

Module 2 - Data Science Topics

Bu notlar, [IBM Data Science Professional Certificate](#) sertifika programının **What is Data Science** eğitimi üzerinden alınan bilgiler temel alınarak hazırlanmıştır. Notların geliştirilmesinde internet kaynaklarından ve yapay zeka araçlarından da yararlanılmıştır.

1. Dijital Dönüşüm (Digital Transformation)

Dijital dönüşüm, organizasyonların dijital teknolojilerin iş süreçlerine ve operasyonlarına entegre edilerek köklü değişiklikler gerçekleştirmesi süreci olarak tanımlanmaktadır. Bu süreç, yalnızca mevcut işlemlerin dijital ortama aktarılmasıyla sınırlı kalmayıp, işleyişin daha verimli, müşteri odaklı ve veri temelli hale getirilmesinin araştırılmasıyla genişletilmektedir.

Amaç ve Faydalari

- **Veri Bilimi'nin Etkisi:** Büyük veri ve veri analitiğinden yararlanılarak rekabet avantajı elde edilmesi sağlanmaktadır.
- **Kültürel Değişim:** Teknolojik dönüşümün, organizasyonel ve kültürel yapının da değiştirilmesi yönünde etkiler yaratması hedeflenmektedir.
- **Değer Yaratma:** Müşterilere sunulan değerin artırılması ve operasyonel mükemmeliyetin yakalanması amaçlanmaktadır.

Dijital Dönüşüm Örnekleri

Dijital dönüşümün, farklı sektörlerde iş modellerinin yeniden yapılandırılmasında nasıl etkili olduğu aşağıdaki örneklerle açıklanmaktadır:

- **Netflix:**
 - Eski model posta yoluyla DVD kiralama üzerine kurulmuşken, veri analitiği ve kullanıcı davranış verilerinin kullanılmasıyla kişiselleştirilmiş öneriler sunan bir video akış hizmetine dönüşümü gerçekleştirilmiştir.
- **Houston Rockets (NBA Takımı):**
 - Özel kameralar aracılığıyla maçlardan ham veriler toplanmış, veri analizi sonucunda hangi atış türlerinin daha verimli olduğu belirlenmiş ve takım stratejisinin istatistiksel verilere dayalı olarak yeniden kurgulanması sağlanmıştır.
- **Lufthansa:**
 - Müşteri davranış verilerinin analiz edilmesiyle hizmet kalitesinin artırılması ve operasyonların veri odaklı hale getirilmesi sağlanmıştır.

1.1 Büyük Veri (Big Data) ve Dönüşüm

Büyük veri, geleneksel veri işleme yöntemleriyle başa çıkmamayacak kadar büyük, hızlı veya çeşitlilik gösteren veri setlerinin ifade edilmesi olarak tanımlanmaktadır. Dijital dönüşümü tetikleyen önemli bir

unsur olarak değerlendirilen büyük veriye ilişkin şu hususlar öne çıkarılmaktadır:

What Is Big Data?

<http://www.weforum.org/>

The World Economic Forum is the International Organization for Public-Private

➡ <https://www.youtube.com/watch?v=eVSfJhssXUA>



• Dijital Dönüşümü Tetiklemesi:

Veri hacimlerinin artması ve veri analizindeki teknolojik gelişmelerin, pek çok sektörde iş modellerinin yeniden şekillendirilmesine zemin hazırlaması sağlanmaktadır. Yeni teknolojiler (örneğin Hadoop, bulut bilişim) sayesinde eskiden "büyük" kabul edilen veri miktarlarının işlenebilir hale gelmesi mümkün kılınmaktadır.

• Dikkat Edilmesi Gereken Noktalar:

Büyük veri kavramının, teknolojik gelişmelerle birlikte zaman içinde değişkenlik göstermesi; veri büyülüğu, çeşidi ve işlenme hızının organizasyonların altyapı yatırımlarını ve stratejilerini etkilediği göz önünde bulundurulmaktadır.

