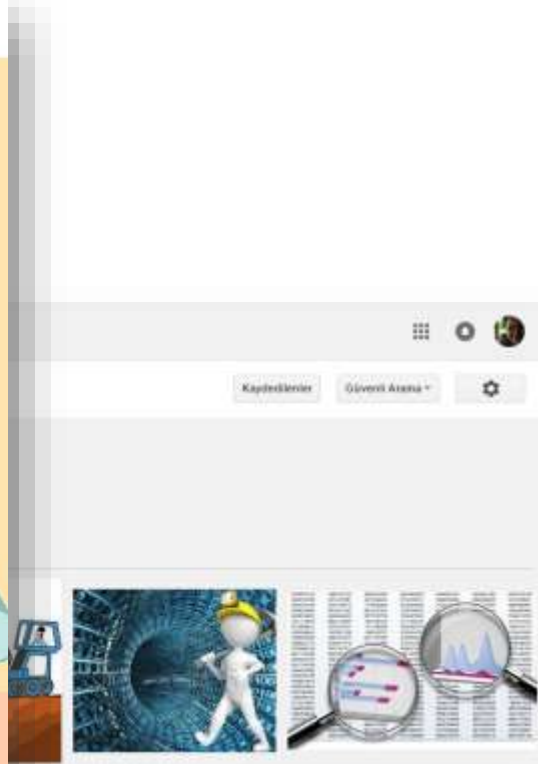
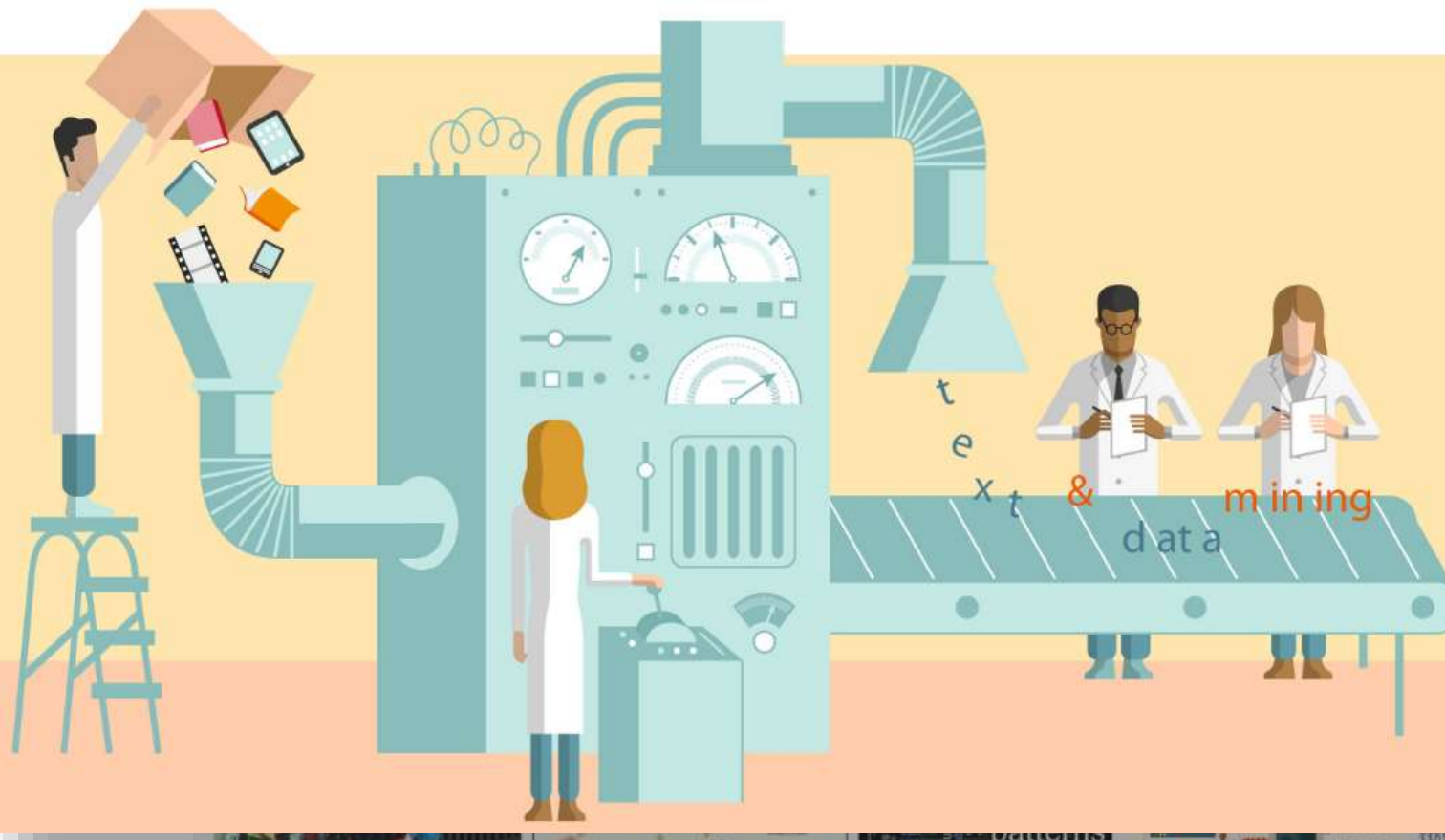
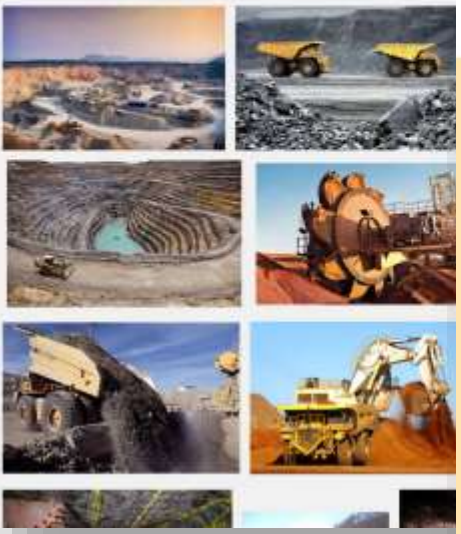


