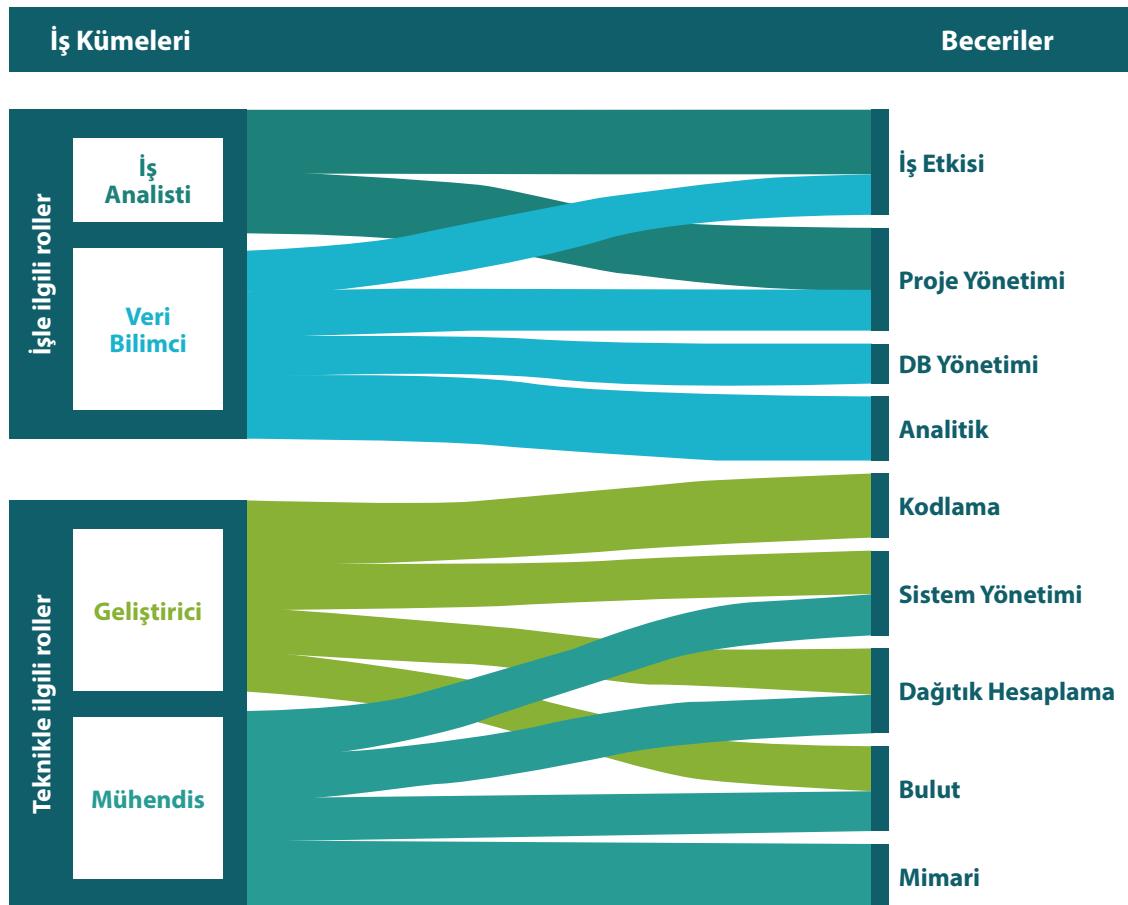
Bir veri bilimcinin disiplinler arası iletişim kurma ve sorunları çözme becerisinin önemi artmaktadır.

düşürmeye yardımcı olabilir. Ayrıca biyomedikal veri biliminde biyoenformatik ve biyoistatistik müfredatın uyumlu olup olmadığı da başka bir konu olarak karşımıza çıkmaktadır. Öğrencilerin birey olarak kendileri için en iyi olanı seçebilmeleri için programların her zaman çeşitliliği olması gerekmektedir. Biyomedikal programlarda veri bilimi dersleri ekip arkadaşlarıyla iletişim kurmak için bir dil oluşturursa da uzmanlık için geniş bir alan sağlamadığı açıkça görülmektedir.

12.3.2 Endüstri

Veri bilimi alanındaki öğrenciler sağlam temellere sahip olmak, iş zekası geliştirmek, verilerin nüanslarını anlamak ve deneyleri inceleyebilmek istemektedir. Günümüzde dahi veri biliminin net bir tanımı bulunmamaktadır, dolayısıyla “veri bilimci” onlarca yıl sürecek bir iş kategorisi olacakmış gibi görülmektedir. Nihayetinde öğrenciler olasılık, istatistik, algoritmalar, doğrusal cebir ve makine öğrenimi hakkında güçlü bir temel anlayışa ihtiyaç duymaktalar ve iş gücünde uzun vadeli başarı elde etmek





Şekil 12.6 Büyük veri uzmanlığında insan kaynağı. (Kaynak: De Mauro, A., Greco M., Grimaldi, M., 2017)

icin daha iyi eleştirel bilimsel düşünme ve problem çözme becerilerine gereksinimleri bulunmaktadır. Öğrencilerin veri ve algoritmalar hakkındaki bilgilerini entegre etmeden önce bir iş problemini nasıl çerceveleyeceklerini öğrenmeleri ve bir argüman oluşturmak için verileri nasıl kullanacaklarını öğrenerek bilmeleri gereklidir. Öğrencilerin ayrıca verilerdeki önyargıyı anlamaları, deneysel sonuçları sorgulamaları ve bunları nasıl çalıştırılacakları yerine hangi araçların işe yaradığını bilmeleri gereklidir. Öğrencilerin daha iyi bir iş zekası oluşturmak için stajlardan ve mentorluklardan yararlanmaları oldukça önemlidir. Öte yandan insanlar herkese açık veri havuzları ve analitiğin yanı sıra, veri bilimi programlarını karşılaştırma ve sıralama yollarını talep etmektedir.

Veri bilimi alanındaki öğrenciler sağlam temellere sahip olmak, iş zekası geliştirmek, verilerin nüanslarını anlamak ve deneyleri inceleyebilmek istemekteler.

Yakın gelecekte yeni işe alınan insanların veri okuryazarı olmasının tercih edileceği öngörülmektedir. Farklı becerileri hedefleyen bireysel kurumların önemi bu doğrultuda daha da artmaktadır. Tüm okulların aynı programları sunması kurumların bu insanları işe alma sürecinde çok fayda sağlayacaktır. Endüstrilerin öğrencilere programlamadan tek önemli beceri olduğu mesajını göndermeye devam etmesi ise gelecekteki iş gücüne dayatılan bir örnek olarak karşımıza çıkmaktadır. Örneğin programlama Amazon'daki grupları işe almak için önemli olmasına rağmen diğer birçok beceri ve nitelik de kendi içinde değer barındırmaktadır.

Bunun yanı sıra veri bilimcileri veri mühendislerinden farklıdır. Kariyer yolunda ilerlerken öğrencilerin bu yolları tartışması oldukça önemlidir. Herkesin veri okuryazarı olması gerekebilir ama herkesin veri bilimi programlarında eğitim alması ilerde bir değer yaratmayabilir. Bu doğrultuda öğrencilerin daha fazla ilgisini ve meraklılığını sağlamak üzere veri biliminin tüm yaşam döngüsünü giriş niteliğindeki bir üniversite kursunda tanıtım mantıklı olabilir.

Veri bilimi programlarının sıralanması konusunda da dikkatli olmak gerekmektedir. Veri alanında iş gücüne ihtiyaç duyan kuruluşlar, adayın hazırlık düzeyini belirlemeye yardımcı olmak üzere görüşmeden önce adayların kurumsal teklifleri hakkında araştırma yapmasını önerebilir.

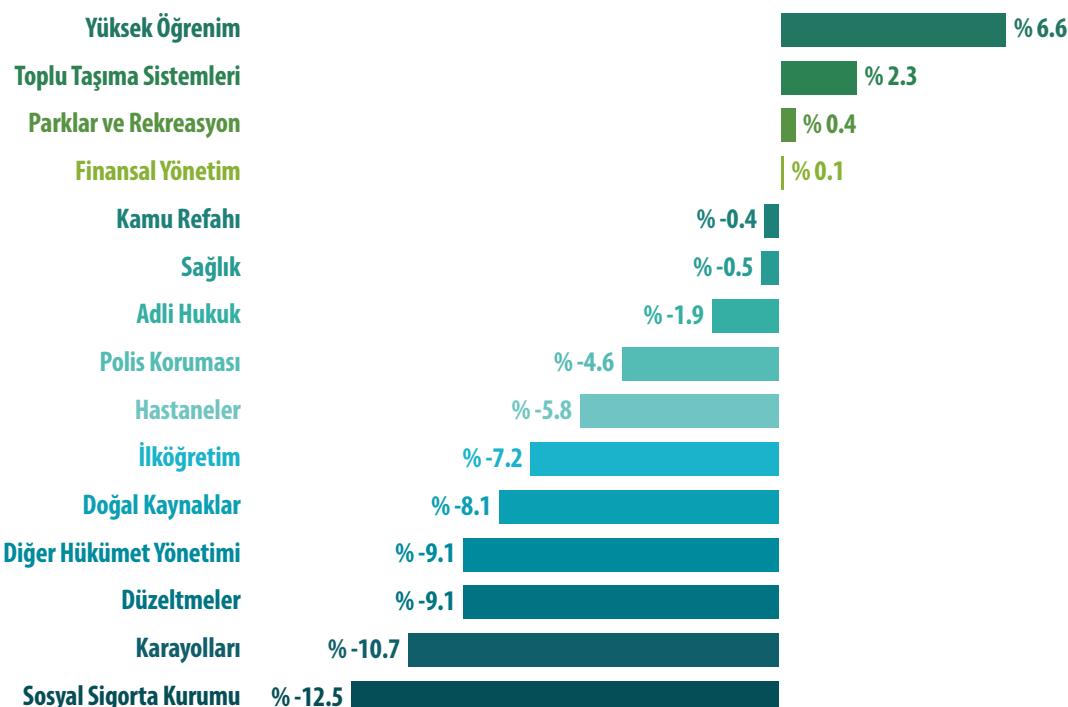
“İş zekasının” “veri zekası”ndan farklı olup olmadığı da bir başka sorunsal olarak kendisini göstermektedir. “İş zekası” üniversitelerin ötesine geçmektedir, çünkü gerçek sorunların çözülmesiyle ilgilidir. Alanda uzman kişiler ise genelde yöntemlere değil, sorumlara odaklanan verinin yoğun olduğu kurslar vermektedir. Bir yüksek lisans veya doktora programında geliştirilen edinme becerileri bu anlamda genellikle daha cazip görünülmektedir fakat kurumların lisans derecesine sahip kişileri işe aldığı görülmektedir.

12.3.3 Hükümet

John Abowd, hükümet bazında çalışmak için öğrencilerin dört beceri geliştirmeleri gerektiğini öne sürmüştür: Tasarlanmış veri metodolojisi, istatistiksel/makine öğrenimi, hiyerarşik modelleme ve kürasyon ve tekrarlanabilirliktir. Tasarlanan veriler anket verileriyle aynı değildir ve bir istatistik kurumunun sahip olduğu her şeyin bir tasarıma sahip olması gerekmesine rağmen verilerin bir anketten olması gerekmektedir. Ayrıca çıkarımın sadece bir öngörü olmadığını da belirtmiştir.

Günümüzde veri analistlerinin genişletilmiş yetkinliklere sahip olma ihtiyacı bulunmaktadır. Lisansüstü ve doktora seviyelerinde, ekonomi veya biyoistatistik gibi bir içerik alanında yoğun bir öğretimden geçmenin gerekliliği ortaya çıkmaktadır. Devlet kurumlarında gerekli olan bilgi işlem kapasitesindeki önemli artışı yönetmek

ABD Eyalet Hükümeti İstihdamı: 2009-2016



Rakamlar, tam zamanlı eşdeğer çalışanlardaki toplu değişiklikleri temsil eder.

Şekil 12.7 ABD Eyalet Hükümeti İstihdamı: 2009-2016. (Kaynak: <https://www.govtech.com/biz/data/Census-Data-Reveals-Where-State-Government-Employment-is-Shrinking.html>)

Bilgisayar bilimi ile istatistiği birleştirmek ve buna veri bilimi adını vermek cazip ve faydalı gibi görünse de birtakım tehlikeleri de bulunmaktadır.

zor olabilir. Veri bilimcileri, hem veri yönetimi hem de altyapı konusunda yardımcı olabilir. Bunun yanı sıra problem çözme çalışanlarda istenen en önemli becerilerden biri olarak kendini göstermektedir.

Bu bilgiler doğrultusunda veri biliminin tanımı ve bir disiplin oluşturup oluşturmadığı hâlâ tam olarak resmileştirilmemiştir. Tanımın esnek tutulması ise bu alanda eğitim görecek ve iş gücü sağlayacak bireyler açısından avantajlı olabilir. Ayrıca veri biliminin evrensel olarak uygulanacak bir dizi araç olduğu da göz önünde bulundurulmalıdır. Bu noktada karşılaşılabilecek zorluklardan biri uzmanlıklarını tanımlamadan veri “okuryazarlığı” oluşturmaktır.

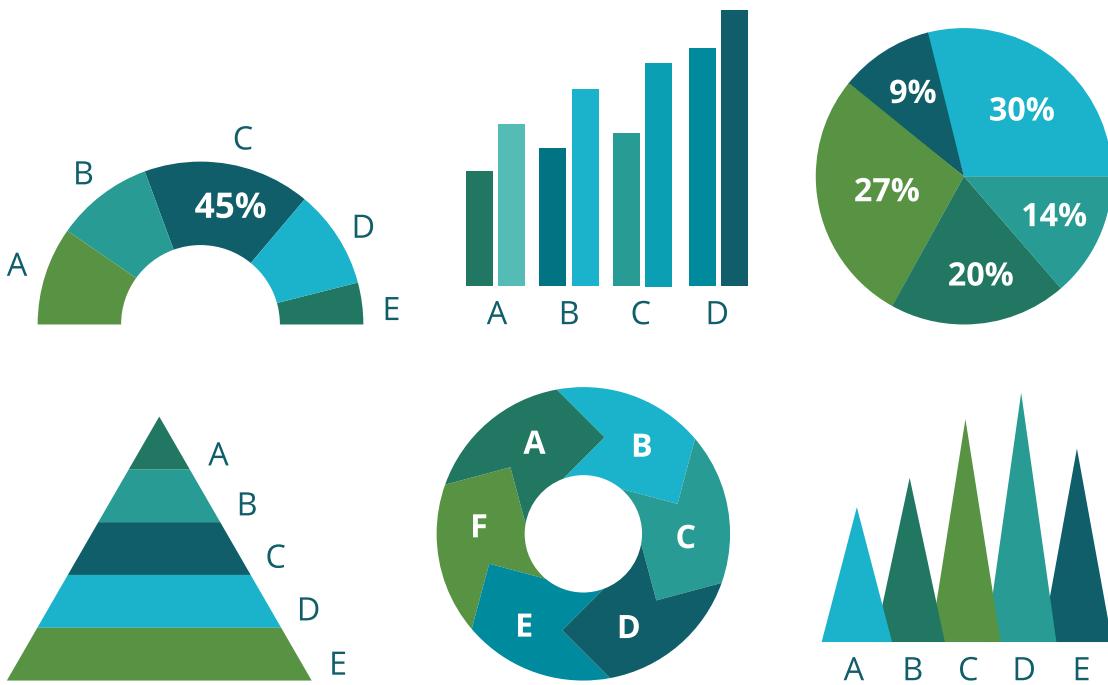
Şu anda tüm mevcut eğitim sistemlerinde veri bilimini öğretmenin yolları tartışılmasına devam etmektedir. Örneğin bilgisayar bilimi ile istatistiği birleştirmek ve buna veri bilimi adını vermek cazip ve faydalı gibi görünse de birtakım tehlikeleri de bulunmaktadır.

12.4 ALAN UZMANLIĞI VE VERİ BİLİMİNİN KESİŞİMİNİN İNCELENMESİ

Veri bilimi eğitim programları, finansman kurumları, profesyonel topluluklar, vakıflar ve endüstriden paydaşların verinin yoğun olduğu alanlarda ortaya çıkan ihtiyaçlar ve fırsatlara ilişkin farlılıklar söz konusudur.

12.4.1 Dil

Bir veri bilimini bir lisans dil müfredatına entegre etmek için hem pedagojik fırsatlar hem de zorluklar olduğu öne sürülmektedir. Fırsatlar kurgusal metinlerde dünyanın farklı bölgelerinin tanımlarının zaman içinde nasıl ve neden değiştiği gibi, edebiyat



eserlerindeki önemli kültürel kalıplar hakkında cevaplanmamış araştırma sorularını keşfetme becerisini içerir. Edebi veriler bol olduğundan ve yeniden üretilmesi nispeten kolay olduğundan, veri bilimi yöntemlerini ve araçlarını müfredata dahil etmek böyle bir soruyu yanıtlamak için güvenilir bir yol sunabilir. Modelleme teknikleri türde veya kitap satışları ile içerik arasındaki ilişkiye dair daha derin bir anlayış geliştirmek için bile kullanılabilir. Ayrıca denetimli sınıflandırma algoritmaları romanlardaki karakterleri sınıflandırmak, cinsiyet temsillerinin zaman içinde nasıl değiştiğine ilişkin soruları ele almak ve yardımcı olmak için kullanılabilir. Bu konuda çalışan insanlar diğer taraftan geçmiş yüzyıllarda kurgudaki sözcüksel eğilimleri daha kolay ve doğru bir şekilde tanımlamak için veri biliminden faydalananabilir.

Bununla birlikte lisans derecesinde dil branşlarının nicel dersler çok nadirdir ve buradaki çoğu alanda veri bilimi yöntemlerini disiplinler arasında uygulamanın değeri anlaşılmaz. Her ne kadar dünyada dijital beşeri bilimler dersleri bazı üniversitelerde ortaya çıksa da genellikle hesaplama yöntemlerine ve nicel akıl yürütmeye göre öncelik vermektedir. O zaman bile, çoğu dil bölümü, nicel temellere fazla dikkat etmeden tek bir ders sunan, yalnızca bir “dijital” eğitmen tutmaktadır. Sonuç olarak alandaki pek çok yeni araştırmacı kendi kendine öğrenmekte ve mevcut öğretim üyelerinin çoğunun bu tür içeriği müfredata dahil etmek için cesaretini kırabilir.

Underwood'a göre, beşeri bilimler öğrencileri genellikle hesaplamayla başladığını ve ardından istatistiğe geçtiğini gözlemlemiştir, bu da öğrencilerin sonuçları nasıl yorumlayacaklarını anlamalarını zorlaştırabilir. Bu istatistik eğitimi eksikliği ise yüksek boyutlu verileri yorumlamayı özellikle zorlaştırır. Dolayısıyla yeniden tasarlanmış müfredat ve daha erişilebilir kurslarla pedagojik bir kanal oluşturmak için yardıma ihtiyaç vardır. Google'dan Peter Norvig, İngilizce departmanlarının veri odaklı problemleri çözmek için bilgi bilimleri bölümünden gelen öğrencilere daha fazla güvendiğini öne sürmektedir. Öğrencilerin disiplinler arası veri bilimi problemleri üzerinde işbirliği içinde çalışmasına izin verecek bir ara yol bulunabilir. Diğer taraftan istatistikçilerin dilbilim ve anlamlı soruların formülasyonundaki becerilerini geliştirmeleri gerekmektedir. Ancak edebiyat öğrencilerinin tür ve tarih hakkında önemli içgörüler ve uzmanlıkları olduğundan, veri bilimi disiplinlerinin beşeri bilimler madenciliği yapmaktansa kendi pedagojik hatlarını geliştirmeleri de bir olasılık olarak görülmektedir. Fakat yine de bu alanlarda yapılan işbirliği sonucunda ortaya çıkan becerilerle beraber, iş gücü olarak talep görüp görmeyecekleri bir sorunsal olarak kendini göstermektedir.

İşverenler yeni işe alınanların iyi yazabilmelerini, net hikâyeler anlatabilmelerini ve verilerin sosyal bileşenlerini keşfetmelerini isterken, birçok öğrenci henüz yeni dersler bulmaya teşvik edilmez. Bunun yanı sıra değişimin ön saflarında dil bölümünün de yer alacağı konusunda iyimser bir görüş birliği de mevcuttur.

12.4.2 Astronomi

Üstün çıkarımsal yetenekler için rekabeti yoğunlaştırarak verilere, bilgi işlem gücüne ve gelişmekte olan teknolojilerle metodolojilere artan erişilebilirliğe güvenilmektedir. En başarılı araştırmacı iyi soru sormayı bilen ve bu sorulara meslektaşlarından daha iyi ve daha hızlı cevap veren kişi olacaktır. Bu başarı sırayla hesaplamalı erişime, çıkarım yöntemlerine, bir anlatının yaratılmasına, yayılmasına ve tekrarlanabilir olmasına dayanmaktadır. Bu rekabetçi ortamın veri bilimi eğitimi ile ilgili müfredat değişiklikleri ihtiyacını ve alan uzmanları ve metodologlar arasında artan işbirliğini güçlendirileceği öngörülmektedir. Bu tür ortaklıklar yüksek etkisi olan keşifleri ve icatları da beraberinde getirecektir.

Veri Aşaması	Astronomi	Twitter	YouTube	Genomik
Edinme	25 zetta bayt/yıl	0,5-15 milyar tweet/yıl	500-900 milyon saat/yıl	1 zetta baz/yıl
Depolama	1 EB/yıl	1-17 PB/yıl	1-2 EB/yıl	2-40 EB/yıl
Analiz	Yerinde veri azaltma	Konu ve duyarlılık madenciliği	Sınırlı gereksinimler	Heterojen veriler ve analiz
	Gerçek zamanlı işleme	Metaveri analizi		Varyant çağrıma, ~2 trilyon merkezi işlem birimi (CPU) saatı
	Büyük hacimler			Tüm çiftler genom hizalaması, ~10.000 trilyon CPU saatı
Dağıtım	Antenden sunucuya özel hatlar (600 TB/sn)	Küçük dağıtım birimleri	Modem kullanıcılarının 4 s bant genişliğinin ana bileşeni (10MB/s)	Birçok küçük (10 MB / s) ve daha az büyük (10 TB/s) veri hareketi

Şekil 12.8 Dört alanın her birinde öngörülen yıllık depolama ve bilgi işlem ihtiyaçları veri yaşam döngüsü boyunca sunuluyor. (Kaynak: <https://journals.plos.org/plosbiology/article/figure?id=10.1371/journal.pbio.1002195.t001>)

Fizik bilimi alanlarındaki öğrencilerin yeni çıkarımlar ve keşifler yapmak için yeni araçlar kullanmak üzere eğitilmeleri gerektiği kabul edilmektedir. Ancak gerçek uzmanlığı geliştirmek için veri odaklı alan eğitiminde (Şekil 12.9) kapsanacak çok fazla içerik mevcuttur. Müfredatı öğrencilere en iyi şekilde hizmet edecek şekilde revize etmek için daha fazla tartışmaya ihtiyaç vardır. Örneğin bilgisayar bilimi yöntemleri üniversite yerine lisede tanıtılması ve öğrencilere verilmesi bu alanda avantaj sağlayabilir.

Tüm bunların yanı sıra yeni yöntemlerin ve araçların sürekli olarak piyasaya sürülmesi insanların hem alan bilgisi hem de metodolojik becerilerde uzmanlık için gereken eğitime ayak uydurmasını giderek zorlaştırmaktadır. 21. yüzyıl eğitimine başarılı bir yaklaşım bir kişiyi derin alan bilgisi veya metodolojik beceriler geliştirecek şekilde eğitmeyi içerebilir. İnsanlar daha sonra çok becerili takımlarda işbirliği yapmaya teşvik edilebilir. Katılımı ve keşfi teşvik etmek için bu tür disiplinler arası ekiplerde tüm taraflar için yenilik ve ödüllerin olması çok önemlidir.



Şekil 12.9 Lise sonrası öğrencilerin bilgisayar bilimleri istatistiklerini ve programlama becerilerini geliştirirken astronomide alan uzmanı olmalarını beklemek bir zorluktur. (Kaynak: Joshua Bloom, University of California, Berkeley, sunum)

12.4.3 Tarih

Sosyal bilimler fizik bilimlerinden farklı şekilde yapılandırılmıştır ve bu fark veri bilimi konularının öğretilme şeklini etkileyebilir. Sosyal bilimlerde kitaplar tipik olarak tercih edilen araştırma ürünleridir, öğretim yükleri genellikle daha ağırdir, dersler genellikle laboratuvarlar yerine seminer ortamlarında ya da amfİlerde verilir. Bu yüzden özellikle sosyal bilimciler ve veri bilimi yöntemlerinde uzmanlaşmış bireyler arasındaki işbirliği fırsatlarını sınırlar ve bu da sonuçta büyük ölçekli problemler üzerinde etkili araştırmaların üretilmesini engeller. Bu tür bir işbirliği büyük miktarda tarihsel verinin genellikle eksik arşivlendiği ve önceden kullanılan nitel yöntemlerin

Kurumların öğrencilerin hem bir alanda hem de veri bilimi metodolojilerinde uzman olmalarını talep etmek yerine disiplinler arasında işbirliğini teşvik etmesi kısa vadede büyük önem arz etmekle birlikte gereklilik de göstermektedir.

belirli çağdaş araştırma sorularını ele almak için uygun olmayabileceği düşünüldüğünde özellikle yararlı olacaktır.

Tarihçiler, tarihsel olayları keşfetmek ve daha iyi anlamak için olay veri setlerinden konu modelleri oluşturabilirler. Tarihçiler, metinlerdeki hükümet politikasını etkileyebilecek kalıpları veya anormallikleri belirlemek için makine öğrenimi yaklaşımlarına güvenebilirler. Bu tür çalışmalarda yer alan veri yoğun olduğundan, doktora seviyesindeki sosyal bilimlerdeki öğrenciler, gerekli analitik beceriler ve hesaplama yöntemlerinde uzmanlaşmak için bir ila iki yıllık ek bir eğitime ihtiyaç duyabilir. Dünyada hesaplamalı sosyal bilimlerin yeni bir alt alanı yavaş yavaş ortaya çıkmakta ancak “hepsini yapabilecek” nispeten az sayıda insan bulunmaktadır. Alan uzmanları ve metodologlar arasındaki gerçek işbirliği, verilerden anlamlı sonuçlar elde etmenin en iyi yolu olabilir.

12.4.4 İşbirliği ve İletişim

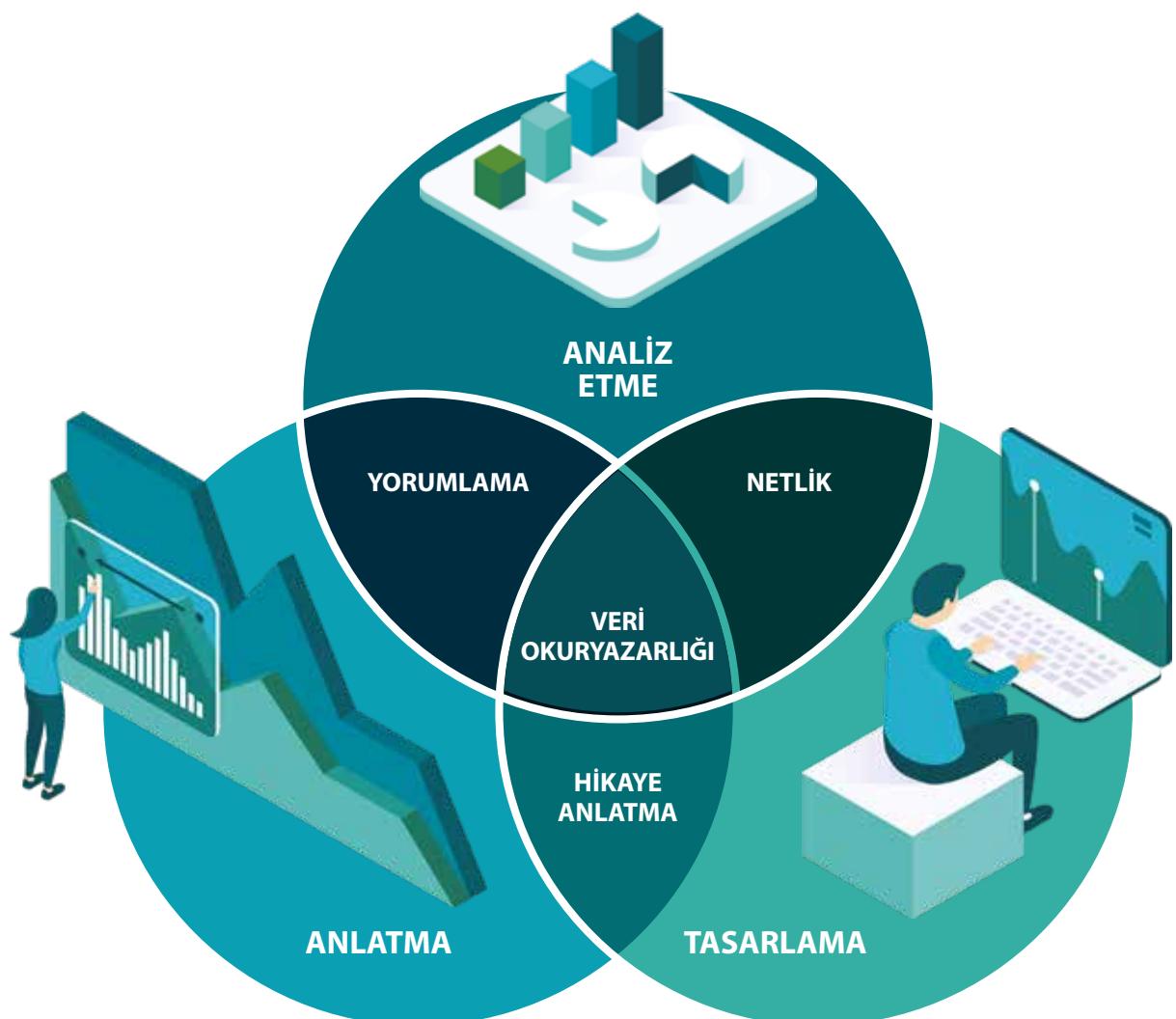
Kurumların öğrencilerin hem bir alanda hem de veri bilimi metodolojilerinde uzman olmalarını talep etmek yerine disiplinler arasında işbirliğini teşvik etmesi kısa vadede büyük önem arz etmekle birlikte gereklilik de göstermektedir. Bununla beraber öğrenciler bulabildikleri her yerde eğitim arayacağı için kurumların öğrencilerin doğru becerileri edinmelerini kolaylaştırmak adına çaba göstermeleri gerekmektedir.

