

Endüstri 4.0

DİJİTAL DÖNÜŞÜM

Giriş



Büyük veri kavramıyla nasıl tanıştık?

Tarihçe

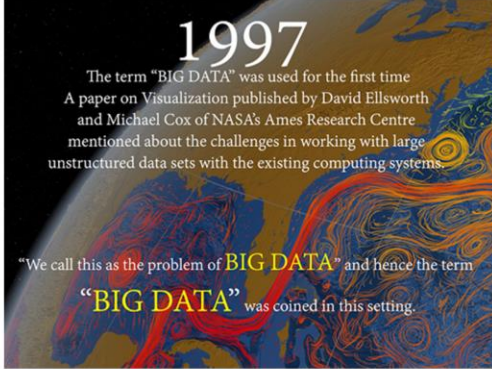


- Wal-Mart'ın 1990'larda Retailer-Link sistemini geliştirerek sattığı her ürünü veri olarak kaydetmesi.
- Amazon'un 1998 yılında kitap satışları tahmini için büyük veri işleyen ve analiz eden algoritmalar kullanmaya başlaması,

Büyük veri kavramı ile ilgili farkındalığın tam olarak 2000 yılından itibaren oluşmaya başladığı iddia edilebilir, ancak bu yıldan önce ismen olmasa da içerik olarak büyük veriye altyapı oluşturabilecek akademik çalışmalar ve iş dünyası uygulamaları görmek mümkündür.

Wal-Mart'ın 1990'larda Retailer-Link sistemini geliştirerek sattığı her ürünü veri olarak kaydetmesi, Amazon.com'un 1998 yılında kitap satışları tahmini için büyük veri işleyen ve analiz eden algoritmalar kullanmaya başlaması ile veri ambarlarından bilgi keşfi yapmaya yönelik veri madenciliği vb. metodların ortaya çıkışı yine aynı yıllara denk gelmektedir.

Tarihçe



- ❑ Büyük veri kavramı, ilk kez Michael Cox ve David Ellsworth tarafından 1997 yılında düzenlenen 8. IEEE Görüntüleme Konferansı'nda (Proceedings of the 8th Conference on Visualization), "Application - Controlled Demand Paging for Out-of-core Visualization" adlı makalede kullanılmıştır.
- ❑ Çalışmada, veri setlerinin çok büyük olduğundan ve bilgisayar sisteminin belleğini, disklerini ve hatta harici diskleri dahi doldurduğundan bahsedilmiş ve karşılaşılan bu soruna "Büyük Veri Problemi" adı verilmiştir

Büyük veri kavramı, ilk kez Michael Cox ve David Ellsworth tarafından 1997 yılında düzenlenen 8. IEEE Görüntüleme Konferansı'nda (Proceedings of the 8th Conference on Visualization), "Application - Controlled Demand Paging for Out-of-core Visualization" adlı makalede kullanılmıştır.

Çalışmada, veri setlerinin çok büyük olduğundan ve bilgisayar sisteminin belleğini, disklerini ve hatta harici diskleri dahi doldurduğundan bahsedilmiş ve karşılaşılan bu soruna "Büyük Veri Problemi" adı verilmiştir

Büyük

According to a report from International Data Corporation (IDC), in 2011, the overall created and copied data volume in the world was 1,8 ZB.

Google processes data of hundreds of Petabyte (PB), Facebook generates log data of over 10 PB per month

Name		Number of Bytes
kilobyte	2^{10}	1,024
megabyte	2^{20}	1,048,576
gigabyte	2^{30}	1,073,741,824
terabyte	2^{40}	1,099,511,627,776
petabyte	2^{50}	1,125,899,906,842,624
exabyte	2^{60}	1,152,921,504,606,846,976
zettabyte	2^{70}	1,180,591,620,717,411,303,424
yottabyte	2^{80}	1,208,925,819,614,629,174,706,176

B y k



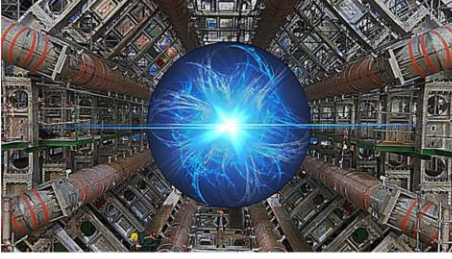
- Bir jet u ađı u tuđu her 30 dakikada 10 terabayt veriyi sens rleri vasıtasıyla toplamaktadır

B y k



- Formula 1 araba yarışında, bir araba  zerinde yer alan 150 sens r vasıtasıyla 20 gigabayt veri  retilmektedir.

B y k



- CERN'de gerekleřtirilen “B y k Hadron arpıřtırıcısı” deneyinde sens rler vasıtasıyla saniyede 1 petabayt b y kl ğ nde veri  retilmiřtir.

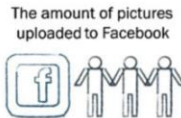
The Phenomenon of Big Data

1.8ZB



Data generated during 2 days in 2011
(larger than the accumulated amount
of data generated from the origin of
civilization to 2003)

750 million



The amount of pictures
uploaded to Facebook

966PB



In 2009, the storage capacity of
American manufacturing industry

209 billion

The number of RFID tags in 2021
(12 million in 2011)



200+TB



Data downloaded during
a computer geek's
2450 thousand hours

200PB

The amount of data generated
by a smart urban project in China



800 billion dollars



Personal location data
in 10 years

300 billion dollars

Medical expense saving by
big data analysis in America

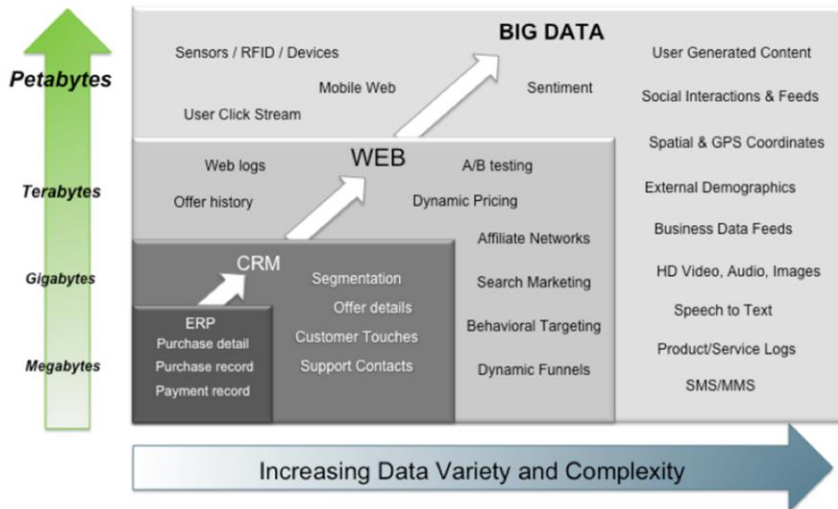


\$32+B



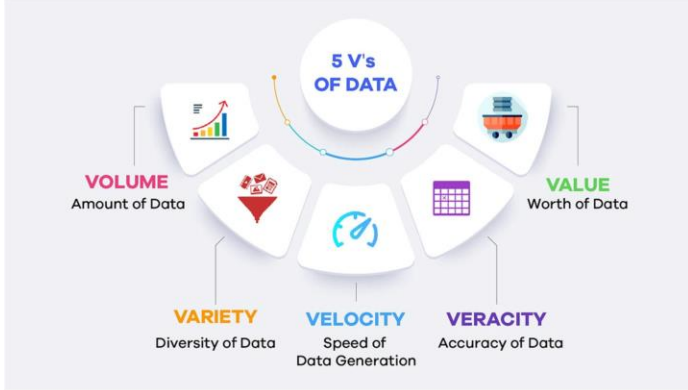
The purchase amount of the
4 big companies since 2010

Big Data = Transactions + Interactions + Observations



Source: Contents of above graphic created in partnership with Teradata, Inc.

5V – 5D



- Devasalık (Volume)
- Değişim Hızı (Velocity)
- Değişkenlik (Variety)
- Değerlilik (Value)
- Doğruluk (Veracity)

Büyük veri oluşumunda 5 bileşen vardır. İngilizce karşılıklarından ötürü 5V olarak adlandırılan bileşenler;

Variety,
Velocity,
Volume,
Verification,
Value.