1.2 Organizasyonel Değişim ve Yönetim

Dijital dönüşümün, bir şirketin tüm kademelerini etkilediği dikkate alınarak, organizasyonel değişim ve yönetim süreçlerinin şu şekilde ele alındığı gözlemlenmektedir:

• Karar Vericiler ve Liderlik:

- En üst yönetim (CEO) desteğinin dönüşüm projelerinin başarısı için kritik olduğu kabul edilmektedir.
- CIO ve CDO gibi pozisyonların, BT altyapısının modernizasyonu ve veri yönetimi stratejilerinin belirlenmesinde anahtar rol üstlendiği vurgulanmaktadır.
- Diğer yöneticilerin de bütçe, personel ve günlük operasyonlar üzerinde kontrol sahibi olmaları gerekmektedir.

• Kültürel Uyum:

Çalışanların, yeni teknolojlere uyum sağlayarak veri temelli düşünce yapısını benimsemeleri; eğitim programları ve iç iletişim faaliyetlerinin bu uyumun sağlanmasında etkili kılındığı belirtilmektedir.

1.3 Sürecin Bütünsel Niteliği

Dijital dönüşüm süreci, operasyonel iyileştirme, veri entegrasyonu ve başarı faktörlerinin dikkate alınmasıyla bütüncül bir yapı içerisinde gerçekleştirilmektedir:

• Operasyonel İyileştirme:

İş süreçleri detaylı olarak incelenmekte, dijitalleştirme aşamasında hangi adımların geliştirilebileceği belirlenmekte ve yeni iş akışları tasarılanarak verimlilik, hız ve müşteri memnuniyeti artırmaktadır.

- **Veri Entegrasyonu:**

Farklı kaynaklardan (video, ses, metin vb.) gelen verilerin saklanması, işlenmesi ve analiz edilmesi konusunda kurumsal stratejiler geliştirilmektedir.

- **Başarı Faktörleri:**

Üst yönetimin güçlü desteği, çalışanların yetkinliği ve motivasyonu, uygun teknolojik altyapı ile veri analitiği araçlarının kullanılması ve veri gizliliği ile güvenliğine verilen önem ön plana çıkartılmaktadır.

1.4 Sonuç (Dijital Dönüşüm)

Dijital dönüşüm sürecinin, yalnızca teknolojik yükseltmelerden ibaret olmadığı; organizasyonların çalışma biçimleri, kültürleri ve stratejik yönelimlerinin derinden etkilenmesiyle kapsamlı bir değişim sağlandığı anlaşılmaktadır. Bu dönüşüm;

- Rekabet ortamında veri odaklı yenilik yaratma zorunluluğu,
- Büyük veri analizleri sayesinde hızlı ve etkili karar alma imkânı,
- Üstten alta tüm kadroların desteğiyle başarının elde edilebileceğibi unsurlarla desteklenmektedir.

2. Bulut Bilişim (Cloud Computing)

Bulut bilişim, internet üzerinden istenildiğinde erişilebilen bilgi işlem kaynaklarının (ağlar, sunucular, depolama, uygulamalar, servisler vb.) hizmet olarak sunulması modelidir. Bu modelde, kullanıcı yalnızca kullandığı kadar ödeme yapar; maliyet avantajı ve esneklik sağlanır.

Temel Özellikler (Five Essential Characteristics)

1. **On-Demand Self-Service (Talep Üzerine Self Servis):**

Kullanıcıların ek müdahaleye gerek kalmadan işlem gücü, depolama veya ağ kaynaklarını ayarlayabilmesi sağlanmaktadır.

2. **Broad Network Access (Geniş Ağ Erişimi):**

Bulut kaynaklarına, internet üzerinden ve çeşitli cihazlarla erişim imkânı sunulmaktadır.