Yapılacak işbirliğinin gerçek bir değer olması gereklidir. Öğretim üyelerinin etkili işbirliğini ve iletişimini, öğretilmesi ve geliştirilmesi gereken açık beceriler olarak görmeleri oldukça

önemlidir. Bu doğrultuda işbirliğinin öğretilmesi gereken bir beceri olduğu konusuna öncelik verilmelidir. Kurumlar sahip olmayı diledikleri yetenek yerine sahip oldukları yetenek için müfredat ve fırsat oluşturmaya yönelebilirler.

12.4.5 Veri Okuryazarlığı ve Ders Tasarımı

Son zamanlarda veri biliminde yüksek lisans öğrencilerinden doktora için artan bir talep gözlemlenmektedir. Son yirmi yılda üniversite kampüslerinde veri biliminin



Şekil 12.10 Veri okuryazarlığı. (Kaynak: <https://medium.com/@isilguler8/veri-okuryazar%C4%B1%C4%9F%C4%B1-data-literacy-ad539055d804>)

popülerliğinin başlıca nedeninin büyük, gürültülü verilerin mevcudiyeti olduğu öne sürülmüştür. Örneklemeye ve tasarım süreçlerinin son yirmi yılda değişmesiyle de veri biliminin cazibesi daha da arṭmıştır.

Bu noktada Ullman henüz bir ana dal seçmemiş olan ve alana daha geniş bir girişten fayda sağlayacak birinci sınıf öğrencilerine belirli bir veri bilimi programını kısaca vermek konusunda endişeliidir. Fox ise nicel olarak düşünen öğrencilere veri bilimi yöntemlerini öğretmeye çalışırken zaten var olan birçok zorluğa dikkat çeker, aynı teknikleri niceliksel becerileri olmayan öğrencilere öğretmeye çalışırken ortaya çıkacak zorluklardan bahsetmez. Öğrencilerin araçların nasıl kullanılacağını değil, ne işe yaradığını anlamalarını sağlayanın önemini pekiştirir. Howe, potansiyel olarak başarılı bir müfredat modeli olarak alanlar tarafından sunulan belirli konuların hafif bir organizasyonunu önermiştir. Choudhary, genel eğitim müfredatının yeniden değerlendirilmesini öne sürmüştür: Veri biliminin temel kavramlarıyla yeni, ayrı dersler oluşturmak yerine genel matematik ve fen derslerine entegre edilebilir mi? Bloom ise öğrencileri muhtemelen üniversite branşlarıyla ilgisi olmayan kariyerlere hazır olmaları için eğitmenin önemini kabul etmektedir. Eğitimin yalnızca mesleki anlamda olmaması tavsiyesinde bulunur; bunun yerine temel kavramların da vurgulanması gereğinin altını çizer. Bu durumda veri okuryazarlığı, basit bilim okuryazarlığından daha verimli bir hedef olabilir. Facebook Yapay Zeka Araştırması'ndan Mark Tygert, katılımcılara endüstrinin hangi becerilere veya yeteneklere ihtiyaç duyduğunu düşünmelerini hatırlatır, örneğin veri biliminin yüzde 95'i veri tartışması gerektirdiğinden, bu öğrencilerin resmi eğitime ihtiyaç duyduğu bir alandır. Bir web yayını katılımcısının beşeri bilimler öğrencileri için veri bilimine geçiş ile ilgili bir sorusuna yanıt olarak Underwood, veri görselleştirmenin ve böyle bir yaklaşımın sonuçlarını iletme becerisinin beşeri bilimler öğrencilerinin geliştirilmesi için önemli bir beceri olduğunu belirtir.

12.4.6 Vaka Araştırması

Washington Üniversitesi

Washington Üniversitesi eScience Enstitüsü on yıldan uzun bir süre önce, entelektüel altyapı tarafından sağlanan veri yoğun bilimin sonunda endüstri ve akademide yay-

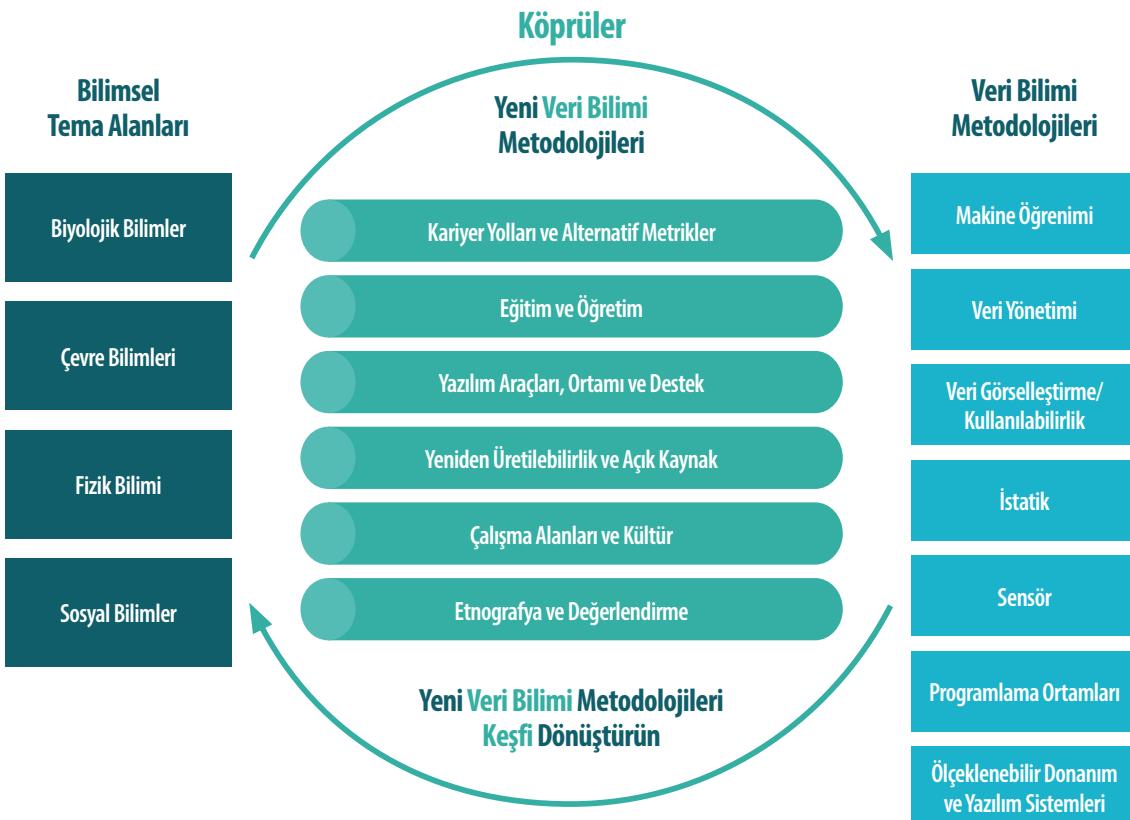
gın olacağı fikrine dayanarak kurulmuştur. eScience Enstitü'nün misyonu, bilimsel tema alanları ile veri bilimi metodolojileri arasındaki boşluğu doldurmak için çalışma grupları kuran bir veri bilimi döngüsü geliştirmektir.

Washington Üniversitesi aşağıdakiler de dahil olmak üzere çeşitli resmi veri bilimi eğitim programlarına sahiptir:

- Veri biliminde profesyonel bir sertifika
- Coursera ile bir veri bilimi kapsamlı açık çevrimiçi kurs (MOOC),
- Lisans ve lisansüstü öğrenciler için Bilgi Okulu veri bilimi dizisi,
- Doktora ile geliştirilmiş bir veri bilimi uzmanlığı,
- Bir lisans veri bilimi uzmanlığı ve
- Disiplinler arası bir veri bilimi yüksek lisans derecesi.

MOOC'un asıl hedefleri, öğrencilerin veri bilimine olan ilgisini gerçek sorunlara yönelik yararlanmak; geniş ölçekte eğitim vermeyi güçlendirmek; birden fazla dersi tek bir giriş dersinde yoğunlaştırmak ve daha geniş veri bilimi tartışmasında veritabanı kavramlarının önemini vurgulamak. Bu 8 haftalık kurs, veri bilimi ortamı, geniş ölçekte veri işleme, analitik, görselleştirme ve özel uygulamalarla ilgili talimatları içerir. En büyük yoğunluk üniversitenin çekmeyi umduğu lisans öğrencilerinden ziyade, profesyonel yazılım mühendisleri olmasına rağmen kursa yaklaşık 9.000 öğrenci kaydolmuştur. Bu deneysel MOOC uygulamasından lisans veri bilimi müfredatı için altı tema ortaya çıkmıştır: Programlama, veri yönetimi, istatistik, makine öğrenimi, görselleştirme ve veri biliminin toplumsal etkileri. Her alan bu temalara farklı şekillerde yaklaşısa da hepsi bu gereksinimleri karşılama kapasitesine sahiptir.

Howe, Washington Üniversitesi'nin tüm birinci sınıf öğrencilerine iki büyük ölçekli ders de sunmuştur: (1) Veri Bilimi Yöntemlerine Giriş ve (2) Veri Bilimi ve Toplumu. Aynı zamanda öğrenme modülleri geliştirmeyi, danışmanlık desteğini artırmayı ve bu kurslar için bir konu incelemeye sürecini başlatmayı planlamaktadır. Nihayetinde üniversite öğrencilere ikna edici argümanlar oluşturmayı ve büyük, gürültülü, heterojen veri kümelerini kullanmayı öğrenmeyi-öğretmeyi umuyor. Öğrencilerin her iki dersin sonunda da uzman olmaları beklenmese de gerçek sorunlar üzerinde çalışarak bu beceriyi geliştirmesi beklenmektedir.



Şekil 12.11 eScience Enstitüsü'nün veri bilimi döngüsü, alan bilimiyle ilgili araştırmaları metodolojik gelişmelere bağlamak için bir araç olarak eğitim ve öğretimi içerir. (Kaynak: Washington Üniversitesi'nden Ed Lazowska ve Moore/Sloan Veri Bilimi Ortamları tarafından üretildi ve Washington Üniversitesi'nden Bill Howe tarafından sunulmuştur.)

Fox ise üniversitede sunulan veri bilimi derslerinin birçok farklı versiyonu olduğundan ders sıralamanın sorunlu hale gelebileceğini belirtmiştir. Howe, öğrencilerin daha ileri düzey dersler için uygun ön koşulları sağlamaları yönünde bu sorunu çözmeye çalışmak için departmanlar arası çalışma grupları bulduğunu belirtmiştir. Stoddern, üniversitenin 5 yıllık planının yanı sıra diğer kurumların programlarından neler öğrenebileceğinin önemini ifade etmiştir. Howe, üniversitenin iş gücü eğitimi ve kurs konusunun iyileştirilmesi hakkında daha fazla çalışılması gerekiğine inanmaktadır. Howe, kurumlarla ilişkilerde üniversitenin genel olarak açık ve işbirliğine dayalı olduğunu ve bu alanda yeniliği desteklediğini kabul etmektedir. Bununla birlikte süreçleri düzene koymak ve bir eğitim çalışma grubu oluşturmak faydalı olabilir.

Öğrencilerin takım tarafından verilen derslere olan ilgisinin güçlü olduğu belirtilmiştir.

Kolombiya Üniversitesi

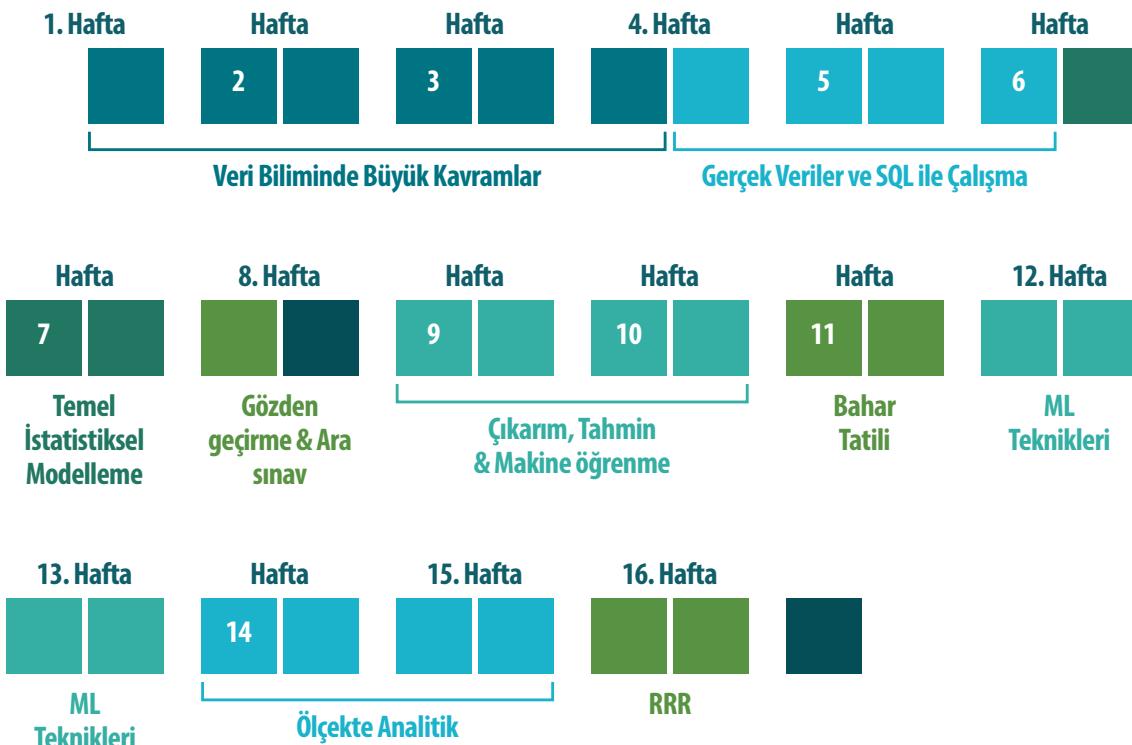
McKeown, 1 yıl önce Columbia Üniversitesi’nde dekanlardan oluşan bir görev gücünün ve bir veri bilimi direktörlüğünün, ekip tarafından öğretilen eğitimin gelişimi engelleyen kurumsal engellerin (ör. Okullar arasındaki öğretim ve fakülte yükü gereksinimleri) nasıl aşılacağını tartışmak için nasıl bir araya geldiğini anlatmıştır. Sonuç olarak veri bilimcilerinin üniversite içindeki okullarda sınıfları takım halinde öğretmek için disiplin uzmanlarıyla ortaklık kurması için finansman sağlayan Columbia Collaboratory kuruldu. En son finansman turunda, kurs önerileri için 18 talep sunulmuş ve dört kurs kabul edilmiştir.

Bu kurslar, yaklaşımları ve bireysel disiplinlerin özel ihtiyaçları göz önüne alındığında, öğrencilerin ne kadar programlama yapacakları bakımından farklılık gösterir, ancak iletişimın değeri üzerinde ortak bir vurgu vardır. Bu dört kursa ek olarak, veri; geçmiş, bugün ve gelecek gibi ek pilot kurslar Collaboratory tarafından finanse edildi. Bu lisans dersi, bir tarihçi ve uygulamalı bir matematikçi tarafından öğretilir ve gelecek yüzyılda verilerin toplumdaki rolünü vurgulayan bir bilgi çekirdeği içerir. Bu ders, öğrencilere ana dallarına çeşitli ödevler ve uygulamalar sunarak teknik ve hümanist olmak üzere iki parça içerir.

Stodden’den gelen bir soruya yanıt olarak McKeown, öğrencilerin takım tarafından verilen derslere olan ilgisinin güçlü olduğunu, ancak finanse edilen bazı derslerin henüz uygulanmadığını belirtmiştir (2017 sonbaharında başlayacaklar). Underwood, okullarda bu tür bir işbirliğinin devam etmesini sağlayacak bir mekanizma olmadığını sormuş ve McKeown, her okulun dekanlarının Collaboratory ile birkaç yıldır çalışmayı taahhüt ettiğini doğrulamıştır.

California Üniversitesi, Berkeley

Carson, Berkeley’deki California Üniversitesi’nin tüm öğrencilerin veri bilimi eğitime yönelik artan öğrenci talebine ve fakülte uzmanlığındaki artan çeşitliliğe yanıt



Şekil 12.12 Veri bilimi yaşam döngüsünden esinlenen, pilot bir Veri Bilimi 100 dersinin müfredatı. (Kaynak: Berkeley, Kaliforniya Üniversitesi'nden Profesör Joseph Gonzalez tarafından üretildi ve Berkeley, Kaliforniya Üniversitesi'nden Cathryn Carson tarafından sunuldu.)

olarak “verilerle yetkin ve kritik bir şekilde ilgilenmesini” sağlamaya çalıştığını ifade etmiştir. Bu hedefe ulaşma çabası içinde üniversite, eğitim geçmişleri veya çalışma alanları ne olursa olsun, tüm öğrencilere temel bir veri bilimi kursu olan Veri 8 (data8.org) sunmaktadır. Şu anda 60 ana dalda 700 öğrenci bu kursa katılmıştır. Bu temel kurs tarayıcı tabanlı bir hesaplama platformundan (Jupyter Not Defterleri) yararlanır ve öğrenciler toplumsal ve etik bağamlarında gerçek verilerle çalışarak hesaplamalı ve çıkarımsal düşünmeyi öğrenirler. Kaydolmak için herhangi bir önkoşul gerekmez ve kurs bilgisayar bilimi, istatistik ve bilgi bölümlerinde çapraz listelenir. Disiplinlerarası bir fakülte ekibi tarafından verilen bu ders, veri bilimi kavramlarını doğrudan öğrencilerin ilgi alanlarına bağlayan ve çeşitli akademik bölümler tarafından sunulan Connector kursları ile birlikte sunulmaktadır. Bu tür kurslar, üniversitenin öğrencilere daha iyi hizmet veren bütünüleyici ve kapsamlı bir müfredatın geliştirilmesine yaklaşır. Bu kurs önerileri şimdiye kadar üniversitenin işbirliğine dayalı ve yenilikçi

California Üniversitesi, eğitim geçmişleri veya çalışma alanları ne olursa olsun, tüm öğrencilere temel bir veri bilimi kursu olan Veri 8 (data8.org) sunmaktadır. Şu anda 60 ana dalda 700 öğrenci bu kursa katılmıştır.

kültürünün bir sonucu olarak mümkün olmuştur. Üniversitedeki veri bilimi eğitimi felsefesi entelektüel, örgütsel ve sosyal değerlere odaklanır ve “Deneyin, Öğrenin ve Hızlı Ölçeklendirin” sloganına dayanır.

Veri 8’i tamamladıktan sonra bu platformda ilerlemek isteyen öğrenciler için fakülte, aşağıdakileri içeren bir dizi başka yeni kurs geliştirmiştir: Veri Bilimi 100: Veri Biliminin İlkeleri ve Teknikleri (Şekil 12.12), İstatistik 28: İstatistiksel Yöntemler Veri Bilimi ve Stat 140: Veri Bilimi Olasılığı.

California Üniversitesi, Berkeley şu anda fakülte için veri bilimi pedagojisi ve uygulaması hakkında daha fazla bilgi edinmeleri için kısa bir kurs sunmaktadır ve bir dizi kurs modülü doğrudan bu çalışmadan gelmiştir. Veri bilimi eğitimi müfredat geliştirme, sosyal yardım ve çeşitlilik ve program altyapısı üzerinde çalışan bir öğrenci ekibi de bulunmaktadır. Carson'a göre, araştırılmaya devam eden temel bir soru, her alandan öğrencinin ve öğretim üyesinin ihtiyaçlarının toplu olarak nasıl karşılanacağıdır.

McKeown, farklı deneyimleri ve eğitim geçmişleri olan Veri 8 öğrencilerine bilgi sunmanın zorluğunu sormuştur. Veri 8’in sonunda çeşitli formatlarda ve farklı seviyelerde sunulmasının gerekip gerekmemiş merak edilmektedir. Carson, üniversitenin dersi gelen öğrencilerin bilgileri açısından farklı tutmak istediğini, ancak bunun önumüzdeki yıllarda ele alınması gereken bir zorluk olduğunu kabul ettiğini söylemiştir. Şu anda kendini Veri 8’e hazır hissetmeyen öğrencilere hazırlık sunan Summer Bridge adında bir hazırlık yaz programı bulunmaktadır. Kursa tüm katılımcıların erişebilmesi için kurs içi uyarlamalar da mevcuttur. Carson farklı öğrencilerin kursun farklı yönleriyle mücadele ettiğini belirtmiştir; örneğin bazı öğrenciler dersin ilk birkaç

haftasında kodlamayı zor bulmaktadır. Şu anda öğrencilerin kurs içeriğine yönelik kabulleri ölçülmektedir, ancak Carson, öğrencilerin gelecekte kursa olan ilgisini ve başarısını ölçmek için kullanılan analizleri görmek istemektedir.

12.5 SONUÇ

Özet olarak liderliği ve finansmanı çevreleyen kurumsal felsefeler veri bilimi girişimlerini yapma veya bozma potansiyeline sahiptir. Berkeley'deki Kaliforniya Üniversitesi'ndeki hub ve bağlayıcı model beşeri bilimler öğrencileri arasında veri biliminin görünürlüğünü artırmak için çekici bir geçit sağlamaktadır. Bunun yanı sıra veri bilimi derslerine olan talebin artmaya devam etmesi halinde profesör eksikliği gibi bir sorunla karşı karşıya kalınacaktır ve bu kurumlardaki endişeyi büyütmektedir. Diğer taraftan alan temelli derslerin fakülte üzerindeki bu yükü hafifletmesi de olası bir çözüm olarak sunulabilir. Bir üniversite organizasyon yapısının karar verme sürecine başka bir boyut eklemektedir. Örneğin üniversitelerin kendi standartları ve bekłentileri vardır. Dolayısıyla özel bir kurumdan çok farklı konuları olan bir devlet kurumunda “orta çıkış” yaklaşımı “aşağıdan yukarıya” yaklaşımından daha etkili olabilir.

Disiplinler arası ve deneysel derslerin tasarılanmasında profesörleri dahil etmenin değeri de veri bilimi öğretiminde önem kazanmaktadır. Örneğin 2016 yılında Stanford Üniversitesi'nde bilgisayar bilimi, beşeri bilimler ve sosyal bilimler kavramlarını başarılı bir şekilde birbirine bağlayan “Büyük Veri, Büyük Hype, Büyük Yanılgılar” başlıklı birinci sınıf semineri verildi. Stanford Üniversitesi bunu gelecek dönemlerde tüm öğrencilere açık olan düzenli bir kurs olarak sunmayı hedefliyor.

Tüm bunların yanı sıra veri bilimi müfredatının sunduğu zorlukların ele alınması için öğretim fakültesi grupları (iş güvenliği ve mesleki gelişim fırsatları ile birlikte) oluşturmaya yönelik kurumsal planların hızlı bir şekilde izlenmesi gerekmektedir. Geleneksel lisans müfredatını gözden geçirmeden önce zaten mevcut olan birçok çevrimiçi veri bilimi dersini çalışmanın değeri de göz önünde bulundurulmalıdır. Herkese uyan tek bir yaklaşım her ne kadar pratik görünmese de kampüslerde açık ve işbirliğine dayalı bir kültür geliştirilebilir. Bir işbirliği kültürü oluşturmaya çalışırken yerel ortaklıklar kurmanın ve yüz yüze etkileşimleri teşvik etmenin değeri de göz önünde bulundurulmalıdır.

Öğrenciler meslek yaşamına ve geleceğe hazırlanırken kullanıcıların işyerinde hangi araçlara ihtiyaç duyacağını tahmin etmeleri de zor olabilir. Eğitimde ve çeşitli büyük iş güçlerini yönetmede önemli zorluklar olduğu aşikârdır. Ayrıca ticari dünyyanın aksine, devlet kurumları yazılımın dağıtılmrasında bürokratik engellerle karşı karşıyadır. Yeni teknolojilere erişmenin daha verimli yollarını bulmak için eğitim sisteminin hem devlet kurumları hem de politika yapıcılar ile işbirliğini artırması yararlı olabilir.

KAYNAKLAR

- ACM (Association for Computing Machinery). 2018. Lighting the Path from Community College to Computing Careers. <https://www.acm.org/binaries/content/assets/education/lighting-the-path-from-community-college-to-computing-careers.pdf>.
- Angwin, J., J. Larson, S. Mattu, and L. Kirchner. "Machine Bias." ProPublica, May 23. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Baker, M. 2016. "1,500 Scientists Lift the Lid on Reproducibility." News Feature. Nature. May 25. <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>.
- Barone, L., J. Williams, and D. Micklos. 2017. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLoS Computational Biology 13(10):e1005755. <https://doi.org/10.1371/journal.pcbi.1005755>.
- Bloom, B.S. 1956. Taxonomy of Educational Objectives: The Classification of Educational Goals. White Plains, NY: Longman.
- Buckheit, J.B., and D.L. Donoho. "WaveLab and Reproducible Research." https://statweb.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf.
- CCC (Computing Community Consortium). 2015. "The Future of Computing Research: Industry-Academic Collaborations." Volume 2. <https://cra.org/ccc/wp-content/uploads/sites/2/2016/06/15125-CCC-Industry-Whitepaper-v4-1.pdf>.
- CCC. 2019. "Evolving Academia/Industry Relations in Computing Research: Interim Report." <https://www.cccblog.org/wp-content/uploads/2019/03/Industry-Interim-Report-w-footnotes.pdf>.
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth.

2000. "CRISP-DM 1.0: Step-By-Step Data Mining Guide." <https://www.the-modeling-agency.com/crisp-dm.pdf>.
- CMU (Carnegie Mellon University). 2017. "Carnegie Mellon University Educational Project Agreement." <https://www.ri.cmu.edu/wp-content/uploads/2017/01/Educational-Project-Agreement.pdf>.
- Conway, D. 2010. "The Data Science Venn Diagram." <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.
- Cramer, C., M. Porter, H. Sayama, L. Sheetz, and S. Uzzo. 2015. "Network Literacy: Essential Concepts and Core Ideas." <http://tinyurl.com/networkliteracy>.
- Dieterich, W., C. Mendoza, and T. Brennan. 2016. "COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity." Northpointe, Inc., Research Department. <https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html>.
- Dwork, C., F. McSherry, K. Nissim, and A. Smith. "Calibrating Noise to Sensitivity in Private Data Analysis." In Theory of Cryptography (S. Halevi and T. Rabin, eds.). TCC 2006. Lecture Notes in Computer Science 3876. Berlin, Heidelberg: Springer.
- EDC (Education Development Center, Inc.). 2017. "Tools for Building a Big Data Career Pathway." <http://oceansofdata.org/sites/oceansofdata.org/files/Tools%20for%20Building%20a%20Big%20Data%20Career%20Path.pdf>.
- FTC (Federal Trade Commission). 1998. Privacy Online: A Report to Congress. <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>.
- Gould, R., R. Peck, J. Hanson, N. Horton, B. Kotz, K. Kubo, J. Malyn-Smith, M. Rudis, B. Thompson, M.D. Ward, and R. Wong. 2018. The Two-Year College Data Science Summit: A Report on NSF DUE-1735199. <https://www.amstat.org/asa/files/pdfs/2018TYCDSFinal-Report.pdf>.
- https://www.researchgate.net/publication/330281868_Adapting_Engineering_Education_to_Industry_40_Vision
- <https://www.semanticscholar.org/paper/The-relationship-between-Big-Data-and-Mathematical-Vecchia/8a015d40858925a36e39996ef22f-9214b250e1f5>
- <https://www.semanticscholar.org/paper/Survey%3A-Big-Data-Application-in-Bio-medical-Research-Bachiller-Busch/4e782caa179ba92d3f53a2841d2aa27f72b2113b>

- <https://www.govtech.com/biz/data/Census-Data-Reveals-Where-State-Government-Employment-is-Shrinking.html>
- <https://journals.plos.org/plosbiology/article/figure?id=10.1371/journal.pbio.1002195.t001>
- <https://medium.com/@isilguler8/veri-okuryazarl%C4%B1%C4%9F%-C4%B1-data-literacy-ad539055d804>
- Ioannidis, J.P.A. 2005. Why most published research findings are false. PLoS Medicine 2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Levy, R., R. Laugesen, and F. Santosa. 2018. Big Jobs Guide. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Meyer, M., A. Cimpian, and S.J. Leslie. 2015. Women are underrepresented in fields where success is believed to require brilliance. Frontiers in Psychology 6:235.
- Microsoft Azure. 2017. “What Is the Team Data Science Process?” <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>.
- NAE (National Academy of Engineering). 2016. Infusing Ethics into the Development of Engineers. Washington, DC: The National Academies Press.
- NASEM (National Academies of Sciences, Engineering, and Medicine). 2018a. Graduate STEM Education for the 21st Century. Washington, DC: The National Academies Press.
- NASEM. 2018b. Data Science for Undergraduates: Opportunities and Options. Washington, DC: The National Academies Press.
- NRC (National Research Council). 2011. A Framework for K-12 Science Education. Washington, DC: The National Academies Press.
- NSF (National Science Foundation). 2015. Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. May.
- ODI (Oceans of Data Institute). 2014. Profile of a Big-Data-Enabled Specialist. Waltham, MA: Education Development Center, Inc.
- ODI. 2016. Profile of the Data Practitioner. Waltham, MA: Education Development Center, Inc.
- Patil, D.J., H. Mason, and M. Loukides. 2018. Ethics and Data Science. Sebastopol, CA: O'Reilly Media.