# Veri ve Metin Madenciliđi

Zehra Tařkın



# Veri Madenciliđi



- Bir kutu toplu iđne
- İçine 3 boncuk düşürdünüz
- Nasıl alacağız?
- Fikirler?

# Veri Madenciliği

- Data – Information – Knowledge
- Veri madenciliği;
  - Büyük yoğunluklu veri/metinler içinden önemli/ilginç örüntüler keşfetme görevleri
  - Toplu iğne örneği

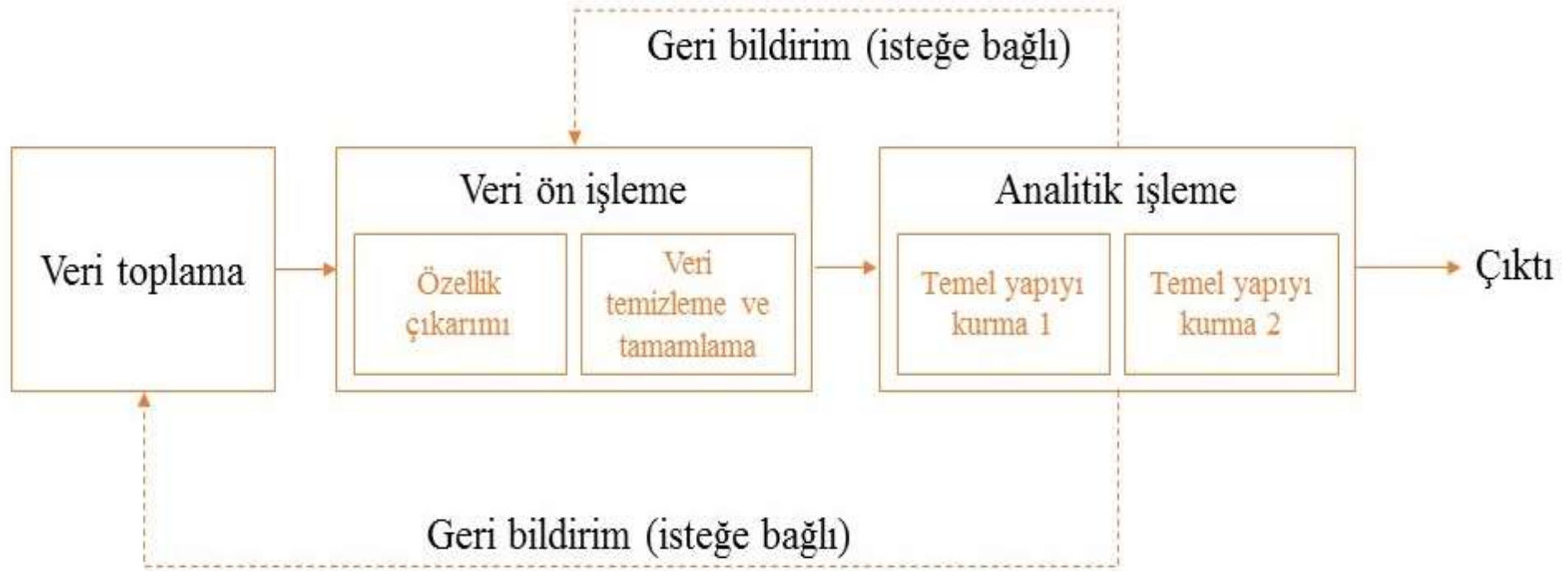
# Veri Madenciliđi

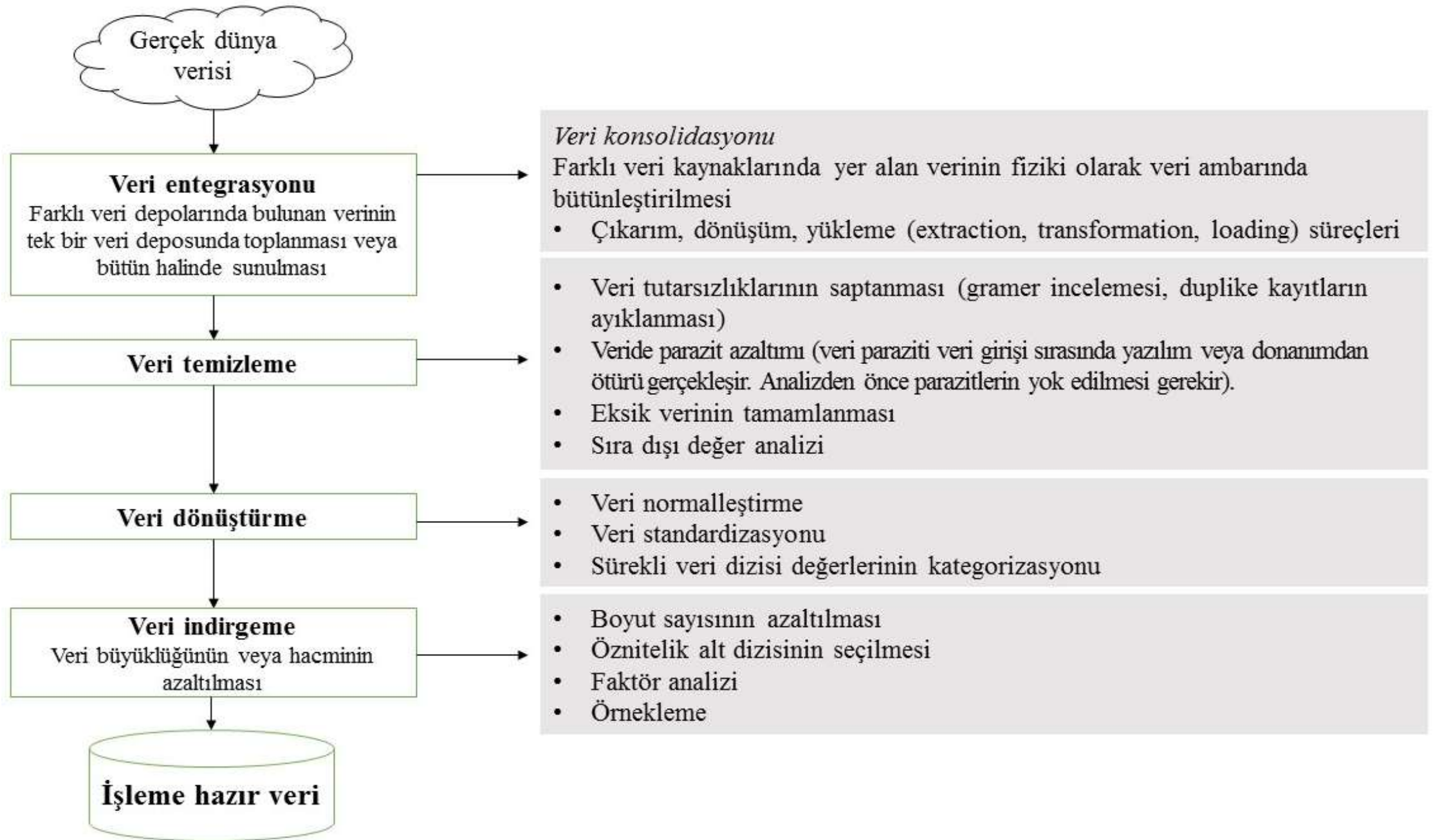
- İlişkili olduđu alanlar;
  - Bilgibilim
