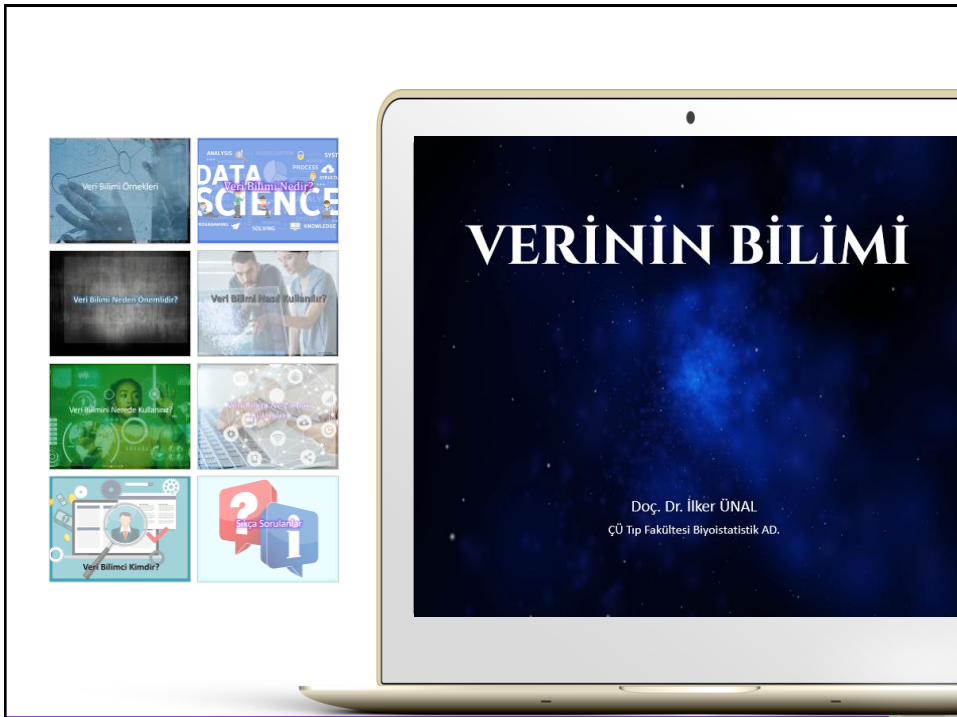


# VERİNİN BİLİMİ

Doç. Dr. İlker ÜNAL  
ÇÜ Tıp Fakültesi Biyoistatistik AD.





# Who's a good dog?

Using Convolutional Neural Networks to classify dog breeds from images


Garrick Chu (May 2018)

Springboard Data Science Career Intensive Track



## Results

Correct Label	Human Prediction*	Model Prediction**
Papillion	Papillion (70.0%) ✓	Papillion (99.50%) ✓
Eskimo Dog	Husky (90.0%) ✗	Saluki (18.06%) ✗
EntleBucher	Greater Swiss Mountain Dog (40.0%) ✗	EntleBucher (86.99%) ✓
Cairn	Norwich Terrier (76.0%) ✗	West Highland Terrier (54.76%) ✗
Malamute	Malamute (46.9%) ✓	Malamute (51.96%) ✓
Gordon Setter	Brittany Spaniel (62.0%) ✗	Gordon Setter (97.43%) ✓
Border Collie	Shetland Sheepdog (50.0%) ✗	Border Collie (45.00%) ✓
Shih-Tzu	Shih Tzu (84.0%) ✓	Shih Tzu (15.16%) ✓
French Bulldog	French Bulldog (80.0%) ✓	French Bulldog (97.15%) ✓
Tibetan Mastiff	Tibetan Mastiff (70.0%) ✓	Tibetan Mastiff (85.87%) ✓
<b>(Overall Score)</b>	<b>Collective: 50% (Average: 44.9%)</b>	<b>80.0%</b>



**David Robinson**  
Principal Data Scientist at Heap, works in R and Python.

- [Email](#)
- [Twitter](#)
- [Github](#)
- [Stack Overflow](#)

Subscribe


[Subscribe to this blog](#)

Recommended Blogs


- R Bloggers
- RStudio Blog
- R4Stats
- Simply Statistics

## Text analysis of Trump's tweets confirms he writes only the (angrier) Android half

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:



**Todd Vaziri** [@tvaziri](#)



Every non-hyperbolic tweet is from iPhone (his staff).

Every hyperbolic tweet is from Android (from him).

**J. Trump @realDonaldTrump**

luck #TeamUSA

ningCeremony #Rig

twitter.com/mS8qsQT

8,3s

8:59 PM

**J. Trump @realDonaldTrump**

ng to New Hampsh

about Hillary sayi

CIRCUITED, and

1,4

via Twi

10:20 PM · Aug 6, 2016

13.3K 9.4K people are Tweeting about this

# 5N1K



Veri nedir?  
Veri bilimi nedir?



Neden Veri bilimi önemli?



Veri bilimi nasıl kullanılır?



Veri bilimi nerede kullanılır?



Veri bilimi ne zaman kullanılır?



Veri bilimci kimdir?