- Patterson, D. 2014. How to build a bad research center. Communications of the ACM 57(3):33-36.
- Rawlings-Goss, R., L. Cassel, M. Cragin, C. Cramer, A. Dingle, S. Friday-Stroud, A. Herron, et al. 2018. “Keeping Data Science Broad: Negotiating the Digital and Data Divide.” https://drive.google.com/file/d/14l_PGq4AxOP9fhJbKqA2necsJZ-gdiKV/view.
- Zweben, S., and B. Bizot. 2018. “2017 CRA Taulbee Survey.” <https://cra.org/crn/2018/05/2017-cra-taulbee-survey-another-year-of-record-undergrad-enrollment-doctoral-degree-production-steady-while-masters-production-rises-again/>.

13. BÖLÜM

ABD'DEKİ ÇOK ULUSLU ŞİRKETLERİN KÜRESEL YAPAY ZEKA AR-GE ETKİNLİĞİNİ HARİTALAMALARI



13.1 GİRİŞ

ABD teknoloji şirketlerinin yapay zeka laboratuvarlarının ve personelinin küresel dağıtımına ilişkin yeni veriler sunulmuştur. Her biri en son yapay zeka araştırma ve geliştirme geçmişine sahip altı şirkete (Amazon, Apple, Facebook, Google, IBM ve Microsoft) odaklanılmıştır. Elde edilen bulgular bu şirketlerin yapay zeka Ar-Ge'yi nerede yürüttüğüne ve neden belirli konumları seçiklerine ışık tutmaktadır.

Ayrıca ABD şirketlerinin yurtdışında Ar-Ge çalışması yapmasının ulusal güvenlik açısından yararları ve risklerilarındaki tartışmalar da bulunmaktadır. Yeni veriler kendi başına bu tartışmaları çözmeyecektir, ancak veriler incelikli politika tartışmaları ve maliyet-fayda analizleri için bir önkosuludur.

Özel sektör, yapay zeka ve makine öğreniminde teknik ilerlemenin arkasındaki itici güçtür ve özel sektör yapay zeka Ar-Ge'sinin büyük bir kısmı sınırlı sayıdaki büyük teknoloji şirketi tarafından gerçekleştirilmektedir. Büyük teknoloji şirketlerinden 35 adedinin araştırıldığı bir McKinsey raporunda, 2016 yılında yapay zeka geliştirme için dahili fonlara 18-27 milyar dolar ve yatırımlarla satın almalar için 8-12 milyar dolar harcadığı ortaya konmuştur.

Amerika Birleşik Devletleri’nde 2010’dan büyük teknoloji devleri 60’tan fazla küçük AI şirketini bünyesine katmıştır. The President’s Council of Advisors on Science and Technology, ABD şirketlerinin 2025 yılına kadar AI Ar-Ge’ye yılda 100 milyar dolardan fazla harcama yapacağını öngörmektedir. Buna kıyasla ABD federal hükümeti sınıflandırılmamış yapay zeka ve makine öğrenimi ile ilgili araştırma ve geliştirmeye yaklaşık 5 milyar dolar harcamayı planlamaktadır.

Dünyanın en büyük AI şirketlerinin çoğu Amerika’da bulunmaktadır. Amazon, Apple, Facebook, Google, IBM ve Microsoft gibi büyük devler bu şirketler arasında yer almaktadır. Bu şirketler birlikte Ar-Ge’ye yılda 76 milyar dolardan fazla harcama yapmaktadır ki toplu piyasa değeri 5 trilyon doların üzerindedir. İş modelleri ise kureselleşmektedir. Apple, Facebook, Google ve IBM’in her birinin toplam gelirinin yarısından azını ABD pazarı oluşturmaktadır. Microsoft ve Amazon sırasıyla yüzde 51 ve yüzde 69 oranında gelir sağlamaktadır.

ABD teknoloji şirketlerinin küresel erişimi Amerika Birleşik Devletleri’ne çeşitli avantajlar sağlamanmaktadır. Dünya çapında faaliyet göstermeleri, yalnızca iç pazarlara odaklandıklarında imkânsız sayılabilen çok büyük ve hızda inovasyon yapmalarına ve dünyanın her yerinden yetenek ve fikirleri benimsemelerini sağlamaktadır. Hem ABD ekonomisi hem de ulusal güvenlik bu durumdan faydalananmaktadır. ABD’nin yurtdışına genişleyen çok uluslu şirketlerinin daha fazla yurt içi istihdama, ihracata ve Ar-Ge’ye katkıda bulunduğu düşünülmekte ve “ABD şirketlerinin yurtdışında daha az yatırım yapmasının ABD ekonomisini zayıflatacağı” iddia edilmektedir.

Ancak bu şirketlerin ölçü ve erişimi riskleri de beraberinde getirmektedir. Ulusal güvenlik açısından, örneğin ticari sektörün küresel varlığı iki ucu keskin bir bıçak olarak düşünülebilir. 2018 Ulusal Savunma Stratejisi’nin ifade ettiği gibi, “birçok teknolojik gelişmenin ticari sektörden geleceği gerçeği, devletin rakiplerinin ve devlet dışı aktörlerin de bunlara erişebileceği anlamına geliyor ki bu da uluslararası alışıktı geleneksel olarak geçme-rekabet etme riskini taşıdığı anlamına gelmektedir.” ABD’deki çok uluslu şirketlerin yurtdışındaki Ar-Ge faaliyetleri, örneğin yerel yetenekleri eğiterek veya fikri mülkiyet ya da işbirliğine dayalı araştırmayı içeren ortak girişimler yoluyla diğer ülkelerin teknoloji ekosistemlerini çeşitli şekillerde artırmaktadır. Bir endüstri uzmanı, “Microsoft, Microsoft Research Asia’yı hiç kurmamış olsaydı, bu Çin’in AI ekosisteminin yükselişini önemli ölçüde geciktirirdi,” şeklinde

Dünyanın en büyük AI şirketlerinin çoğu Amerika'da bulunmaktadır ve Amazon, Apple, Facebook, Google, IBM ve Microsoft gibi büyük devler bu şirketler arasında yer almaktadır. Bu şirketler birlikte Ar-Ge'ye yılda 76 milyar dolardan fazla harcamaktadır ki toplu piyasa değeri 5 trilyon doların üzerindedir.

ifade etmiştir. 2019'da Müşterek Kurmay Başkanları'nın o zamanki başkanı, Google'ın Çin'deki çalışmalarının "dolaylı olarak Çin ordusuna fayda sağladığını" belirtmiştir. Microsoft ve Google dahil diğer teknoloji devleri ise bu tür iddialara itiraz etmektedir.

Bu farklı faydalalar ve endişeler, ABD'li politika kurucuların ABD şirketlerinin yurtdışındaki AI faaliyetlerini düzenleyip düzenlemeyecekleri ve nasıl düzenleyeceklerini düşünürken karmaşık durumlarla karşı karşıya kaldıkları anlamına gelmektedir. Bu karmaşık durumu çözmek de yine veri gerektirmektedir. Yeni politikaların sonuçlarını veya statükonun devamını değerlendirmek için politika kurucuların ABD'deki çok uluslu şirketlerin nerede aktif olduğunu, yurtdışında hangi faaliyetlerde bulunduklarını ve nedenlerini bilmeleri gerekecektir. Siyasi yelpazedeki analistler ve karar vericiler uzun zamandır bu ve ilgili sorular hakkında daha iyi verilere ihtiyaç olduğunu savunmaktadır. Bu bağlamda Amazon, Apple, Facebook, Google, IBM ve Microsoft'un küresel AI Ar-Ge faaliyetlerinin haritalandırılması gerekmektedir.

13.2 METODOLOJİ VE VERİ

Bu bölümde AI laboratuvarları ve altı büyük çok uluslu ABD teknoloji şirketinin personeli hakkında yeni veriler sunulmaktadır: Amazon, Apple, Facebook, Google, IBM ve Microsoft. Bu şirketler Ar-Ge harcamaları, AI başlangıç satın almaları, patent

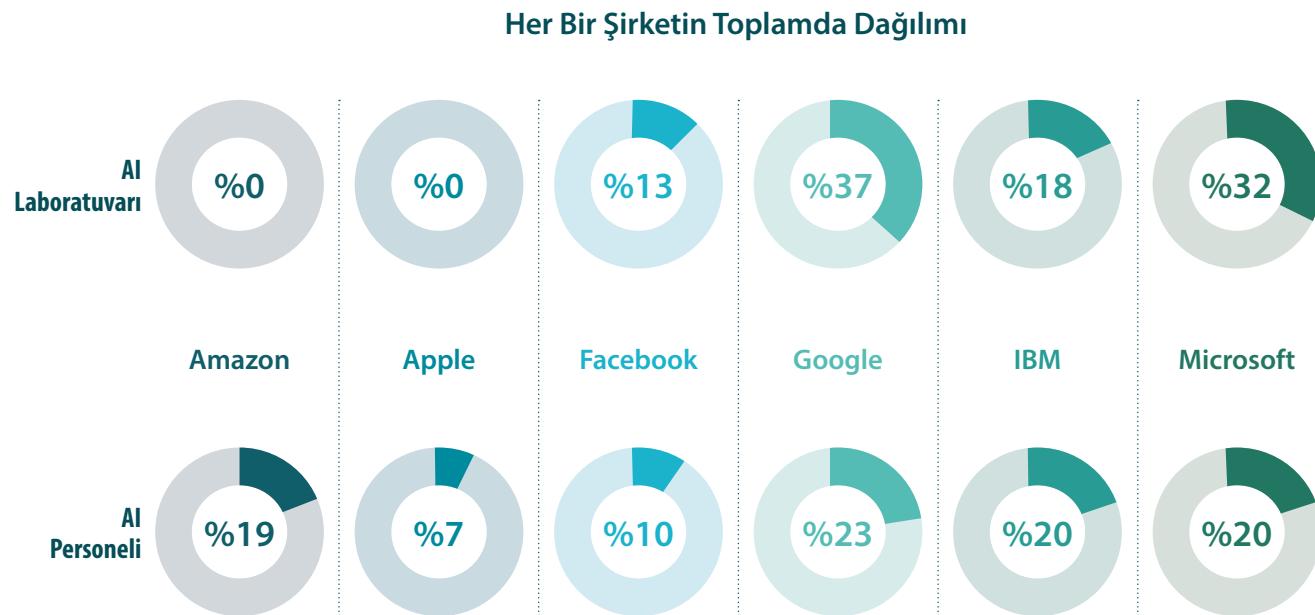
sahiplikleri ve profesyonel AI organizasyonlarına katılım dahil olmak üzere çeşitli AI Ar-Ge ölçümlerinde yüksek puanlarına göre seçilmiştir.

Üzerinde durulan ilk Ar-Ge ölçüsü, şirketlerin yapay zeka Ar-Ge'sinin medya ve endüstri analizlerinde belirgin bir şekilde öne çıkan "laboratuvarları"nı içermektedir. Şirketlerin kendi web siteleri ve haber raporları aranılarak laboratuvar listeleri derlendi ve listelerin doğruluğu şirket temsilcileri aracılığıyla teyit edildi. Amazon ve Apple'da herkese açık AI laboratuvarı bulunamamıştır; Amazon 2014 yılında AI araştırma ekibini ürün odaklı ekiplere entegre etmeye başlamıştır, Apple ise AI Ar-Ge faaliyetlerini kamuya açık bir şekilde gerçekleştirmemektedir. Facebook, Google, IBM ve Microsoft olmak üzere kalan dört şirketin birçoğunda Google tarafından işletilen toplam 62 laboratuvar bulunmuştur (Şekil 13.1).

Bakılan ikinci ölçü personel sayılarıdır. "Yapay zeka personelini" LinkedIn profillerinde "bilgisayar görüşü" (mevcut bir yapay zeka araştırması kategorisi) veya "PyTorch" (belirli bir makine öğrenimi kitabı) gibi bir veya daha fazla "Yapay Zeka becerisi"nin listelendiği herhangi bir şirket çalışanı olarak sayılmaktadır. Bilindiği üzere LinkedIn, yapay zeka becerileri ve küresel kapsama sahip işler hakkında tek veri kaynağıdır. Ancak verileri mükemmel değildir; LinkedIn kullanım oranları ülkeye farklılık göstermekte ve veriler yalnızca ilgili şirket toplamları ve konuma özgü toplamlar olarak mevcuttur.



Şekil 13.1 Şirketlere göre herkese açık AI laboratuvarlarının sayısı. (Kaynak: CSET AI laboratuvar verileri. Amazon ve Apple, bir "laboratuvar" yapısı kullanmadığı veya laboratuvarlarını kamuya açık bir şekilde yapılandırmadıkları için listede yer almamıştır.)



Şekil 13.2 AI laboratuvarlarının ve personelinin şirketlere göre küresel dağılımı. (Kaynak: CSET AI laboratuvar verileri; LinkedIn'den AI personeli verileri.)

Şekil 13.2'de şirketlerdeki yapay zeka personelinin yüzdesi yapay zeka laboratuvarlarının yüzdesiyle karşılaştırılmaktadır ve iki göstergenin göreceli dağılımının şirketler arasında ölçüde farklılık gösterdiği ortaya çıkmaktadır. Örneğin Google tüm laboratuvarların üçte birinden fazlasına sahipken, tüm personelin yalnızca yüzde 23'üne sahiptir. Personel verileri Mayıs 2020 itibarıyla günceldir.

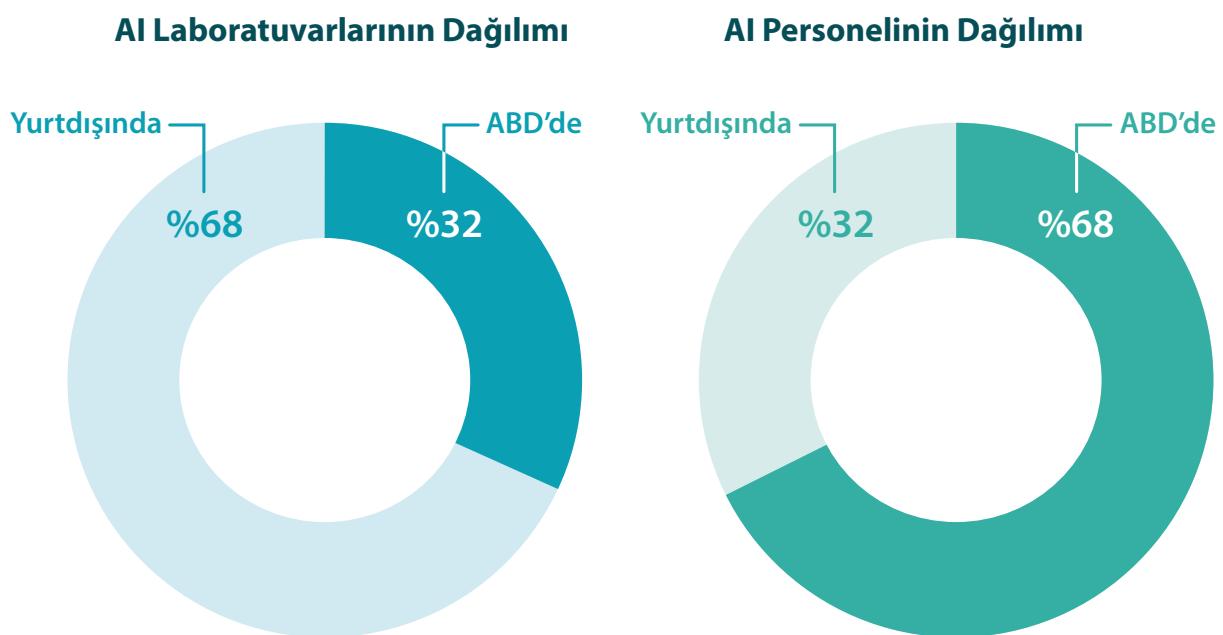
Yapay zeka Ar-Ge faaliyetinin ölçümleri olarak hem laboratuvar hem de personel ölçümlerinin güçlü ve zayıf yönleri mevcuttur. Örneğin bir AI laboratuvarının olması AI Ar-Ge'nin gerçekleştiği anlamına gelmez, bununla birlikte bir laboratuvarın olmaması da Ar-Ge'nin olmadığı anlamına gelmez. Personel sayımlarının anlık görüntü ölçümleri mevcut faaliyet için faydalı bir göstergedir, ancak bir şirketin belirli bir konumda büyümeyi planlayıp planlamadığını veya ne üzerinde çalıştığını açıkça göstermez. Örneğin Paris'teki Facebook AI Research'ün kuruluşunda sadece altı araştırmacı vardı, ancak şirket en başından beri laboratuvarı genişletmeyi planladığını ve sonunda on kattan fazla büyüğünü ifade etmektedir. Her iki ölçümde de bazen veri mevcudiyetinin eksikliği sıkıntılıya neden olmaktadır ve bir ölçümdeki sonuçlar diğerindeki boşlukları doldurmaya yardımcı olabilir.

13.3 BULGULAR

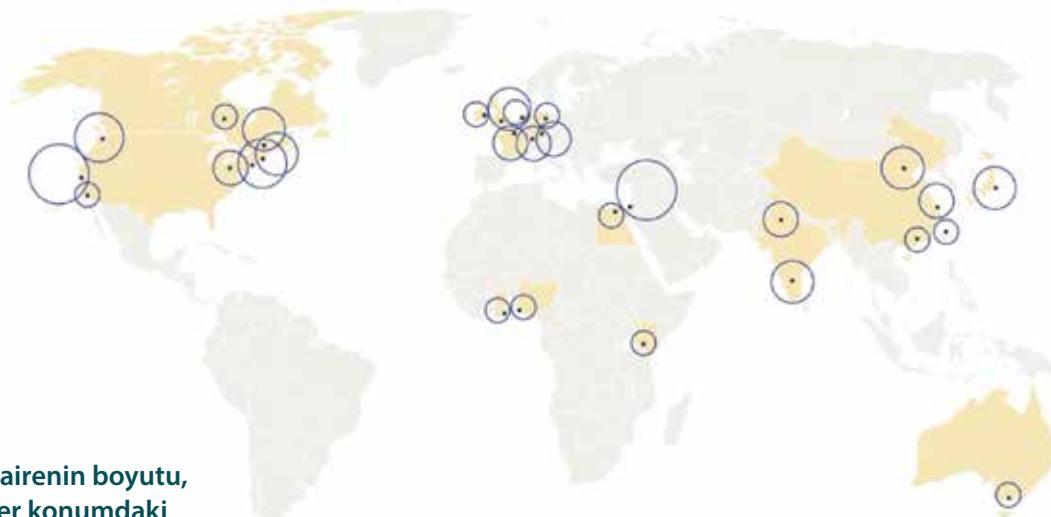
Bu bölümde ilk olarak Amazon, Apple, Facebook, Google, IBM ve Microsoft'un yapay zeka Ar-Ge faaliyetlerinin nerede konumlandırdıkları tartışılara hem laboratuvarlar hem de personeldeki veriler analiz edilecektir. İkinci olarak AI laboratuvarlarına sahip dört şirkette farklı konumlarda ne tür AI Ar-Ge yaptıkları analiz edilecektir. Son olarak da şirketlerin yapay zeka Ar-Ge'sini belirli yerlerde konumlandırmaya karar vermelerinin nedenleri hakkında önceki araştırmaların ve şirket beyanlarının neler önerdiği gözden geçirilecektir.

13.3.1 AI Laboratuvarları ve Personelinin Bulunduğu Yer

ABD'deki çok uluslu şirketlerin AI Ar-Ge faaliyetleri haritalandırılırken iki ölçüt incelenmektedir: AI laboratuvarları ve AI personeli. Şekil 13.3 bu iki ölçümün farklı



Şekil 13.3 ABD dışındaki konumlarına göre ABD'ye göre ABD MNC AI laboratuvarlarının ve personelinin yüzdesi. (Kaynak: CSET AI laboratuvar verileri; LinkedIn'den AI personel verileri. Personel için dahil edilen şirketler Amazon, Apple, Facebook, Google, IBM ve Microsoft'tur; Amazon ve Apple için mevcut laboratuvar verileri mevcut değildir.)



Dairenin boyutu,
her konumdaki
laboratuvarların
sayısına karşılık gelir.

ABD	AVRUPA	AFRİKA	KANADA
6 San Francisco/Bay Area	3 London/Cambridge	1 Accra	3 Montreal
4 Seattle	2 Paris	1 Lagos	1 Toronto
4 New York City	2 Zürih	1 Nairobi	
3 Boston	2 Münih		
1 Los Angeles	1 Berlin		
1 Pittsburgh	1 Amsterdam		
	1 Dublin		
JAPONYA	AVUSTRALYA	ÇİN/TAYVAN	ORTA DOĞU
3 Tokyo	1 Melbourne	3 Beijing	1 Kahire
		2 Şanghay	6 Tel Aviv/Haifa
		1 Shenzhen	
		1 Taipei	
			HINDİSTAN
			3 Bangalore
			2 Delhi

Şekil 13.4 AI laboratuvarlarının konumlarına göre küresel dağılımı. (Kaynak: CSET AI laboratuar verileri.)

şekilde dağılım gösterdiğini ortaya çıkarmaktadır: AI laboratuvarlarının yalnızca üçte biri (yüzde 32) Amerika Birleşik Devletleri'ndedir, AI personelinin ise yaklaşık üçte ikisi (yüzde 68). Aşağıda ilk olarak laboratuvarların coğrafi dağılımına daha yakınandan bakılacak ve ardından personel dağılımına yönelik daha fazla ayrıntı verilecektir.

Şekil 13.4'te 17 ülke ve 29 metropol alanda olmak üzere 41 şehirde bulunan veri setimizdeki 62 AI laboratuvarının uluslararası dağılımı gösterilmektedir. Amerika Birleşik Devletleri'nde bulunan 20 laboratuvardan altısı Silikon Vadisi'nde, dördü New York City'de veya yakınında ve diğer dörtü Seattle'da veya yakınılarında bulunmaktadır. Amerika Birleşik Devletleri dışında en popüler ülkeler İsrail, Çin ve

Hindistan'dır. Teknolojik bağlantı noktalarına sahip ülkelerin dışında Afrika'da dört laboratuvar bulunmaktadır, bunlar Mısır, Gana, Kenya ve Nijerya'dadır. Özellikle Latin Amerika'da bulunan herhangi bir laboratuvar mevcut değildir.

Ülkeler içinde laboratuvarlar ağırlıklı olarak Londra, Pekin ve İsrail'in batı kıyısı bölgesi gibi tanınmış "teknoloji kümelerinde" yer almaktır ve genellikle "Silikon Vadisi" olarak adlandırılmaktadır. Ar-Ge kümelemesiyle ilgili ekonomik araştırmaya paralel olarak en çok laboratuvara ev sahipliği yapan pazar alanları genellikle aktif AI start-up ekosistemlerine ve yüksek dereceli AI programlarına sahip üniversitelere ev sahipliği yapmaktadır. Aslında şirketler bazen bu girişimleri satın alarak veya üniversitelerle ortaklık kurarak laboratuvarlar da kurarlar (Kutu 1'de açıklanacaktır). Bunu

Kutu 13.1

Şirketler AI Laboratuvarlarını Nasıl Kurar?

Veri kümemizdeki laboratuvarlar üç yoldan biriyle oluşturulmuştur: Personelin işe alınması veya yerinin değiştirilmesi, yerel kurumlarla ortak olunması ya da işbirliği yapılması veya start-up'ların satın alınması.

Şirket içi işe alma ve yer değiştirme açık arayla en yaygın kuruluş şeklidir. Bazen bir şirket, özellikle yapay zeka üzerinde çalışmak için bağımsız bir konum oluşturur. Bu oluşuma bir örnek olarak dünya çapındaki bir düzineden fazla Google Brain grubundan biri olan Google Brain Tokyo gösterilebilir. Diğer durumlarda şirketler mevcut (AI olmayan) Ar-Ge tesislerini AI odaklı projeleri içerecek şekilde genişletmektedir.

Şirketler ayrıca ortak Ar-Ge projeleri veya tesisleri oluşturmak için üniversiteler veya diğer kuruluşlarla ortak olabilir. Facebook AI Research'ün Carnegie Mellon Üniversitesi profesörleriyle Pittsburgh ofislerini kurarken yaptığı gibi, bir şirketin prestijli bir üniversitede bir profesörle ortaklık kurması AI içindeki yaygın bir yaklaşımdır. Kurumsal ortaklıklar da bir seçenektedir. Massachusetts Teknoloji Enstitüsü (MIT) tek başına iki ortak laboratuvara ev sahipliği yapmaktadır: MIT-IBM Watson AI Lab ve Microsoft Research Lab-AI. Bunun yanı sıra Şangay'da Microsoft, bölgesel yapay zeka ürünü geliştirmeyi gerçekleştirdiği bir Yapay Zeka İnovasyon Merkezi oluşturmak için devlete ait INESA ile işbirliği yapmaktadır.

Son olarak şirketler zaman zaman yeni şirketler satın alarak laboratuvarlar kurmaktadır. 2010'dan bu yana en az 635 AI şirketi satın alınmış ve bunların kabaca onda biri burada sözü edilen altı şirket tarafından gerçekleştirılmıştır. Microsoft'un 2018'de Montreal'de yapay zeka Ar-Ge'sinin önemli ölçüde genişlemesinin temelini oluşturduğu Montreal merkezli Maluuba'yı satın alması bu duruma bir örnektir. Start-up bir laboratuvarın açık bir örneği de DeepMind'dır. Google'ın 2014 yılında şirketi satın almasından bu yana, DeepMind yalnızca genel merkezini yüzlerce çalışan için genişletmekle kalmamış, aynı zamanda beş küresel ofis de kurmuştur. Ancak her start-up satın alma işlemi yeni bir laboratuvar oluşumu anlamına gelmez. Start-up personeli ve fikri mülkiyeti, genellikle bunun yerine mevcut ürünlere veya ekiplere katılır, bu durumda da bağımsız "laboratuvar" olmaktan çıkarlar.