  - Makine öğrenme, bilgisayar bilimleri
  - Veri tabanı teknolojileri
  - Görselleştirme
  - İstatistik
  - ...

# Neden Veri Madenciliđi

- İhtiyacım olan veriyi bulamıyorum!
- İhtiyacım olan veriyi buluyorum ama elde edemiyorum!
- Bulduğum veriyi anlamıyorum!
- Anladığım veriyi kullanamıyorum!
- ...

# Veri Madenciliği Süreçleri







# Veri Madenciliği Araçları

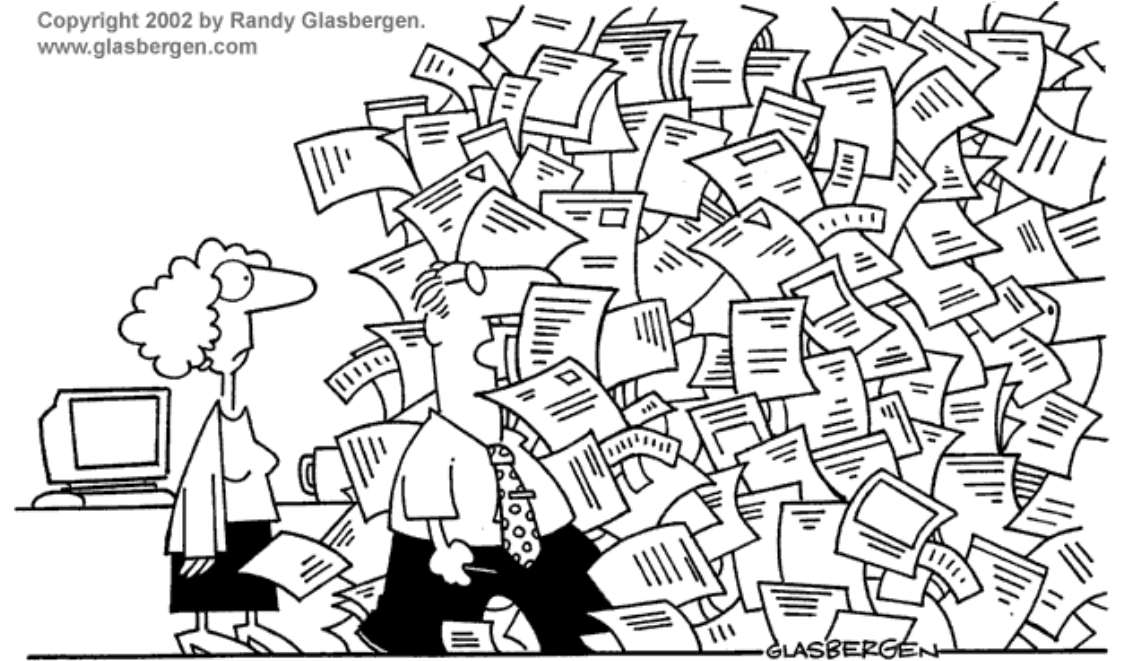
- Nooj
- Weka
- KNIME
- Angoss Knowledge Studio
- Matlab
- ...