Variety: Üretilen verinin %80' i yapısal veri değildir. Her veri kaynağının ürettiği veri farklı teknolojiler içerecektir. Dolayısıyla, «veri tipi» problemleri ile uğraşılması gerekmektedir.

Velocity (hız): Hızlı üreyen veri, o veriye muhtaç olan işlem sayısının ve çeşitliliğinin de aynı hızda artması sonucunu doğurmaktadır.

Volume (veri büyüklüğü): IDC istatistiklerine göre, 2020' de ulaşılabilecek veri miktarı 2009' un 44 katı olacak. Kurumun veri arşivleme, işleme, bütünleştirme, saklama vb. teknolojilerinin bu büyüklükteki veri hacmiyle nasıl başa çıkacağının kurgulanması gerekmektedir.

Verification (doğrulama): Verinin akışı sırasında, doğru katmandan, olması gerektiği

güvenlik seviyesinde izlenmesi, doğru kişiler tarafından görünebilir veya gizli kalması gerekiyor.

Value (değer): En önemli bileşen ise değer yaratmasıdır. Büyük veri' nin, veri üretim ve işleme katmanlarınızdan sonra kurum içi bir artı değer yaratıyor olması gerek. Karar veriş süreçlerine anlık olarak etki etmesi, doğru kararın verilmesinde hemen el altında olması gerekmektedir.

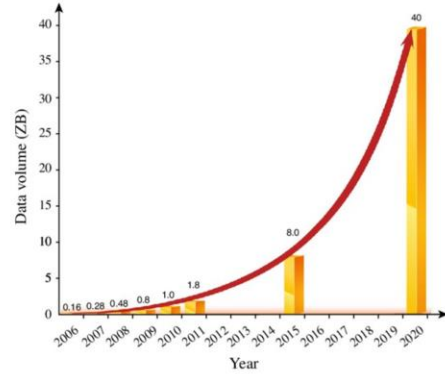
Türkçeye uyarladığımızda ise bu 5 V yerine 5 D diyebiliriz:

Bunun yanı sıra büyük veriye ek bir özellik olarak Devamlılık da eklenebilir (Volatility)

- **Değişkenlik** özelliğinden kasıt, büyük verinin görüntü, ses, harita, metin gibi değişik format ve kayırlardan bir araya getiriliyor olmasıdır.
- **Doğruluk** ise toplanan verinin yanlış, sahte vs. gibi özelliklere sahip olmamasıdır. Büyük verinin önemli bir kısmı kullanıcı tarafından oluşturulmuş veriler (User Generated Data) olduğu için doğruluk özelliği son derece dikkatle incelenmesi gereken bir özelliktir.
- Ayrıca, bu veriler içinden yararlı bilgilerin çıkartılabileceği **değerli** veriler olmalıdır.
- Büyük verinin ortaya çıkmasında en önemli etken sayısal ortamda verilerin büyük bir hızla çoğalmasıdır. Bu da büyük veriye **değişim hızı** özelliğini kazandırmaktadır.
- Bu değişim hızıyla devasa boyutlara ulaşan verilerin büyük veri olarak tanımlanabilmesi için bu verilerin **devamlı** olması gerekmektedir. Yani veriler bir zaman diliminde üretilmiş ve tekrarlanmayan veriler olmamalı. Bu verilerden çıkarılacak olan örüntü ve

Volume

2003'e kadar insanlık tarihinde üretilmiş toplam veri miktarı 5 eksabayt hacmindeyken, günümüzde aynı miktardaki veri sadece 2 gün içerisinde üretilmektedir.

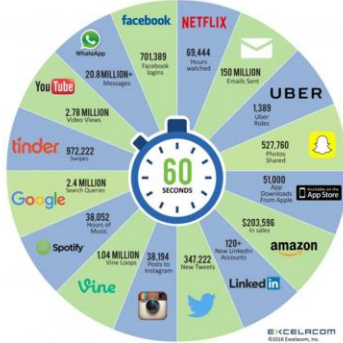


Büyük veri problemlerinin başında hacim boyutu gelmektedir. Çünkü veriyi depolama ve veriye erişim için yenilikçi araçlar gerekmektedir. Örneğin bir doktorun hastasıyla ilgili tuttuğu not birkaç kilobaytlık metin dosyası olarak saklanabilecekken, tetkik için üretilen ham hâldeki görüntü dosyaları birkaç megabayt, manyetik rezonans gibi daha gelişmiş tanılama araçlarının sonuçları ise birkaç gigabayt olabilmektedir. Hastanede gerçekleştirilen ilave test sayısı kadar bu hacmin çoğalacağı düşünüldüğünde, terabayt ve hatta petabayt seviyeleriyle başa çıkmak gerekecektir. Şayet hastanın geçmiş verisinin de analizlerde kullanılması gerekli olursa veri hacminin eksabayt seviyesine gelmesi kaçınılmaz olacaktır (Cyganek ve diğerleri, 2016, s. 498).

Büyük veri, artık mevcut veri tabanlarına sığamayacak, geleneksel veri analiz teknikleriyle işlenemeyecek, terabayt ve petabaytlardan çok daha yüksek hacimsel seviyelere ulaşmıştır. Hayatımızın bir parçası ve vazgeçilmezi hâline gelen akıllı telefonlar, IP (Internet Protocol) temelli uzaktan kontrol edilebilen cihazlar ve akıllı sayaç sistemleri gibi birçok donanım, sensörleri vasıtasıyla ürettikleri veriyi çeşitli uygulamalara aktarmaktadır. Dolayısıyla üretilen, depolanan ve iletilen veri miktarında üstel bir artış olmaktadır.

Velocity

2016 What happens in an INTERNET MINUTE?



Veri sürekli hareket hâlinindedir. Bu bağlamda veri akışının analizi, veri bilimcileri için önemli konulardan biri olmaya başlamıştır. Büyük verinin üretilme hızı çok yüksektir ve her geçen gün bu hız daha da artmaktadır.

Bu açıdan bakıldığında veriyi işleyecek, analiz edecek süreçlerin de büyük verinin üretimiyle aynı hızda olması gerekmektedir.

Veri sürekli hareket hâlinindedir. Bu bağlamda veri akışının analizi, veri bilimcileri için önemli konulardan biri olmaya başlamıştır (Cyganek ve diğerleri, 2016, s. 499). Büyük verinin üretilme hızı çok yüksektir ve her geçen gün bu hız daha da artmaktadır. Sadece büyük veri için değil, tüm iş süreçleri için de hız önemli bir faktördür. Bu açıdan bakıldığında veriyi işleyecek, analiz edecek süreçlerin de büyük verinin üretimiyle aynı hızda olması gerekmektedir. Büyük verinin ne kadar hızda üretildiğine örnek olarak,

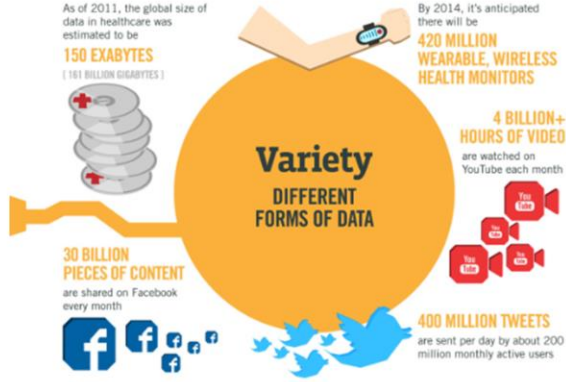
Facebook'ta günde 2,7 milyar kez beğenme tıklaması ve yorum paylaşılması (Gerhardt, Griffin ve Klemann, 2012, s. 3), Twitter üzerinden dakikada 350 bin ve günde 500 milyon tweet'in atılması (*Twitter Usage Statistics*, t.y.), WhatsApp'ta günde 50 milyar mesajın işlenmesi (Burnham, 2014) ve Dünya genelinde günlük 200 milyarın üzerinde e-posta gönderim ve alımının gerçekleştirilmesi (Radicati ve Levenstein, 2015, s. 4) verilebilir.

Sosyal medya haricinde de veri üretiminin hızına örnek teşkil edecek birçok çalışma alanı mevcuttur.