Doç. Dr. İlker ÜNAL

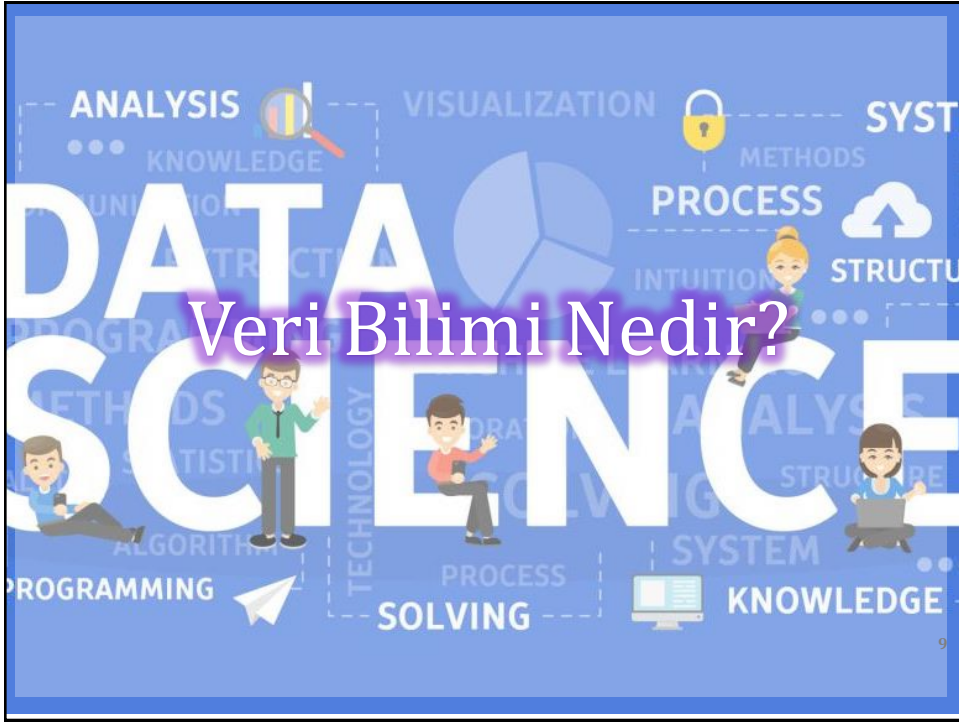
24%







91%


8





Veri Bilimi Nedir?

Veri bilimi, görünmeyen kalıpları bulmak, anlamlı bilgiler elde etmek ve iş kararları almak için modern araçlar ve teknikler kullanarak çok miktarda veriyle ilgilenen çalışma alanıdır.

Veri bilimi, tahmine dayalı modeller oluşturmak için karmaşık makine öğrenimi algoritmaları kullanır.

Verinin Bilimi

10

Doç. Dr. İlker ÜNAL

## Tarihçesi

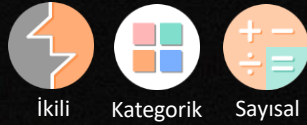
1962	John Tukey Veri analizi
1974	Peter Naur Bilgisayar bilimi için alternatif
1985	Jeff Wu Çin Bilimler Akademisi
1992	İstatistik sempozyumuna katılanlar yeni bir disiplin
2001	William S. Cleveland İstatistiğin teorisinin ötesinde teknik alanlara genişlemesi
2008	DJ Patil ve Jeff Hammerbacher Veri bilimcisi