3. **Resource Pooling (Kaynak Havuzu):**

Kaynakların çoklu kullanıcı modeliyle havuzlanarak, talebe göre dinamik olarak paylaştırılması sağlanmaktadır.

4. **Rapid Elasticity (Hızlı Esneklik):**

İhtiyaç duyulan kaynakların hızla artırılıp azaltılabilmesi mümkün kılınmaktadır.

5. **Measured Service (Ölçümlenebilir Hizmet):**

Kullanım net bir şekilde ölçülp raporlanmakta ve yalnızca kullanılan kaynaklar için ödeme yapılmaktadır.

2.1 Bulut Dağıtım Modelleri

Bulut altyapısının nerede barındırıldığı, kimin sahip olduğu ve nasıl erişildiği temel alınarak üç ana dağıtım modeli uygulanmaktadır:

- **Public Cloud (Genel Bulut):**

Altyapının, bulut sağlayıcısına ait olup kamuya açık internet üzerinden sunulması.

- **Private Cloud (Özel Bulut):**

Altyapının, yalnızca tek bir kuruluşun kullanımına özel olarak tahsis edilmesi; yüksek güvenlik ve kontrol gerektiren işletmelerde tercih edilmektedir.

- **Hybrid Cloud (Hibrit Bulut):**

Genel ve özel bulut altyapılarının entegre bir biçimde kullanılmasıyla esnek yapı sunulmaktadır.

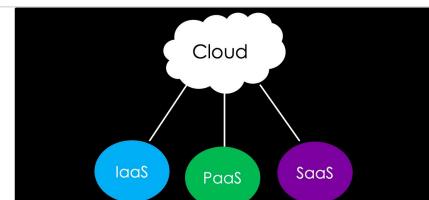
2.2 Bulut Hizmet Modelleri

Bulut bilişim, üç temel hizmet modeli aracılığıyla altyapıdan uygulama katmanına kadar çözümler sunmaktadır:

The main DIFFERENCES between IaaS, SaaS and PaaS explained...

This Tec2Check video explains the main differences between Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)

➡ <https://www.youtube.com/watch?v=cuOnIPnzGyo>



1. Infrastructure as a Service (IaaS):

Sunucular, depolama, ağ donanımları gibi temel altyapı bileşenlerinin çevrimiçi olarak erişilebilir kilinması sağlanmaktadır.

2. Platform as a Service (PaaS):

Uygulamaların geliştirilmesi, test edilmesi ve dağıtılması için gerekli platformun sunulması; altyapı yönetiminin servis sağlayıcı tarafından gerçekleştirilmesi amaçlanmaktadır.

3. Software as a Service (SaaS):

Tamamen kullanıma hazır yazılım ve uygulama hizmetlerinin abonelik modeliyle sunulması; bakım ve güncellemelerin servis sağlayıcı tarafından yapılması sağlanmaktadır.

Kullanıcı Avantajları

- **Maliyet Etkinliği:** Kurulum ve bakım maliyetlerinin düşürülmesi, yalnızca kullanılan kaynaklar için ödeme yapılması.
- **Güncellilik:** Uygulama ve altyapıların sürekli güncellenmesi.
- **Depolama Tasarrufu:** Yerel depolama alanından tasarruf edilmesi.

- **İşbirliği:** Gerçek zamanlı dosya paylaşımı ve eşzamanlı çalışma imkânının sunulması.
 - **Esneklik ve Ölçeklenebilirlik:** Kaynakların ihtiyaca göre kolayca genişletilip daraltılması.
-

2.3 IaaS, PaaS, SaaS – Günlük Hayattan Örnekler

1. Infrastructure as a Service (IaaS)

- **Örnek Senaryo:** Yetersiz yerel işlem gücü veya depolama kapasitesi durumunda, AWS EC2 veya Azure Virtual Machines gibi hizmetlerden sanal sunucu kiralanması sağlanmaktadır.
- **Gündelik Hayat Benzeri Durum:** Evde büyük bir parti düzenlemek için eksik mobilya veya ekipman kiralamasına benzetilebilmektedir.