Örneğin, bir jet uçağı uçtuğu her 30 dakikada 10 terabayt veriyi sensörleri vasıtasıyla

toplamaktadır (Sun ve Heller, 2012, s. 4). Benzer şekilde Formula 1 araba yarışında, bir araba üzerinde yer alan 150 sensör vasıtasıyla 20 gigabayt veri üretilmektedir (George, Haas ve Pentland, 2014, s. 321). Büyük verinin üretim hızına bir başka örnek olarak, CERN’de (Conseil Européen pour la Recherche Nucléaire - Avrupa Nükleer Araştırma Kuruluşu) gerçekleştirilen “Büyük Hadron Çarpıştırıcısı” deneyinde sensörler vasıtasıyla saniyede 1 petabayt büyüklüğünde verinin üretilmiş olması verilebilir (Ohlhorst, 2013, s. 13).

Variety



- Büyük veri, geniş bir yelpazede herhangi bir türde ve formatta üretilebilmekte ve bu karışık veri türleri arasında standart bir dizi veya kural bulunmamaktadır. Veri yapısal, yarı yapısal ve yapısal olmayan olarak üç türde ortaya çıkmaktadır.
- Diğer bir ifadeyle çeşitlilik, bir veri kümesindeki yapısal heterojenliği belirtmektedir. Bu heterojen yapının %95'ini ise yapısal olmayan veri oluşturmaktadır.

Büyük veri, geniş bir yelpazede herhangi bir türde ve formatta üretilebilmekte ve bu karışık veri türleri arasında standart bir dizi veya kural bulunmamaktadır. Veri yapısal, yarı yapısal ve yapısal olmayan olarak üç türde ortaya çıkmaktadır. Diğer bir ifadeyle çeşitlilik, bir veri kümesindeki yapısal heterojenliği belirtmektedir. Bu heterojen yapının %95'ini ise yapısal olmayan veri oluşturmaktadır. Nitekim günümüzde üretilen verinin büyük çoğunluğu yapısal olmayan türde olup Facebook, Twitter ve video içeriği gibi çeşitli kaynaklardan beslenmektedir.

Value



- Büyük veri, değere dönüşüm için belirli teknoloji ve analitik yöntemler gerektiren, yüksek hacim, hız ve çeşitlilik ile karakterize edilen bilgi varlıklarıdır. Bu tanımdan hareketle büyük verinin değeri, verinin organizasyonlar ve toplum için ekonomik değer yaratabilecek bir anlayışa dönüşmesi ile ortaya çıkacaktır.
- Büyük veri, genellikle “düşük yoğunluklu değeri” ifade etmektedir. Diğer bir ifadeyle orijinal formda alınan veri, genellikle ilk etapta hacmine göre düşük bir değere sahiptir. Ancak bu tür geniş hacimli verinin analiz edilmesiyle yüksek bir değer elde edilebilmektedir.

Büyük veri, değere dönüşüm için belirli teknoloji ve analitik yöntemler gerektiren, yüksek hacim, hız ve çeşitlilik ile karakterize edilen bilgi varlıklarıdır. Bu tanımdan hareketle büyük verinin değeri, verinin organizasyonlar ve toplum için ekonomik değer yaratabilecek bir anlayışa dönüşmesi ile ortaya çıkacaktır (Mauro, Greco ve Grimaldi, 2016, s. 131). Büyük veri, genellikle “düşük yoğunluklu değeri” ifade etmektedir. Diğer bir ifadeyle orijinal formda alınan veri, genellikle ilk etapta hacmine göre düşük bir değere sahiptir. Ancak bu tür geniş hacimli verinin analiz edilmesiyle yüksek bir değer elde edilebilmektedir (Gandomi ve Haider, 2015, s. 139). Büyük verinin değeri, birbiriyle ilişkili bireyler ve gruplar hakkındaki veya bilginin kendi yapısıyla alakalı veri parçaları arasında bağlantı kurularak ulaşılabilen örüntülerden elde edilmektedir (Boyd ve Crawford, 2011, s. 2).

Büyük verinin, üretim ve analiz aşamalarından sonra organizasyon iş süreçlerine artı bir değer katıyor olması gerekmektedir (Naik ve Joshi, 2017, s. 117). Büyük verinin, karar veriş süreçlerine anlık olarak etki etmesi, doğru kararın verilmesinde hemen ulaşılabilir olması, değer bileşeni açısından çok önemlidir. Büyük verinin iş süreçlerine katacağı maliyet avantajının, büyük veriden yararlanabilmek için organizasyonların yükleneceği maliyetten fazla olması büyük veriyi değerli kılacaktır (Narasimhanve Bhuvaneshwari, 2014, s. 351).

Veracity

VOLUME

How much data?

VARIETY

What kind of data?

VELOCITY

How frequent or
real-time is the data?

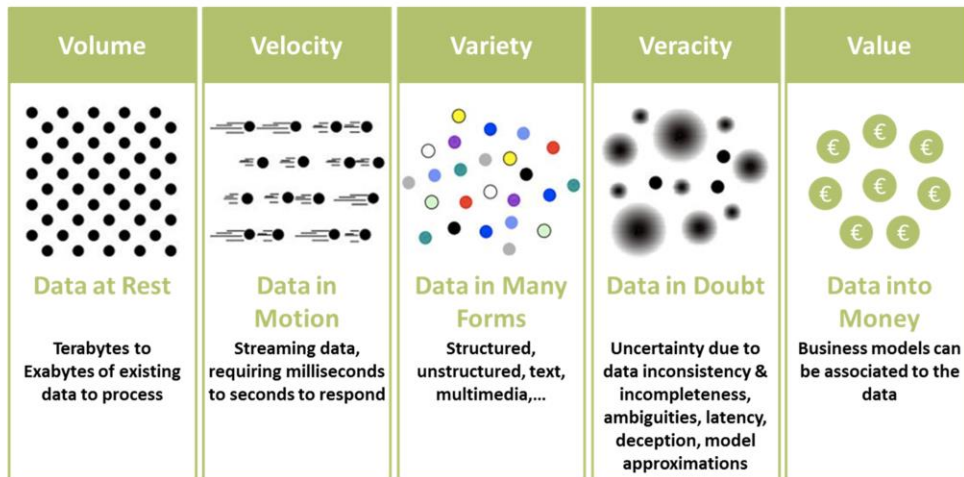
VERACITY

How accurate & applicable
is the data?

- Gerçeklik, büyük verinin ne kadar doğru olduğunu veya güvenilirliğini göstermektedir. Veri, iş kararlarında kullanılabilecek ölçüde güvenilir olmalıdır.

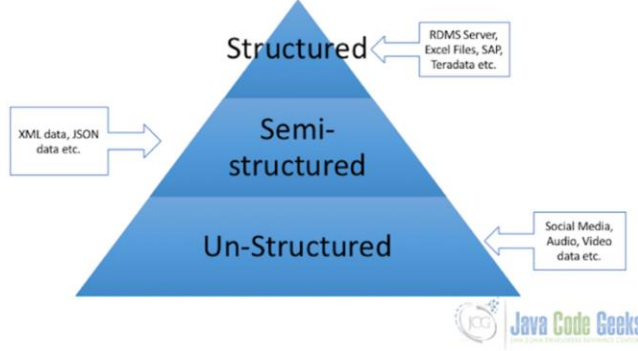
Gerçeklik, büyük verinin ne kadar doğru olduğunu veya güvenilirliğini göstermektedir. Veri, iş kararlarında kullanılabilecek ölçüde güvenilir olmalıdır. Büyük verinin yüksek çeşitlilikte olması, analiz edilen verinin kalitesini ve güvenilirliğini sağlama sürecini zorlaştırmaktadır (Chandra, Ray ve Goswami, 2017, s. 49). Büyük verinin çok önemli bir boyutu olan gerçeklik ile veri kalitesi değerlendirilmektedir. Çünkü güvenilir modeller, yalnızca yüksek kaliteli veri ile üretilebilmektedir. Maalesef, verinin çoğu etki altında kalmakta ya da belirli bir gürültü seviyesinde olabilmektedir. Diğer bir ifadeyle veride aykırı veya eksik değerlerin varlığı gibi anomaliler saptanabilmektedir (Cyganek ve diğerleri, 2016, s. 499). Bu anomaliler, bazı veri kaynaklarına özgü güvenilmezlik durumunun var olması ile ilişkilendirilmektedir. Örneğin, sosyal medyada yansıtılan müşteri duyguları, her ne kadar değerli bilgiler içermekteyse de kişisel yargı gerektirdiği için kendi içinde belirsizlikleri barındırabilmektedir. Dolayısıyla kesin olmayan ve belirsizlik içeren veriyle başa çıkma ihtiyacı, belirsiz verinin yönetimi ve veri madenciliği için geliştirilen araçların ve analiz yöntemlerinin kullanılmasıyla ele alınması gereken büyük verinin başka bir yönünü yansıtmaktadır (Gandomi ve Haider, 2015, s. 139). Bu noktada toplanan verinin doğruluğu ve geçerliliği son derece önem taşımaktadır. Doğru veya geçerli olmayan büyük miktardaki veri, analizler için hem temel teşkil etmeyecek, hem de yanlış yorumlamalara yol açabilecektir (Gahi ve diğerleri, 2016, s. 953). Toplanan büyük veri istatistiksel hatalara ve yanlış