Verinin Bilimi

11

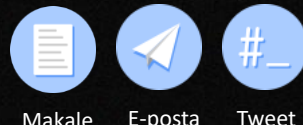
Doç. Dr. İlker ÜNAL

## Yeni Veri Tipleri



İkili Kategorik Sayısal

### Geleneksel



Makale E-posta Tweet

### Metin



Zaman bazlı Kayıtlar

### Kayıtlar



Coğrafi konum Ağlar Sensör İmajlar

### Diğerleri

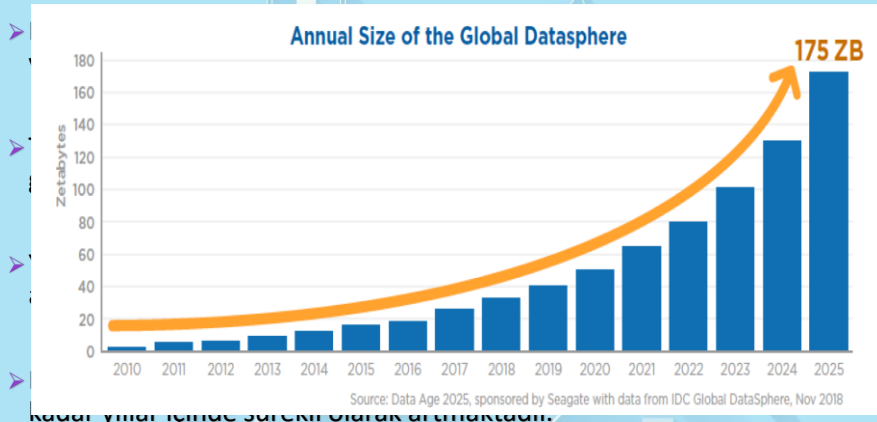
Doç. Dr. İlker ÜNAL



12

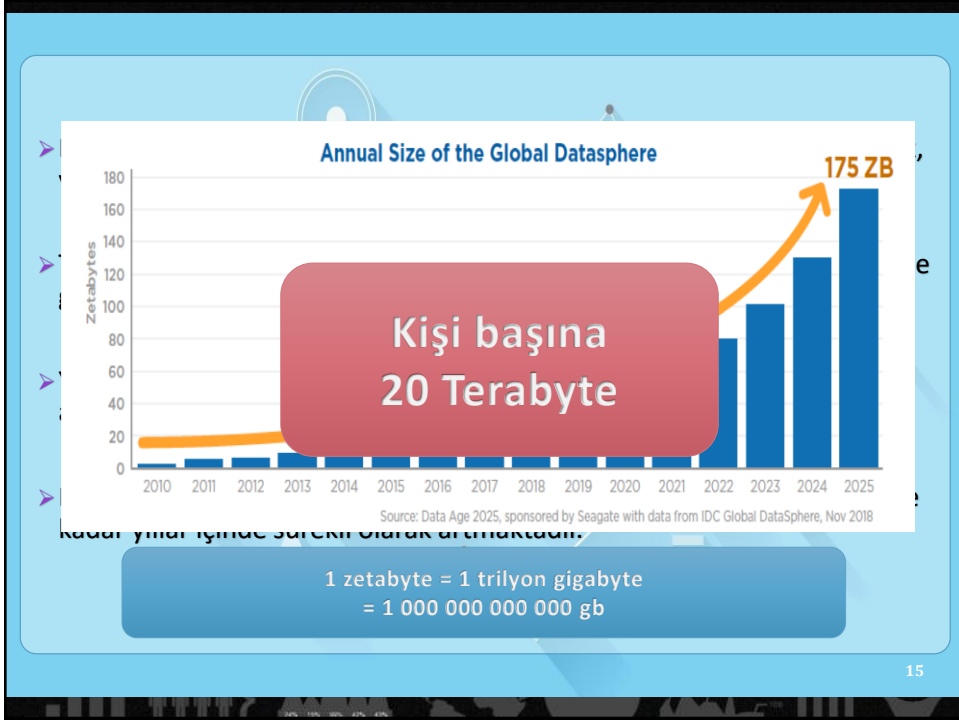
- Büyük veriler, büyük boyutlu veri kümeleridir, bu veri kümelerini işlemek, yaygın olarak kullanılan yazılımların yeteneğinin ötesindedir.
- Terim 1990'lardan beri kullanılmaktadır (John Mashey terimi popüler hale getirdi).
- Yapılandırılmamış, yarı yapılandırılmış ve yapılandırılmış verileri kapsar, ancak asıl odak noktası yapılandırılmamış verilerdir.
- Büyük verilerin boyutu, birkaç düzine terabayttan birçok zettabayt veriye kadar yıllar içinde sürekli olarak artmaktadır.

13



1 zetabyte = 1 trilyon gigabyte  
= 1 000 000 000 000 gb

14



**Büyük veri aşağıdaki özelliklerle açıklanabilir**

**Hacim**



Üretilen ve depolanan verilerin miktarı

**Çeşitlilik**



Verilerin türü ve doğası

**Hız**



Üretilme ve işlenme hızı

**Doğruluk**



Verilerin kalitesi

big data

**Verinin Bilimi**

Doç. Dr. İlker ÜNAL

16

Fotoğraflar, Bilinmeyen Yazar, CC BY altında lisanslanmıştır



## Büyük Veri Ne Kadar Büyük Olmalı?

➤ Veri çok büyük. Hepsini gerekli mi?



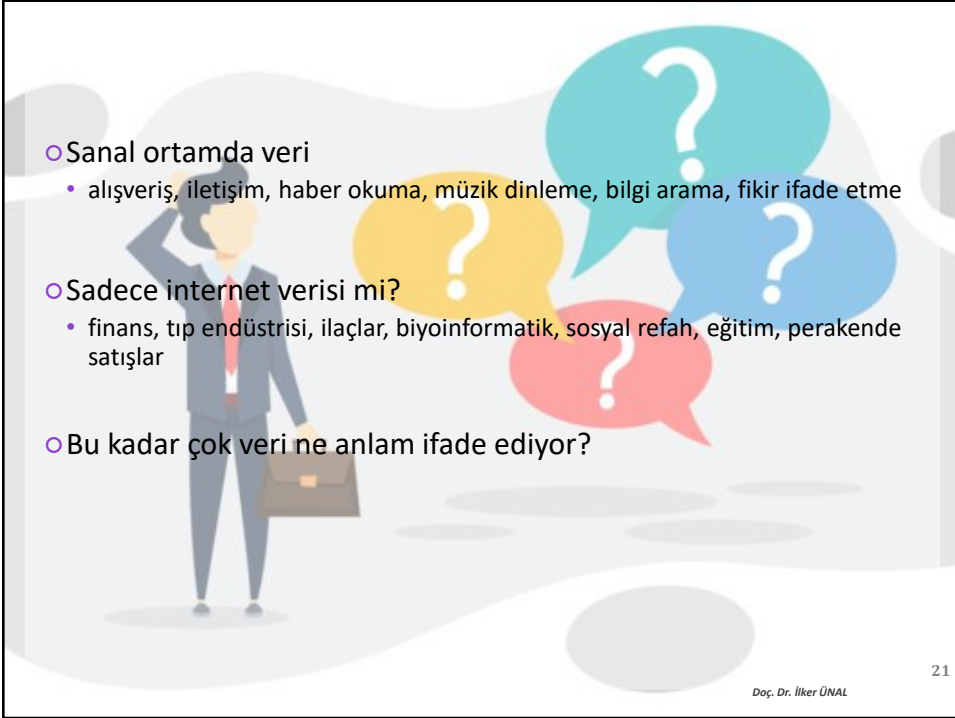
## Örnekleme

➤ Ne kadar veri gerekli?





Veri Bilimi Neden Önemlidir?



○ Sanal ortamda veri

- alışveriş, iletişim, haber okuma, müzik dinleme, bilgi arama, fikir ifade etme

○ Sadece internet verisi mi?

- finans, tıp endüstrisi, ilaçlar, biyoinformatik, sosyal refah, eğitim, perakende satışlar

○ Bu kadar çok veri ne anlam ifade ediyor?