2. Platform as a Service (PaaS)

- **Örnek Senaryo:** Yazılım geliştirme sürecinde veritabanı, işletim sistemi ve sunucu yönetimi gibi detayların servis sağlayıcı tarafından üstlenilmesi, geliştiricilerin yalnızca uygulama koduna odaklanmasılığını sağlamaktadır.
- **Gündelik Hayat Benzeri Durum:** Taşınmaya hazır, temel tesisatı tamamlanmış bir dairenin kiralanması; temel altyapı ile uğraşmadan yerleşim ve dekorasyona odaklanması örneğin gösterilebilmektedir.

3. Software as a Service (SaaS)

- **Örnek Senaryo:** Gmail, Google Docs veya Microsoft 365 gibi uygulamaların internet tarayıcısı üzerinden kullanılması; yazılım güncellemelerinin ve altyapı yönetiminin servis sağlayıcı tarafından yapılması sağlanmaktadır.
 - **Gündelik Hayat Benzeri Durum:** Netflix üzerinden dizi/film izlenmesi veya Spotify üyeliği gibi, altyapı ve güncelleme işlemlerinin kullanıcının sorumluluğunda olmadığı hizmetler örneği verilebilmektedir.
-

3. Büyük Veri (Big Data) Tanımı ve Temel Unsurlar

Büyük veri, insanların, makinelerin ve araçların ürettiği devasa, hızlı ve çeşitli veri kümeleri olarak tanımlanmaktadır; bu verilerin toplanması, saklanması ve analiz edilmesi için yenilikçi teknolojiler ve ölçeklenebilir altyapılar kullanılmaktadır.

What is Big Data | Big Data In 3 Minutes | Big Data Analytics

Curious about Big Data? Watch this What is Big Data video to learn the basics of Big Data in 3 minutes! Discover what Big Data is, how it's used, and its impact on businesses. From data collection to analysis, this video covers the essentials of Big Data.

➡ <https://www.youtube.com/watch?v=g4zUJPhzMzs>



5V Yaklaşımı

1. Velocity (Hız):

Verinin çok hızlı üretilmesi ve akışının devamlılığı sağlanmaktadır.

2. Volume (Hacim):

Veri miktarının sürekli artması göz önünde bulundurulmaktadır.

3. Variety (Çeşitlilik):

Yapilandırılmış, yarı-yapilandırılmış ve yapılandırmamış verilerin farklı kaynaklardan elde edilmesi sağlanmaktadır.

4. Veracity (Doğruluk):

Verinin kalitesi, güvenilirliği ve bütünlüğü değerlendirilmekte, hatalı verilerin analiz sürecini olumsuz etkilemesinin önüne geçilmeye çalışılmaktadır.

5. Value (Değer):

Büyük veri analizinden elde edileBILECEK işlevsel veya toplumsal fayda ortaya çıkarılmaya çalışılmaktadır.

3.1 Büyük Veri İşleme Teknolojileri

Büyük veri işleme teknolojileri, yapılandırılmış, yarı-yapilandırılmış ve yapılandırmamış veri kümeleri üzerinde değer elde edebilmek amacıyla ölçeklenebilir ve dağıtık sistemler sunmaktadır. Aşağıda, üç temel açık kaynak teknoloji incelenmektedir:

1. Apache Hadoop

What is Apache Hadoop?

Explore watsonx.data: an open, hybrid and governed data store →
<https://ibm.biz/BdmCUX>

➡ <https://www.youtube.com/watch?v=JWX5lnb--ig&t=21s>



• Tanım:

Büyük veri kümelerinin dağıtık depolama ve işlenmesi amacıyla oluşturulan, Java tabanlı açık kaynaklı çerçeve olarak kullanılmaktadır.