yorumlamalara sebebiyet verebileceğinden, değerli bilginin elde edilebilmesi adına verinin gerçekliği kritik bir öneme sahiptir. Gerçekliğin tesisi; elde edilen verinin onaylanması, gürültü seviyesinin indirgenmesi, ilişki dizisinin ortaya çıkarılması ve aldanma seviyesinin tespiti aşamalarından oluşmaktadır (Goes, 2014, s. v). Dolayısıyla veri tutarsızlığı, eksikliği ve aldanma gibi çeşitli faktörlere bağlı olarak veride oluşacak belirsizlikler gerçeklik bileşeni ile ortaya konmaktadır (Debattista ve diğerleri, 2015, s. 92). Bu aşamada, amaç ve hedeflere uygun olan verinin elde edilmesinde bazı arındırma kontrolleri tesis edilmelidir. Bununla birlikte çok büyük miktarda veri ve heterojen yapıdaki veri kaynağı, girdilerin bütünlüğünün ve değerinin garanti edilmesi adına, hem toplanan verinin organizasyonunda hem de çapraz denetiminde sıkı kurallara tabi olmalıdır (Gahi ve diğerleri, 2016, s. 953).



Adapted by a post of Michael Walker on 28 November 2012

Veri türleri



Büyük veri, geleneksel veri tabanı tekniklerinin kullanılması suretiyle işlenmesi mümkün olmayan, farklı hacimlerdeki heterojen veriyi tanımlayan yeni bir kavramdır ve çeşitli dijital içeriklerden oluşmaktadır (Gahi, Guennoun ve Mouftah, 2016, s. 953):

1. *Yapısal veri*: Yapısal veri, modellenmesi, girdi olarak sokulması, saklanması, sorgulanması, işlenmesi ve görselleştirilmesi kolay olan tüm veri türlerini ifade etmektedir. Genel olarak, belirli tür ve boyutlarda önceden tanımlı alanlarda sunulmakta, ilişkisel veri tabanlarında veya tablolarda yönetilebilmektedir. Katı bir yapıya sahip olan bu veri türünde, süreçlerin yüksek performanslı yetenekler veya paralel teknikler gerektirmemesinden dolayı faydalı bilgilerin elde edilmesi diğer veri türlerine kıyasla daha kolaydır.

2. *Yarı yapısal veri*: Yarı yapısal veya kendi kendini açıklayan (self-describing) veri, yapısal bir veri türünü yansıtmakla birlikte özünde sadece katı bir modeli barındırmamaktadır. Diğer bir ifadeyle yarı yapısal veri, yapısallığın tanımlandığı modellerin yanı sıra belirli öğeleri ve verideki farklı alanların hiyerarşik bir gösterimini tanımlamak adına kullanılan etiketler ve işaretler gibi çeşitli meta modelleri de bulundurmaktadır. Yarı yapısal verinin en çok bilinen örnekleri arasında XML (Extensible Markup Language) ve JSON (JavaScript Object Notation) programlama dilleri yer almaktadır.

3. Yapısal olmayan veri: Yapısal olmayan veri, tanımlı bir format haricinde sunulan ve depolanan kayıt türleridir. Genellikle, kitaplar, makaleler, belgeler, e-postalar gibi serbest formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından oluşmaktadır. Bu türdeki verinin katı bir şekilde sunulmasının zor olması, veri işleme süreçlerinde NoSQL (Not only SQL) gibi yeni mekanizmaların ortaya çıkmasına neden olmuştur.

Yapısal veri



Yapısal veri, modellenmesi, girdi olarak sokulması, saklanması, sorgulanması, işlenmesi ve görselleştirilmesi kolay olan tüm veri türlerini ifade etmektedir. Genel olarak, belirli tür ve boyutlarda önceden tanımlı alanlarda sunulmakta, ilişkisel veri tabanlarında veya tablolarda yönetilebilmektedir.

Yapısal veri, modellenmesi, girdi olarak sokulması, saklanması, sorgulanması, işlenmesi ve görselleştirilmesi kolay olan tüm veri türlerini ifade etmektedir. Genel olarak, belirli tür ve boyutlarda önceden tanımlı alanlarda sunulmakta, ilişkisel veri tabanlarında veya tablolarda yönetilebilmektedir. Katı bir yapıya sahip olan bu veri türünde, süreçlerin yüksek performanslı yetenekler veya paralel teknikler gerektirmemesinden dolayı faydalı bilgilerin elde edilmesi diğer veri türlerine kıyasla daha kolaydır.

Yarı yapısal veri



Yarı yapısal veya kendi kendini açıklayan (self-describing) veri, yapısal bir veri türünü yansıtmakla birlikte özünde sadece katı bir modeli barındırmamaktadır.

Yarı yapısal veya kendi kendini açıklayan (self-describing) veri, yapısal bir veri türünü yansıtmakla birlikte özünde sadece katı bir modeli barındırmamaktadır. Diğer bir ifadeyle yarı yapısal veri, yapısallığın tanımlandığı modellerin yanı sıra belirli öğeleri ve verideki farklı alanların hiyerarşik bir gösterimini tanımlamak adına kullanılan etiketler ve işaretler gibi çeşitli meta modelleri de bulundurmaktadır. Yarı yapısal verinin en çok bilinen örnekleri arasında XML (Extensible Markup Language) ve JSON (JavaScript Object Notation) programlama dilleri yer almaktadır.

Yapısal olmayan veri

Unstructured data types

 Text files and documents	 Server, website and application logs	 Sensor data	 Images
 Video files	 Audio files	 Emails	 Social media data

Yapısal olmayan veri, tanımlı bir format haricinde sunulan ve depolanan kayıt türleridir. Genellikle, kitaplar, makaleler, belgeler, e-postalar gibi serbest formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından oluşmaktadır.

Yapısal olmayan veri, tanımlı bir format haricinde sunulan ve depolanan kayıt türleridir. Genellikle, kitaplar, makaleler, belgeler, e-postalar gibi serbest formatlardaki metinlerden ve resim, ses ve video gibi medya dosyalarından oluşmaktadır. Bu türdeki verinin katı bir şekilde sunulmasının zor olması, veri işleme süreçlerinde NoSQL (Not only SQL) gibi yeni mekanizmaların ortaya çıkmasına neden olmuştur.

Structured Data



0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.529	0.340	0.829
0.153	0.485	0.560	0.428	0.628

Unstructured Data



Unstructured data

The university has 5600 students.
John's ID is number 1, he is 18 years old and already holds a B.Sc. degree.
David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Semi-structured data

```
<University>
  <Student ID="1">
    <Name>John</Name>
    <Age>18</Age>
    <Degree>B.Sc.</Degree>
  </Student>
  <Student ID="2">
    <Name>David</Name>
    <Age>31</Age>
    <Degree>Ph.D. </Degree>
  </Student>
  ....
</University>
```

Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

Structured Data vs Unstructured Data

Can be displayed
in rows, columns and
relational databases



Numbers, dates
and strings



Estimated 20% of
enterprise data (Gartner)



Requires less storage



Easier to manage
and protect with
legacy solutions



Unstructured Data



Cannot be displayed
in rows, columns and
relational databases



Images, audio, video,
word processing files,
e-mails, spreadsheets



Estimated 80% of
enterprise data (Gartner)



Requires more storage



More difficult to
manage and protect
with legacy solutions

Tanım

- ▶ VERİ ne anlama gelir.
- ▶ Bilgi nedir.