21

Doç. Dr. İlker ÜNAL

**Verileştirme, "hayatın tüm yönlerini alıp veriye dönüştürme" sürecidir.**

**Bu, herhangi bir şeyi veriye dönüştürdüğünüzde, bu nesnelere verileştirdiğiniz anlamına gelir.**

**Örneğin, Twitter başıboş düşünceleri veri dosyası haline getirir. LinkedIn, profesyonel ağları veri dosyası haline getirir**

**«Bir şeyleri verileştirdikten sonra, amaçlarını anlayabilir ve bilgiyi yeni değer biçimlerine dönüştürebiliriz»**

KN Cukier and VM Schoenberger, (2013) «The rise of Big Data» Foreign Affairs

LinkedIn

## 2020 Emerging Jobs Report

Yapay zeka uzmanı	Robotik mühendisi	Veri bilimci
Web geliştirme uzmanı	Sistem güvenilirliği mühendisi	Müşteri başarı uzmanı
Satış geliştirme temsilcisi	Veri mühendisi	Davranışsal sağlık teknisyeni
Siber güvenlik uzmanı	Arkayüz yazılımcısı	Ana gelir yöneticisi
Bulut mühendisi	JavaScript yazılımcısı	Ürün geliştiricisi

Verinin Bilimi

Doç. Dr. İlker ÜNAL

23

LinkedIn

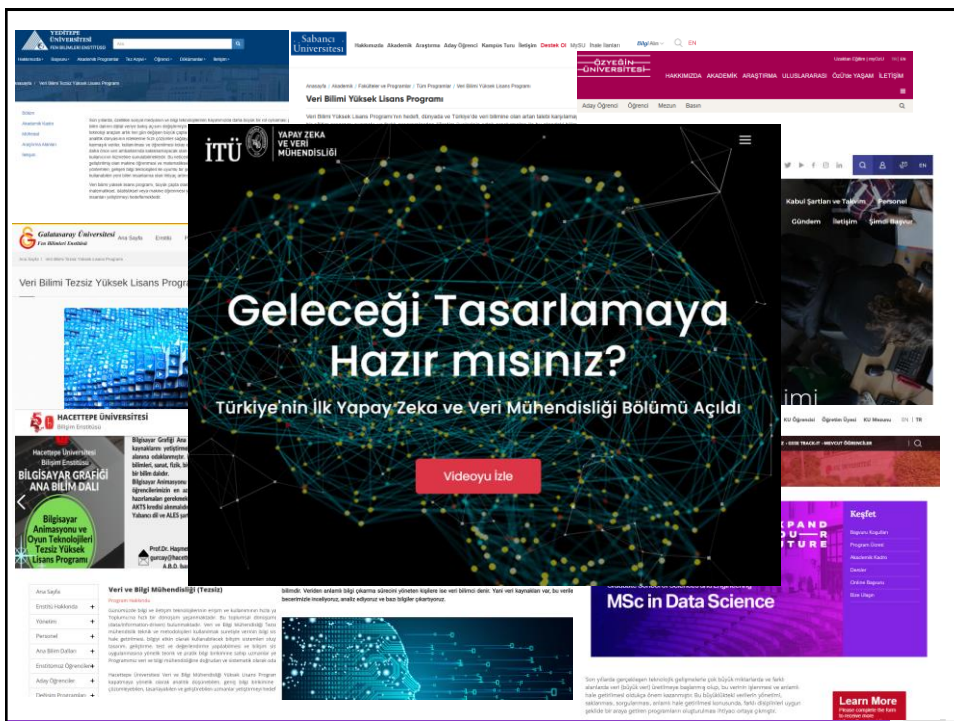
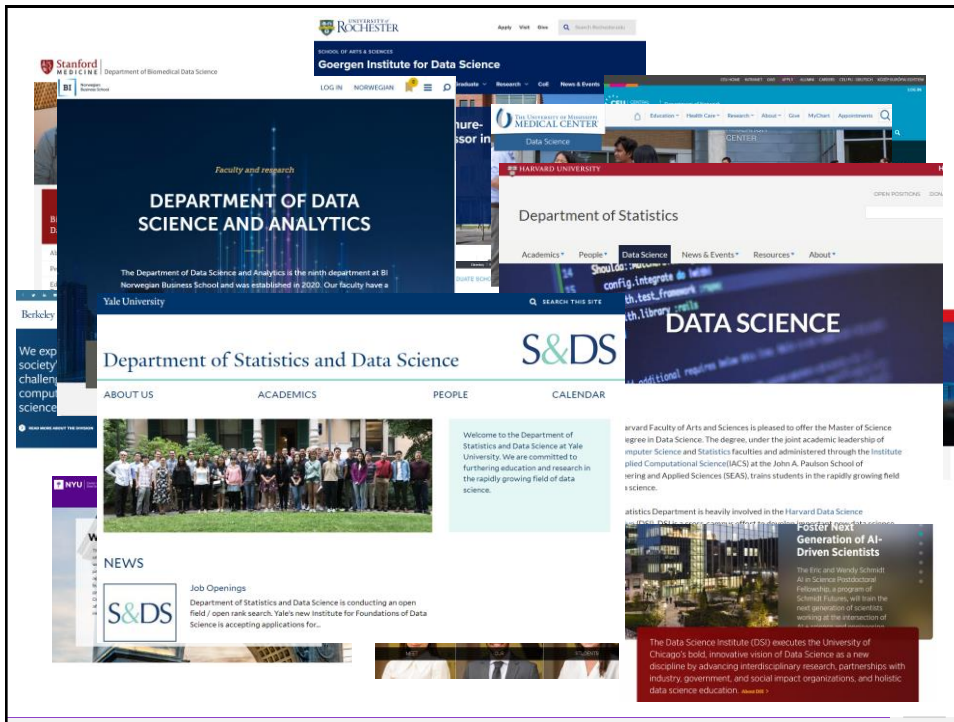
## 2021 Jobs on the Rise