• Önemli Bileşen – HDFS (Hadoop Distributed File System):

- Verilerin, parçalara bölünerek farklı düğümlerde depolanması sağlanmakta;
- Her parçanın varsayılan olarak en az üç kopya şeklinde tutulmasıyla veri kaybı riski azaltılmakta;
- Veri işlemlerinin, verinin bulunduğu düğüme yakın gerçekleştirilmesi (data locality) sayesinde ağ trafiği azaltılarak işleme hızı artırılmaktadır.

• Hadoop'un Katkıları:

Yapilandırılmış, yarı-yapilandırılmış ve yapılandırmamış verilerin tek bir ortamda toplanması, "soğuk" verinin kurumsal veri ambarlarından taşınarak maliyet optimizasyonunun sağlanması ve farklı veri formatlarının saklanarak işlenmesi mümkün kılınmaktadır.

2. Apache Hive

- Tanım:**

Hadoop üzerinde çalışan açık kaynaklı veri ambarı yazılımı olarak, SQL benzeri sorgular (HiveQL) kullanılarak büyük veri kümelerine erişim, raporlama ve analiz imkânı sunulmaktadır.

- Özellikleri:**

- Hadoop'un sıralı tarama mantığından yararlanması nedeniyle sorgu yanıt sürelerinin yüksek olabileceği;
- OLTP senaryoları için uygun olmayıp, temel kullanım alanının ETL, raporlama ve veri analizi olduğu göz önünde bulundurulmaktadır.

3. Apache Spark

What Is Apache Spark?

Learn more about Apache Spark → <https://ibm.biz/BdPfYS>

Check out IBM Analytics Engine → <https://ibm.biz/BdPmmv>

Unboxing the IBM POWER E1080 Server → https://youtu.be/dlZsBL4I_9k

▶ <https://www.youtube.com/watch?v=vVZ7EHLdrVo0>

What is
Apache
Spark?



- Tanım:**

Bellek içi (in-memory) veri işleme yaklaşımını esas alan, dağıtık veri işleme ve analitik platformu olarak yüksek hacimli verilerin gerçek zamanlı işlenmesi sağlanmaktadır.

- Çalışma Biçimi ve Kullanım Alanları:**

- Verilerin öncelikli olarak bellekte tutulması, yalnızca bellek yetersizliğinde diske yazılması (spilling to disk) sayesinde yüksek performans elde edilmekte;
- Standalone veya Hadoop altyapıları üzerinde çalışabilme, HDFS, Hive gibi çeşitli veri kaynaklarından veri çekebilme özellikleriyle;
- Makine öğrenmesi (Spark MLlib), gerçek zamanlı analiz (Spark Streaming) ve etkileşimli analiz imkânlarının sunulması sağlanmaktadır.

Özet ve Karşılaştırma

- Hadoop:**

Büyük verinin saklanması ve temel işlenmesi için altyapı sağlanmakta; HDFS ile veriler dağıtık şekilde saklanmaktadır.

- Hive:**

Hadoop ekosistemi üzerinde veri ambarı işlevi üstlenilerek SQL benzeri sorgularla analiz yapılmaktadır.

- Spark:**

Bellek içi işlem avantajı sayesinde gerçek zamanlı ve ileri analitik çözümler sunulmakta; makine öğrenmesi ve etkileşimli analiz alanlarında performans üstünlüğü sağlanmaktadır.

4. Veri Madenciliği Süreci ve Aşamaları

Veri madenciliği süreci, hedef belirleme aşamasından sonuçların değerlendirilmesi ve sürekli geri besleme döngüsüyle tamamlanacak çok aşamalı bir süreç olarak ele alınmaktadır.

1. Veri Madenciliği Hedeflerinin Belirlenmesi

- **Anahtar Sorular ve Hedefler:**

Veri madenciliğinden hangi soruların cevaplanması beklentiği ve elde edilecek sonuçların doğruluk ve yararlılık seviyesinin belirlenmesi sağlanmaktadır.