Veri nedir?

Tanım

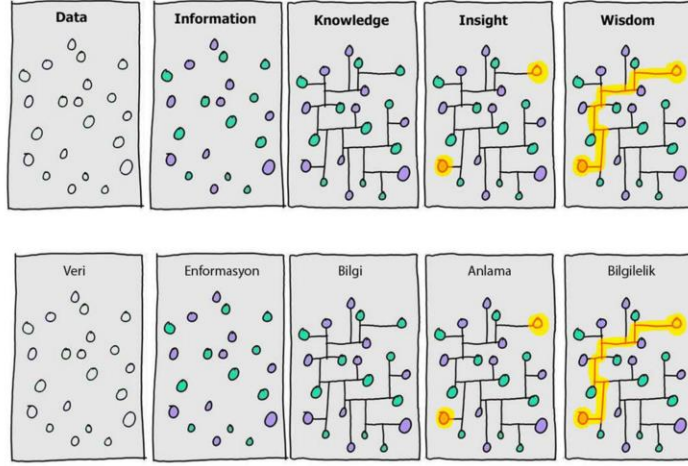
► **Veri:**

- Bir araştırmanın, bir tartışmanın, bir muhakemenin temeli olan ana öge, muta, done.
- Olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi.

► **Bilgi:**

- İnsan aklının erebileceği olgu, gerçek ve ilkelerin bütünü, bili, malumat.
- Öğrenme, araştırma veya gözlem yolu ile elde edilen gerçek, malumat.
- İnsan zekâsının çalışması sonucu ortaya çıkan düşünce ürünü, malumat.

Vermekten veri, bilmekten bilgi.



Veriden Bilgelığe (Bilmenin Seviyeleri)

Data: Veri (ham, işlenmemiş) – Dijital platformdaki tüm filmler ve abone bilgileri.

Information: Malumat (biçimlendirilmiş, kategorize edilmiş, derlenmiş veri) – Aksiyon, romantik, korku filmleri / genç, yaşlı, eğitim seviyesi aboneler.

Knowledge: Bilgi (deney, tecrübe, yorum ve analiz ile zenginleştirilmiş hali) – Hangi aboneler ne tür filmler seyrediyor ve beğeniyor.

Insight: Kavrayış, İlgörü (Bilgiyi kullanarak problem hakkında derin kavrayış sağlar) – Filmlerdeki özellikler ve abone özellikleri arasında bağlantı kuruyor. (Çocuk, Harry Potter) / (Yaşlı, Al Pacino)

Wisdom: Bilgelik, İrfan, Feraset, karar verme. – Ahmet bu filmi sever, ona tavsiye et.

In this diagram '**Data**' is represented by a series of random dots that could mean something – or nothing.

Next is '**Information**', which is where meaning or relationship is applied to the raw data. This is indicated by applying different colours to the dots.

'**Knowledge**' is gained when we are able to memorise the information, for i.e. standard multiplication tables or sunrise & sunset times in a given month. As we gain knowledge we begin to make sense of things and draw connections between different pieces of information.

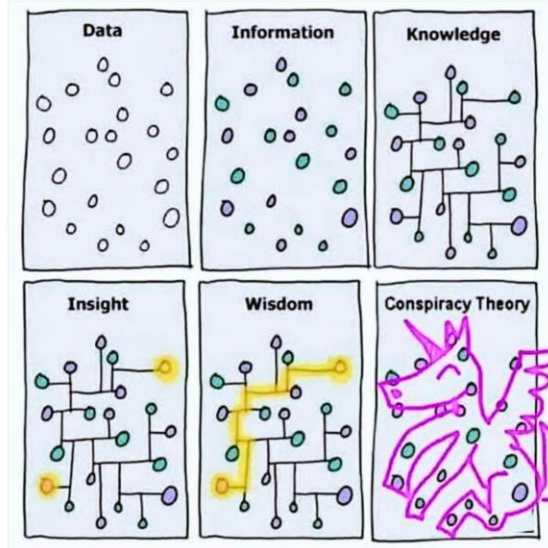
However, it's at the '**Insight**' level where data becomes seriously useful. Insight is the

ability to synthesise knowledge in order obtain a deep understanding of a problem. With insight comes the prospect of '**Wisdom**' – the ability to use insight to facilitate informed decision making.

1- Veri, ham (işlenmemiş) gerçek enformasyon parçasığına verilen addır. **Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir.** Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler **nicel veriler**, sayısal bir değer bildirmeyen veriler de **nitel veriler** olarak adlandırılmaktadır. Her sembolik gösterim gibi, veri de belirli bir nesne, birey ya da olguya ilişkin bir soyutlamadır. Ancak enformasyon ve bilginin soyutluk düzeyleri ile karşılaştırıldığında, verilerin soyutluk düzeyi daha düşüktür. **Bir verinin tek başına bir anlamı ve işlevi bulunmamaktadır.** Veriler toplandıktan sonra gruplanarak, sıralanarak ve özetlenerek, elle ya da bilgisayarla işlenip enformasyona dönüştürüldüklerinde anlam kazanmakta; ait oldukları bağlamı açıklama gücüne kavuşmaktadır. Problem çözme ya da karar verme gibi bir amaca hizmet edebilecek duruma gelmektedir.

2- Enformasyon, en genel anlamda belirli ve görece dar kapsamlı bir konuya (bağlama) ilişkin, derlenmiş bilgi parçasıdır. **Derleme süreci ölçüm, deney, gözlem, araştırma ya da haber toplama (istihbarat) bulgularının özetlenmesi biçimini almaktadır.** Bulgular, onların biçimlendirilmesi ve sunulmasında kullanılan sembollerin genel olarak kabul görmüş bir yaklaşımla yorumu ile anlamlandırılmaktadır. **Belirli bir konuda zamanla biriken enformasyon, ayıklanıp sınıflandıktan ve düzenlendikten sonra, genelliği ölçüsünde bilgiye dönüşmektedir.** Bu açıdan, enformasyon özel bir konunun anlaşılmasına ya da özel bir problemin çözümüne hizmet ederken, bilgi görece daha genel bir konunun anlaşılması veya belirli türden problemlerin tümünün çözülmesi için kullanılmaktadır.

3- Bilgi, enformasyon haline dönüştürülmüş gerçeklerin **analiz edilmesi ve sentezlenmesi** sonucu **karar vermeye yönelik olarak elde edilen** daha üst seviyeli gerçekleri içerir. Bilgi, enformasyonun deney, tecrübe, yorum, analiz ve bağlam ile zenginleştirilmiş halidir.



Komplo teorilerine her zaman yer var.

BIG DATA

BIG DATA

Data that is too large or too complex to be managed using traditional data processing, analysis, and storage techniques.



Big Data

- In addition, big data also brings about new opportunities for discovering new values, helps us to gain an in-depth understanding of the hidden values.



Derinlerdeki saklı değerlerin anlaşılması...
Kaostan değer çıkartmak...

Big Data

- Büyük veri kümeleri üzerinde çalışılabilmesine imkan veren teknoloji, bu büyük veri setlerinden ekonomik, sosyal, teknik vb. konularda örüntüler çıkarabilen analiz teknikleri ve bu analizler sonucunda daha önce elde edilemeyen ölçüde gerçek, objektif ve doğru bilgi elde edileceğine olan inancın birleşmesiyle ortaya çıkan teknik, akademik ve kültürel bir olgudur"

1. Donanım Teknolojileri
2. Analiz Teknikleri
3. Yeni Bilgi Eldesine İnanç

Büyük Veri kavramı ile ilgili yapılan pek çok farklı tanımlamada verinin kendisinin büyüklüğünden çok klasik metotlarla işlenmesinin dışına çıkışına vurgu yapılmaktadır. "Büyüklük" kelimesinin göreceli olmasından dolayı kavramı tek başına ifade edememektedir. Büyük veri ifadesi tek bir sunucuya sığmayacak kadar büyük, yapılandırılmış/yapılandırılmamış (resim, video, eposta, işlem verileri ve sosyal medya etkileşimleri vb.) veriler için kullanılmaktadır. Bu veriler statik bir veri ambarına sığmayacak şekilde sürekli akmakta ve geleneksel metotlarla analiz edilememektedir

Big Data – Kullanım



Büyük veri, çeşitli sektörlerde birçok farklı kullanım alanı bulunmaktadır.