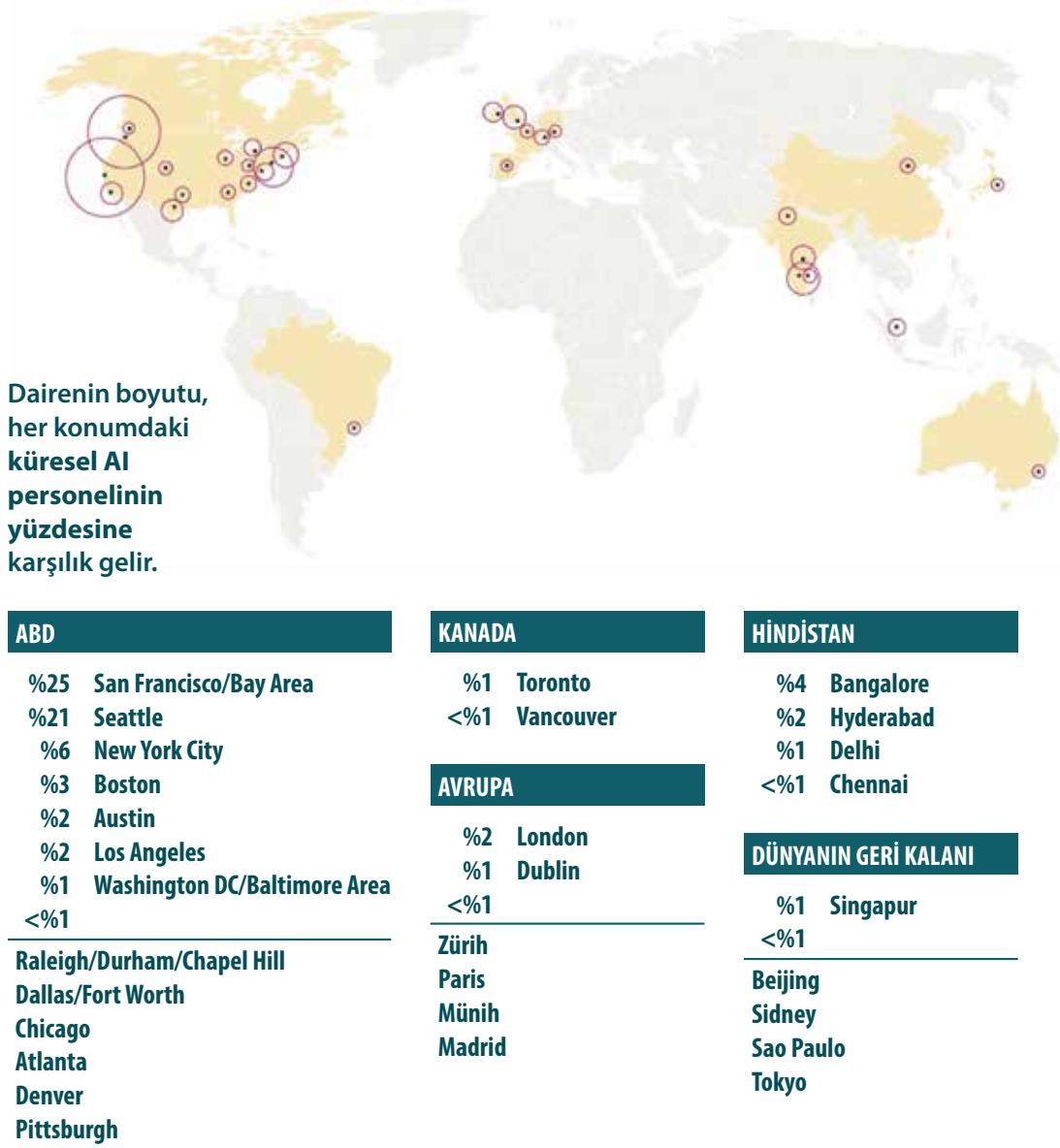
	HER KONUMDAKİ TOPLAM LABORATUVAR SAYISI	LABORATUVARLARIN ORANI	ŞİRKETLERE GÖRE HER KONUMDAKİ LABORATUVARLARIN SAYISI			
			Facebook	Google	IBM	Microsoft
ABD	20	%32	4	9	3	4
Avrupa	12	%19	2	6	1	3
Çin	6	%10	0	1	1	4
İsrail	6	%10	1	2	1	2
Hindistan	5	%8	0	1	2	2
Kanada	4	%6	1	2	0	1
Diger	9	%15	0	2	3	4
Toplam	62		8 Laboratuvar	23 Laboratuvar	11 Laboratuvar	20 Laboratuvar

Şekil 13.5 Şirketlere göre farklı yerlerdeki AI laboratuvarlarının sayısı. (Kaynak: CSET AI laboratuar verileri.)

yapmasalar bile, şirketler genellikle orada bulunmaktan fayda sağlar, çünkü kümeler büyük yetenek havuzlarına ve yeni teknik fikirlere ev sahipliği yapmaktadır.

Şekil 13.5'te laboratuvarları olan dört şirkette yapay zeka laboratuvarlarının konumuna daha yakından bakılmaktadır. Amerika Birleşik Devletleri tüm şirketler arasında en yaygın konumdur, ancak her birinin Avrupa ve İsrail'de laboratuvarları vardır ve dört şirketin üçünün Kanada, Çin ve Hindistan'da laboratuvarları bulunmaktadır. Genel olarak IBM ve Microsoft, Kuzey Amerika ve Avrupa dışında, Google ve Facebook'tan orantılı olarak daha aktif görülmektedir.

Onceki iki rakam laboratuvarların sayısını ve konumunu açıklıyor, ancak çoğu şirket laboratuvarlarının ne kadar büyük olduğuna dair kamuya açık verileri yayınlamamaktadır. Üstelik yukarıda belirtildiği gibi, Amazon yapay zeka Ar-Ge'sini laboratuvarlarda yapılandırmamaktadır ve Apple, Ar-Ge faaliyetlerine ilişkin bilgileri kamuya açıklamamaktadır ve geri kalan dört şirket de laboratuvarlarının dışında da yapay zeka Ar-Ge'si yapabilmektedir. Bu nedenle bu bulguları AI faaliyetinin başka bir göstergesi olan personel sayılarıyla karşılaştırmak gerekmektedir.



Şekil 13.6 Şirketlerin kümülatif olarak 500'den fazla AI personeline sahip olduğu yerler. (Kaynak: LinkedIn. Personeli dahil olan şirketler Amazon, Apple, Facebook, Google, IBM ve Microsoft'tur.)

Şirketlerin AI personelinin yüzde 68'inin Amerika Birleşik Devletleri'nde bulunduğu Şekil 13.3'te açıkça görülmektedir. Şekil 13.6'da ise altı şirketin kümülatif olarak 500 veya daha fazla yapay zeka personeli istihdam ettiği 15 ülkedeki 30 metropol alanı haritalandırılmaktadır. Yapay zeka personelinin yaklaşık yarısı (yüzde 46) sadece iki metropol bölgesinde bulunmaktadır: San Francisco ve Seattle. Austin, Boston, Los Angeles ve New York dahil diğer dört ABD şehri de ilk on konumdadır.

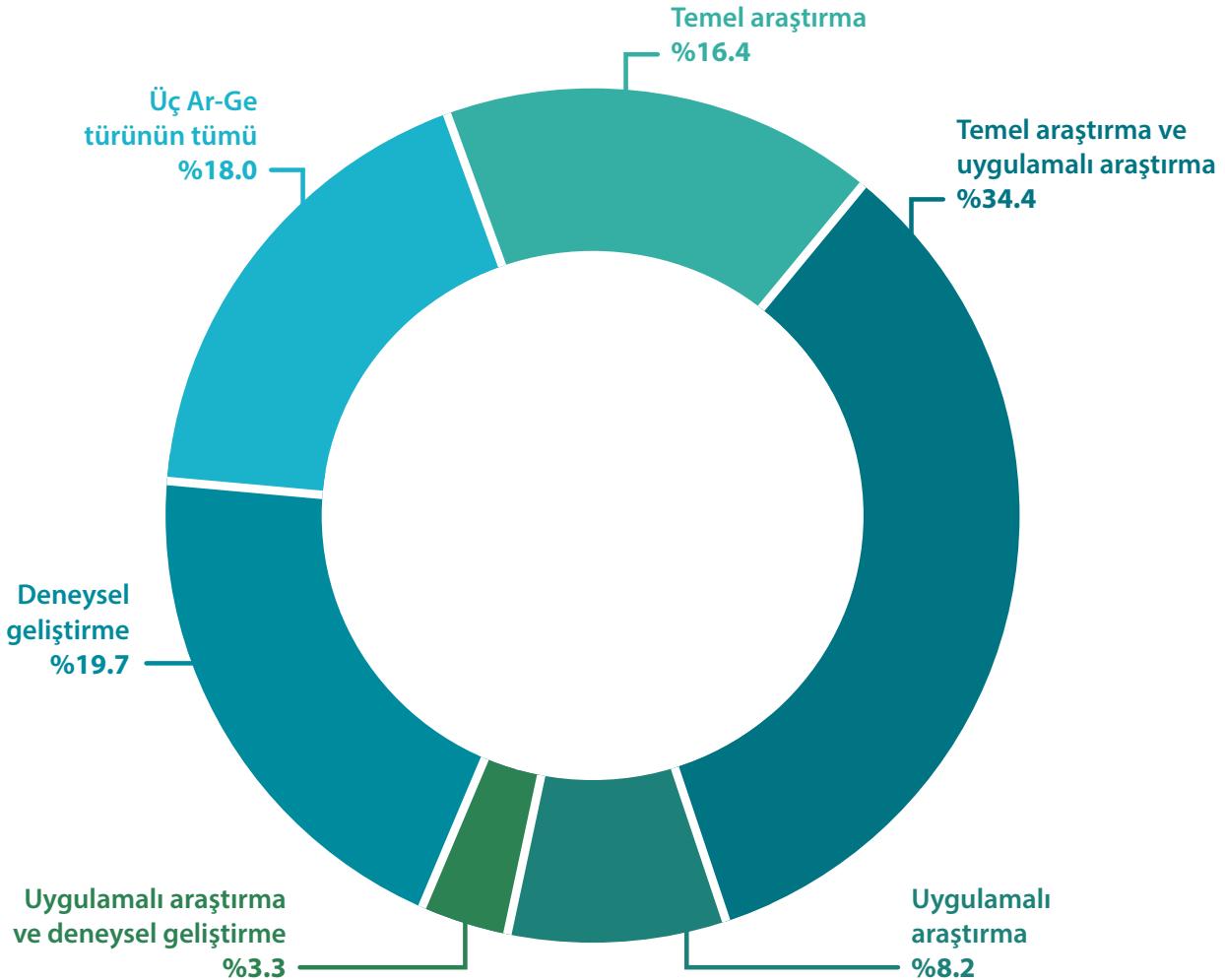
Amerika Birleşik Devletleri dışında altı şirket Avrupa, Hindistan ve Asya-Pasifik bölgesinde çoğu AI personelini istihdam etmektedir. Pek çok yerde yapay zeka laboratuvarları bulunmaktadır, ancak Çin (Şanghay, Shenzhen ve Pekin içinde veya yakınında) ve İsrail'deki çeşitli yerler dahil olmak üzere 500'den az yapay zeka personele sahip olduğu görülmektedir. Yapay zeka personeli için ana kaynağımız LinkedIn olduğundan, bu bulguların dikkatle yorumlanması gerekmektedir; bazı farklılıklar farklı bu ülkelerdeki çok uluslu çalışanlar arasında farklı LinkedIn kullanım oranları ile açıklanabilir.

13.3.2 AI Laboratuvarları Ne Tür Ar-Ge Yürüttür?

Bu bölümde şirketlerin yapay zeka laboratuvarlarında yürüttükleri Ar-Ge türlerine degenilmektedir. Ar-Ge faaliyetleri temel araştırma, uygulamalı araştırma veya deneysel geliştirme olarak kategorize edilebilir. Bu bağlamda “temel araştırma” yapay zekada temel atılımların peşinde koşmak, AI araştırmasının bilinen sorunlara uygulanması olarak “uygulamalı araştırma” ve AI prototiplerinin ve ürünlerinin oluşturulması olarak “deneysel geliştirme” olarak tanımlanabilir. AI laboratuvarlarının Ar-Ge faaliyetlerini bu üç kategoriye ayırmak için haberlere, basın bültenlerine ve diğer kurumsal bildirilere bakıldı. Birçok laboratuvar birden fazla türde Ar-Ge yürütmektedir.

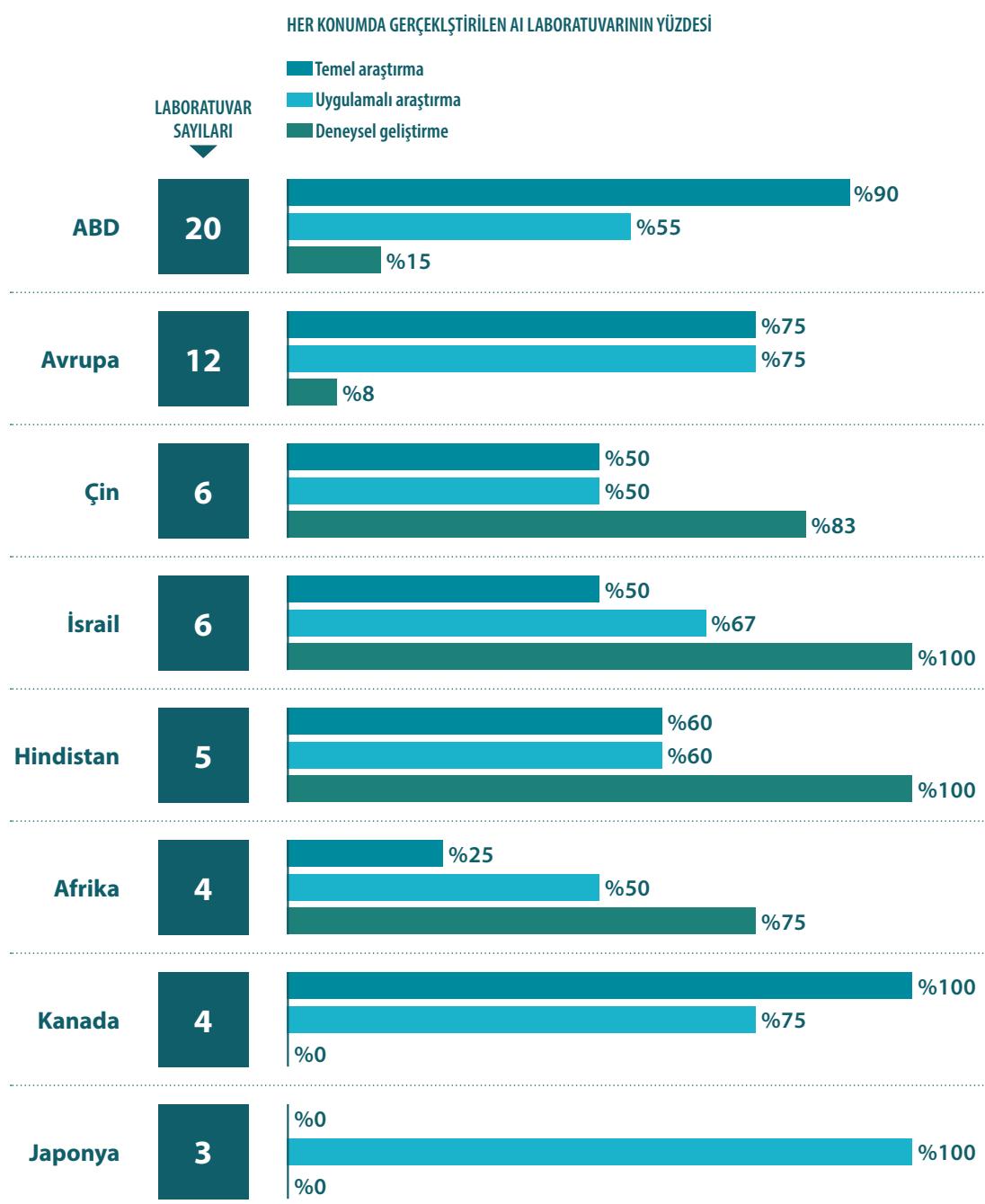
Şekil 13.7 her bir Ar-Ge ürünü yürüten AI laboratuvarlarının oranını göstermektedir. Yapay zeka laboratuvarlarının yaklaşık yüzde 80'i bir çeşit araştırma yapmaktadır. Yapay zeka laboratuvarları genellikle birden fazla türde AI Ar-Ge'si yürütse de çoğu AI laboratuvarı bazı temel araştırmalar yürütmektedir. Laboratuvarların yarısından fazlası en az iki tür AI Ar-Ge gerçekleştirmektedir.

Şekil 13.8, ülkeye veya bölgeye göre her bir Ar-Ge ürünü gerçekleştiren AI laboratuvarlarının bölümlemesini göstermektedir. Kuzey Amerika ve Avrupa'daki laboratuvarlar ağırlıklı olarak araştırma odaklıken, diğer bölgelerde bulunan laboratuvarlar daha az temel araştırma ve daha az deneysel geliştirme yapmaktadır. Google'ın AI Çin Merkezi, en azından lansman sırasında dikkate değer bir istisnaydı; basın bülteninde, “temel yapay zeka araştırmalarına odaklanılacağı” belirtilmiştir. Google'ın Gana'daki AI laboratuvarı da hem temel hem de uygulamalı araştırmalar yürütmektedir.



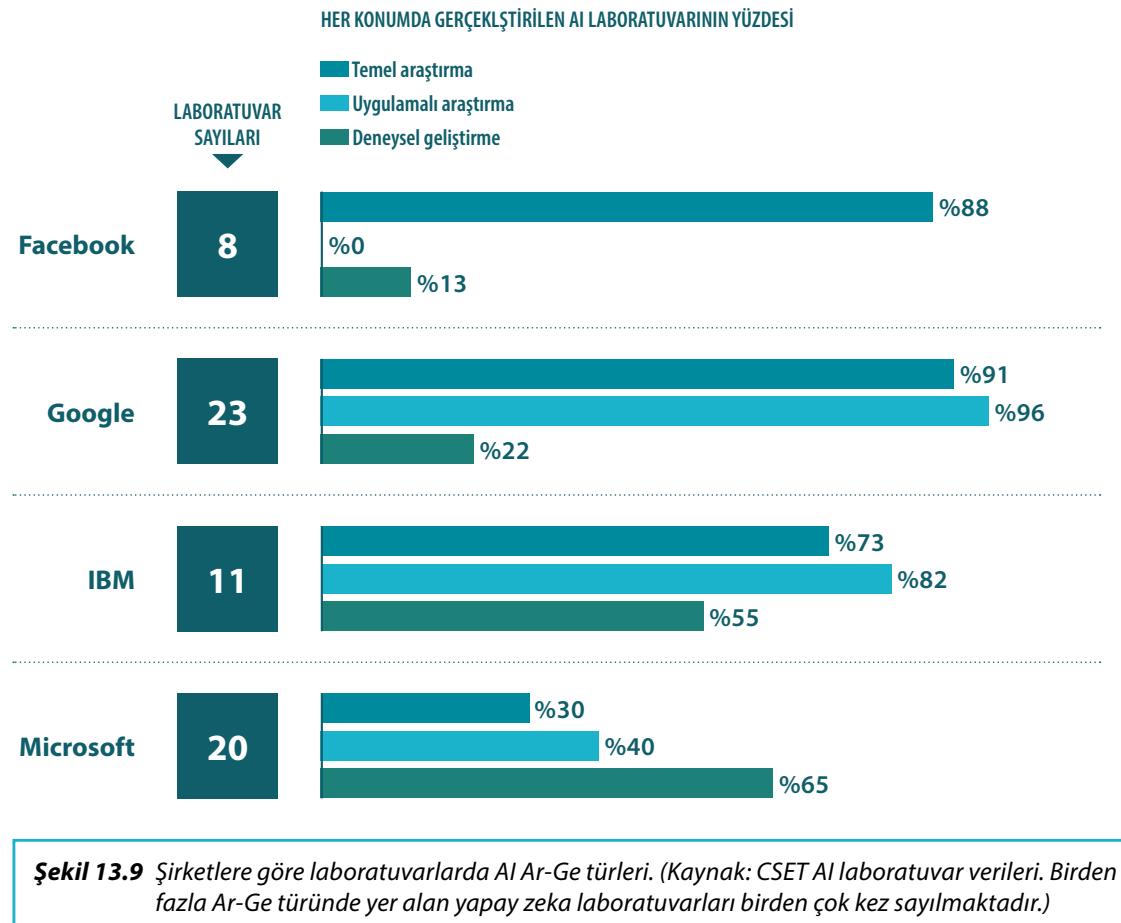
Şekil 13.7 Ar-Ge kategorisine göre AI laboratuvarlarında yapılan Ar-Ge türleri. (Kaynak: CSET AI laboratuvar verileri.)

Şekil 13.9'da her bir şirketin yapay zeka laboratuvarlarının her tür Ar-Ge'yi yürüten bölümü gösterilmektedir. Laboratuvarlarının gerçekleştirdiği Ar-Ge türleri açısından şirketler arasında önemli farklılıklar bulunmaktadır. Facebook AI Araştırma (FAIR) laboratuvarları neredeyse yalnızca temel araştırmalarla ilgilenir, Google'in laboratuvarları temel ve uygulamalı araştırmaların bir karışımını yürütmektedir, Microsoft laboratuvarlarının deneysel geliştirmeye odaklanması olasılığı daha yüksektir ve IBM'in laboratuvarları üç Ar-Ge türünün bir karışımını yürütmektedir. Bu, teknoloji şirketlerinin Ar-Ge faaliyetlerini farklı şekilde yapılandırdığını ve bununla birlikte "laboratuvar" terimini farklı şekilde kullanabileceğini ortaya koymaktadır.



Şekil 13.8 Ülkeye veya bölgeye göre AI laboratuvarlarındaki Ar-Ge türleri. (Kaynak: CSET AI laboratuar verileri. Tek bir AI laboratuvari birden fazla Ar-Ge türünde yer alabilir. Her biri yalnızca bir laboratuvara ev sahipliği yapan Avustralya ve Tayvan sayının dışında tutulmuştur.)

Şirketlerin AI Ar-Ge faaliyetlerindeki farklılıklar, nerede kurulduklarına ve nasıl performans gösterdiklerine göre de görülebilir. Yapay zeka laboratuvarlarının esas olarak temel ve uygulamalı araştırmalara odaklandığı Seattle'da, Facebook'un yerel laboratuvar yöneticisi Washington Üniversitesi'nde bilgisayar bilimleri profesörlü-



günü sürdürmektedir ve güçlü üniversite bağlarını kapsamaktadır. Buna karşın İsrail’deki diğer şirketlerin laboratuvarları uygulamalı araştırmalara ve deneysel geliştirmeye odaklanma eğilimindedir. Örneğin İsrail Geliştirme Merkezi ve İsrail İleri Teknoloji Laboratuvarı gibi oradaki Microsoft AI laboratuvarları, start-up kurucularıyla yakın bir şekilde çalışıyor ve daha güçlü bir mühendislik odağına sahiptir.

13.4 KONUM KARARLARINI NE YÖNLEDİRİR?

Bu bölümde şirketlerin neden belirli yerlerde laboratuvar ve ofisler kurduğu ele alınmaktadır. Bu kararların tipik olarak çeşitli gerekliliklerle verildiği görülmektedir, ancak ekonomi ve iş alanındaki araştırmalarla birlikte şirketlerin açıklamaları üç ana neden üzerine yoğunlaşmaktadır: Yeteneklere erişim, pazara erişim veya ürün uyarlaması ile maliyet tasarrufu.

En az bir AI laboratuvarına sahip bölgelerin yüzde 38'i aynı zamanda dünyanın en iyi 50 AI üniversitesinden birine veya daha fazlasına ev sahipliği yapmaktadır.

13.4.1 Yeteneklere Erişim

Şirketler neredeyse her zaman yerel bir yetenek havuzunun varlığını konum seçimleri için baskın bir neden olarak göstermektedir. Bir dizi yeni Facebook AI laboratuvarından bahseden CTO Mike Schroepfer basitçe şunu ifade etmektedir: "Temelde yeteneğin olduğu yere gidiyoruz." Bir Google sözcüsü ise, "Yapay zeka araştırma ofisi konumlarını seçmemizin ana sebebi, ya zaten bölgede yaşayan ya da oraya taşınmak isteyecek yeteneklerin mevcudiyetidir," diye belirtmiştir. Microsoft Research Asia'nın genel müdürü, şirketin Çin'de bir laboratuvar açma kararını, "ABD'de araştırma laboratuvarına sahip olamadıklarına çünkü yeterince insan işe almayacaklarını," kavramaları sonucuna dayandırmaktadır. Araştırma yapan zeki ve en tutkulu tüm insanların hepsinin [Birleşik Devletler'e] geleceğine inanmadıklarını belirtiyor. Bu argümanlar şirketlerin gerekli AI becerilerine sahip küresel bir insan kıtlığı olduğu yönündeki algısıyla pekiştirilmekte ve bu da yetenek kümelerinde konumlandırmayı özellikle ilgi çekici kılmaktadır.

Laboratuvarlar ve ofisler genellikle üniversitelerin yakınında bulunmaktadır; AI laboratuvarları veritabanında en az bir AI laboratuvarına sahip bölgelerin yüzde 38'i aynı zamanda dünyanın en iyi 50 AI üniversitesinden birine veya daha fazlasına ev sahipliği yapmaktadır. Şirketlerin yerel ofisleri yeteneklere erişmek amacıyla genellikle üniversitelerle yakın bağlar kurar. Örnekler arasında Google'ın Münih Teknik Üniversitesi ile ilişkileri ve Microsoft'un Cambridge Üniversitesi ile ilişkisi yer almaktadır. İsrail'de Google'ın mühendislerinin yarısını Tel Aviv Üniversitesi'nden sağladığı bildirilmektedir. Şirketler genellikle üniversiteler tarafından istihdam edilen profesörleri ve araştırmacıları işe almakla ilgilenmektedir. Önemli bir örnek olarak Uber,

Pittsburgh laboratuvarı kurulduktan kısa bir süre sonra 2015 yılında Carnegie Mellon Üniversitesi'nden yaklaşık 50 personel işe alınmıştır. Üniversiteler ayrıca teknoloji şirketlerinin potansiyel ortaklıklar veya satın almalar için göz önünde bulundurmak istediği start-up'ları da kuluçkaya yatırmaktadır.

Bazı yerler daha geniş bir bölgeden gelen yetenekler için iyi bir buluşma noktasıdır, çünkü ya gelecek vadeden merkezler ya da dostça göçmen politikaları vardır. Örneğin Google AI Gana ve Microsoft Tayvan, yalnızca kendi ülkelerinden değil, daha geniş bölgelerden de iş alım yapmak üzere kurulmuş gibi görülmektedir. Vatandaşlar için dahili iş gücü hareketliliğinin kısıtlanmadığı Avrupa Birliği içindeki laboratuvarlar da genellikle bölgesel yetenekler için merkez görevi görmektedir.

13.4.2 Pazar Erişimi ve Ürün Uyarlaması

Ekonomistler genellikle yurtdışındaki “bilgi arayışı” veya “yetenek arayan” yatırımları “pazar odaklı” yatırımlarla karşılaştırmaktadır. Yabancı laboratuvarlar iki yoldan biriyle pazarlara erişim sağlamaya yardımcı olabilir. İlk olarak bazı ülkeler şirketlerin pazara erişim karşılığında yerel tesisler veya ortaklıklar kurmasını zorunlu kılar. Ancak bu tür gereksinimlerin olmadığı ülkelerde bile şirketler, müşterilerine yakın olmak ve ürünlerini yerel koşullara uyarlamak için genellikle yerel ofisler kurarlar.

Pazar erişimi, veri kümelenmesinde yeni bağımsız yapay zeka laboratuvarları kuran şirketler için baskın bir motivasyon gibi görünmemektedir. Ancak AI Ar-Ge konumu için pazara erişim dolaylı olarak önemli olabilir. Örneğin Çin'de bir çalışma “temel araştırmayı yürüten bir avuç Ar-Ge merkezinin çوغunun güçlü ürün geliştirme işlevleri kurana kadar bunu yapmaya başlamadığını” belirtmektedir. Diğer araştırmalarda Hindistan ve Çin dahil, yeterli Ar-Ge yetenek havuzuna sahip ülkelerde benzer yerel “değer zinciri tırmanışı” bulunmaktadır.

Ev sahibi ülkenin yabancı teknolojiye olan eğilimi de çok uluslu şirketlere mal yetirebilmektedir.

Ev sahibi hükümetler bazen, örneğin düşük kira veya kira-ücretsiz ofis alanı veya sübvansiyonlu elektrik sağlayarak maaşsız teşvikler yoluyla çok uluslu şirketleri cezbetmektedir.

13.4.3 Maliyet Tasarrufu

Firmalar maliyetten tasarruf etmek için yurt dışında da Ar-Ge çalışmaları yapmaktadır. AI yeteneğinin pahalı olduğu genel olarak kabul edilen bir gerçektir; kesin rakamlar çalışmalar arasında farklılık gösterse de çoğu Amerika Birleşik Devletleri'nde ortalama maaşı 100.000 doların çok üzerinde bulunmaktadır. Benzer şekilde ayrıntılı AI maaş istatistikleri diğer ülkeler için mevcut değildir, ancak National Science Foundation tarafından 2013 yılında yapılan bir ankete göre ABD bilgi ve iletişim teknolojileri şirketlerinin yurtdışında gerçekleştirilen Ar-Ge'ye kıyasla yurt içinde yürütülen Ar-Ge için çalışan başına ortalama iki kat daha fazla harcama yapmak zorunda kaldığı ortaya konulmuştur. Bu, maaşlarda maliyet tasarrufunun önemli olabileceğini göstermektedir.

Ev sahibi hükümetler bazen, örneğin düşük kira veya kira-ücretsiz ofis alanı veya sübvansiyonlu elektrik sağlayarak maaşsız teşvikler yoluyla çok uluslu şirketleri cezbetmektedir. ABD vergilerini ödemekten kaçınma veya ev sahibi ülke tarafından sağlanan vergi teşviklerinden yararlanma arzusu da şirketler için motive edici olabilmektedir. Ev sahibi hükümetler bu tür teşvikler sağlar çünkü nüfuslarının, işletmelerinin ve ekonomilerinin fayda sağlayacağına inanmaktadır. Ancak ev sahibi bir ülkenin yabancı teknolojiye olan eğilimi de çok uluslu şirketlere maliyet getirebilmektedir. Bu maliyetler fiziksel veya fikri mülkiyet hırsızlığı riskini ve bu tür risklere karşı koruma masraflarını da içermektedir.

Şirketlerin kendileri genellikle yurtdışında ofis kurma kararlarında maliyetin bir rol oynadığını kabul etme konusunda isteksizdir, ancak uzmanlar bunun önemli bir faktör olduğunu kabul etmektedir. Örneğin Taipei merkezli bir endüstri analisti Mic-

rosoft'un Tayvan'daki yapay zeka Ar-Ge bileşenini içeren kuruluşunu değerlendirirken, düşük maaşlar nedeniyle şirketin Tayvan'da daha yüksek yatırım getirisine sahip olacağuna işaret etmektedir. Ancak analistler daha düşük maliyetlerin faydalı olmakla birlikte, bir şirketin yeni bir yer açması için yeterli bir neden olmadığını savunmaktadır. Yeterince yüksek kalitede yerel yetenek arzı olmadan şirketler, maliyetleri önemli ölçüde daha düşük olsa bile, önemli teknik çalışmaları içeren girişimler başlatma konusunda isteksiz olacaktır.