# Veri Madenciliği Uygulamaları

- Bilgi erişim (information retrieval)
- Bilgi çıkarımı (information extraction)
- Makine çevirisi (machine translation)
- Özetleme (summarization)
- Metin kategorizasyonu (text categorization)

# Bilgi Eriřim

- İnsanların herhangi bir paragrafın, kitabın ya da büyük yoğunluklu metnin içinden ihtiyaç duydukları herhangi bir kısma erişimlerinin sağlanabilmesi
- Nasıl yapılabilir?



**“I have some paperwork to catch up. If I’m not back in two days, organize a search and rescue team!”**

# Bilgi Çıkarımı

- Büyük yoğunluklu bir metnin içinden temel anahtar unsurların (kişiler, kurumlar, konular, ülkeler gibi) tanımlanması, etiketlenmesi ve çıkarımı
- Bilgi çıkarımı yapınca işlem biter mi?
- Nooj programında bir veri çıkarımı uygulaması
  - Elde edilen verileri anlamlandırmak için ne yapılabilir?

# Makine Çevirisi

- Bir dilden diğer dile otomatik çeviri
- Algoritmalar nasıl yapılandırılabilir?

Çeviri

Anında çeviriyi kapat



Türkçe İngilizce Rusça Dili algıla



İngilizce Türkçe Rusça

Tercüme et

bir berber bir berbere bre berber gel beraber bir berber dükkanı açalım demiş



A barber a barber bre a barber come together a barber shop open



Düzenleme önerin

Türkçe İngilizce Rusça Dili algıla



İngilizce Türkçe Rusça

Tercüme et

bir berber bir berbere bre berber gel beraber bir berber dükkanı açalım demiş.



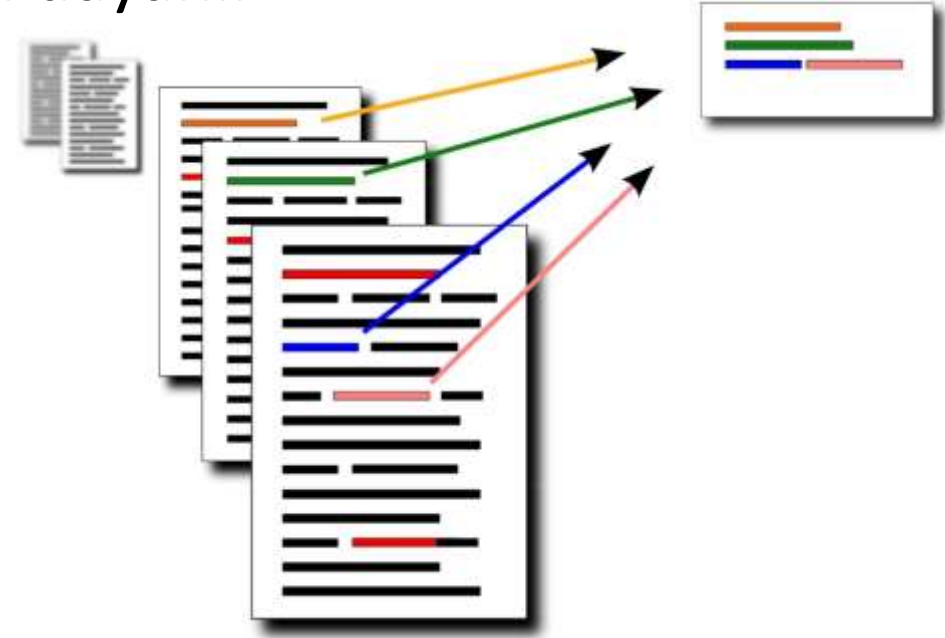
A barber, a barber, a barber, a barber, let's open a barber shop.