Önplan e-ticaret çalışanı	Kiralama ve mortgage uzmanı	Sağlık destek personeli
İş geliştirme ve satış profesyoneli	İşyeri çeşitliliği uzmanı	Dijital market uzmanı
Hemşire	Eğitim profesyoneli	Dijital içerik geliştirici
Profesyonel ve kişisel koçluk	Uzmanlaşmış mühendisler	Akıl sağlığı uzmanları
Kullanıcı deneyimi profesyoneli	Veri bilimi uzmanı	Yapay zeka uzmanı

Verinin Bilimi

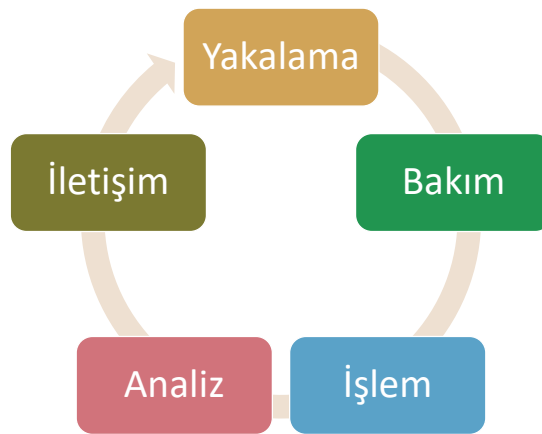
Doç. Dr. İlker ÜNAL

24

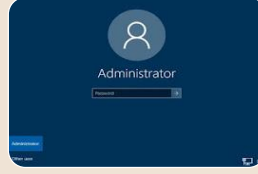




## Veri Bilimi Yaşam Döngüsü



## Sürecin Denetlenmesi



### İşletme yöneticisi

Veri bilimi eğitim yöntemini denetler

### BT yöneticisi

Çalışma yönteminin verimli ve güvenilir olduğunu denetler

### Veri bilimi yöneticisi

Ekibin günlük etkinliklerini yönetir ve prosedürleri denetler

Verinin Bilimi

29

Fotoğraflar, Bilinmeyen Yazar, [CC BY](#) altında lisanslanmıştır

Doç. Dr. İlker ÜNAL

## Veri Biliminde Kullanılabilecek Araçlar



### Veri Analizi

- R, Python
- SAS, Jupyter, Matlab



### Veri Tabanı

- SQL (MS SQL, MySQL, PostgreSQL)
- Apache Cassandra, Couchbase, Amazon Redshift



### Veri Görselleştirme

- Tableau, QlikView, Power BI
- Datawrapper, Plotly, Excel



### Makine Öğrenmesi

- Azure ML Studio, Weka, Python-Keras
- Scikit-learn, PyTorch, Apache Mahout

Verinin Bilimi

30

Fotoğraflar, Bilinmeyen Yazar, [CC BY](#) altında lisanslanmıştır

Doç. Dr. İlker ÜNAL



## Veri Bilimini Nerede Kullanırız?



Sağlık hizmetleri



Oyun



Görüntü tanıma



Tavsiye sistemleri



Lojistik



Dolandırıcılık tespiti



İnternet araması



Konuşma tanıma



Hedefli reklam



Havayolu rota planlaması



Artırılmış gerçeklik

Verinin Bilimi

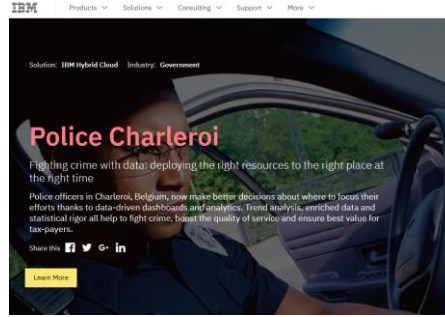
33



## Kullanım Örnekleri

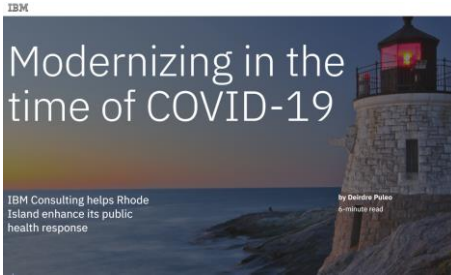
Suçu önlemek için polisin nerede ve ne zaman konuşlandırılması?

Öncelikleri anlamak ve bunlara odaklanmak ve zaman içinde bu önceliklerle başa çıkmadaki etkinliğini ölçmek için kuruluş, operasyonlarının içgüdülerden ziyade gerçekliğe dayanmasını sağlamak için veri odaklı bir yaklaşım geliştirmiştir.



Pandemide vaka ve temaslı takibi nasıl yapılabilir?

Rhode Island eyaleti, vaka ve temaslı takibini hızlandırmak için, vatandaşlardan gelen çok sayıda endişeli çağrıyı koordine edebilen ve az sayıda bir personel ile çalışan bir çağrı merkezi kurdu.



Verinin Bilimi

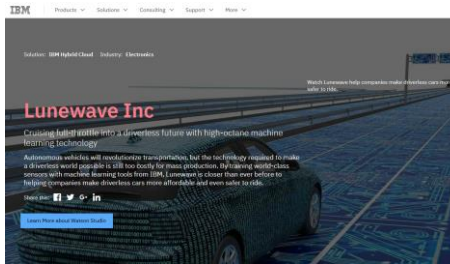
34

Doç. Dr. İlker ÜNAL

## Kullanım Örnekleri

Sürücüsüz araç sensörlerinin eğitimi

Birinci sınıf sensörleri makine öğrenimi araçlarıyla eğiten şirket, şirketlerin sürücüsüz otomobilleri daha uygun fiyatlı ve hatta sürüşü daha güvenli hale getirdi.



Netflix, kitlelerin izlemeyi tercih ettiği temaları ve kategorileri tanımak için kullanıcısının davranışını ve kalıplarını analiz etmek için veri bilimini kullanır.