13.5 SONUÇ

Sonuç olarak ABD'deki çok uluslu şirketlerin yapay zeka Ar-Ge laboratuvarlarıyla ilgili analiz doğrultusunda, Facebook, Google, IBM ve Microsoft'un tüm dünyaya yayılmış 62 laboratuvarı olduğu görülmektedir; 20'si Amerika Birleşik Devletleri'nde ve geri kalan 42'si de 16 ülkede yer almaktadır. Ancak AI laboratuvarlarının aksine, çoğu şirketin AI personelinin ABD'de yoğunlaşmış bir durumda olduğu görülmektedir.

Kuzey Amerika ve Avrupa'da yapay zeka Ar-Ge'sinin çoğu araştırmaya odaklanmışken, başka yerlerdeki laboratuvarlar geliştirmeyi vurgulama eğilimindedir. Şirketler birçok nedenden ötürü yurtdışında AI Ar-Ge'si yapmaya karar vermektedir, ancak küresel yeteneklere erişim genellikle itici faktör olarak gösterilmektedir.

Diğer taraftan ABD göçmenlik reformlarının ABD şirketlerini yurt içinde daha fazla Ar-Ge yapmaya teşvik edeceğini öne süren geçmiş araştırmaları güçlendirir niteliktedir, ancak yerel yetenek arzını artırmanın şirketlerin Ar-Ge'yi tamamen durdurması veya tersine çevirmesi olası görünmemektedir. İkincisi, daha fazla yerli Ar-Ge'yi teşvik etmek için kullanılabilecek diğer araçlar için bu veriler politika yapıcıların farklı politika yaklaşımlarının onde gelen ABD AI şirketlerini nasıl etkileyeceğini değerlendirmelerine yardımcı olabilir. Bu tür etki değerlendirmeleri Ar-Ge'nin küreselleşmesinin karmaşık faydalarını ve risklerini deneğemenin önemli bir parçasıdır. Üçüncüsü, çok yönlü Ar-Ge girişimleri ve bir "ittifak inovasyon üssü" çağrıları önceden mevcut olan Ar-Ge bağlantılarının kurulmasını önermektedir.

Bu çalışma aynı zamanda analistlerle ilgili dersler ve kaynaklar sağlar. Örneğin yapay zeka laboratuvarları ve yapay zeka personeli için yerli-yabancı oranlarının

neredeyse ayna görevi gördüğü bir gerçektir. Bu noktadan hareketle ABD'deki çok uluslu şirketlerin Ar-Ge'sinin küreselleşmesinin kapsamı incelenirken yurtdışında yapay zeka Ar-Ge'sinin birden fazla ölçümüne bakmak önemlidir.

KAYNAKLAR

Bronwyn H. Hall, "The Internationalization of R&D" (University of California, Berkeley, August 2010), 10, https://eml.berkeley.edu/~bhhall/papers/BHH10_RND_international_August.pdf.

Cade Metz, "Facebook Adds A.I. Labs in Seattle and Pittsburgh, Pressuring Local Universities," The New York Times, May 4, 2018, <https://www.nytimes.com/2018/05/04/technology/facebook-Artificial-intelligenceresearchers.html>.

Dario Gil, "How to Ensure the U.S.'s Quantum Future," Scientific American, August 20, 2020, <https://www.scientificamerican.com/article/how-to-ensure-the-us-quantum-future/>.

Department of Defense, Summary of the 2018 National Defense Strategy (Department of Defense, 2018), 3, <https://www.hsdl.org/?View&did=807329>.

Fei-Fei Li, "Opening the Google AI China Center," Google: The Keyword (blog), December 13, 2017, <https://www.blog.google/topics/google-asia/google-ai-chinacenter/>.

Gary Clyde Hufbauer, Theodore H. Moran, Lindsay Oldensi, and Martin Vieiro, Outward Foreign Direct Investment and US Exports, Jobs, and R&D: Implications for US Policy, Peterson Institute for International Economics, August 2013, <https://www.piie.com/bookstore/outward-foreign-direct-investment-and-us-exports-jobsand-rd-implications-us-policy>.

<https://www.computerworld.com/article/3181020/apple-is-pumping-money-into-creating-an-international-chain-of-randd-enters.html>.

Jacques Bughin, Eric Hazan, Sree Ramaswamy, Michael Chui, Tera Allas, Peter Dahlström, Nicholaus Henke, and Monica Trench, "Artificial Intelligence: The Next Digital Frontier?" (McKinsey Global Institute, June 2017), 10, <https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>.

Microsoft Asia News Center, “Microsoft establishes Microsoft Research Asia-Shanghai and a new AI Innovation Center,” Microsoft Blog, September 17, 2018, <https://news.microsoft.com/apac/2018/09/17/microsoft-establishes-microsoftresearch-asia-shanghai-and-a-new-ai-innovation-center/>.

Omri Wallach, “How Tech Giants Make Their Billions,” Visual Capitalist, July 6, 2020, <https://www.visualcapitalist.com/how-big-tech-makes-their-billions-2020/>; Vaishali Ladha, “IBM: Revenue by Region (2020),” Business Quant, <https://businessquant.com/ibm-revenue-by-region-2>.

Rani Molla, “Amazon spent nearly \$23 billion on R&D last year - more than any other U.S. company,” Vox, April 9, 2018, <https://www.vox.com/2018/4/9/17204004/amazon-research-Development-rd>.

Recommendations for Strengthening American Leadership in Industries of the Future (Washington, DC: Office of Science and Technology Policy, June 2020), 5, https://science.osti.gov/-/media/_/pdf/about/pcast/202006/PCAS_T_June_2020_Report.pdf?la=en&hash=019A4F17C79FDEE5005C51D3D6CAC81FB31E3ABC.

Ryan Browne, “Top US general says Google ‘is indirectly benefiting the Chinese military,’” CNN Politics, March 14, 2019, <https://www.cnn.com/2019/03/14/politics/dunfordchina-google/>.

Steven Levy, “Inside Amazon’s Artificial Intelligence Flywheel,” WIRED, February 1, 2018, <https://www.wired.com/story/amazon-artificial-intelligence-flywheel/>

Synced, “Google Boosting its AI Research in Tokyo,” Medium, April 27, 2018, <https://medium.com/syncedreview/google-boosting-its-ai-research-in-tokyo-608f8ba11c9>.

“The Race For AI: Here Are The Tech Giants Rushing To Snap Up Artificial Intelligence Startups,” CB Insights, September 17, 2019, <https://interactives.cbinsights.com/artificialintelligence-acquisitions-by-famga/>.

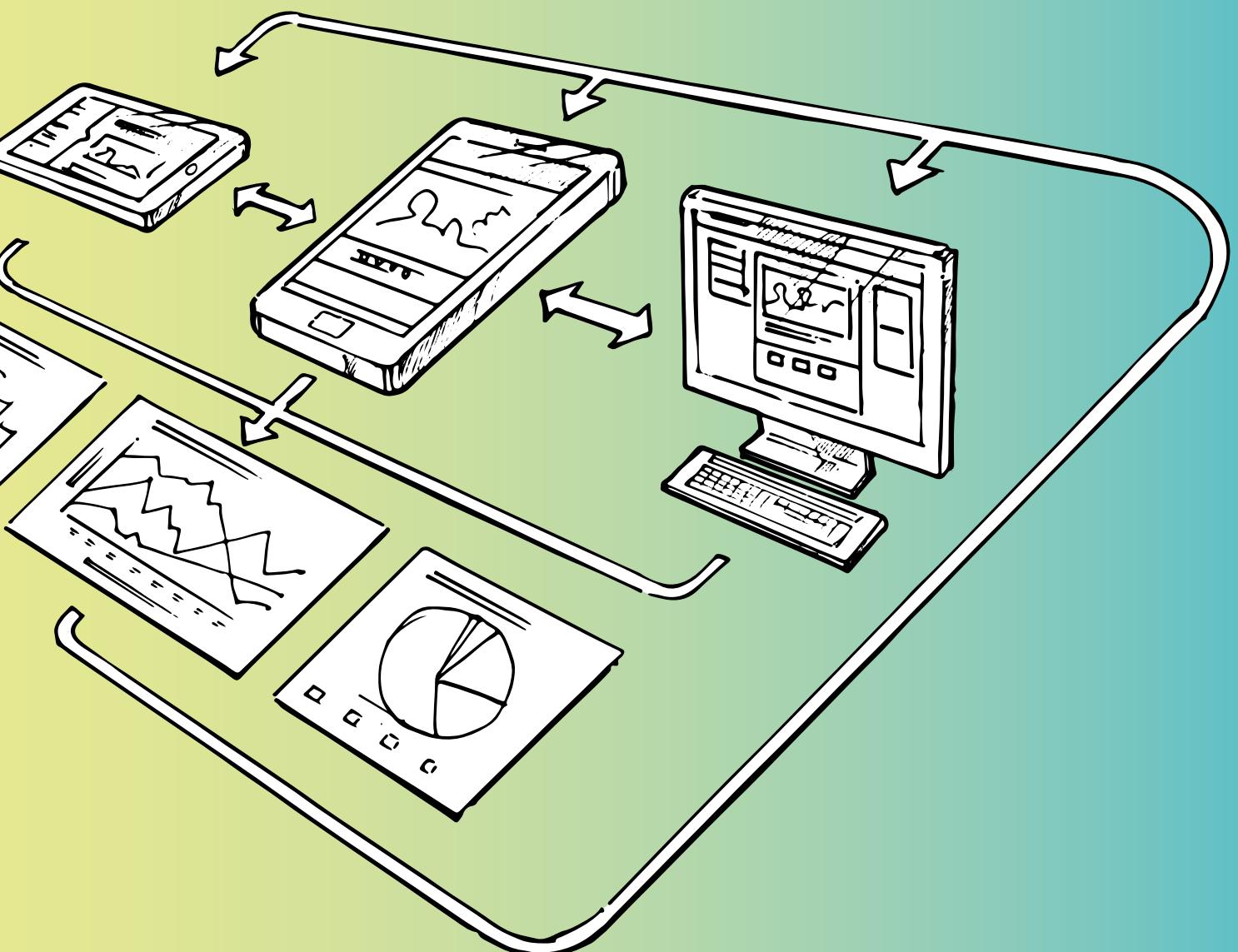
William R. Kerr and Frédéric Robert-Nicoud, “Tech Clusters,” The National Bureau of Economic Research, no.27421 (June 2020), <https://www.nber.org/papers/w27421>.

Xiaohong Quan, “Multinational Research and Development Labs in China: Local and Global Innovation,” Ph.D. dissertation, University of California, Berkeley, 2005, ProQuest (3211492).

Zwetsloot ve diğ., “Strengthening the U.S. AI Workforce,” 6-7.

14. BÖLÜM

İŞ YERİNDE VERİ BİLİMİ EĞİTİMİ VE ALTERNATİF MEKANİZMALAR



14.1 DEVLETTE VERİ BİLİMİ UYGULAMASI

ABD Sayım Bürosu'ndan Ron Prevost, ABD Nüfus Sayım Bürosu tarafından üretilen verilerin tarafsız, istatistiksel olarak doğru, düşük maliyetle hızlı bir şekilde sunulması, nedenselliği belirlemede yararlı, tekrarlanabilir, şeffaf ve korumalı olmasının beklediğini açıklamıştır. İstatistik kurumları bu bekłentileri karşılamaya çalışırken, anketler için beklenenden daha yüksek maliyetler ve beklenenden düşük yanıt oranları, karmaşık bilgi talepleri, veri ürünleri arasındaki rekabet ve ürün geçerliliği soruları, yeni veri kaynakları ve metodolojileri ve politika gereksinimleri dahil olmak üzere birçok zorlukla karşı karşıyadır.

Nüfus Sayımı Bürosu, anket verilerini başka kaynaklardan yeniden tasarlanan verilerle desteklemeyi ummaktadır. Ancak bu veri entegrasyonunun şeffaf ve güvenilir olması, kalite ölçütlerinden faydalananması ve ideal olarak model tabanlı tahmin ve veri kaynağı edinme ve entegrasyon süreçlerini içermesi gerekmektedir. Prevost'a göre, Sayım Bürosu, (1) iş süreçlerini ve sistemleriyle genelleştirilmiş çözümleri konsolide etmek, (2) mevcut iş süreçlerini yeni süreçlerle desteklemek, (3) yeni ürünler geliştirmek (4) yeni yetenekler oluşturmak ve (5) mevcut iş süreçlerini optimize etmek dahil olmak üzere birkaç kritik adım atarak bu paradigma değişikliğini ilerletebilir.

Prevost kurumsal, bütçe ve güvenlik engellerinin bu tür büyük ölçekli dönüşümü sınırlayabileceğini belirtmektedir. Örneğin bu dönüşümü kapsamlı bir şekilde ele almak için bilgi teknolojisi departmanlarındaki personelin verileri ve meta verilerini yönetmesi, iyileştirmesi ve kullanmak için yeni süreçleri öğrenmesi, yeni yazılımın devlet güvenlik protokollerine uygun olmasını sağlaması ve gerçek verileri genellikle “korumalı alan” olarak adlandırılan bir ortak çalışma ortamında organize etmesi, keşfetmesi ve test etmesi gerekmektedir.

Amerika'da birçok federal kurum, disiplinler arası ekip çalışması ve profesyonel gelişim için artan fırsatların, çalışanları yeni bilgi işlem teknikleri ve yeni metodolojiler gerektiren işler için nasıl daha iyi donatabileceğini araştırmaktadır. Hem mevcut hem de gelecekteki çalışanlar tarafından hangi beceri setlerine ihtiyaç duyulacağını ve projelerin nasıl finanse edileceğini belirlemek için program kullanım örnekleri değerlendirebilebilir. Mevcut bilgi boşlukları arasında veri bilimi, iş/veri analitiği, yeniden üretilabilirlik, yazılım tasarıımı ve mühendisliği, veri depolama ve erişim modelleri ve operasyon araştırması bulunmaktadır. Bu yönde kapsamlı bir araştırmmanın ardından, Nüfus Sayımı Bürosu, Washington, D.C.'de veya yakınında bulunan veya gerekli içerik alanlarında dereceler, sertifikalar veya kısa kurslar sunan çevrimiçi olarak erişilebilen 80 programdan (toplam 600 kurs içeren) bir katalog oluşturdu. Buna ek olarak devasa açık çevrimiçi kurslar (MOOC'lar) uygun maliyetli bir alternatif olabilir ve bu daha geleneksel eğitim programlarının tamamlayıcısı olabilir.

Çalışanların işlerinde farklı veri türlerini kullanmanın benzersiz avantajlarını ve dezavantajlarını anlamaları için eğitime ihtiyaç duyulmaktadır. Bunun dışında veri etiği ve siber güvenlik konularının çalışan eğitimi için çok önemli olduğu alanlardır. Fakat Stanford Üniversitesi'nden Jeffrey Ullman, eğitim ile gerçek dünya sorunları arasında temel bilgisayar bilimi becerilerinin daha da geliştirilmesiyle ortadan kaldırılabilcek bir kopukluk olduğunu öne sürmüştür. Bu alanda beceri anlamında eksikliklerin olduğu aşikardır, ancak bilgi teknolojisi uzmanlarından çok araştırma analistleri için eğitime odaklanması gerekmektedir.

Diğer taraftan bir işyerinde resmi, standartlaştırılmış bir eğitim politikası etkili olmayabilir, çünkü personel eğitimi görev açısından kritik faaliyetler ve hızla değişen teknolojiler etrafında gelişmektedir. Devletlerde yaratıcı eğitim girişimleri geliştirme, eğitim programları sunma veya yeni personel işe alma konusunda sınırlı



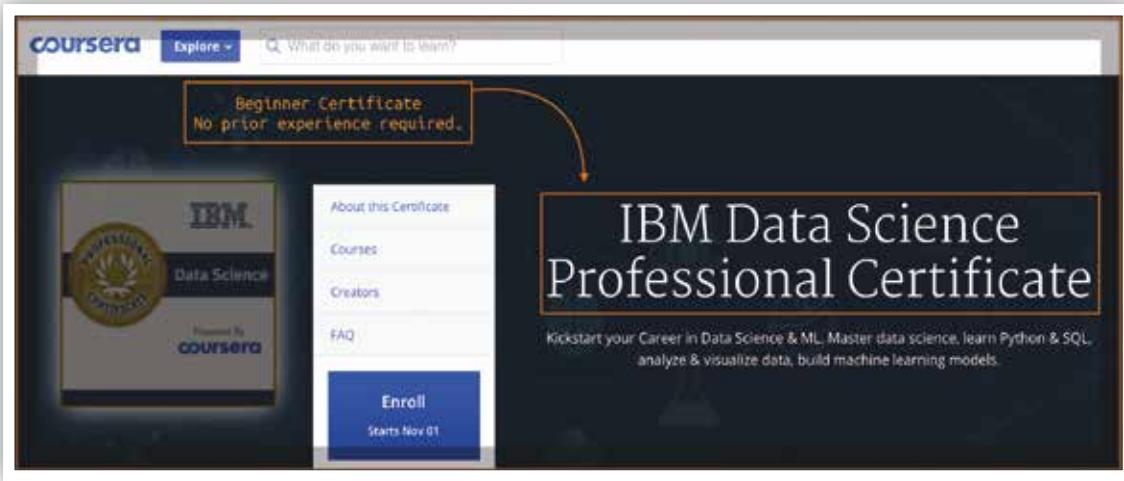
finansman veya yetkisi olan federal kurum yöneticileri için çok önemlidir. Doğru veri bilimi becerilerinin nasıl bir araya getirileceğini değerlendirilirken çalışanların ve yöneticilerin göz önünde bulundurması gereken bu iki model yararlı olabilir: (1) bir çalışanın bir görevi tamamlamak için gereken tüm becerilere sahip olduğu bir “unicorn” model veya (2) çeşitli becerilere sahip kişilerin bir görevi tamamlamak için birlikte çalıştıkları bir “X-Men” modeli.

Ayrıca veri bilimi konusunda hükümet dışı eğitimler de faydalı olabilir. Hükümet dışı kurumlar da devletle birlikte doğacak ihtiyaçları belirleyerek eğitmenlerle bu yönde programlar geliştirebilir ve çalışanlar bu programlara katılabılır. Bunun dışında başka bir sorunsal olarak belirli sorunları çözmek için çalışanları eğitmenin mi, o alanda danışman işe alınan mı kararının verilmesi ortaya çıkmaktadır. Buradan hareketle mentorluk programlarının da geliştirilmesinin önemi göz önünde bulundurulmalı. Fakat bu bağlamda da finansal desteği olan ihtiyacın artacağı yadsınamaz bir gerçek olarak varlığını sürdürmektedir.



14.2 Teknoloji Sektörü

Veri bilimciler niceł eğitime, etkili iletişim becerilerine, iş zekasına ve çeşitli veri ve analitik araçlarına ve programlama dillerine güvenen ve sorunları çözen “öncüler”dir. Veri bilimi bilişsel bilgi işlem çağına yanıt olarak gelişmeye devam etmektedir. Veri bilimcilerin artık hibrit analitik, veri akışı, yapay zeka, uygulama programı arayüzü tabanlı analitik hizmetler ve bulut tabanlı çözümlerde becerilere ihtiyacı bulunmaktadır. Veri bilimciler genellikle işverenlerinden, gelişen alana ayak uydurmak için teknik ve iş becerilerini geliştirmelerine yardımcı olmalarını bekler, bu nedenle işverenlerin bir sertifika yol haritası oluşturması yararlı olabilir. Bu, yani işyeri veri bilimi eğitimi sunmak yalnızca çalışan performansını iyileştirmekle kalmaz, aynı zamanda çalışanları elde tutma oranını da artırabilir. Örneğin IBM, veri bilimcilerini becerilerini geliştirmeye devam etmeye teşvik etmek için bir Veri Bilimi Mesleği oluşturdu; “beceri etiketlerini” ve başarıları temsil eden meta verileri içeren “açık rozetleri”



Coursera veri bilimine dair çeşitli sertifika programları sunmaktadır.

hem çalışanların becerilerini teyit etmek hem de meslektaşlar arasında sosyal bağlantıları geliştirmek için kullanılıyor.

IBM çalışanları için veri bilimi eğitimi fırsatları şunları içerir:

- **Veri Bilimi Eğitim Kampı:** Yeni veri bilimi çalışanları, çeşitli veri bilimi kavramları hakkında farkındalık geliştirebilir ve 8 gün içinde diğer uygulayıcılarla ağlar oluşturabilir.
- **Veri Bilimi Deneyimi:** Bilim insanları sorunları çözmek için veri analitiğini kullanarak işbirliği yapar.
- **Büyük Veri ve Analistik Üniversitesi:** Katılımcılar, üç uzmanlık seviyesinden birinde sanal veri bilimi kurslarına kaydolur.
- **Analistik Ürün Kursu:** Kısa kurslar, mevcut IBM ürünlerine genel bakış sağlar.
- **Geliştirme Faaliyetleri:** Çalışanlar, aylık temel bilgiler kursları için konular seçer.
- **Analistik Eğitim Dizisi:** Çalışanlar, doğal dil işleme veya zaman-mekansal analistik gibi konularda IBM uzman derslerinin 30'dan fazla 1 saatlik videosu arasından seçim yapar.
- **Bilişsel Akademi:** Veri bilimcileri, veri görselleştirme veya makine öğrenimi gibi alanlarda eğitim alır.
- **Kurum Çapında Analistik - IBM, Büyük Veri ve Analitikten İş Değerini Nasıl Gerçekleştirir:** Ders kitabı, veri analitiği kullanılarak çözülen sorunlara ilişkin 32 örnek olay içerir.



Ayrıca IBM'de yeni işe alınan pek çok kişinin bilişsel bilgi işlem konusunda bilgi eksiği olduğu ve iş sorunlarını çözmek için yapılandırmamış verileri daha iyi kullanma becerilerine sahip olduğu belirtilmektedir ve bu noktada da daha güçlü nicel temellerden, daha iyi iletişim becerilerinden ve daha fazla merak ve sabırdan faydalananmayı ummaktalar. Her gün değişen dünyada veri bilimi gelişmeye devam edeceğinden, veri bilimi eğitimindeki bilgi eksikliği konusu uzun bir süre gündemde kalmaya devam edeceğe benzemektedir. IBM bu soruna yönelik gelecekte bünyesine bir tür çıraklık programları dahil edebilir. Bunun yanı sıra paydaşlar tarafından veri bilimi adına kuruluşların işyeri eğitimi için fikirlerini paylaşabileceği bir halk eğitim sistemi oluşturulması da yararlı olabilir.

Farklı kuruluşlarda da deneysel tasarımda yeni işe alınan kişiler arasında önemli bir bilgi eksikliği söz konusudur. Dolayısıyla bu alandaki eğitimin önemi giderek artmaktadır. IBM bu konuya ilgili olarak meslektaşlarına deneysel tasarım konusunda yardımcı olmak için deneysel fizikçileri işe almaktadır ve ayrıca Altı Sigma isimli bir kursta deney tasarımını öğretmektedir. Bunun yanı sıra mevcut bilgisayar bilimleri lisans programlarındaki eksikliklerle ilgili olarak yapay zeka, derin öğrenme ve doğal dil işlemede daha fazla hazırlık yapılması bu konudaki eksiklikleri gidermede etkili olabilir.

Diger taraftan çalışan eğitiminde bilgi eksikliğini giderme ihtiyacının veri bilimi alanına özgü bir sorun değildir. Veri bilimi eğitimi olmayan çalışanların da araçları verimli bir şekilde nasıl kullanacağını ve verileri nasıl analiz edip paylaşacağını öğrenmesi gerekmektedir. Veri bilimindeki uzmanlık alanlarında ise çalışanların iletişim kurma, hikaye anlatma, müşterilerle çalışma, şirkette bir ekiple çalışma, hedef kitleyi anlamak ve doğru yaklaşımı seçmek gibi “danışmanlık becerilerini” öğrenmesi yönünde ciddi bir ihtiyaç söz konusudur. Bu konuda Booz Allen Hamilton, danışmanlık firmalarında çalışanların vasıflandırılmasında benzersiz altyapı kısıtlamaları olduğundan, çalışanlarına ekip çalışması ve sunum becerilerinde gerekli bilgi ve ihtiyaçları içeren çeşitli eğitim programları sunmaktadır:

- **Data Science Bowl:** Dünyanın dört bir yanından yaklaşık 2.000 ekip, sosyal fayda için bu 90 günlük çevrimiçi hackathon'a katılıyor.
- **Tech Tank:** Matematik ve bilgisayar bilimleri bölümünü içeren, bir yüksek lisans sertifikası programına benzer şekilde, bir süpervizörün adayı gösterilmesi üzerine bilimsel geçmiş olan çalışanlara işe alındıktan sonraki 2 yıl içinde 12 ay boyunca 160 saatlik eğitim sunulur. Teknik eğitime ek olarak katılımcılar iletişim becerileri ve mentorluk konusunda eğitim (kişilik testi sonuçlarına dayalı olarak) alırlar.



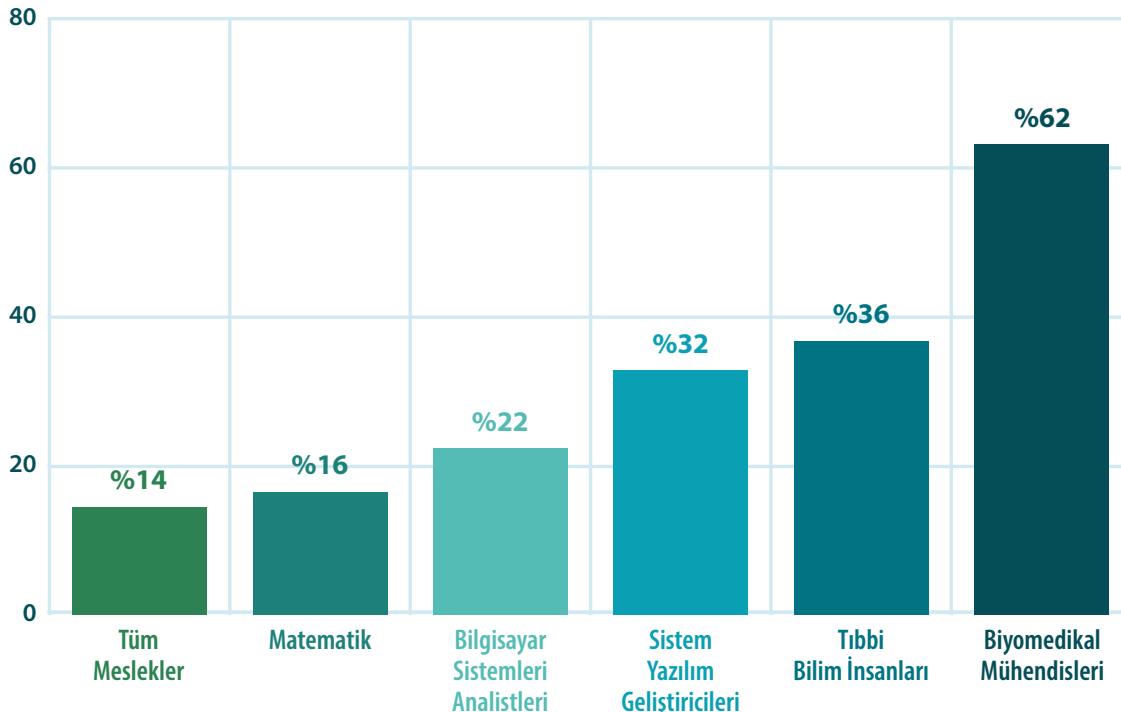
Booz Allen, Orta Gürcistan'da bir STEM ekosistemi inşa ediyor. (Kaynak: <https://www.wbcsd.org/Overview/News-Insights/3blnews?mid=305416&pgno=1&fdpgno=1>)

- **Staj Programı (Yaz Oyunları):** Yaklaşık 300 lisans stajyeri STEM odaklı sorunlar üzerinde çalışır ve 9 haftalık bir oturumda Booz Allen Hamilton liderliğine adım atar.
- **Data Science 5K Challenge:** Tech Tank'a benzer, tek fark eğitimin Booz Allen Hamilton liderliği yerine, harici bir satıcı tarafından verilmesidir. Bu, daha fazla kişinin uygun düzeyde katılmasına olanak tanır ve şirketin personeldeki toplam veri bilimcisi sayısını artırmasına yardımcı olur.

Booz Allen Hamilton ayrıca çalışanların veri bilimiyle daha fazla ilgilenmeleri için esnek yollar sunarak bir veri bilimi kitap kulübü, Yammer grupları, iki ayda bir Hackathonlar, seçkin bir konuşmacı dizisi, ara sıra eğitim kampları ve bir atölye dizisi gerçekleştiriyor. Booz Allen Hamilton stajyerlerinin ve çalışanlarının müşterileri daha iyi anlamasına veya çözmeye yardımcı olduğu ek veri bilimi problemleri hakkında şu projelerin altını çizmektedir: (1) kalp fonksiyonunu değerlendirmek için kardiyolojiye analitik uygulama, (2) hayvan barınaklarındaki benimseme oranlarını artırmak için veri analitiğini kullanmak, (3) Amerika Birleşik Devletleri'nde insan ticaretini daha iyi anlamak için ağ analizini kullanmak ve (4) evlerin elektrik şebekesinden çıkışına yardımcı olmak için veri analizi ve teknolojiyi kullanmak. Booz Allen Hamilton, programın değerini ölçmek için Tech Tank katılımcılarının faturalandırılabilir saatlerini ve promosyonlarını takip eder, ancak bu bilgiler dışarıda paylaşılmaz.