- **Maliyet-Fayda Analizi:**

Yüksek doğruluk talebinin projenin maliyetini artırabileceği; azalan getiri prensibinin göz önünde bulundurulması gerekmektedir.

2. Veri Seçimi

- **Veri Kalitesi ve Erişilebilirlik:**

Projede kullanılacak verinin mevcut kaynaklardan temin edilebilmesi ya da yeni veri toplama girişimlerinin gerekli olup olmadığını belirlenmesi sağlanmaktadır.

- **Maliyet Etkisi:**

Verinin türü, boyutu ve toplama sıklığının toplam maliyeti doğrudan etkilediği göz önünde bulundurulmaktadır.

3. Veri Ön İşleme (Preprocessing)

- **Temizlik ve Düzeltme:**

Hatalı veya alakasız verilerin tespit edilip çıkarılması, veri bütünlüğünün sağlanması amacıyla eksik veya yanlış bilgilerin düzeltilmesi gerçekleştirilmektedir.

- **Eksik Veri Yönetimi:**

Rastgele eksik verilerin basit yöntemlerle doldurulması; sistematik eksik verilerde ise özenle yöntem seçimi yapılması sağlanmaktadır.

4. Veri Dönüştürme (Transforming Data)

- **Nitelik Azaltma (Feature Reduction):**

Gereksiz veya tekrarlı özelliklerin elenmesi; örneğin Principal Component Analysis (PCA) gibi yöntemlerle bilgiyi koruyan daha az bileşene indirgenmesi sağlanmaktadır.

- **Veri Türü Dönüşümleri:**

Sürekli değişkenlerin kategorik hale getirilmesi veya farklı kaynaklardan gelen bilgilerin birleştirilerek tek bir değişken oluşturulması gerçekleştirilmekte, böylece modelin performansı ve yorumlanabilirliği artırılmaktadır.

5. Veri Depolama

- **Depolama Formatı ve Erişim İmkanları:**

Veri madenciliği sırasında oluşturulan yeni değişkenlerin hızlı okuma/yazma imkânına sahip, güvenli, ölçeklenebilir ve merkezi depolama sistemleri tercih edilmektedir.

- **Veri Güvenliği ve Gizlilik:**

Hassas verilerin korunması, erişim haklarının ve şifreleme yöntemlerinin uygun şekilde yönetilmesi sağlanmaktadır.

6. Veri Madenciliği (Mining Data)

- **Analiz Yöntemleri ve Algoritmalar:** Parametrik ve non-parametrik yöntemler, makine öğrenmesi teknikleri (sınıflandırma, kümeleme, regresyon, derin öğrenme vb.) uygulanmakta; etkileşimli görselleştirme ile ön keşif analizleri desteklenmektedir.

7. Madencilik Sonuçlarının Değerlendirilmesi (Evaluating Mining Results)

- **Model Performansı ve Geçerlilik:**

İç örnek tahmini ve dış örnek veriler kullanılarak modelin performansının ölçülmesi sağlanmaktadır.

- **Paydaş Geri Bildirim ve İyileştirme:**

Sonuçların ilgili iş birimleriyle paylaşılması ve geri bildirimler doğrultusunda modelde sürekli iyileştirmeler yapılmaktadır.

5. Büyük Veri, Bulut Bilişim ve Veri Madenciliği: Genel Bir Özeti

- **Büyük Veri:**

Hacim, hız, çeşitlilik, doğruluk ve değer boyutları göz önüne alınarak, geleneksel yöntemlerin ötesinde ölçeklenebilir çözümler gerektiren devasa veri kümeleri oluşturulmaktadır.