Bu veriler, Umbrella Academy, Orange Is the New Black ve Queen's Gambit gibi şovlar üretmek için kullanılır.

Veri analizi, Netflix'in izleyicilerinin izlemek istediklerini bilmeden önce bile izlemek istedikleri içeriği bulmalarına yardımcı oluyor.

Verinin Bilimi

35

Doç. Dr. İlker ÜNAL

## Kullanım Örnekleri

Amazon, önleyici bir önlem olarak, şirket her sipariş için geçmiş ve gerçek zamanlı veriler toplar.

Hileli olma olasılığı daha yüksek olan işlemleri bulmak için Makine öğrenimi algoritmalarını kullanır. Bu proaktif önlem, şirketin aşırı sayıda ürün iadesi yapan müşterileri elemesinde yardımcı oldu.



Airbnb, kişilerin eğilimlerini anlamak, kullanıcı deneyimlerini derecelendirmek için çok sayıda müşteri yorumlarını, ev sahibi girdilerini kullanır. Yorumları ve bunların arkasındaki duyguları anlamak için doğal dil işlemeyi kullanır. NLP modelleri Evrişimli sinir ağları kullanılarak geliştirilmiştir.

Mükemmel bir eşleşme bulmak için misafirin önceki konaklamalarını ve alan bilgilerini dikkate alan modeller oluşturmak için derin sinir ağlarını kullanır.



## Kullanım Örnekleri

Pfizer'in yapay zeka uygulaması, 44.000 kişi üzerinde yapılan COVID-19 klinik denemesinin milyonlarca karmaşık verisindeki yapıları hızla belirledi.

Doğal dil işleme ve hasta kayıtlarının açıklayıcı veri analizi, klinik çalışmalar için uygun hastaların belirlenmesine yardımcı oldu.

Makine öğrenimi modelleri, üretim adımlarını otomatikleştirerek ve optimize ederek verimli tedarik sistemlerinin belirlenmesine yardımcı oldu.

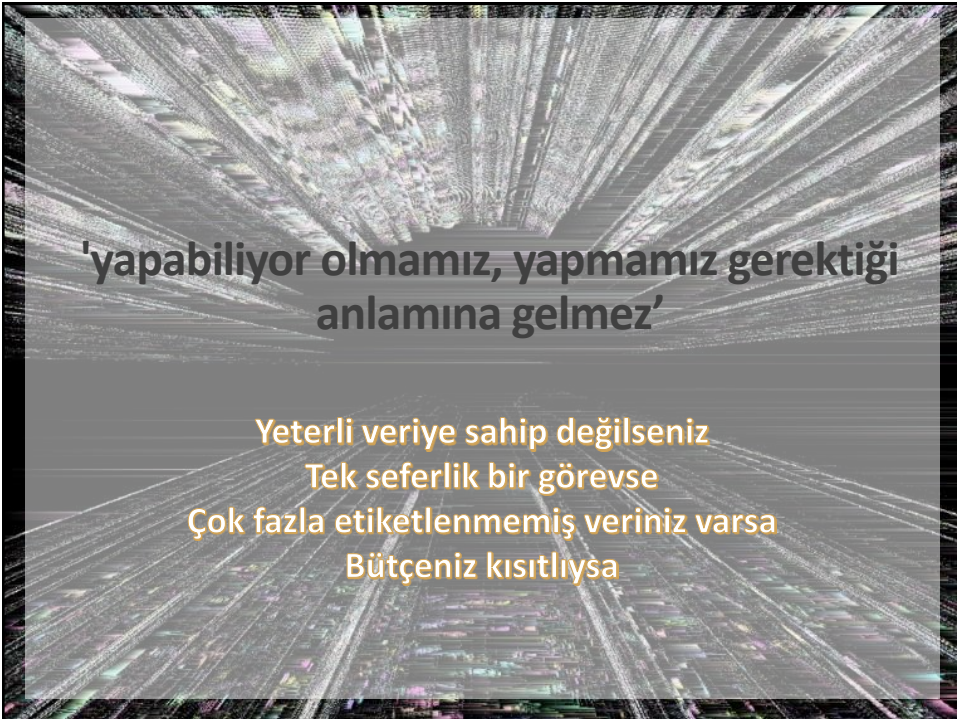


**LinkedIn**  
RECRUITER



LinkedIn işe alım uzmanı, doğrusal olmayan korelasyonları dahil etmek için Gradient Boosted karar ağaçlarına kullanır. Bu modellere ek olarak, kişiselleştirilmiş sonuçlar vermek için Genelleştirilmiş Doğrusal Karma modelini de kullanır.

Bir üye LinkedIn'i her ziyaret ettiğinde, makine öğrenimi algoritmaları, yayınlar arasında sıralama yaparak ve en alakalı sonuçları en üstte sıralayarak akışta görüntülenecek en iyi imkanları belirler.





### Veri Bilimcisi Kimdir?

- «Bilim adamı» sözü akla bilimi getirir.
- Akademide veri bilimci var mı?
- Bir veri bilimcisi, istatistik, programlama, makine öğrenimi ve ilgili alan hakkında ayrıntılı bilgi ile verilerden nasıl anlam çıkarılacağını ve yorumlanacağını bilen kişidir.
- Kimse her konuda iyi olmadığından, farklı kişilerden (alanlardan) farklı becerileri (profilleri) içeren bir veri bilimi ekibi olmalıdır.