Bu değerlendirmeler katılımcıların çalışmaları için sosyal medya "rozetleri" veya ödül puanları kazanma fırsatından çok, gerçek veri setlerini kullanarak gerçek sorunları çözme fırsatıyla daha fazla teşvik edildiğini ortaya koymuştur.

Nugent, şirketler ve üniversiteler arasında özellikle öğrenci beceri değerlendirmeleri açısından artan işbirliğini teşvik etti, böylece şirketler en uygun çalışanları işe alıyor. Berkeley, California Üniversitesi'nden Deborah Nolan'ın beceri geliştirme zaman çizelgesilarındaki sorusuna yanıt olarak Lanier, yeni işe alınanların iletişim, liderlik ve sunum becerilerini hemen geliştirmeye başladığını belirtti. Bunu yapmak, yeni çalışanların hangi projelerle uyumlu olması gerektiğini belirlemeye de yardımcı olur. Lanier, Ullman'ın mevcut bilgisayar bilimleri lisans programlarındaki boşluklarla ilgili bir sorusuna yanıt olarak, makine öğrenimi ve sunum becerilerinde daha fazla hazırlık görmek istediğini söyledi. Concord Konsorsiyumu'ndan William



Şekil 14.1 STEM İşlerinde Öngörülen Artışların Yüzdesi: 2010-2020. (Kaynak: https://www.researchgate.net/figure/Projected-percentage-increases-in-STEM-jobs-2010-2020-1_fig1_315351764)

Finzer, yeni ortaya çıkan veri bilimi eğitimi araştırma alanının eğitimindeki zorlukları ele alıp alamayacağını sordu. Lanier, Booz Allen Hamilton’ın şu anda dahili problem çözmeyi kolaylaştırmak için araştırma işbirliği oturumları, hızlı yenilik atölyeleri ve tasarım odaklı düşünme egzersizlerini kullandığını belirtti.

14.3 İŞYERİNDE VERİ BİLİMİ EĞİTİMİ: YÖNETİCİ EĞİTİMİ

Johns Hopkins’ın Data Science Specialization programı Coursera aracılığıyla “Veri Bilincisinin Araç Kutusu,” “R Programlama,” “Veri Alma ve Temizleme,” “Keşif Veri Analizi,” “Tekrarlanabilir Araştırma,” “İstatistiksel Çıkarım,” “Regresyon Modeleri,” “Pratik Makine Öğrenimi,” “Veri Ürünü Geliştirme” isimli çeşitli dersler sunmaktadır.

Dersler büyük miktarda içerik içerdiginden eksiksiz bir veri bilimi müfredatı sunmaya çalışması açısından benzersiz görünülmektedir; R markdown’da GitHub’daki tüm ders notlarını sağlar ve neredeyse yalnızca R’yi kullanır; Etkileşimli R Öğrenme



John Hopkins Coursera aracılığıyla veri bilimi uzmanlığı programları sunmaktadır.

(Swirl5) ile istatistikleri kullanır; Leanpub aracılığıyla ücretsiz ders kitabı indirmeye izin verir ve mezunların programı tamamladıktan sonra bağlantı kurmaları için bir LinkedIn alanı sunar.

Dünya genelinde şirketler endüstri yöneticilerinin genellikle alt düzey çalışanlarından daha az teknik beceriye sahip olduğunu gördüklerinden, yöneticileri ekiplerini yönetmek için doğru becerilerle donatmak için özel bir eğitim programına acil ihtiyaç olduğunu belirtmişlerdir. Johns Hopkins, veri bilimi yönetimine genel bir bakış sağlayan Yönetici Veri Bilimi Uzmanlığını oluşturmak için Veri Bilimi Uzmanlığını uyardı. Yönetici Veri Bilimi Coursera müfredatı, her biri yalnızca 1 haftada tamamlanacak şekilde tasarlanmış farklı içerikten oluşan dört ders içerir:

1. Veri Biliminde Hızlandırılmış Kurs: Örneklerle istatistiğin üst düzey genel görünümü; makine öğrenimi, veri bilimi için yazılım mühendisliği, veri bilimi deneylerinin çıktıları, başarı tanımları ve veri bilimi araç kutusu.

2. Bir Veri Bilimi Ekibi Oluşturma: Veri bilimcileri ve veri mühendisleri arasındaki farklılıklara ve birlikte nasıl etkili bir şekilde çalışabileceklerine dair genel bir bakış.

3. Veri Analizini Yönetme: Veri bilimcileri tarafından sorulan soru türlerine, sağlam soru oluşturan niteliklere, keşif analizlerine, çıkarımlara, tahminlere, yorumlamaya, modellemeye ve iletişime genel bakış.

4. Gerçek Hayatta Veri Bilimi: Temiz veri çıkarımları, dikkatlice tasarlanmış deneyler ve net sonuçlar dahil olmak üzere veri analizi için ideal hedefler ve kararlarının net olmadığı veya veri ürünlerinin etkisiz olduğu durumlar için stratejiler.

Orijinal programa benzer şekilde, yönetici programı aktif öğrenmeyi vurgular ve dersin tamamlanmasının ardından bir Bitirme Projesi sunar. Geçtiğimiz yıl 2.020 kişi, yürütme programındaki Bitirme Projesi'ni tamamlamıştır ve yüzde 99'u olumlu derecelendirmelerle ödüllendirilmiştir. İçerikler özellikle yöneticiler düşünülerek tasarlanmıştır, ancak kursa yöneticilerin gerçekten kayıt olup olmadığı teyit edilmez. Kurs videoları ve materyalleri ücretsiz olarak izlenebilmesine rağmen öğrencilerin kursu tamamladıktan sonra sertifika almak için ödeme yapması gerekmektedir. Diğer taraftan yüksek maliyetli çevrimiçi kurslara katılanların genel olarak yalnızca yüzde 30'u katıldığı programı tamamlama oranına sahiptir.

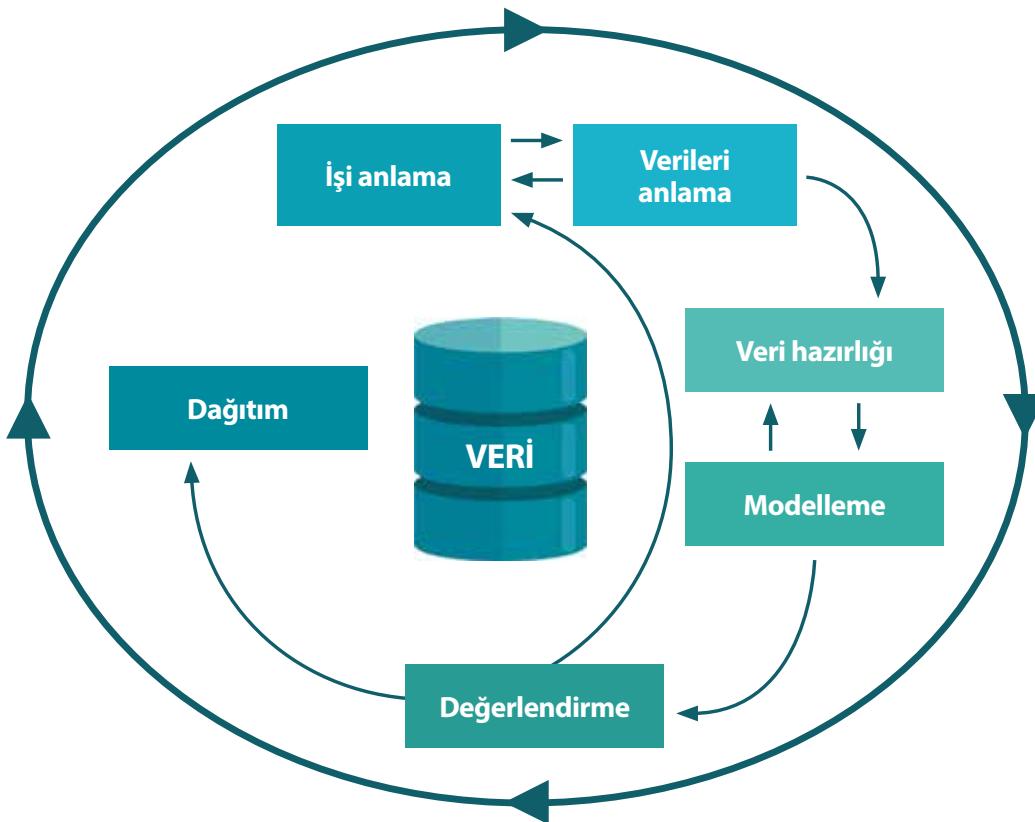
Bunun yanı sıra çevrimiçi kurs içeriğini etkili bir şekilde sunmak isteyen öğretim üyeleri için profesyonel gelişim de bir gereklilik olarak önem arz etmektedir. Çalışanlar için de bir performans incelemesinin bir bileşeni veya bir para ödülü çalışanı bir yönetici kursunu tamamlaması için teşvik ve motive edici bir araç olabilir. Ayrıca maliyetli ve teşvikin olmadığı veya yer almadığı durumlar için de insanları katılabileceği çok düşük maliyetli kursların gerekliliği göz önünde bulundurulmalıdır.



Coursera veri bilimi alanında birçok kurs vermektedir.



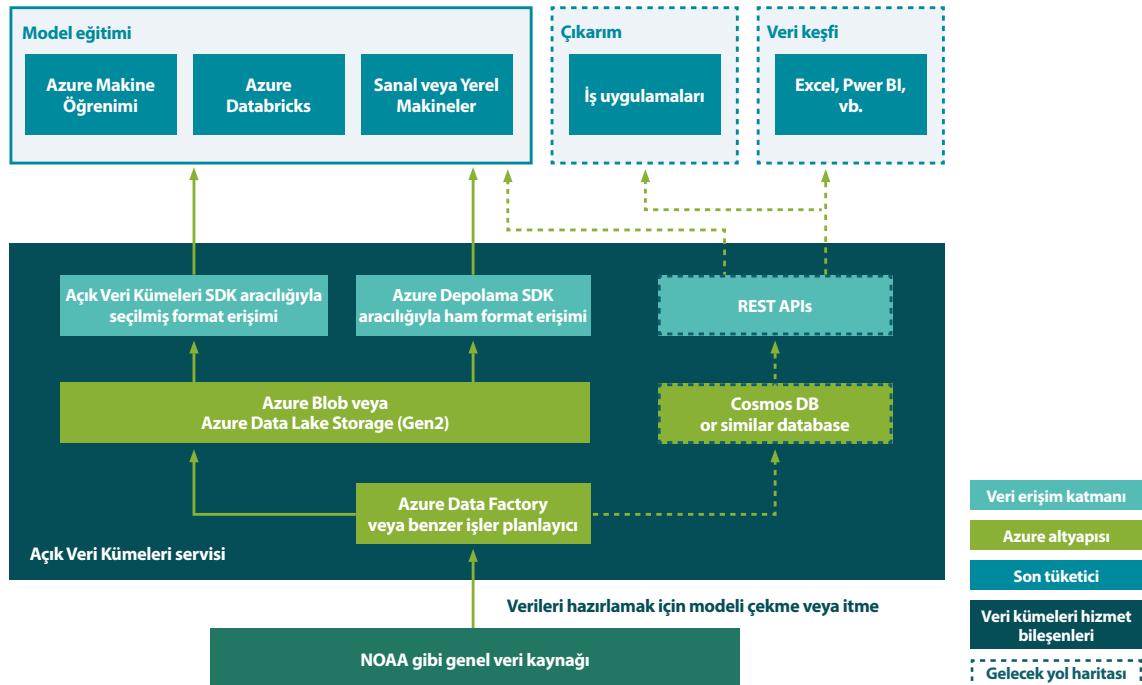
Perlich, New York Üniversitesi’nde sunduğu İş Zekası için Veri Madenciliği başlıklı, işletme yönetimi (MBA) öğrencileri için iki yol sunan bir kursu anlatmaktadır: Teknik ve yönetimsel. Teknik kısım Veri Bilimi Merkezi ve bilgisayar bilimi bölümü ile işbirliği içinde verilir ve yalnızca Python öğretilir. Yönetimsel kısım ise genellikle herhangi bir programlama becerisi olmayan ancak veri bilimini nasıl yöneteceğini öğrenmek isteyen öğrencilere yönelikdir. Bu derste veri bilimi (1) terminolojisini; (2) yöntemlerini (örneğin denetimli ve denetimsiz öğrenme, model değerlendirme, veri işleme); (3) uygulamalarını (örneğin vaka çalışmaları, Weka6) ve (4) yönetimini (örneğin dağıtım, işe alma, görüşme ve teklif değerlendirme) anlatılır. Bu içerik haftalık dersler, konuk konuşmacılar, ev ödevleri, final sınavı ve final ekibi projesi aracılığıyla verilir. Kurs projesi öğrencilerin bir sorunu tanımlamasını, veri bulmasını, sorunu çözmесini, iş değerini göstermesini, yazılı bir rapor sunmasını ve sınıfı sunmasını gerektirir ki bunların hepsi önemli veri bilimi becerileridir (Şekil 14.2).



Şekil 14.2 New York Üniversitesi İş Zekası için Veri Madenciliği dersinde, veriler aracılığıyla ele alınabilecek bir problemin formülasyonunu vurgulayan yinelemeli bir veri bilimi sürecini izler. (Kaynak: Kenneth Jensen, 2012, "File:CRISP-DM Process Diagram.png," courtesy of CC-BY-SA-3.0.)

Perlich'in görüşüne göre, öğrenciler genellikle tahmine dayalı bir modelleme problemini tanıtmakta, temel ana hatların değerini anlamakta, bir modeli eyleme dönüştürmekte, veri hazırlama için zaman ayırmakta güçlük çekiyor. Bununla birlikte yönetimsel kısmın tamamlanmasının ardından, öğrencilerden bazı şeyler beklenmektedir. Bunlar:

- Performansı artırmak için veri analitiğini kullanarak iş sorunlarına dikkatlice yaklaşma ve bir veri bilimciyi nasıl işe alacağını öğrenme;
- Veri hazırlığının zaman almasına rağmen gerekli olduğunun farkına varma;
- Tüm problemlerin veri bilimi problemleri olmadığını kabul etme;
- Veriden ileri doğru değil, sorundan geriye doğru düşünme;
- Veri madenciliği süreçleri, algoritmaları ve sistemlerinin temellerini bilme;
- Madencilik verileriyle ilgili uygulamalı deneyime sahip olma.



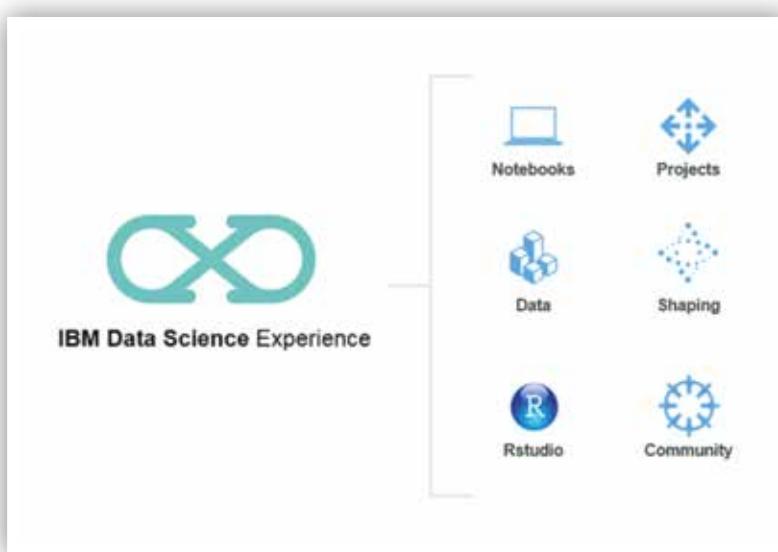
Şekil 14.3 Açık Veritabanı Servisi: Bloklar kurma. (Kaynak: <https://docs.microsoft.com/en-us/azure/open-datasets/overview-what-are-open-datasets>)

Veri bilimine dair teknik donanımın yanı sıra öğrencilerin işverenlerin istediği becerileri kazanmalarına yardımcı olma konusunda, onlara iletişim becerileriyle ekip çalışmasını öğretmeninin gerekliliği de yadsınamaz bir gerçek olmasına rağmen bu becerileri kazandırmak oldukça güçtür. Programlama olmadan bir ders sunmak her ne kadar değerli olsa da veri bilimi müfredatındaki teknik içerikte bir düşüş yaşanması ihtimali de söz konusu. Veri bilimci çalışanları arasında genel olarak veri okuryazarlığı, verileri paylaşma ve departmanlar arasında iletişim kurma konusunda zorluklar yaşamaktadır. Özellikle kamu sektöründeki işe alma fırsatlarının özel sektörde göre daha sınırlıdır. Bu sorunu aşmak adına kuruluşların kendilerine hem çekirdek çalışanların hem de çevrelerindeki çalışanların becerilerini geliştirmek için neye ihtiyaç duyulacağını sormaları ve personel eğitimiyle ilgili yeni bir sorun ortaya çıktığında uygulanabilecek bir “ürün-eğitim-süreç” döngüsü geliştirmeleri uygun olabilir. Stajyerler de kısa vadede belirli veri bilimi ihtiyaçlarını karşılayabilir.

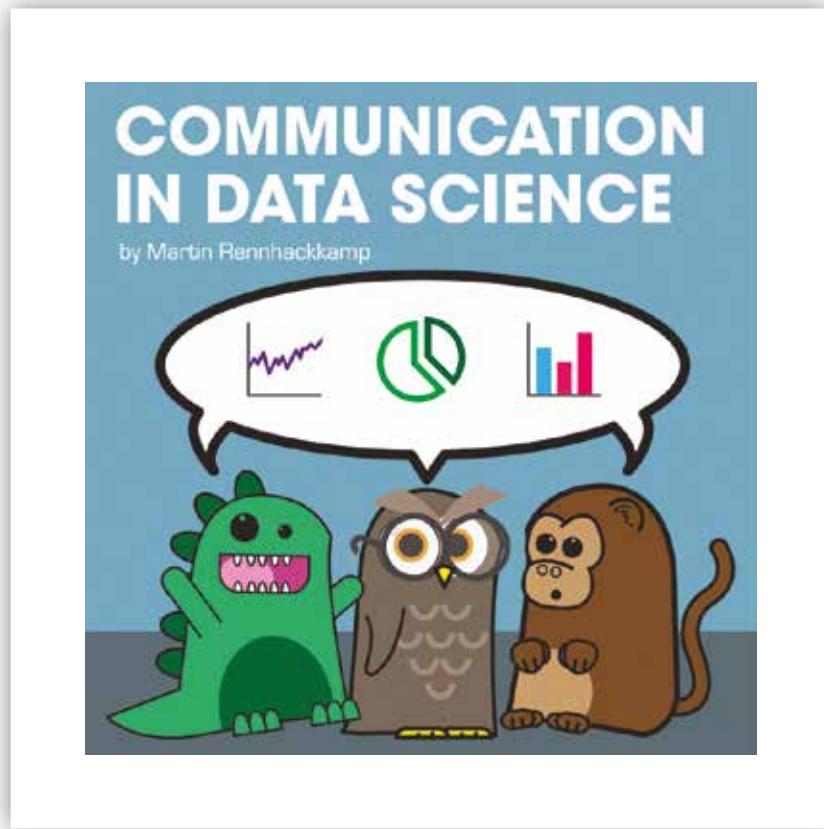
College Park, Maryland Üniversitesi’nden David Levermore ise öğrenciler için uygulamalı projeler oluşturmaya çalışan öğretim üyeleri için verilere erişim eksik-

liğinin sorun olduğunu bildirmiştir. Endüstri ve hükümetin verilerini öğrencilerin ders çalışmaları için üniversitelere sunmasının ne kadar yararlı olacağını belirtmiştir. Buna rağmen öğretim üyelerine sınıflarında verileri kullanmaları için açık erişim vermek yetersiz olabilir. Fakültelerde çok zaman alan bir süreç olan müfredat hedeflerine uygun hale getirmek için önce veriler analiz edilmeli ve hazırlanmalıdır. Bazı küçük IBM veri kümeleri halka açık kullanım için bir korumalı alanda paylaşılmaktadır, ancak IBM'den Laura Haas veri lisanslamasının zor olabileceğini ifade etmiştir; çoğu şirketin telif hakkıyla korunan verileri yanlışlıkla erişilebilir hale getirmek için yasal işlem yapma riskini almış istedğini belirterek bunun da verilerin üniversitelerle paylaşılmasında önemli bir engel oluşturduğuna dechinmiştir. Devlet kurumlarının veri paylaşımında benzer engellerle karşılaşabileceğini, çünkü açık olduğu varsayılan bazı verilerin aslında telif hakkıyla korunan materyalleri içerebileceğini öne sürmüştür.

Boston Üniversitesi'nden Eric Kolaczyk, birden fazla kuruluşa yayılan sanal alanların bir ekipteki uzmanlarla yinelemeli olarak çalışma konusunda bütünsel bir deneyim sunduğunu açıklamıştır; tek başına veri havuzlarının kullanımı yetersizdir. Başarılı sanal alan deneyimleri oluşturmak pahalı olabilir, enerji ve zaman gereklidir ve paydaşlar arasında kurulan ilişkilere bağlı olabilir. Ancak ekiplere verileri



IBM Veri Bilimi Deneyimi. (Kaynak: https://medium.com/@armand_ruiz/welcome-to-the-data-science-experience-250fa2e784a2)



Kaynak: <https://www.martinsights.com/?p=1256>

depolamak ve işbirliği yapmak için bir yer sağladığı için IBM'in Data Science Experience'ın ücretsiz beta sürümü göz önünde bulundurulabilir. Kullanıcıların IBM verilerine ve IBM ekip üyelerine erişimi olması halinde bu daha da kullanışlı bir platform olabilir.

14.4 SONUÇ

Tüm bunların yanı sıra veri bilimcileri için iletişim becerilerini benzersiz kıyan şeyin ne olduğu da bir başka sorunsal olarak kendini göstermektedir. Google'dan Patrick Riley geleneksel teknik alanlarda çalışan insanlar ağırlıklı olarak diğer teknik alanlardaki insanlarla konuşurken, veri bilimcilerinin teknik alanlarda çalışmayan kitlelere zor kavramları açıklayabilmeleri gerektiğini ifade etmektedir. Bu noktadan hareketle öğrencilerin teknik alanda olmayan insanlar için problemleri net bir şekilde çerçevellemeyi ve açıklamayı öğrenmeleri gerekmektedir. Öğrenciler analizlerinin özetlerini

çeşitli kitlelere sunarak gereken alıştırma fırsatından yararlanabilir. Diğer taraftan bilimsel fikirlerin başkalarına nasıl iletileceğine yönelik yöntemler herhangi bir veri bilimi müfredatına entegre edilebilir.

Ayrıca eğitimcilerin sadece farklı bilgi seviyelerine değil, farklı geçmişlere ve bakış açılarına sahip ekipler oluşturmaya odaklanması da önem arz eden başka bir konu olarak varlığını göstermektedir. Geniş ekip çalışması becerilerinin öğretildiği ve bu becerinin kazandırıldığı bir üniversite öğrencileri çeşitli çalışma ortamlarına en iyi şekilde hazırlayabilir ve bunu yapmanın bir yolu da öğrencilere bireysel olarak çözülmesi imkansız problemler vermek olabilir. Öğrencileri deneyimsel öğrenmeye dahil etmek ve sorunları tanımlamakla iyileştirmek için teknik ve teknik olmayan kişileri bir araya getirmek bu bağlamda oldukça önemlidir. Öğrencileri disiplinlerarası ekiplerde daha işbirliği içinde çalışmaya hazırlamak için farklı disiplinlerdeki benzersiz kelime dağarcığına ve yaklaşımlara maruz bırakmanın da değeri göz ardı edilmemelidir. Sanat becerileri (örneğin eleştirel düşünme, soyutlama) daha iyi veri bilimcileri yetiştirmeye konusunda büyük fayda sağlayabilir. Ayrıca bu tür yaklaşımlar ve değerlerin yükseköğretim öncesi çocuklara kazandırılması da bu alanda büyük avantaj sağlayabilir.

KAYNAKLAR

- ACM (Association for Computing Machinery). 2018. Lighting the Path from Community College to Computing Careers. <https://www.acm.org/binaries/content/assets/education/lighting-the-path-from-community-college-to-computing-careers.pdf>.
- Angwin, J., J. Larson, S. Mattu, and L. Kirchner. "Machine Bias." ProPublica, May 23. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Baker, M. 2016. "1,500 Scientists Lift the Lid on Reproducibility." News Feature. Nature. May 25. <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>.
- Barone, L., J. Williams, and D. Micklos. 2017. Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. PLoS Computational Biology 13(10):e1005755. <https://doi.org/10.1371/journal.pcbi.1005755>.

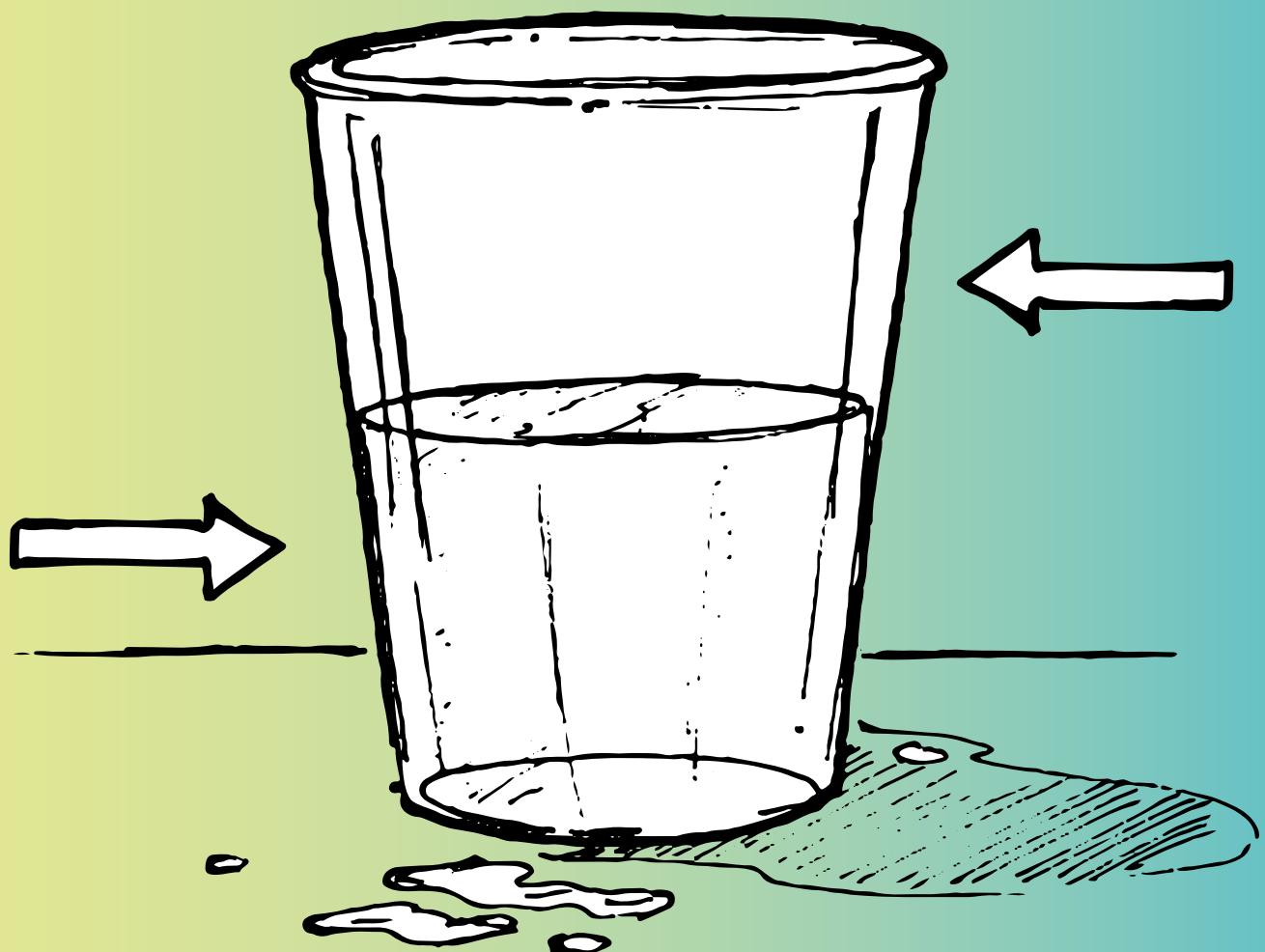
- Bloom, B.S. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. White Plains, NY: Longman.
- Buckheit, J.B., and D.L. Donoho. “WaveLab and Reproducible Research.” https://statweb.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf.
- CCC (Computing Community Consortium). 2015. “The Future of Computing Research: Industry-Academic Collaborations.” Volume 2. <https://cra.org/ccc/wp-content/uploads/sites/2/2016/06/15125-CCC-Industry-Whitepaper-v4-1.pdf>.
- CCC. 2019. “Evolving Academia/Industry Relations in Computing Research: Interim Report.” <https://www.cccblog.org/wp-content/uploads/2019/03/Industry-Interim-Report-w-footnotes.pdf>.
- Chapman, P., J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. “CRISP-DM 1.0: Step-By-Step Data Mining Guide.” <https://www.the-modeling-agency.com/crisp-dm.pdf>.
- CMU (Carnegie Mellon University). 2017. “Carnegie Mellon University Educational Project Agreement.” <https://www.ri.cmu.edu/wp-content/uploads/2017/01/Educational-Project-Agreement.pdf>.
- Conway, D. 2010. “The Data Science Venn Diagram.” <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.
- Cramer, C., M. Porter, H. Sayama, L. Sheetz, and S. Uzzo. 2015. “Network Literacy: Essential Concepts and Core Ideas.” <http://tinyurl.com/networkliteracy>.
- Dieterich, W., C. Mendoza, and T. Brennan. 2016. “COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity.” Northpointe, Inc., Research Department. <https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html>.
- Dwork, C., F. McSherry, K. Nissim, and A. Smith. “Calibrating Noise to Sensitivity in Private Data Analysis.” In *Theory of Cryptography* (S. Halevi and T. Rabin, eds.). TCC 2006. Lecture Notes in Computer Science 3876. Berlin, Heidelberg: Springer.
- EDC (Education Development Center, Inc.). 2017. “Tools for Building a Big Data Career Pathway.” <http://oceansofdata.org/sites/oceansofdata.org/files/Tools%20for%20Building%20a%20Big%20Data%20Career%20Path.pdf>.
- FTC (Federal Trade Commission). 1998. *Privacy Online: A Report to Congress*.