Düzenleme önerin

# Otomatik Özetleme

- Büyük yoğunluklu metinlerde yer alan cümle ya da paragraflar içinden,
- Dilsel veya istatistiksel bazı yöntemleri kullanarak,
- En önemli sözcükleri ya da cümleleri seçmeye,
- Metni temsil eden anlamlı bir özet çıkarmaya dayanır

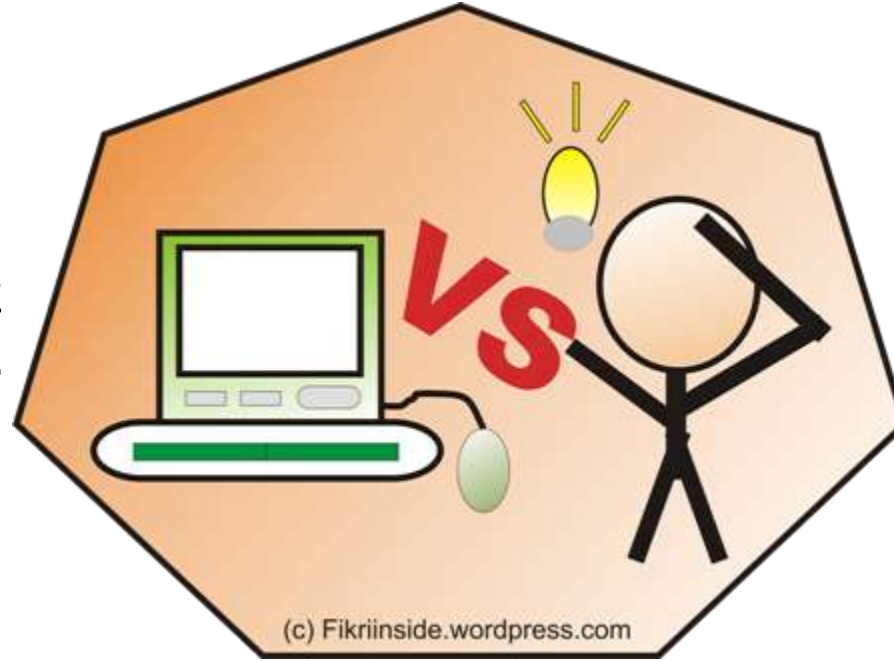


# Metin Kategorizasyonu

- Tahminleyici bir metot
- Neleri tahmin edebiliriz?
  - Hava nasıl olacak?
  - Patron kim olacak?
  - Eurovision'u kim kazanacak?
  - Golü kim atacak?
- 2 soru
  - Nasıl tahmin edecek?
  - Nasıl başarıya ulaşacak?

# Gözetimli vs. Gözetimsiz Metin Kategorizasyonu

Gözetimsiz  
Tamamen bilgisayar yardımı ile



Gözetimli  
Uzman gözetiminde

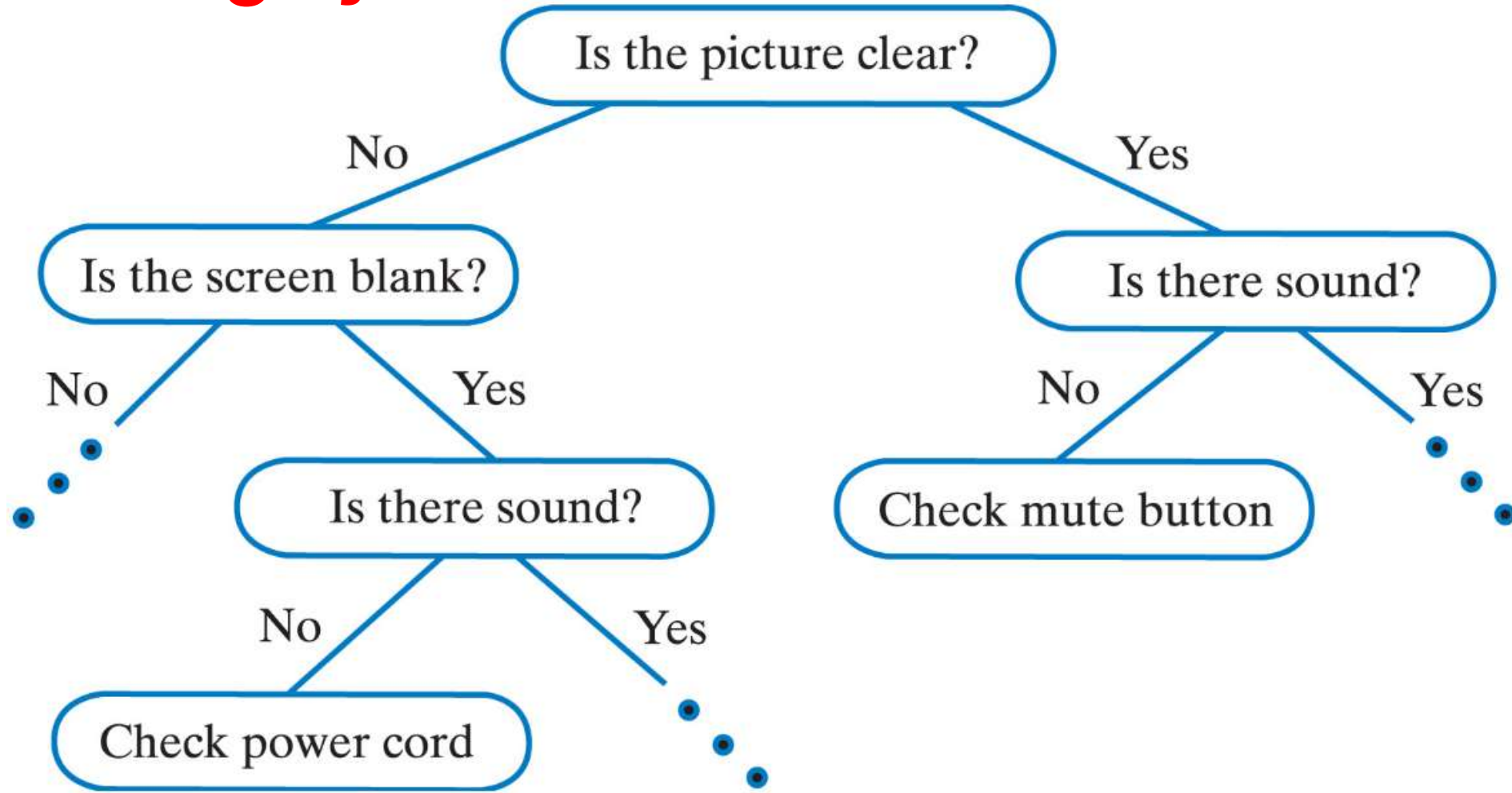
Her ikisinin de avantaj ve dezavantajları neler olabilir?