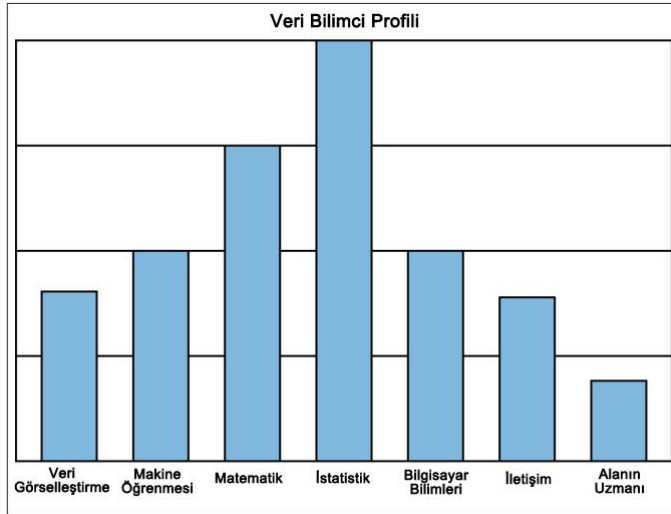
## Veri Bilimcisi Kimdir?

«Bilim

Akade

Bir ver  
alan h  
ve yor

Kimse  
farklı k



e ilgili  
ıcağını

dan)  
ır.

Verinin Bilimi

42

Doç. Dr. İlker ÜNAL

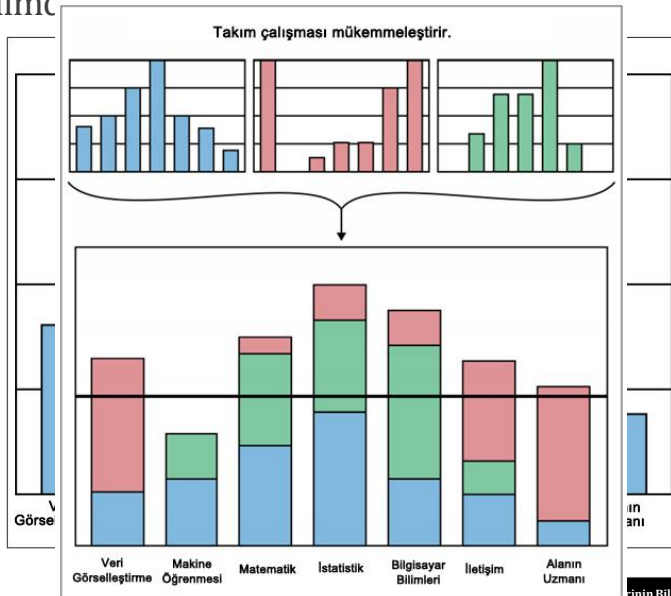
## Veri Bilimcisi Kimdir?

«Bilim

Akade

Bir ver  
alan h  
ve yor

Kimse  
farklı k



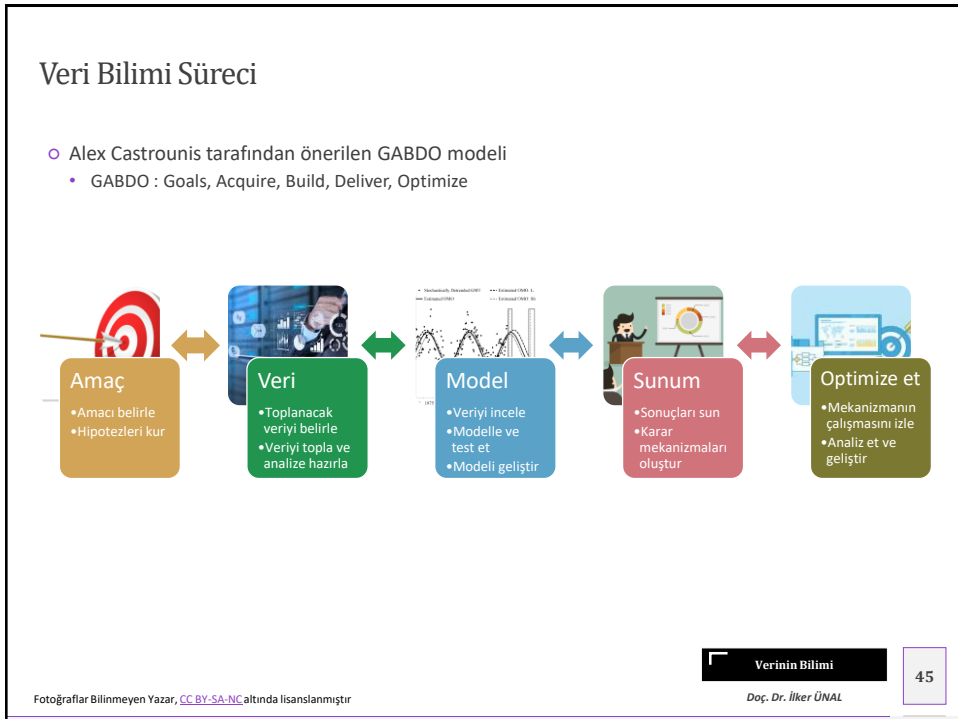
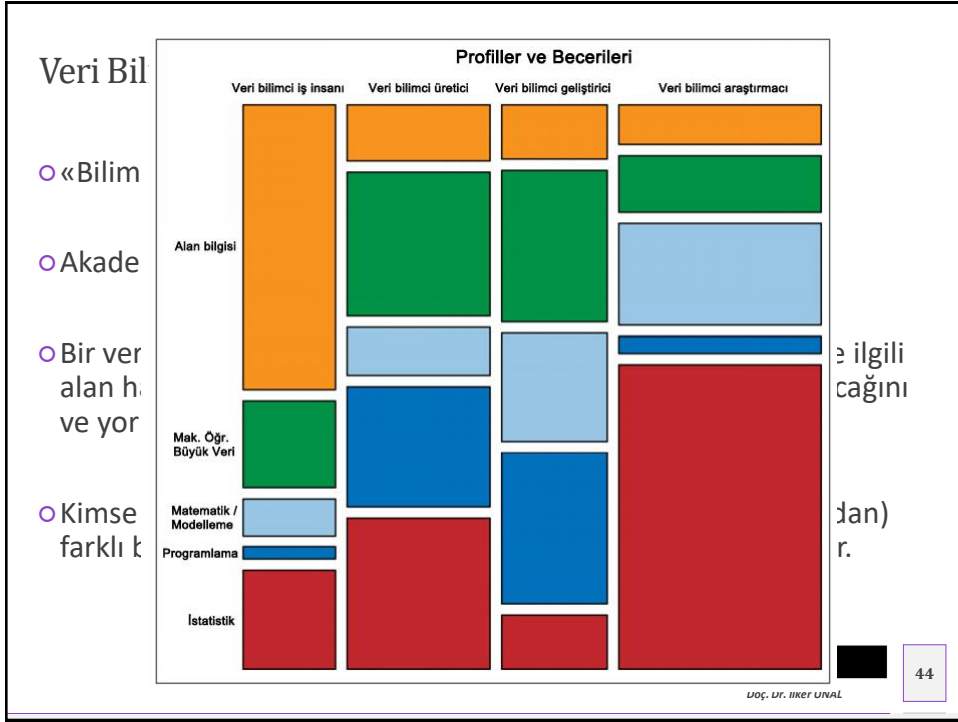
e ilgili  
ıcağını