İşletmeler, araştırmacılar, hükümetler ve diğer kuruluşlar, büyük veri analizi yoluyla daha iyi kararlar almaya, trendleri ve gizli kalıpları keşfetmeye ve işletmelerinin performansını artırmaya yardımcı olabilirler. Bazı büyük veri kullanım alanları şunlardır:

Pazarlama: Büyük veri, işletmelerin müşterilerinin ihtiyaçlarını ve tercihlerini daha iyi anlamalarına yardımcı olabilir. Pazarlama kampanyalarının etkililiğini artırmak için kullanılan birçok farklı teknik vardır, bunlar arasında kişiselleştirilmiş reklamlar, segmentasyon ve davranışsal analiz yer almaktadır.

Sağlık: Büyük veri, sağlık sektöründe hastalıkların teşhis ve tedavisinde kullanılabilir. Örneğin, bir hastanın tıbbi kayıtlarının analizi, hastalıkların teşhisinde ve tedavi planlarının oluşturulmasında yardımcı olabilir.

Finans: Büyük veri, finansal kuruluşların dolandırıcılık veya riskleri tespit etmelerine yardımcı olabilir. Ayrıca, finansal işlemleri optimize etmek, portföyleri yönetmek ve risk yönetimini geliştirmek için kullanılabilir.

Lojistik: Büyük veri, lojistik sektöründe tedarik zinciri yönetiminde kullanılabilir. Tedarik zincirindeki verilerin analizi, üretim planlamasını ve envanter yönetimini optimize etmek için kullanılabilir.

Tarım: Büyük veri, tarım sektöründe üretim verimliliğini artırmak için kullanılabilir. Veriler, tarım ürünlerinin üretim planlamasını, sulama ve gübreleme gibi tarım uygulamalarını yönetmek için kullanılabilir.

Enerji: Büyük veri, enerji sektöründe enerji kullanımını optimize etmek ve sürdürülebilir enerji üretimini artırmak için kullanılabilir. Veriler, enerji tüketimini izlemek, yenilenebilir enerji kaynaklarını optimize etmek ve enerji verimliliğini artırmak için kullanılabilir.

Big Data - Konsept



Verilerin toplanması: Verilerin toplanması, işlenmesi ve depolanması için uygun bir altyapıya sahip olmak önemlidir. Veri kaynakları, veri toplama araçları ve yöntemleri, verilerin depolanması ve yedeklenmesi için uygun bir veri tabanı sistemine sahip olunması gereklidir.

Verilerin temizlenmesi: Büyük verilerin temizlenmesi, verilerin eksik, yanlış veya tekrarlı olması durumunda yapılan bir işlemdir. Bu adım, verilerin doğru ve güvenilir bir şekilde analiz edilebilmesi için önemlidir.

Verilerin analizi: Verilerin analizi, verilerin keşfedilmesi ve yorumlanması işlemidir. Verilerin analizi, farklı veri analitiği araçları ve teknikleri kullanılarak yapılabilir. Bu teknikler arasında veri madenciliği, yapay zeka, makine öğrenmesi ve istatistiksel analiz bulunur.

Sonuçların yorumlanması: Analiz edilen verilerin sonuçları yorumlanarak, verilerin ne anlama geldiği ve ne tür eylemlerin alınabileceği belirlenir. Bu adım, büyük verilerin kullanımının en önemli aşamalarından biridir.

Karar verme ve uygulama: Büyük verilerin kullanımı sonucu elde edilen bilgiler

doğrultusunda kararlar alınır ve uygulanır. Bu adım, verilerin gerçek dünyada kullanılabilecek somut sonuçlar elde etmesi açısından önemlidir.

HACKING HUMAN - Duygu & düşünce



"Bu konuda ne hissediyorsunuz?"

"Ne düşünüyorsunuz?"

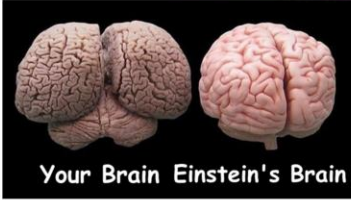
Referandum ve seçimler her zaman insanların duygularıyla ilişkilidir, mantıklarıyla değil. Demokrasi mantıklı tercihler yapmaya ilişkin bir mesele olsaydı, herkese eşit oy hakkı tanınmanın hatta belki de herhangi bir oy hakkı tanınmanın hiçbir mantıklı gerekçesi olmazdı.

Brexit oylamasında her bir İngiliz vatandaşına, "Bu konuda ne hissediyorsunuz?" diye sorulan bir referandum düzenledi. "Ne düşünüyorsunuz?" yerine, "Ne hissediyorsunuz?" diye sorulmuş olmasına itiraz edebilirsiniz ama böyle bir yaklaşım genel bir yanılsamaya düşmek olur. Referandum ve seçimler her zaman insanların duygularıyla ilişkilidir, mantıklarıyla değil. Demokrasi mantıklı tercihler yapmaya ilişkin bir mesele olsaydı, herkese eşit oy hakkı tanınmanın hatta belki de herhangi bir oy hakkı tanınmanın hiçbir mantıklı gerekçesi olmazdı.

Bazı insanların diğerlerinden, bilhassa da belli ekonomik ve siyasi sorular söz konusuysa daha bilgili ve mantıklı oldukları yönünde bolca delil var.

Brexit referandumunun ardından ünlü biyolog Richard Dawkins, İngiliz halkının, kendisi de dahil, büyük bir kısmından böyle bir referandumda oy kullanmasını istemenin yanlış olduğunu çünkü gerekli ekonomi ve siyaset bilimi altyapısına sahip olmadıklarını ifade etmişti. Einstein'ın cebir hesaplamalarının doğruluğuna halk oylamasıyla karar vermekten ya da pilotun hangi piste ineceğini yolcuların oyuna bırakmaktan farksız bir şey.

HACKING HUMAN - Özgür İRADE



Demokrasiye göre insan duyguları gizemli ve derin bir kavram olan "özgür iradeyi" yansıtır, bu "özgür irade" otoritenin temel kaynağıdır ve kimi insanlar diğerlerinden daha akıllı olsalar da tüm insanlar eşit derecede özgürdür.

Okuma yazma bilmeyen bir hizmetçinin de Einstein ve Dawkins kadar özgür iradesi vardır.

Fakat öyle ya da böyle seçim ve referandumlar ne düşündüğümüzle değil nasıl hissettiğimizle alakalıdır. Ve iş hislere gelince Einstein ve Dawkins'in hiç kimseden farkı yoktur. Demokrasiye göre insan duyguları gizemli ve derin bir kavram olan "özgür iradeyi" yansıtır, bu "özgür irade" otoritenin temel kaynağıdır ve kimi insanlar diğerlerinden daha akıllı olsalar da tüm insanlar eşit derecede özgürdür. Okuma yazma bilmeyen bir hizmetçinin de Einstein ve Dawkins kadar özgür iradesi vardır. Seçim günü geldiğinde de verdiği oya yansıyan hisleri herkesinki kadar geçerlidir.

HACKING HUMAN - Yüreğinin sesi



"Siyasi hayatımın her aşamasında kendime tek bir soru sormuşumdur: 'Doğrusu nedir? Yüreğin ne diyor?'"

Duygular sadece seçmenleri değil liderleri de yönlendiriyor. 2016'da düzenlenen Brexit referandumunda ayrılma taraftarı kampanyanın Boris Johnson ile Michael Gove bulunuyordu. David Cameron istifa ettikten sonra Gove başbakanlık görevi için önce Johnson'ı desteklerken, son dakika-da Johnson'ın bu göreve uygun olmadığını beyan edip kendi adaylığını koyma niyetini açıkladı. Gove'un Johnson'ın şansını yok eden bu davranışı Machiavellvari bir siyasi suikast şeklinde tanımlandı. Bunun karşısında Gove duygularından dem vurarak, "Siyasi hayatımın her aşamasında kendime tek bir soru sormuşumdur: 'Doğrusu nedir? Yüreğin ne diyor?'" sözleriyle kendi tutumunu savundu. Gove'a göre Brexit için canını dişine takmasının, vaktiyle müttefiki olan Boris Johnson'ı arkadan bıçaklamaya mecbur hissetmesinin ve liderlik koltuğuna kendini aday göstermesinin sebebi hep buydu; yüreğinin sesini dinlemişti.