# Metin Kategorizasyonunda Kullanılan Teknik ve Algoritmalar

- Karar ağaçları
- İstatistiğe dayalı algoritmalar
- Mesafeye dayalı algoritmalar
- Yapay sinir ağları

# Karar Ağaçları



# Karar Ağaçları

- Sınıflandırma için bir ağaç oluşturulur (kök)
- Daha sonra her bir kayıt bu ağaca uygulanır (yaprak/düğüm)
- Çıkan sonuca göre kayıtlar sınıflandırılır
  
- Çeşitli algoritmaları var
  - ID3, C4.5, CART

# İstatistiğe Dayalı Algoritmalar

- Verilerin önceden sınıflara ayrılması
- Gelecekte elde edilen sonuçların tahmini
  - Yani sınıfların tahmini

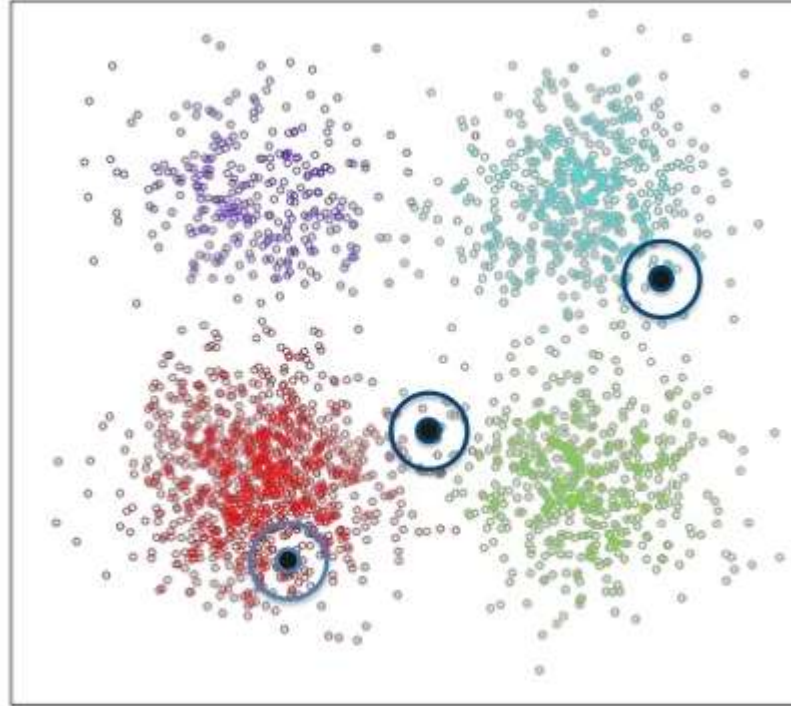
Öğrenci mezun olduysa Araştırma Yöntemleri dersinden geçmiştir.  
Yüksek notla geçtiyse istatistiğe dayalı algoritmaları daha kolay anlar.