- **Bulut Bilişim:**

Büyük veriyle başa çıkmak için gerekli ölçeklenebilirlik, erişim kolaylığı ve maliyet verimliliği sağlayan; talep üzerine servis, geniş ağ erişimi, kaynak havuzu, hızlı esneklik ve ölçümlenebilir hizmet gibi temel özelliklerle desteklenen bir model olarak kullanılmaktadır.

- **Veri Madenciliği:**

Hedef belirleme, veri seçimi, ön işleme, dönüştürme, madencilik ve sonuçların değerlendirilmesi adımlarıyla, veriden anlamlı içgörüler elde edilmesi amacıyla sürekli tekrarlanan bir süreç olarak uygulanmaktadır.

6. Temel Veri Bilimi ve Yapay Zekâ Kavramları

Veri bilimi, makine öğrenmesi, derin öğrenme, yapay sinir ağları ve yapay zekâ gibi kavramların arasındaki ilişkiler şu şekilde özetlenebilmektedir:

1. **Büyük Veri (Big Data):**

Geleneksel analiz yöntemlerinin yetersiz kaldığı devasa veri kümelerinin, dağıtık bilgi işlem güçleri ve yeni analiz araçları sayesinde işlenebilir hale getirildiği ifade edilmektedir.

2. **Veri Madenciliği (Data Mining):**

Veriler içerisinde gizli kalıpların, ilişkilerin ve eğilimlerin otomatik veya yarı otomatik yöntemlerle ortaya çıkarılmasına yönelik süreç uygulanmaktadır.

3. **Makine Öğrenmesi (Machine Learning):**

Algoritmaların, veriden ögrenerek geleceğe yönelik tahminler yapabilmesi amacıyla, açıkça programlama gerektirmeksızın veri odaklı yaklaşım benimsenmektedir.

4. **Derin Öğrenme (Deep Learning):**

Çok katmanlı yapay sinir ağları kullanılarak, karmaşık veri setlerinden yüksek seviyede soyutlamaların elde edilmesi sağlanmaktadır; verinin hacmi büyündükçe performansın artması beklenmektedir.

5. **Yapay Zekâ (Artificial Intelligence) ve Veri Bilimi (Data Science):**

Yapay zekânın, makinelerin insan benzeri kararlar almasını sağlayan geniş bir alan olarak; veri biliminin ise uçtan uca veri işleme süreçlerini (temizleme, dönüştürme, analiz, görselleştirme) kapsayarak iş ve toplumsal fayda yaratmayı amaçladığı ifade edilmektedir.

AI, Machine Learning, Deep Learning and Generative AI Explained

Want to learn about AI agents and assistants? Register for Virtual Agents Day here
→ <https://ibm.biz/BdaAVa>

Want to play with the technology yourself? Explore our interactive demo →

➡ <https://www.youtube.com/watch?v=qYNweeDHiyU&t=356s>



7. Generative AI ve Veri Bilimi

Generative AI, mevcut veriyi analiz etmek yerine yeni veri veya içerik üretebilen modelleri kapsamaktır; veri bilimi ile birleştiğinde çeşitli avantajlar sağlanmaktadır.

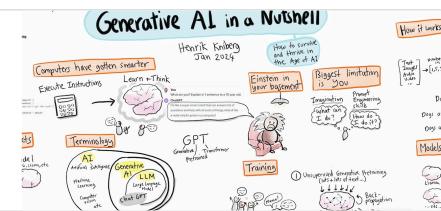
What is generative AI and how does it work? – The Turing Lectures with Mirella Lapata

How are technologies like ChatGPT created? And what does the future hold for AI language models?

➡ https://www.youtube.com/watch?v=_6R7Ym6Vy_I



Generative AI in a Nutshell - how to survive and thrive in the age of AI
Basically a full day AI course crammed into 18 mins of drawing & talking. Target audience: Everyone.
NEWS: Book is now available: <https://leanpub.com/ainutshell>
➡ <https://www.youtube.com/watch?v=2IK3DFHRFw&t=180s>



Generative AI Nedir?