HACKING HUMAN



Pekin ya da San Francisco'da biri çıkar da insan yüreğine erişip onu yönlendirebilecek teknolojiyi geliştirirse, demokratik siyaset duyguların oynatıldığı bir kukla gösterisine dönüşebilir.

Yüreğinin sesini dinlemek liberal demokrasinin zayıf noktası olmaya namzet. Zira Pekin ya da San Francisco'da biri çıkar da insan yüreğine erişip onu yönlendirebilecek teknolojiyi geliştirirse, demokratik siyaset duyguların oynatıldığı bir kukla gösterisine dönüşebilir.

Özelliklerin (Olgu-Algı-Duygu) eşleştirilmesiyle duyguların yönlendirilmesi belki de deneniyor.

REKLAM

Veri toplama yarışı başladı. Bedava bilgi, hizmet ve eğlence sunarak ilgimizi çekiyorlar ve bu ilgiyi reklamcılara satıyorlar.



İlkelleri başladı bile.

REKLAM - Milyonluk Bilgi



Algoritmalar sizin sevdiğiniz popüler ikonun kim olduğunu ürettiğiniz veriden çıkartabilir. Sizin haberiniz bile olmaz. Ama onların olur ve bu bilgi milyon değerindedir.

Çekip gitseniz, hatta kendinizden ve sınıf arkadaşlarınızdan saklanmaya devam etseniz de Amazon'dan, Alibaba'dan ya da gizli polisten saklanamazsınız. İnternette gezerken, Youtube'dan bir şey izleyip sosyal medyada paylaşılanları okurken algoritmalar gizliden gizliye sizi gözetler, analiz eder ve size gazlı içecekler satmak isteyen Coca-Cola'ya sizin sevdiğiniz popüler ikonun rol aldığı reklamları kullanmasının daha iyi olacağını söyler. Sizin haberiniz bile olmaz. Ama onların olur ve bu bilgi milyon değerindedir.

Reklam – GEREK YOK!

Google kendisine her şeyi sorabileceğimiz ve dünyanın en iyi cevabını alabileceğimiz bir aşamaya gelmek istiyor. Bu haliyle geleneksel reklam endüstrisini tehdit ediyor.



Benim için en uygun araba hangisi?

Sorusunun algoritma tarafından cevaplandığı dünyada artık reklamlara ihtiyaç kalmayabilir.

Deneyimlerimiz bize kolayca yönlendirilebilir duygularımız yerine Google'a güvenmemiz gerektiğini söyleyebilir.

KARAR - En sevdiğiniz film



Hacklenmek ciddi bir tehdit.

Ama belki her şey apaçık ortada yaşanır ve insanlar daha iyi tavsiyeler alabilmek ve algoritmanın nihayetinde kendileri adına karar alabilmesini sağlamak için bilgilerini memnuniyetle paylaşırlar.

Ama belki her şey apaçık ortada yaşanır ve insanlar daha iyi tavsiyeler alabilmek ve algoritmanın nihayetinde kendileri adına karar alabilmesini sağlamak için bilgilerini memnuniyetle paylaşırlar. Bu tarz şeyler hangi filmi izleyeceğinize karar vermek gibi basit işlerle başlar. Bir grup arkadaşla keyifli bir gece geçirmek için ekran başına geçtiğinizde, önce hangi filmi izleyeceğinize karar vermeniz gerekir. Elli yıl önce bir seçeneğiniz yoktu ama günümüzde istediğinizi seçip izleyebildiğiniz kanalların yaygınlaşmasıyla binlerce film seçeneğiniz var. Ortak bir karara varmak zor olabilir çünkü siz bilimkurgu ve gerilim filmlerini seviyorken, Arda romantik komedileri tercih ediyor, Ceyda da oyunu sanatsal Fransız filmlerinden yana kullanıyor olabilir. Sonuçta herkesi hayal kırıklığına uğratacak ikinci sınıf vasat bir film izlemeye razı olabilirsiniz.

Algoritmanın bu konuda yararı dokunabilir. Herkes önceden izleyip gerçekten beğendiği filmleri söyler ve algoritma devasa istatistiksel veritabanını baz alarak grup için en uygun filmi bulabilir.

KARAR - Şaibeli beyanlar



İnsanların kendileri hakkındaki beyanları, bilindiği üzere, gerçek tercihlerini anlamak için son derece güvenilir bir ölçüt olduğundan, iyi tasarlanmamış bir algoritma kolayca yanlış yönlendirilebilir.

Maalesef insanların kendileri hakkındaki beyanları, bilindiği üzere, gerçek tercihlerini anlamak için son derece güvenilir bir ölçüt olduğundan, böylesine üstünkörü bir algoritma kolayca yanlış yönlendirilebilir. Sık sık bir sürü insanın bir filmi şaheser diye övdüğünü duyar, kendimizi filmi izlemeye mecbur hisseder ve yarısında uyuyakalsak da kültürsüz görünmek istemediğimizden herkese ne kadar muhteşem olduğunu söyleriz.

Oysa bu tür sorunlar, algoritmanın şaibeli beyanlarımıza maruz kalması yerine biz filmleri izlerken gerçek zamanlı veri toplanmasına izin verdiğimiz takdirde çözülecektir. Öncelikle, algoritma hangi filmlerin tamamını izledik, hangilerini yarıda bıraktık gözlemleyebilir. Tüm dünyaya Rüzgar gibi Geçti çekilmiş en iyi filmidir desek de algoritma esasen filmin ilk saatini çıkaramadığımızı ve Atlanta'nın yandığını falan görmediğimizi bilir. Ama algoritma bundan çok daha ileriye de gidebilir. Mühendisler insanların duygularını göz hareketlerine ve yüz kaslarına bakarak tespit edebilen bir yazılım geliştiriyor şu sıralar. Televizyona iyi bir kamera eklememiz durumunda algoritma hangi sahnelerle güldüğümüzü, hangi sahnelerde üzüldüğümüzü ve hangi sahnelerde sıkıldığımızı anlayabilir. Sonra algoritma biyometrik sensörlere bağlanırsa her bir karenin kalp atışımızı, tansiyonumuzu ve beyin faaliyetimizi nasıl etkilediğini anlayabilir. Mesela Tarantino'nun Pulp Fiction filmini izliyoruz diyelim; Vincent

Markin'i kazara vurduğunda suçlu suçlu sırttığımızı ve Big Kahuna Burger esprisini anlamasak da salak görünmemek için güldüğümüzü kaydedebilir. Kendimizi gülmeye zorladığımızda, gerçekten komik olduğunu düşündüğümüz bir şeye gülerken kullandığımızdan farklı beyin devreleri kullanıyoruz.

İnsanlar aradaki farkı çoğunlukla anlamıyor. Ama biyometrik bir sensör anlayabilir.

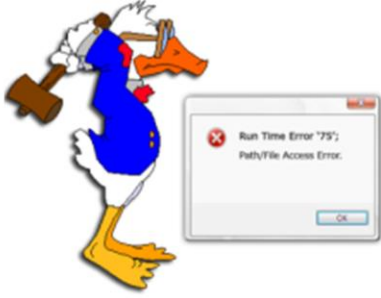
Karar - algoritma VERİR!



Algoritmalar bize en uygun filmleri seçmemize yarayabileceği gibi ne okuyacağımız, nerede çalışacağımız ve kiminle evleneceğimiz gibi hayatımızın önemli kararlarını da bizim adımıza alabilir.

Tarantino'nun bütün filmlerini izledikten sonra çoğu aklınızdan çıkıp gidebilir. Ama Netflix, Amazon ya da televizyon algoritmasının sahibi her kimse, o kişi bizim kişilik tipimizi ve ruhumuzun tellerine nasıl dokunacağını bilecek. Böyle bir veri Netflix ve Amazonun tüyler ürpertici bir doğrulukla bize en uygun filmleri seçmesine yarayabileceği gibi ne okuyacağımız, nerede çalışacağımız ve kiminle evleneceğimiz gibi hayatımızın önemli kararlarını da bizim adımıza almalarını sağlayabilir.