dan)  
ır.

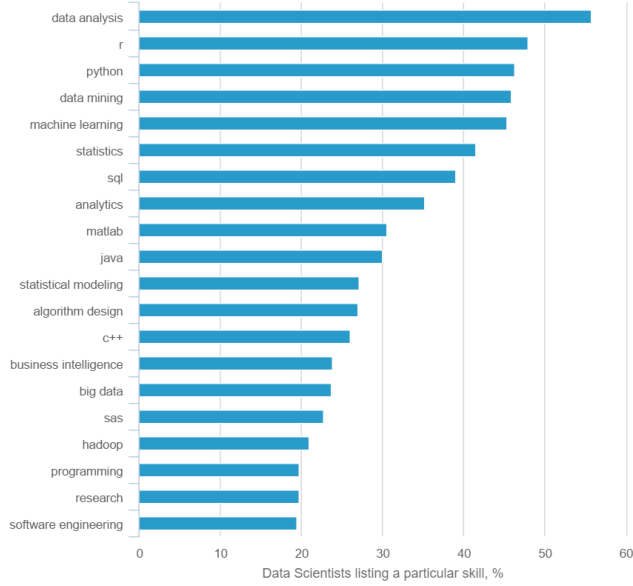
Verinin Bilimi

43

Doç. Dr. İlker ÜNAL



## Veri Bilimcinin Sahip olması gereken en önemli 20 özellik



Verinin Bilimi

46

RJMETRICS

Doç. Dr. İlker ÜNAL

## Veri Analisti - Veri Bilimcisi Farkı

Kaç tane beceriye sahipsiniz?

Ortak kısımları matematik – istatistik bilgisi ve alan bilgisi

Bilgisayar bilimleri olmadan veri bilimcisi olunmuyor.



Verinin Bilimi



### **Veri bilimi, yapay zeka ve makine öğrenimi arasındaki fark nedir?**

Yapay Zeka, bir bilgisayarın insan gibi davranmasını/düşünmesini sağlar. Veri bilimi AI alt kümesidir ve verilerden bilgi ve anlam çıkarmak için bilimsel analiz ve istatistiksel yöntemleri kullanır. Makine öğrenimi, bilgisayarlara verilerden bir şeyler öğrenmeyi öğreten AI'nın bir alt kümesidir.

### **Basit kelimelerle Veri Bilimi nedir?**

Her verinin içinde var olan bilginin çıkarılması ve yorumlanması için kullanılan analiz ve yöntemleri içeren bilimdir.



**Veri Bilimcisi ne iş yapar?**

Bir veri bilimcisi, anlamlı bulgular elde etmek için iş verilerini analiz eder ve yorumlar.

**Veri Bilimine bir örnek verebilir misiniz?**

Örneğin, finans şirketleri, kredibilitiyi ve kredi riskini değerlendirmek için bir müşterinin bankacılık ve fatura ödeme geçmişini kullanabilir.

**Veri Bilimini kendi başıma öğrenebilir miyim?**

Veri bilimi, birçok zor teknik gereksinimi olan karmaşık bir alandır. Yapılandırılmış bir öğrenme programının yardımı olmadan veri bilimi öğrenmeyi denemeniz önerilmez.

Verinin Bilimi

Dinlediğiniz için çok teşekkürler

## Kaynaklar

- Robert Chen, «Eat, Rate, Love -- A Proposal for Modifying Yelp's Rating System», [https://github.com/rchen314/SpringboardCapstone\\_YelpAnalysis](https://github.com/rchen314/SpringboardCapstone_YelpAnalysis)
- Garrick Chu, «Who's a Good Dog?» [https://github.com/thegarrickchu/Springboard-Dog\\_Breed\\_Classifier/blob/master/Capstone\\_2\\_Final\\_Deck.pdf](https://github.com/thegarrickchu/Springboard-Dog_Breed_Classifier/blob/master/Capstone_2_Final_Deck.pdf)
- Fighting crime with data: deploying the right resources to the right place at the right time <https://www.ibm.com/case-studies/police-charleroi>
- Modernizing in the time of COVID-19 <https://www.ibm.com/case-studies/state-of-rhode-island/>
- Cruising full-throttle into a driverless future with high-octane machine learning technology <https://www.ibm.com/case-studies/lunewave-inc-watson-ai-building-sensors-for-autonomous-vehicles>