- <https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf>.
- Gould, R., R. Peck, J. Hanson, N. Horton, B. Kotz, K. Kubo, J. Malyn-Smith, M. Rudis, B. Thompson, M.D. Ward, and R. Wong. 2018. The Two-Year College Data Science Summit: A Report on NSF DUE-1735199. <https://www.amstat.org/asa/files/pdfs/2018TYCDSFinal-Report.pdf>.
- Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Medicine* 2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Levy, R., R. Laugesen, and F. Santosa. 2018. *Big Jobs Guide*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Meyer, M., A. Cimpian, and S.J. Leslie. 2015. Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology* 6:235.
- Microsoft Azure. 2017. “What Is the Team Data Science Process?” <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>.
- NAE (National Academy of Engineering). 2016. *Infusing Ethics into the Development of Engineers*. Washington, DC: The National Academies Press.
- NASEM (National Academies of Sciences, Engineering, and Medicine). 2018a. *Graduate STEM Education for the 21st Century*. Washington, DC: The National Academies Press.
- NASEM. 2018b. *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: The National Academies Press.
- NRC (National Research Council). 2011. *A Framework for K-12 Science Education*. Washington, DC: The National Academies Press.
- NSF (National Science Foundation). 2015. Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science. Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. May.
- ODI (Oceans of Data Institute). 2014. *Profile of a Big-Data-Enabled Specialist*. Waltham, MA: Education Development Center, Inc.
- ODI. 2016. *Profile of the Data Practitioner*. Waltham, MA: Education Development Center, Inc.
- Patil, D.J., H. Mason, and M. Loukides. 2018. *Ethics and Data Science*. Sebastopol, CA: O'Reilly Media.

- Patterson, D. 2014. How to build a bad research center. Communications of the ACM 57(3):33-36.
- Rawlings-Goss, R., L. Cassel, M. Cragin, C. Cramer, A. Dingle, S. Friday-Stroud, A. Herron, et al. 2018. “Keeping Data Science Broad: Negotiating the Digital and Data Divide.” https://drive.google.com/file/d/14l_PGq4AxOP9fhJbKqA2necsJZ-gdiKV/view.
- Zweben, S., and B. Bizot. 2018. “2017 CRA Taulbee Survey.” <https://cra.org/crn/2018/05/2017-cra-taulbee-survey-another-year-of-record-undergrad-enrollment-doctoral-degree-production-steady-while-masters-production-rises-again/>.

15. BÖLÜM

SONUÇ



Yapılan bu çalışmanın ışığında, büyük veri için teknik altyapısal gereksinimlerin ve yetkin personel konusunun hassas bir önemi olduğu görülmüştür. Yönetimlerin konuya benimseyiş biçimleri, büyük verinin bazı fırsatlar getirmesinin yanında zorlukları da beraberinde getirmesi gibi konularda literatür verileri ile analiz verilerimiz ciddi anlamda örtüşmektedir.

Sonuçlar göstermiştir ki veri miktarı çok hızlı bir şekilde artmasına rağmen büyük veri daha yavaş bir şekilde gelişmektedir, fakat onun varlığı farklı sektörlerde ve çeşitli şekillerde operasyonları etkilenen kurumların departmanlarında hissedilmektedir. Kurumların yöneticileri büyük verinin potansiyel kullanımlarının farkındalar ve bir bütün olarak “büyük veri” kavramını kullanma niyetindeler, sadece sektördeki hedefleri için değil, rakiplerine karşı bir avantaj kazanmak için de bir girişim içindedirler.

Büyük verinin kurumlar üzerindeki en büyük etkisi, ağırlıklı olarak müşteri analizi ve ürün geliştirme konularında hissedilmektedir. Artan veri miktarı müşterilerin artık değişen çeşitli gruplarının azalması demektir. Bunun yerine, bir müşterinin bilgisi, örneğin önceki satın almaları ve sosyal medya beğenileri kişiselleştirilmiş bir profil oluşturmak için korelasyon olabilir. Elbette, kurumlar her bir müşteri için kişiselleştirilmiş kampanyalar ve teklifler oluşturamayabilir. Ancak daha fazla bireysel düzeyde reklam belirlemenin yanı sıra, yeni ürün geliştirmede ve kişiselleş-

Katılımcılar büyük veriyi ne sektörlerinin “kurtarıcısı” olarak görmekte ne de basit bir şekilde ilgilenilmesi gereken bir sorun ya da bir engel olarak görmektedirler.

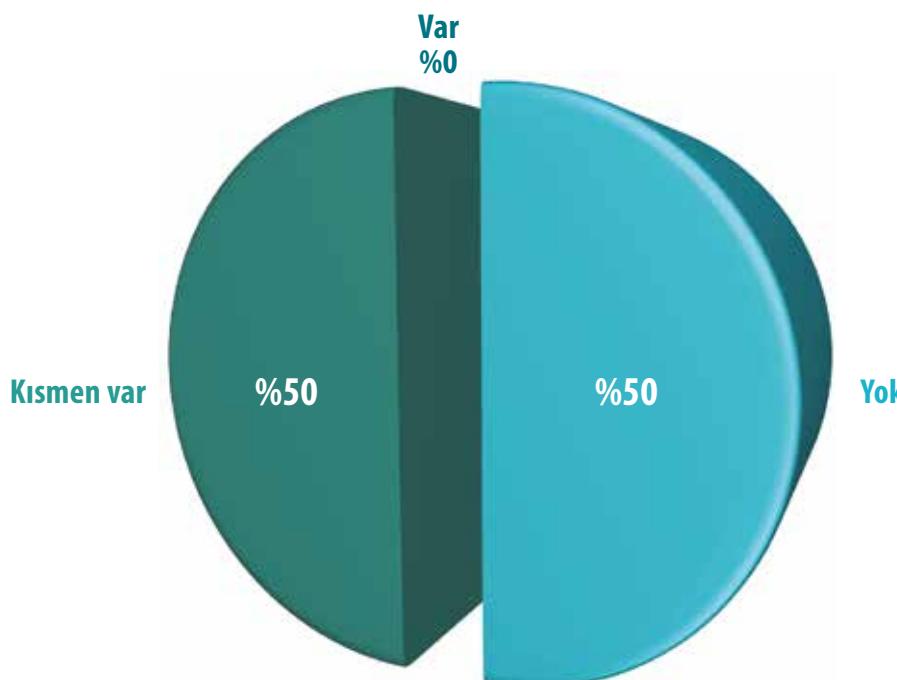
tirilmiş kampanyalar oluşturmada kullanabilirler. Bu durum büyük veri analizinin yatırımlar üzerindeki olumlu etkisi ve fırsat olarak değerlendirilmektedir.

Bunların yanı sıra katılımcıların şu noktayı da gözden kaçırmadıkları görülmüşdür. Büyük verinin gerçeğini, yani onun potansiyelleri kadar zorluklarının da olduğunu fark etmişlerdir. Onlar büyük veriyi ne sektörlerinin “kurtarıcısı” olarak görmekte ne de basit bir şekilde ilgilenilmesi gereken bir sorun ya da bir engel olarak görmektedirler. Bu da bize göstermektedir ki basındaki abartıların çoğundan ayrı olarak yöneticiler aslında işleriyle ilgili olarak günlük bazda büyük veri ile ilgilenmeye başlamıştır, onlar bu konuya gerçekçi ve pratik bir şekilde yaklaşmaktadır.

Kurumlar şirketin iç yeniden yapılanmasını önemli ölçüde değiştirmemektedir ve bu yüzden büyük verinin konularıyla başlıca olarak genelde IT departmanları ilgilenmekte ve pazarlama gibi departmanlara bilgi desteği sağlamaktadır.

Önemli ölçüde teknik birkaç konuya da katılımcılar tarafından değinilmiştir; veri depolama maliyetinin kontrolü, veri sızıntısı ve veri güvenliği konuları endişe duyulan taraflardır. Ayrıca hukuki ve gizlilik konusu da büyük veri konusunun onde gelen kaygılarındanandır. Bunlar büyük verinin potansiyel zorlukları olarak ortaya konulmuştur.

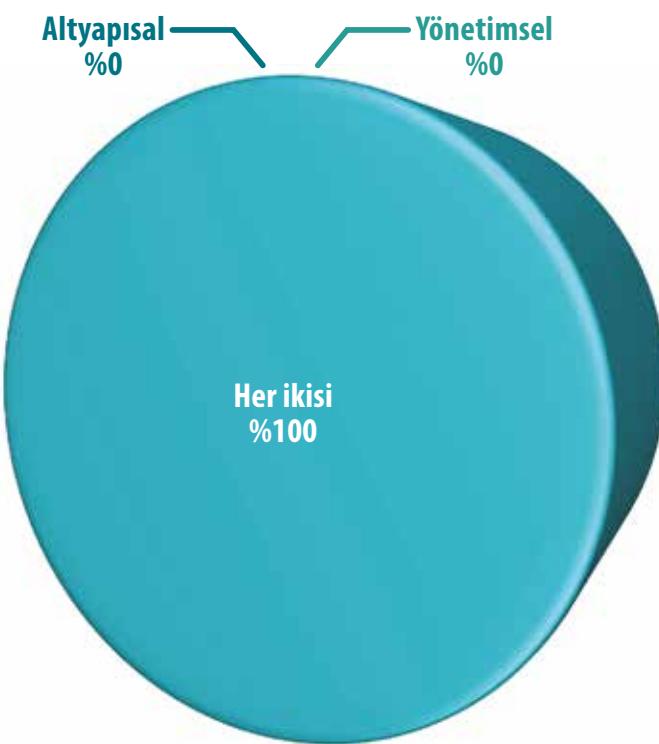
Büyük verinin yatırımlar üzerindeki etkisi sorumuzda “Kısmen” vardır şeklinde cevap veren “%50”lik grup içerisindeki bir katılımcı, “Kısmen etkili olması konunun önemsizliği değil, aksine bu kadar önemli bir konuda yapılan az çalışma yüzündendir. Önümüzdeki dönemde bahsetmiş olduğum konuların gerçekleştirilmesine yönelik yoğun bir büyük veri çalışmasının olmayacağı düşünüyorum. Yapılan çalışmaların da belirli bir konuya özel ve parça halinde olacağını, kendimce bütünsel bir yorum için yeterli olmayacağı kanaatindeyim. Yatırım için bir düşünce



Şekil 15.1 Türkiye'de büyük veri için altyapısal yeterlilik var mıdır? (Kaynak: Analiz çıktıları, 2015)

söz konusu olduğunda uzun bir süre daha büyük veri sonuçlarından ziyade alfa numarik veri ve değerlere ulaşmanın daha kolay olacağı, büyük veriden beklediğim fonksiyonların yine kişiler ve insan deneyimleri temasında gerçekleştirileceğini düşünüyorum. Kendi adıma günümüz şartlarında tutmuş olduğumuz verilerin yorumlanması konusunda kaybedilen zamanın deneyimler ile kısa sürede söylemiş bir karara kıyasla çok daha az efektif olduğunu düşünüyorum. Günümüz koşulları birçok durumda büyük veri kullanımının deneyim kullanmadan daha pahalı olduğunu bana düşündürmektedir. Yapılacak çalışmaların artması ve topluma iletilen bir veri akışının söz konusu olması durumunda kullanılacak düşük maliyetli çözümler

Büyük Veri için teknik altyapısal gereksinimlerin ve yetkin personel konusunun hassas bir önemi olduğu görülmüştür.



Şekil 15.2 Büyük veri yönetimine olan ihtiyaç, size göre altyapısal mı, yoksa yönetimsel bir konu mudur? (Kaynak: Analiz çıktıları, 2015)

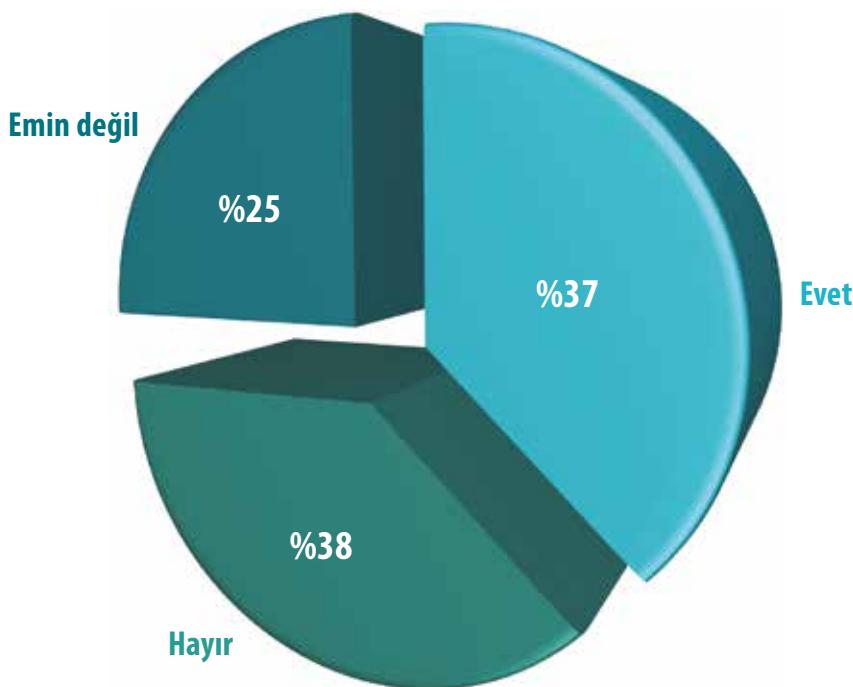
gelinceye deðin büyük veri tarafı söz konusu olası yatırımlarında ikinci planda kalacaktır. Teknik yetersizlik,” şeklinde ifade etmiştir.

Katılımcılar ile yaptığımız görüşmelerden ve aldığımız cevaplardan kapsamlı olarak şu neticeyi çıkarmaktayız:

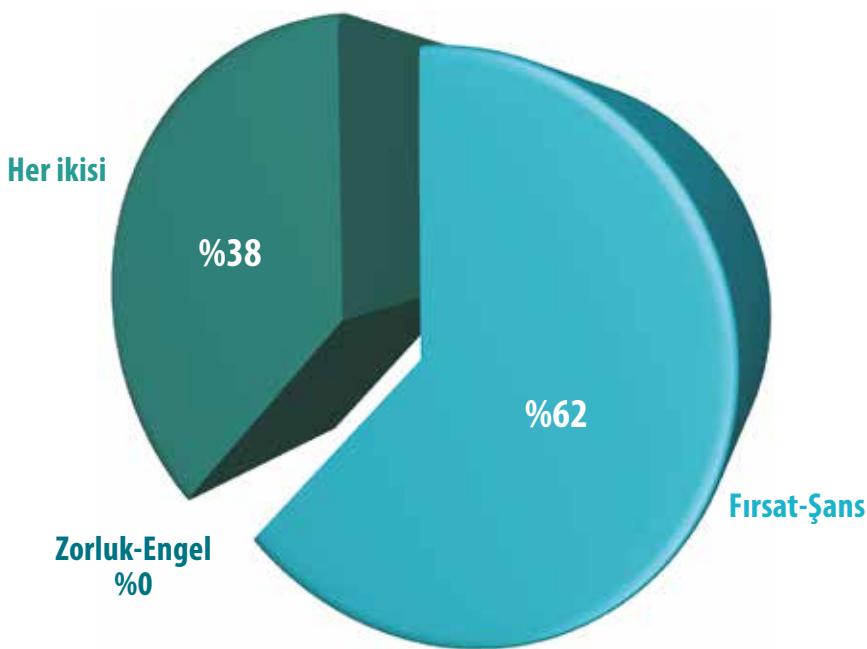
Katılımcılar büyük veri nedir, neler yapabilir, zorlukları ve fırsatları nelerdir farkındadırlar. Bununla birlikte genel kısımlarda da bahsedilen verinin işlenmesi zorluğu, teknik kısıtlar, deneyimli personel eksikliği vb. gibi sebeplerle de büyük veri konusuna gayet temkinli ve gerçekçi yaklaşmaktadır.

Nihai olarak görüşme analiz tablosunda da gördüğümüz sonuçlar ve karşılıklı edinimler üzerine şunları söyleyebiliriz:

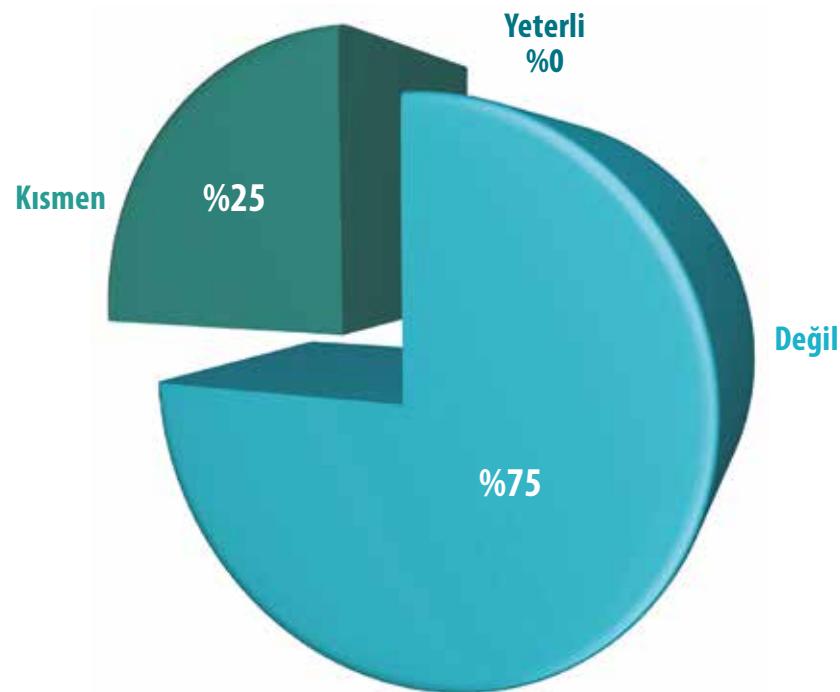
Kişi ve kurumların büyük veri kullanımını eğiliminde olduğu, fakat altyapısal ve yönetimsel yetersizlikler nedeniyle tam olarak faydalananamadığı, kökten bir değişiklik yerine “etkili ilave bir kaynak” olarak düşünüldüğü ve sınırları şu anda öngörülemeyen ve daha çok fırsat-şans olarak görülen bir paradigma olduğu sonucuna ulaşmıştır.



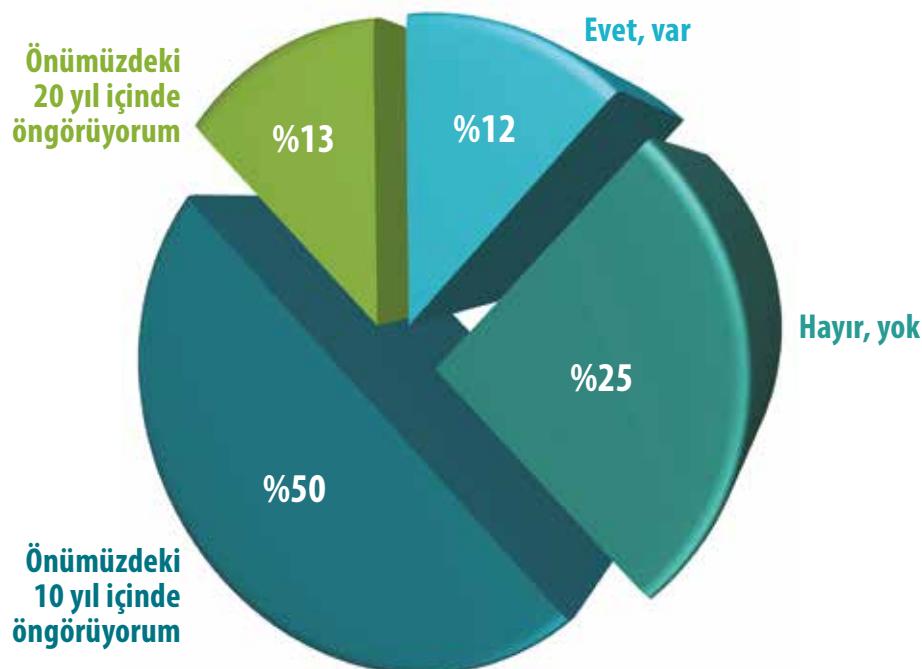
Şekil 15.3 Büyük verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz?
(Kaynak: Analiz çıktıları, 2015)



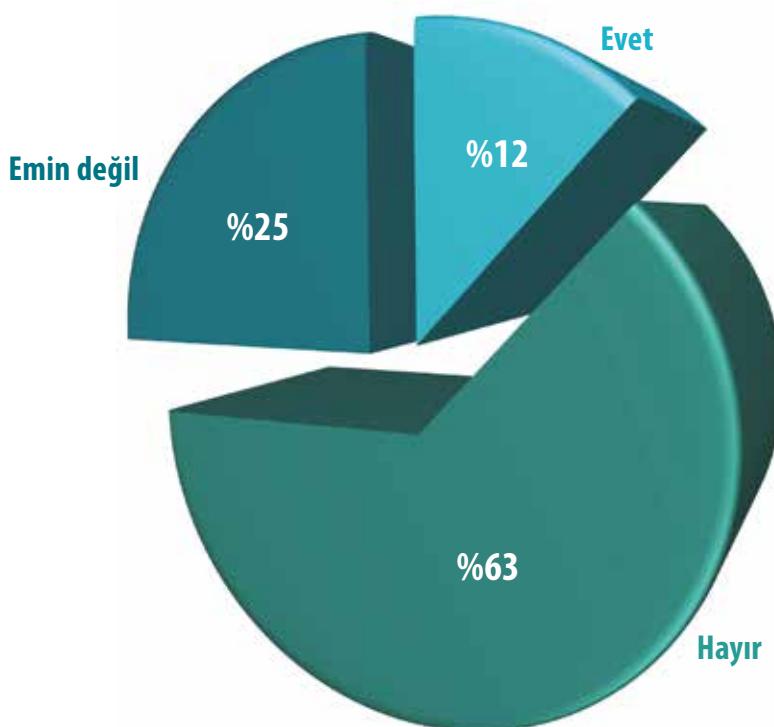
Şekil 15.4 Sizce büyük veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır?
(Kaynak: Analiz çıktıları, 2015)



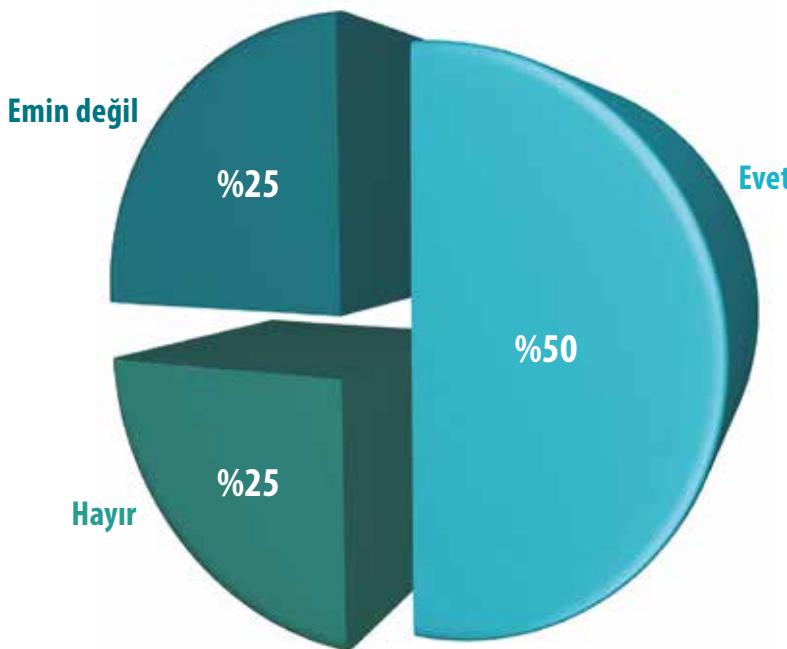
Şekil 15.5 Türkiye'de büyük veriden faydalama seviyeleri ile ilgili neler söylemek istersiniz? (Kaynak: Analiz çıktısı, 2015)



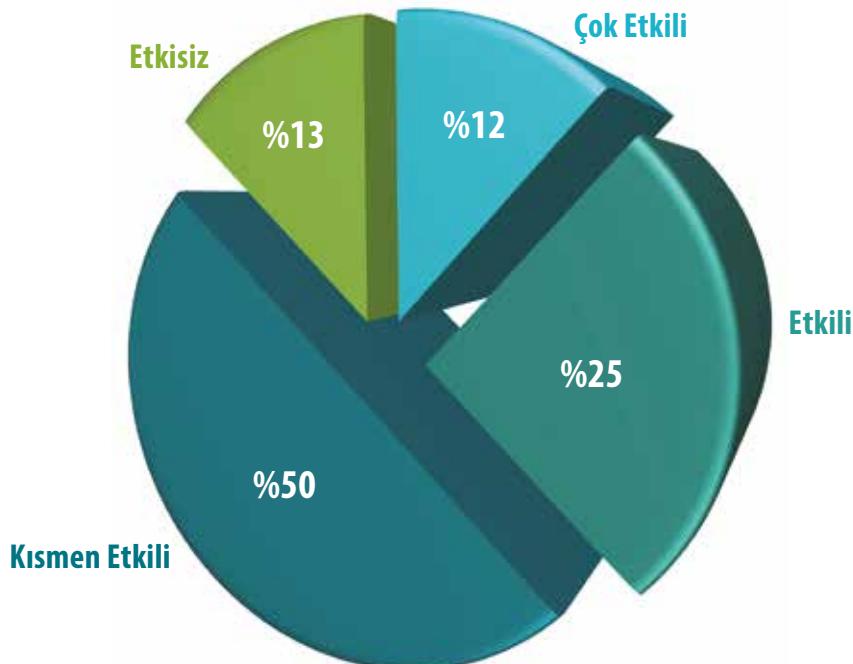
Şekil 15.6 Sizce Türkiye'de Google, Facebook, Amazon vb. gibi büyük veriden ciddi anlamda faydalanan benzeri firmalar var mıdır? (Kaynak: Analiz çıktısı, 2015)



Şekil 15.7 Büyük verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz?
(Kaynak: Analiz çıktıları, 2015)



Şekil 15.8 Büyük verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? (Kaynak: Analiz çıktıları, 2015)



Şekil 15.9 Büyük veri yatırımlarınız üzerinde ne kadar etkili olacaktır? (Kaynak: Analiz çıktısı, 2015)

Bunlara ilaveten Türkiye açısından kısmi altyapı yeterliliği, faydalananma derecesinin henüz başlangıç seviyesinde olduğu ve kurumlar tarafından ağırlıklı olarak muhtemelen önümüzdeki 10 yıl içerisinde daha efektif kullanılacağı düşünülmektedir.

Büyük veri “her derde deva bir kurtarıcı” değildir, fakat o rekabet dünyasında, elde edilmek istenen sonuçlara ulaşmada kullanılması gereken yeni bir olgudur.

Katılımcıların yatırımlara olan etkisine bakışı ise büyük veri literatür araştırmalarındaki uluslararası çaptaki veri uzmanlarının düşüncelerinden çok uzak düşmeyecek şekilde gerçekçi çıkmıştır.

Katılımcılar % 12,5'luk dilimler ile iki ucta çok etkili ve etkisiz olarak değerlendirmiştir, % 25'lik bir dilimde etkili ve % 50'lik kısmında düşünce ise kısmen etkili olarak tespit edilmiştir.