Mesela;



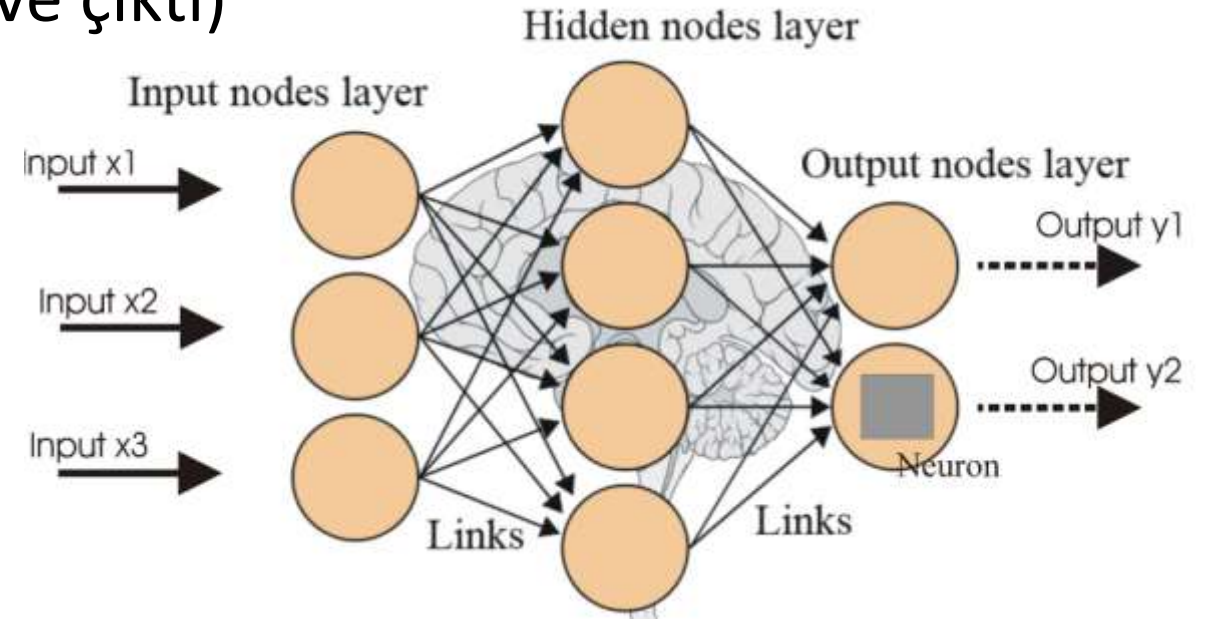
# Mesafeye Dayalı Algoritmalar

- Eldeki verilerin birbirlerine olan uzaklığı veya benzerliği kullanılarak sınıflandırma yapılır



# Yapay Sinir Ağları

- Biyolojik sinir ağlarından esinlenerek geliştirilmiştir
- Yapay sinir hücrelerinin birbirleriyle çeşitli şekilde bağlanmasından oluşur
- Genel olarak katmanlar şeklinde düzenlenir
- 3 temel katman vardır (girdi, gizli ve çıktı)



# Performans Değerlendirme

- Metodolojik değerlendirme
  - Test ve eğitim seti
  - Çapraz doğrulama
- Sayısal Değerlendirme
  - Başarımın sayısal ölçümü
  - Anma?
  - Duyarlık?
  - f değeri?
  - ROC eğrisi

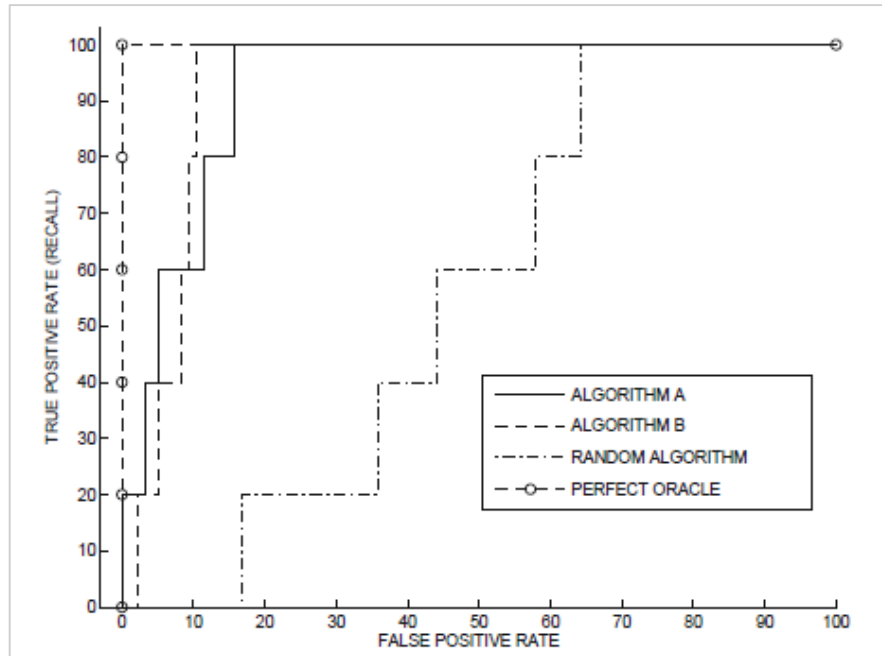
# Performans Değerlendirme

	Gerçek kategori	
	Kategori 1	Kategori 2
Tahmin edilen kategori	Kategori 1	Doğru pozitif (TP)      Yanlış pozitif (FP)
	Kategori 2	Yanlış negatif (FN)      Doğru negatif (TN)

$$\text{Anma} = \text{TP} / \text{TP} + \text{FN}$$

$$\text{Duyarlılık} = \text{TP} / \text{TP} + \text{FP}$$

$$f = \frac{2 * \text{anma} * \text{duyarlılık}}{(\text{anma} + \text{duyarlılık})}$$



→ ROC Eğrisi



# peki bu bilgi gerçek hayatta ne işime yarayacak



şükela: tümü | bugün

başlıkta ara

taşı

takip et

1

4



1

egitim sistemimizdeki antin kuntin gediklerden biridir.

beynimizi gereksiz bilgiyle bombalarlar,işimize yaramayacak şeyler öğretirler.

yani şimdi mesela kristof colomb'un amerikayı bulduğunda orayı hindistan zannetmesi gerçekte ne işimize yarar bir düşünelim.hintlilere gidip 'yaa kardeş bak böyleyken böyle,bizim kiristo amerikayı sizin mekan sanmış' falan mi diyecez?

ayip şeyler bunlar.bombalamayın ulan beynimizi.