- Mevcut verilerin ötesinde yeni veri veya içerik üretebilen modellerin, özellikle derin öğrenme teknikleri kullanılarak geliştirildiği ifade edilmektedir.
- **Öne Çıkan Model Türleri:**
 - **Generative Adversarial Networks (GANs):** Üreteci ve ayırt edici ağlar arasındaki rekabet yoluya gerçekçi içerik üretilmektedir.
 - **Variational Autoencoders (VAEs):** Verinin olasılık dağılımını öğrenerek benzer örnekler üretilmekte ve veri boyutunun küçültülmesinde kullanılmaktadır.

Generative AI Uygulama Alanları

- **Doğal Dil İşleme (NLP):** GPT-3 gibi modellerin insan benzeri metin üretimi gerçekleştirdiği;
- **Sağlık:** Yapay tıbbi görüntülerin üretilmesiyle eğitim materyallerinin çeşitlendirildiği;
- **Sanat ve Tasarım:** Özgün sanat eserleri ve moda tasarımlarının ortaya konulmasında kullanıldığı;
- **Oyun Geliştirme:** Gerçekçi oyun ortamları ve karakterlerin oluşturulmasıyla tasarım süresinin azaltıldığı;
- **E-ticaret ve Kişiselleştirme:** Kullanıcı beğenilerine dayalı ürün önerileri ve reklam kampanyaları üretilmektedir.

Veri Biliminde Generative AI Kullanımı

- **Sentetik Veri Oluşturma:**
Gerçek veriye benzer dağılım ve örüntüler üretilerek model eğitimi ve test süreçlerinin zenginleştirilmesi sağlanmaktadır.
- **Kod ve Model Geliştirme:**
Otomatik kod yazımı ile düşük seviyedeki kodlama işlerinin bir kısmının otomasyona devredilmesi mümkün kılınmaktadır.
- **Kapsamlı İçgörü ve Raporlama:**
Gerçek zamanlı veri güncellemeleriyle anlık raporlar ve iş öngörüleri oluşturulmakta; manuel analizde gözden kaçabilecek kalıplar tespit edilebilmektedir.
- **Örnek Araç:**
IBM Cognos Analytics gibi araçların, doğal dil işleme yetenekleriyle kullanıcının sorularına dayalı veri analizi gerçekleştirdiği belirtilmektedir.

Gelecek ve Önemli Noktalar

- Karar verme süreçlerine hızlı yansıyan içgörülerin, rekabet avantajı yaratma potansiyeli;
 - Etik, gizlilik ve model şeffaflığı konularının sürekli takip edilmesi gerekliliği;
 - Sürekli öğrenme ve güncelleme ile generative AI'nın potansiyelinin genişletilmesi sağlanmaktadır.
-

8. Sonuç (Genel Değerlendirme)

Veri biliminin ve dijital dönüşümün, büyük veri, bulut bilişim, veri madenciliği, makine öğrenmesi, derin öğrenme ve generative AI gibi kavramlar çerçevesinde ele alınmasıyla;

- Dijital dönüşümün, organizasyonların kültüründen işleyişine kadar geniş kapsamlı değişikliklere neden olduğu,
- Bulut bilişimin ölçülebilirlik, erişim kolaylığı ve maliyet etkinliği ile modern işletmelerin temel ihtiyaçlarını karşıladığı,
- Büyük verinin ve veri madenciliğinin, veriden anlamlı içgörüler elde edilmesi sürecinde kritik rol oynadığı,
- Generative AI'nın ise sentetik veri üretimi, otomatik kodlama ve kapsamlı raporlama gibi alanlarda veri bilimine yeni boyutlar kazandırdığı; ortaya konulmaktadır. Tüm bu teknolojiler ve süreçler, veri odaklı karar alma kültürünün temel taşları olarak modern organizasyonların rekabet gücünü artırmak amacıyla sürekli gelişmekte ve entegre edilmektedir.