Karar - Algoritmada hata yapar!



Algoritmaların mükemmel olması şart değil. Biz insanların ortalamasından daha iyi olması yeter.

İnsanlar yetersiz verinin, hatalı programlanmanın (genetik ve kültürel), bulanık hedef tanımlarının ve hayatın kaotik doğasının cefasını algoritmalarından daha çok çekiyor.

Elbette Amazon her daim doğru isabet ettiremeyecek. Bu imkânsız. Algoritmalar yetersiz veri, hatalı programlama, bulanık hedef tanımlamaları ve hayatın kaotik doğası gibi sebeplerle tekrar tekrar hata yapabilir. Ama Amazon'un mükemmel olması şart değil. Biz insanların ortalamasından daha iyi olması yeter. Ve bu da çok zor değil çünkü pek çok insan hayatlarının en önemli kararlarını verirken sıklıkla korkunç hatalar yapıyor. İnsanlar yetersiz verinin, hatalı programlanmanın (genetik ve kültürel), bulanık hedef tanımlarının ve hayatın kaotik doğasının cefasını algoritmalarından daha çok çekiyor.

100



Her yıl milyonlarca gencin üniversitede ne okuyacağına karar vermesi gerekiyor. Bu son derece önemli ve çok zor bir karar. Her biri farklı menfaatler ve fikirler taşıyan anne babaların, arkadaşların ve öğretmenlerin baskısı altındasınız. Ayrıca başa çıkmanız gereken korkularınız ve hayalleriniz var. Kafanız Hollywood yapımı gişe filmleriyle, ucuz romanlarla ve alengirli reklam kampanyalarıyla karıştırılıyor. Farklı mesleklerde başarılı olmak için ne yapmak gerektiğini gerçekten bilmediğiniz, zayıf ve güçlü yanlarınız hakkında ille de gerçekçi bir imgeye sahip olmadığınız için akıllıca bir tercih yapmanız bilhassa zor. Nasıl başarılı bir avukat olunur? Baskı altında nasıl bir performans sergilerim? İyi bir takım oyuncusu muyum?

Bir öğrenci kendi becerileri hakkında yanlış fikirlere kapıldığı ve avukatlığın esasen neler gerektirdiğine dair daha da çarpık bir görüşe sahip olduğu için hukuk fakültesine gidebiliyor (bütün gün çarpıcı konuşmalar yapıp "İtiraz ediyorum, Sayın Hakim!" diye bağırılmıyor). Bu esnada bir arkadaşı da gerekli kemik yapısına ya da disipline sahip olmadığı halde çocukluk hayalinin peşinden koşup profesyonel balerin olmaya karar veriyor. Yıllar sonra ikisi de tercihlerinden pişman. Gelecekte bu tarz tercihlerimizi Google'a bırakabiliriz. Google bana hukuk fakültesinde ya da bale okulunda vaktimi boşa harcayacağım' ama harika (ve çok mutlu) bir psikolog ya da tesisatçı olabileceğimi söyleyebilir.

KARAR - Oto pilot



- Algoritmalar nasıl karar alındığına dair daha derin bir anlayış edindikçe insanlara daha cazip gelmeye başlayabilir. Büyük veri algoritmaları daha güvenilir olmakla kalmayıp, eş zamanlı olarak insan duygularını da daha az güvenilir yapacak.
- Fikir ve duygularımızın bu kadar kolayca yönlendirilebilmesi karşısında, vertigo atağı geçiren bir pilotun kendi duyularının kılavuzluğunu bir kenara bırakıp tamamen uçağın aksamalarına güvenerek yol alması gibi, sırtımızı algoritmalara dayamak zorunda kalacağız.

Algoritmaları kuşatan sorunları bir bir sıralayıp insanların asla algoritmalara güvenmeyeceği sonucuna varabilirsiniz. Fakat bu bir nebze, demokrasinin eksikliklerini listeleyip aklı başında hiç kimsenin böyle bir sistemi desteklemeyeceği sonucuna varmaya benzer.

Winston Churchill'in meşhur bir sözü var: Demokrasi, diğer tüm sistemleri saymazsak, dünyanın en kötü siyasi sistemidir. İnsanlar, doğru ya da yanlış, büyük veri algoritmaları konusunda da aynı sonuca varabilir, birçok aksaklıkları var ama daha iyi bir seçeneğimiz yok diyebilirler.

Bilimadamları insanların nasıl karar aldığına dair daha derin bir anlayış edindikçe algoritmalara itimat etmek daha cazip gelmeye başlayabilir. İnsanların karar alma süreçlerine erişim büyük veri algoritmalarını daha güvenilir kılmakla kalmayacak, eş zamanlı olarak insan duygularını da daha az güvenilir yapacak.

Devletler ve şirketler insanların işletim sistemine erişim sağlayınca yönlendirme, reklam ve propagandadan oluşan hassas güdümlü bir bombardımana tutulacağız.

Fikir ve duygularımızın bu kadar kolayca yönlendirilebilmesi karşısında, vertigo atağı

geçiren bir pilotun kendi duyularının kılavuzluğunu bir kenara bırakıp tamamen uçağın aksamalarına güvenerek yol alması gibi, sırtımızı algoritmalara dayamak zorunda kalacağız.

Dikkat



İnsan zihnini keşfetmek adına yeterince yatırım yapmıyor, onun yerine internet bağlantımızın hızını ve büyük veri algoritmalarımızın etkinliğini artırmaya yoğunlaşıyoruz. Dikkatli olmazsak geliştirilmiş bilgisayarlar kullanarak kendine ve dünyaya büyük zararlar veren geriletilmiş insanlara dönüşeceğiz.

İnsanlar bir açıdan diğer evcil hayvanlara benziyor. Yüklü miktarda süt veren ama onun haricinde vahşi atalarının çok aşağısında kalmış uysal inekler yetiştirdik. O denli çevik, meraklı ya da becerikli değiller. Şimdi de yüklü miktarda veri üreten ve dev bir veri işleme mekanizmasının son derece etkin çipleri olma fonksiyonunu yerine getiren yumuşak başlı insanlar yetiştiriyoruz ama bu veri ineklerinin insanın potansiyelini üst seviyeye taşıdığı söylenemez. Esasen insanların azami potansiyelinin ne olduğu konusunda bir fikrimiz yok çünkü insan zihni hakkında çok az şey biliyoruz. Ve buna rağmen insan zihnini keşfetmek adına yeterince yatırım yapmıyor, onun yerine internet bağlantımızın hızını ve büyük veri algoritmalarımızın etkinliğini artırmaya yoğunlaşıyoruz. Dikkatli olmazsak geliştirilmiş bilgisayarlar kullanarak kendine ve dünyaya büyük zararlar veren geriletilmiş insanlara dönüşeceğiz.

Dikkat

Algoritmalar özgür oldukları ileri sürülen toplumlarda bile otorite kazanabilir çünkü çok çeşitli konularda onları dikkate almayı deneyimlerimiz sonucu öğrenip sonunda yavaş yavaş kendi adımıza karar verme yetimizi kaybedebiliriz.



Bazı ülkelerde ve bazı durumlarda, insanlara bir seçenek tanınmayıp büyük veri algoritmalarının kararlarına uyma zorunluluğu getirilebilir. Ancak algoritmalar özgür oldukları ileri sürülen toplumlarda bile otorite kazanabilir çünkü çok çeşitli konularda onları dikkate almayı deneyimlerimiz sonucu öğrenip sonunda yavaş yavaş kendi adımıza karar verme yetimizi kaybedebiliriz. Milyarlarca insanın yirmi yıl kadar kısa bir süre içinde, tüm zamanların en önemli görevlerinden biri olan güncel ve güvenilir bilgi arama konusunda nasıl da Google'ın arama motoruna güvenir hale geldiğini düşünün. Artık bilgiyi aramıyoruz; Google'lıyoruz. Ve cevaplar için Google'a gitgide daha çok bel bağladığımızdan kendi kendimize bilgi edinme yetimiz azalıyor. "Hakikat" şimdiden, Google aramalarında en üstte çıkan sonuçlarla belirleniyor.