Bu değerlendirmenin Türkiye'de henüz büyük veri konusunun çok yeni olmasından ve yeterince fayda sağlanmadığından dolayı olduğu ifade edilebilir. Büyük veri analizi hizmeti yapan şirketler ve de bu analizleri uygulayan kurumlar arttıkça muhtemelen bu oranlar ciddi anlamda artacaktır.

Tahmini olarak bir perakendeci analizin gücünü kullanarak operasyon marjını %60'a kadar artırabilir. Yeni fırsatlar kullanılarak yeni gelirler için önemli oranda başı çekmektedir.

Belki de büyük veri analizlerinin en gözle görülür uygulamaları iş dünyasında-
dır. Tahmini olarak bir perakendeci analizin gücünü kullanarak operasyon marjını
% 60'a kadar artırabilir. Yeni fırsatlar kullanılarak (örneğin yer-farkındalıklu ve
yer-tabanlı hizmetler) yeni gelirler için önemli oranda başı çekmektedir.

Kapsamlı bir analitik çerçeve müşteri işlemleri, envanter yönetimi, mağaza ta-
banlı video verileri, reklam ve müşteri ilişkileri, müşteri özellikleri ve fikirleri,
satış yönetimi altyapısı, finansal veri ve diğerlerini içeren çok modlu geniş mik-
tarda veriler toplayan işletmelerin tedarik zinciri yönetimi, müşteri yönetimi, satış
sonrası destek, reklam vb. entegrasyonunu gerektirebilir.

Büyük perakendeciler için tüm verilerin toplanması kolay bir şekilde exabaytlar
seviyesine yaklaşmaktadır. Envanter izlemek için RFID'nin kapsamlı yayılması,
tedarikçi veri tabanlarına bağlanma ile, müşteri özellikleri ve profilleri ile enteg-
rasyon (mağaza sadakat programlarıyla) ve finansal sistemlere tam entegrasyonu
ile gelişmiş verimlilik potansiyeli muazzamdır ve şimdi sadece farkındalık başla-
maktadır.

Ayrıca büyük verinin uluslararası alanda ne denli ilgiyle takip edildiği ve önem
verildiğini ifade etmesi açısından şu dikkat çekici bilgiye değinmekte de fayda
vardır:

[https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_ may_1_2014.pdf](https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf) adresinden tamamina ulaşılabilen, büyük veri ile ilgili olarak
ABD Başkanının talimatıyla Beyaz Saray'da 1 Mayıs 2014 tarihli "Big Data: Sei-
zing Opportunities, Preserving Values" - Büyük Veri: Fırsatları Yakalama, Değerleri
Koruma - adlı, Başkan Danışmanı John Podesta, Ticari Sekreter Penny Pritzker,
Enerji Sekreteri Ernest J. Moniz, Bilim Ofisi ve Teknoloji Politikası Direktör John

Büyük perakendeciler için tüm verilerin toplanması kolay bir şekilde exabaytlar seviyesine yaklaşmaktadır.

Holdren ve Ulusal Ekonomi Konseyi Direktör Jeffrey Zients imzasıyla detaylı bir rapor hazırlanmıştır.

Raporun giriş kısmında şu ifadelere yer verilmiştir. “Sevgili Başkan,

Sosyal, ekonomik ve teknolojik devrimin tam ortasındayız. Nasıl iletişim kuracağımız, nasıl sosyalleşeceğimiz, nasıl zaman harcayacağımız ve işimizi nasıl yöneteceğimiz internet üzerine taşındı. İnternet telefonlarımıza, evlerimizde ve şehirlerimizde cihazlara ve endüstriyel ekonominin gücü olan fabrikalara taşındı. Veri ve keşif patlamasının sonucu dünyamızı değiştiriyor.

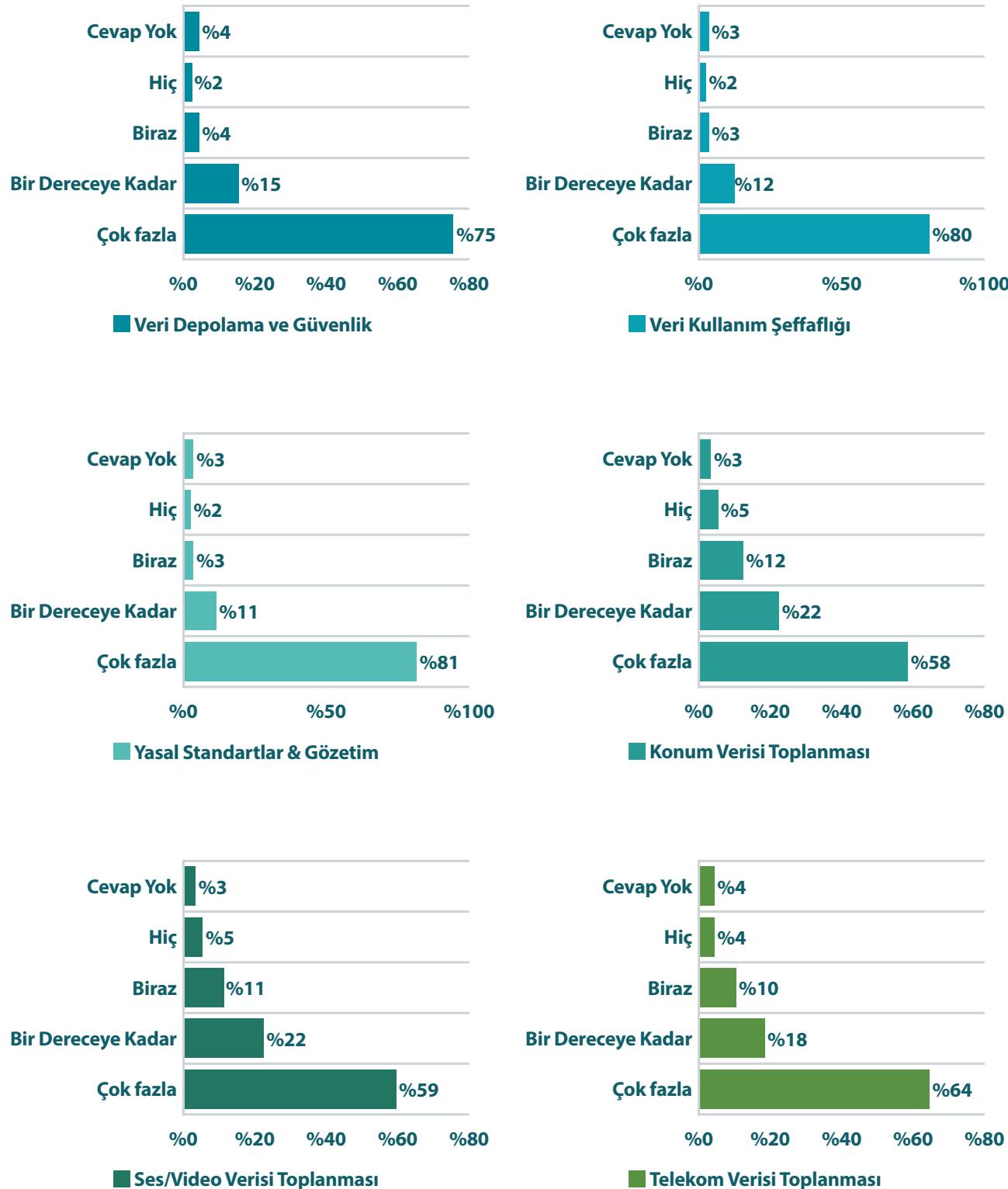
Ocak ayında, büyük verinin yaşamımızı, devlet ilişkilerini, vatandaşlarımızı, işletmeleri ve tüketicileri nasıl dönüştüreceği ile ilgili bir çalışma istediniz. Bu inceleme kamu ve özel sektörün büyük verinin risklerini minimize ederken, faydalarını nasıl maksimize edebileceğی üzerine odaklanmıştır.

Aynı zamanda bu çalışma, ekonomimizi büyütmek, sağlık ve eğitimi geliştirmek, ulusal güvenliğimizi sağlamak ve daha fazla enerji verimliliği oluşturmak için büyük veri fırsatlarını açıklamaktadır.”

Raporda büyük veri nedir, veriden farkları nelerdir, zorlukları ve fırsatları, riskleri, endişeler, hangi alanlarda nasıl kullanılabilir, kamuda ve özel sektörde veri yönetimi vb. sorulara cevaplar verilmiştir.

Yine bu raporda gösterilen birkaç ankete de burada degeinmek önemli görünmektedir. Beyaz Saray tarafından 4 hafta boyunca 24.092 kişiden toplanan yanıtlarla elde edilen veriler, tam bir istatistiksel veri olarak düşünülmese de Beyaz Saray tarafından rapora yansıtılmış ve önem verilmiştir.

Veri uygulamaları ile ilgili sorulan sorulara verilen cevapların oranları aşağıdaki grafiklerde görülmekte olup, katılımcıların ortalaması % 61 oranında “çok fazla endişeli” oldukları sonucu elde edilmiştir. Katılımcıların % 80’i veri kullanımının şeffaflığı konusunda “çok fazla endişeli” görüşündeler.



Şekil 15.10 Büyük veri ile ilgili ABD Beyaz Saray anketi. (Kaynak: White House Report, Big Data: Seizing Opportunities, Preserving Values, Executive Office of the President, 2014)

“Bilgi, 21. yüzyılın petrolü olacaktır!” (Peter Sondergaard, Gartner, 2011)

Aslında bu rapor büyük verinin uluslararası çapta ne denli önemli bir yere sahip olduğunu göstermesi açısından oldukça mühimdir.

Şunu da ifade etmek gerekiyor ki bilişim dünyasının bu derece geliştiği bu çağ “Bilgi Çağı” olarak ifade de edilmektedir. Bilişimin ve aygıtların “son kullanıcı” olan insan hayatının içeresine girmiş olmasından dolayı büyük veri kavramı çok önemli bir hal almıştır. Kısaca “Bilgi, 21. yüzyılın petrolü olacaktır!” (Peter Sondergaard, Gartner, 2011).

İnsanlar artık ifade etmek istediklerini, karşılıklı diyalog şeklinde müşteri temsilcilerine, kurum danışmanlarına, dilekçe ile ya da başka bir tekil şekilde sadece karşısındaki muhatabının bileceği şekilde değil, herkesin rahatça okuyabileceği, görebileceği, üzerine yorumlar yazabileceği, öneri verebileceği bir şekilde ortak platformlar üzerinden paylaşmaktadır.

İnsanların paylaşımları bilerek/isteyerek ya da bilmeden/istemdişi her an olmaktadır. Mobil telefonlar aracılığı ile konum bilgisi, arama motorlarındaki anahtar kelimelerin saklanması, web logları vb. gibi bilmeden/istemdişi bilginin yanında Facebook, Twitter, Google+, forumlar vb. gibi bilerek/isteyerek yapılan paylaşımalar aracılığıyla bilgi türemektedir.

Bunlar gibi literatür kısmında bahsedilen onlarca kaynaktan oluşan devasa veri “büyük veri yönetimi” ile ele alınmalı ve kontrol edilmelidir. Çünkü kontrol edilmeyen ve işletmenin/kurumun haricinde büyümeye devam eden veri o işletmeye/kuruma, hatta yöneticilerine negatif etkiler oluşturabilir. Özellikle rekabetçi piyasada doğru şekilde analiz edilmeyen müşteri düşünce ve talepleri işletmelerin kısa ve uzun vadede negatif etkilenmesine, iyi bir şekilde analiz edilen veriler ise, hem kısa hem uzun vadede literatür kısmında da bahsedildiği örneklerde görüldüğü gibi pozitif kazanımlar kazanmasına neden olacaktır.

Türkiye’deki en belirgin örneklerden birisi olan *Hürriyet* gazetesinin yaptığı büyük veri çalışması bize bunu göstermesi açısından somut bir örnek ortaya koymak-

tadır: Hürriyet Gazetecilik Yönetim Kurulu Başkanı Vuslat Doğan Sabancı, büyük verinin yayıncılığı hızlı bir şekilde değiştirdiğini ifade etti. *Hürriyet* gazetesinin 20 yıl önce internetle tanıştığını söyleyen Vuslat Doğan Sabancı, ilk önce yayın yönetmenliğinin yanı sıra bir de trafik editörlüğü oluşturduklarını anlattı. Vuslat Doğan Sabancı, "Trafığımızı çok daha farklı ölçebilmeye başladık. Yayınlarımızdaki etkiyi tırajlardan veya ertesi günlerde görmek yerine gerçek zamanlı olarak tespit etmeye başladık. Birdenbire erişimimiz o kadar çok arttı ki biz bile inanamadık. Her zaman doğru bir şekilde yatırım yaptık ve buna istikrarlı bir şekilde devam ettik. Ayda 1 milyon erişimden 15 milyon erişime ulaştık. Aylık sayfa gösterimimiz 1 milyara ulaştı. Müthiş bir trafik. Hâlâ da bu yolda gitmeye devam ediyoruz," dedi. Vuslat Doğan Sabancı, büyük verinin de yaklaşık 1.5 yıl önce Hürriyet Dünyası'nın hayatına girdiğini belirtti. Vuslat Doğan Sabancı, "Promosyonlar gazetelerin önemli bir harcamasıdır. Sadece 2-3 aylık bir çalışma ile aynı satışlarını tutmuş olmamıza rağmen yüzde 40 maliyetten tasarruf ettik. Bu hayal bile edemeyeceğimiz bir şeydi" diye konuştu. Büyük veriyi okur tatminini artırmak ve reklam verenlere daha iyi hedeflenmiş reklam envanteri sunabilmek için kullandıklarını söyleyen Vuslat Doğan Sabancı sözlerine şu şekilde devam etti: "En doğru yerde, en doğru zamanda, en doğru cihazla hizmet vermeye çalışıyoruz. Hürriyet Dünyası'nın okunduktan sonra çıkan haberlerini hesaplıyoruz. İçerik öneri motoruyla siteden çıkma oranını yüzde 10 düşürdük. Ayrıca çoğunlukla sadece erkeklerin tercih ettiği otomobil modeli veya markasının reklamını sadece erkekler göstererek yüzde 20 daha doğru reklam hedeflemesi yaptı." (<http://www.hurriyet.com.tr/buyuk-veri-cagi-29115095>, 2015).

Son olarak şunu da söyleyebiliriz ki kurumlar daha iyi hizmet verebilmek, işletmeler daha iyi ürünler ve hizmetler sunmak, eksik yönlerini tespit edebilmek, pazar

**Büyük veri olmadan alınacak stratejik kararlar
her zaman eksik kalacaktır. Büyük veriyi "yeni
çağın yeni danışmanı" olarak tanımlamak hiç de
abartı olmayacaktır.**

Son olarak şunu da söyleyebiliriz ki kurumlar daha iyi hizmet verebilmek, işletmeler daha iyi ürünler ve hizmetler sunmak, eksik yönlerini tespit edebilmek, pazar paylarını ve kâr marjlarını artırmak vb. işletme faaliyetlerinde üstünlük sağlayabilmek için büyük veriden yararlanmak zorundadırlar.

paylarını ve kâr marjlarını artırmak vb. işletme faaliyetlerinde üstünlük sağlayabilmek için büyük veriden yararlanmak zorundadırlar.

Bütün bu değerlendirmeler sonrası bir yerde yatırım söz konusu ise; “Büyük Veri olmadan alınacak stratejik kararlar her zaman eksik kalacaktır” diyebiliriz. Büyük veriyi “yeniçağın yeni danışmanı” olarak tanımlamak hiç de abartı olmayacağı.

KAYNAKLAR

Akgül, M. K., 2014, Büyük (Mega) Veriler ve Süper İşlemler Çağı-1, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-1/526>, [Ziyaret Tarihi: 22 Şubat 2015].

Akgül, M. K., 2014, Büyük (Mega) Veriler ve Süper İşlemler Çağı-2, <https://anahtar.sanayi.gov.tr/tr/news/buyuk-mega-veriler-ve-super-islemler-cagi-2/548>, [Ziyaret Tarihi: 25 Nisan 2015].

Baran, A., 2013, ‘Big Data’: Büyük Veri Bileşenleri ve Pazarlama Değişimleri, <http://www.pazarlamasyon.com/2013/04/big-data-buyuk-veri-bilesenleri-ve-pazar-lama-degisimleri/>, [Ziyaret Tarihi: 28 Şubat 2015].

Barriball K.L. ve White, A., 1994, Collecting Data Using a Semi Structured Interview: A Discussion Paper. Journal of Advanced Nursing, 19, 328-335.

Boyacigiller, N. A. ve Adler, N. J., 1991, The parochial dinosaur: Organizational science in a global context. Academy of management Review, 162, 262-290.

- Boyd, D. ve Crawford, K., 2011, Six Provocations for Big Data A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, <http://ssrn.com/abstract=1926431> or <http://dx.doi.org/10.2139/ssrn.1926431>, [Ziyaret Tarihi: 9 Eylül 2014].
- Bryman, A., 2012, Social research methods, Oxford University press, Oxford, ISBN: 978- 0-19-958805-3, 157-590.
- Callery, C. A., 2012, “Data-Driven Decision Making In Community Colleges: An Integrative Model For Institutional Effectiveness” Dissertations, <http://digitalcommons.nl.edu/diss/59>, [Ziyaret Tarihi: 22 Kasım 2014].
- Chen, C. L. ve Zhang, C., 2014, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Information Sciences, <http://www.sciencedirect.com/science/article/pii/S0020025514000346>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Chen, M., Mao, S. ve Liu, Y., 2014, Big Data: A Survey,, <http://link.springer.com/article/10.1007%2Fs11036-013-0489-0>, [Ziyaret Tarihi: 4 Mayıs 2015].
- Cisco, 2009, Cisco Basın Bültenleri, <http://www.cisco.com/web/TR/news/press/archive/2009/070209.html>, [Ziyaret Tarihi: 16 Eylül 2015].
- Cisco, 2011, Cisco Basın Bültenleri, <http://www.cisco.com/web/TR/news/press/archive/2011/020611.html>, [Ziyaret Tarihi: 16 Eylül 2015].
- Davenport, T. ve Dyché J., 2013, Big Data in Big Companies, https://www.sas.com/content/dam/SAS/en_us/doc/research1/big-data-big-companies-executive-summary-106462.pdf, [Ziyaret Tarihi: 1 Mayıs 2015].
- Desouza, K. C. ve Smith, K. L., 2014, Big Data for Social Innovation, http://ssir.org/articles/entry/big_data_for_social_innovation, [Ziyaret Tarihi: 6 Mayıs 2015].
- Dobre, C. ve Xhafa, F., 2013, Intelligent services for Big Data science, <http://www.sciencedirect.com/science/article/pii/S0167739X13001593>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Fielding, N. G. ve Raymond, M. L., 2002, New patterns in the adoption and use of qualitative software, Field Methods, 14, 197-216.
- George, G., Haas, M. R. ve Pentland, A., 2014, Big Data and Management, <http://dx.doi.org/10.5465/amj.2014.4002>, [Ziyaret Tarihi: 4 Mayıs 2015].
- Göksu, C., 2014, “Big Data” nedir? Geleneksel “Veri Yönetimi”ne etkisi ne olur?, <http://datawarehouse.gen.tr/big-data-nedir-geleneksel-veri-yonetimine-etkisi-ne-olur/>, [Ziyaret Tarihi: 28 Şubat 2015].

- Gülle, M. T., 2013, Büyük Veri ya da İçgörü, <http://www.tk.org.tr/index.php/tk/article/view/955>, [Ziyaret Tarihi: 22 Şubat 2015].
- Hansen, M. M., Miron-Shatz, T., Lau, A. Y. S. ve Paton, C., 2014, Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4287084/>, [Ziyaret Tarihi: 30 Eylül 2015].
- Hürriyet Gazetesi, 2015, Büyük veri çağrı, <http://www.hurriyet.com.tr/buyuk-veri-cagi-29115095>, [Ziyaret Tarihi: 23 Ekim 2015].
- Jao, J., 2015, Why Big Data Is A Must in Ecommerce, <http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce>, [Ziyaret Tarihi: 23 Ekim 2015].
- Kambatla, K., Kollias, G., Kumar, V. ve Grama, A., 2014, Trends in Big Data Analytics, Journal of Parallel and Distributed Computing, 74 (7), 2543-2686.
- Kuş-Khalilov, M. C. ve Gündebahar, M., 2014, Bankacılıkta Büyük Veri Uygulamaları: Bir İnceleme, Kuveyt Türk Katılım Bankası, <http://ab.org.tr/ab14/bildiri/191.pdf>, [Ziyaret Tarihi: 25 Nisan 2015].
- LaborComm 2014 Bildiriler Kitabı, 2014, LaborComm 2014 - 5. Uluslararası İşçi ve İletişim Konferansı, http://laborcomm.org/wp-content/uploads/2014/08/LaborComm2014-bildiriler_kitabi.pdf, [Ziyaret Tarihi: 1 Mart 2015].
- Laney, D. ve Sussin, J., 2014, Customer Analytics and the Art of the Possible With Big Data, <https://www.gartner.com/doc/2703617/customer-analytics-art-possible-big>, [Ziyaret Tarihi: 25 Nisan 2015].
- Lester, S., 1999, An introduction to phenomenological research, Stan Lester Developments. Lewins, A. ve Silver, C., 2007, Using software in qualitative research: a step-by-step guide, Sage Publications Ltd., Hamburg, ISBN: 978-0761949220.
- Lindlof, T. R. ve Taylor, B. C., 2002, Qualitative communication research methods, Sage Publications Ltd., Hamburg, ISBN: 978-1412974738.
- Luo, S., Wang, Z. ve Wang, Z., 2013, Big-Data Analytics: Challenges, Key Technologies and Prospects, http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2013/2/articles/201307/t20130724_402922.html, [Ziyaret Tarihi: 1 Mayıs 2015].
- Microsoft Blog, 2012, Büyük Veri Nedir?, <http://blog.microsoft.com/tr/buyuk-veri-nedir.html>, [Ziyaret Tarihi: 22 Şubat 2015].
- Orb, A., Eisenhauer, L. ve Wynaden, D., 2001, Ethics in qualitative research, Journal of nursing scholarship, 331, 93-96.

- Orlikowski, W. J. ve Baroudi, J. J., 1991, Studying information technology in organizations: Research approaches and assumptions, *Information systems research*, 21, 1-28.
- Özdoğan, O., 2014, Big Data Kavramı Nedir? - Oracle Big Data Appliance, <http://www.bilgicozumleri.com/big-data-kavrami-nedir-oracle-big-data-appliance>, [Ziyaret Tarihi: 1 Mart 2015].
- Parkhe, A., 1993, “Messy” research, methodological predispositions, and theory development in international joint ventures, *Academy of Management review*, 182, 227-268.
- Paul K., Radhika K., ve Udo S., 2013, Finding Big Value in Big Data: Unlocking the Power of High-Performance Analytics, *Big Data and Business Analytics*, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.
- Richards, N. M. ve King, J. H., 2014, Big Data Ethics, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2384174, [Ziyaret Tarihi: 5 Mayıs 2015].
- Rouse, M., 2014, Data mart (datamart) definition, <http://searchsqlserver.techtarget.com/definition/data-mart>, [Ziyaret Tarihi: 25 Ekim 2015].
- Rouse, M., 2015, Data silo definition, <http://searchcloudapplications.techtarget.com/definition/data-silo>, [Ziyaret Tarihi: 25 Ekim 2015].
- Rubin, H.J. ve Rubin, I.S., 2005, Qualitative interviewing: The art of hearing data, Sage Publications Ltd., USA, ISBN: 978-1-4129-7837-8.
- Stan Lester Developments, 2012, An introduction to phenomenological research, http://www.academia.edu/1936094/An_introduction_to_phenomenological_research, [Ziyaret Tarihi: 17 Eylül 2014].
- Streubert, H. J. C., 1999, Qualitative Research in Nursing Advancing the Humanistic Imperative, Wolters Kluver, China, ISBN: 978-0-7817-9600-2.
- Tambe, P., 2014, Big Data Investment, Skills, and Firm Value, <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.1899>, [Ziyaret Tarihi: 6 Mayıs 2015].
- Taylan, E., 2012, Big Data konusunda bilinmesi gereken 6 araç, <http://webrazzi.com/2012/12/26/big-data-konusunda-bilinmesi-gereken-6-arac/>, [Ziyaret Tarihi: 25 Nisan 2015].
- Tewksbury, R., 2009, Qualitative versus Quantitative Methods: Understanding Why Qualitative Methods are Superior for Crimonology and Criminal Justice, *Journal of Theoretical and Philosophical Criminology*, 1 (1), 38 - 50.

- Topal, H. C., 2011, 2012'nin modası Big Data, <http://kodcu.com/2011/12/2012nin-modasi-big-data/>, [Ziyaret Tarihi: 20 Kasım 2014].
- Trajman, O., 2013, The Intristic Value of Data, Big Data and Business Analytics, In: Liebowitz, J., 5, CRC Press, Boca Raton, ISBN: 978-1-4822-1851-0.
- Tucker, D. S. ve Wellford, H. B., 2014, Big Mistakes Regarding Big Data, The Antitrust Source, http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/dec14_tucker_12_16f.authcheckdam.pdf, [Ziyaret Tarihi: 6 Mayıs 2015].
- Ularu, E., Puican, F., Apostu, A. ve Velicanu, M., 2012, Perspectives on Big Data and Big Data Analytics, http://dbjournal.ro/archive/10/10_1.pdf, [Ziyaret Tarihi: 2 Mayıs 2015].
- Vaitheeswaran, G. ve Arockiam, L., 2014, Big Data for Education in Students' Perspective, <http://research.ijcaonline.org/icaccthpa2014/number4/ icaccthpa6041.pdf>, [Ziyaret Tarihi: 28 Eylül 2015].
- World Economic Forum, 2012, Big Data, Big Impact: New Possibilities for International Development, http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf, [Ziyaret Tarihi: 25 Nisan 2015].

EKLER

EK 1.**“STRATEJİK BÜYÜK VERİNİN YATIRIMLAR ÜZERİNDEKİ ETKİLERİ”
ARAŞTIRMASI GÖRÜŞME REHBERİ.**

Kurum Adı :

Kurumun Sektörü :

Görüşülen Kişi :

Pozisyonu :

1. Büyük Veri sektörünüzde kullanılıyor mu?

- Evet
 - Hayır
 - Yeni yeni
-

2. Nasıl? Açıklar misiniz?

3. Büyük Veri kurumunuzda kullanılıyor mu?

- Evet
 - Hayır
 - Yeni yeni
-

4. Nasıl? Açıklar misiniz?

5. Büyük Verinin sektörünüzdeki operasyonları kökten değiştireceğine inanıyor musunuz? Nasıl? Açıklar misiniz?

- Evet
 - Hayır
 - Emin değil
-

6. Büyük Verinin kurumunuzdaki operasyonları kökten değiştireceğine inanıyor musunuz? Nasıl? Açıklar misiniz?

- Evet
 - Hayır
 - Emin değil
-

7. Sizce Büyük Veri sektörünüzde ne kadar önemli bir gelişme oluşturabilir? Açıklar misiniz?

- Çok önemli
- Önemli
- Kısmen önemli
- Önemsiz

8. Büyük Verinin kurumunuzda elde edilmek istenen sonuçlara gerçekten etkisi olduğuna inanıyor musunuz? Nasıl daha faydalı olabilir? Neden daha faydalı olabilir?

- Evet
 - Hayır
 - Emin değil
-

9. Büyük Veri Yönetimine olan ihtiyaç, size göre altyapısal mı yoksa yönetimsel bir konu mudur? Açıklayınız?

- Altyapısal
 - Yönetimsel
 - Her ikisi
-

10. Türkiye'de Büyük Veri için altyapısal yeterlilik var mıdır?

- Var
- Yok
- Kısmen var

11. Büyük Veri ve Veri Yönetiminin sınırlarının olacağını tahmin edebiliyor musunuz? Açıklar mısınız?

- Evet
 - Hayır
-

12. İş dünyasında Büyük Veri Yönetimi yeni bir gelişme ya da yeni bir paradigma olması ile ilgili sizin düşünceniz nedir? Açıklar mısınız?

- Evet, Yeni bir paradigma veya olgu
 - Hayır, Verinin genişleyen devamı
-

13. Sizce Büyük Veri bir kurum için fırsat/şans olmak yerine daha çok zorluk/engel mi olacaktır? Açıklar mısınız?

- Fırsat-Şans
- Zorluk-Engel
- Her ikisi

14. Türkiye'de Büyük Veri' den faydalanan seviyeleri ile ilgili neler söylemek isteriniz?

- Yeterli
 - Yeterli değil
 - Kısmen yeterli
-

15. Sizce Türkiye'de Google, Facebook, Amazon vb. gibi Büyük Veri' den ciddi anlamda faydalanan benzeri firmalar var mıdır?

- Evet, var
 - Hayır, yok
 - Önümüzdeki 10 yıl içinde öngörüyorum
 - Önümüzdeki 20 yıl içinde öngörüyorum
-

16. Nihai olarak, yaptığınız açıklamalar ve değerlendirmeler doğrultusunda, Büyük Veri yatırımlarınız üzerinde ne kadar etkili olacaktır? Bu etkilerin nedenlerini açıklayabilir misiniz? Eğer yanıtınız olumlu ise, Büyük Verinin yatırımlar üzerinde ne tür etkileri olacaktır? Açıklar misiniz?

- Çok etkili
- Etkili
- Kısmen etkili
- Etkisiz