5

#20024 22.07.1999

rotting horse on the deadly ground



2

ogrenci atasozu



6

#37182 25.09.1999

body



3

(bkz: trigonometri)



#73000 18.12.1999 16:59

peche



Want Lifehacker's email newsletter? ✕

Add your email address

[Subscribe](#)

You may also like



io9

What the Hell Is *The Walking Dead* Even Doing

Yesterday 7:55am

# The Top 50 Companies That Mine and Sell Your Data (and How to Opt Out)



Melanie Pinola

12/13/13 9:30am - Filed to: PRIVACY



157.2K



38



15

## Categories of Data

Here are the categories of marketing data used. To view, edit or remove data, click a category



We all know that our personal data is very valuable to marketers and other agencies—and that it's collected and sold by data brokers.

StopDataMining.me is a master list of opt-out links to stop these data brokers from collecting information about your online and offline activities.



[HOME](#)

[OPT OUT LIST](#)

[BLOG](#)

[OUR MISSION](#)

[CONTACT](#)



#### LATEST BLOG POSTS

---

---

## “Data Brokers: A Call For Transparency and Accountability” by the Federal Trade Commission

*Posted on May 29, 2014 by [Joel Winston](#) in [Consumers](#)*

On May 27, 2014, the Federal Trade Commission (FTC) issued a report titled, “[Data Brokers: A Call for Transparency and Accountability](#)”. In the report, the FTC concludes that data brokers operate with a fundamental lack of transparency. In addition, the FTC recommends that Congress enact legislation to make data broker practices more visible to consumers and to give consumers greater control over the immense amounts of personal information about them collected and shared by data brokers.

The report, “[Data Brokers: A Call for Transparency and Accountability](#)” is the result of a study of nine data brokers, representing a cross-section of the industry, undertaken by the FTC to shed light on the data broker industry. Data brokers obtain and share vast amounts of consumer information, typically behind the scenes, without consumer knowledge. Data brokers sell this information for marketing campaigns and fraud prevention, among other purposes. Although consumers benefit from data broker practices which, for example, help enable consumers to find and enjoy the products and services they prefer, data broker practices also raise privacy concerns.

“The extent of consumer profiling today means that data brokers often know as much – or even more – about us than our family and friends, including our online and in-store purchases, our political and religious affiliations, our income and socioeconomic status, and more,” said FTC Chairwoman Edith Ramirez. “It’s time to bring transparency and accountability to bear on this industry on behalf of consumers, many of whom are unaware that data brokers even exist.”

### SEARCH



### SUBSCRIBE

We are constantly adding new Data Brokers to our list. Be the first to know about new data reporting companies and get other useful tips on how to keep your information private.



### RECENT POSTS

“Data Brokers: A Call For Transparency and Accountability” by the Federal Trade Commission

StopDataMining.me is Featured on Lifehacker.com

Data Broker Axiom Offers Consumers a Peek Behind the Curtain

Master List of Data Broker Opt Out Links

### TAGS



Takip ediliyor

## Kimola Analytics

@kimolaanalytics

Official account for @kimola101's analytics services that provide #businessintelligence and #dataanalytics for #SocialTV, #Brands, #Trends & #Politics.

Ankara, Türkiye

kimola.com

Ocak 2014 tarihinde katıldı

Tweetie

Tanıdığın diğer 8 takipçi



1.700 Fotoğraf ve video



Tweetler

Tweetler ve yanıtlar

Medya



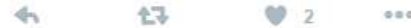
Kimola Analytics @kimolaanalytics - 1 sa.

Sosyal TV haftaya nasıl başladı? Detaylar için [kimo.la/gIIR9y7](http://kimo.la/gIIR9y7) adresine uğra ;) #İçerde #KırgınÇiçekler #BanaSevmeyiAnlat



Pazartesinin Sosyal TV Karnesi – KimolaAnalytics

Sosyal TV bu haftanın ilk günü de İçerde idi :)  
medium.com



Kimola Analytics @kimolaanalytics - 5 sa.

28 Kasım Marka Endeksi  
1- Digiturk (TV Platformu)  
2- THY (Havayolları)  
3- Mercedes (Otomotiv)  
4- Iphone (Akıllı Telefonlar)



Kimi takip etmeli · Yenile · Tümünü görüntüle



melissa terras @melissa... x

Takip et



Dücane Cündioğlu @ducane... x

Takip et



Martin Paul Eve @martin\_eve x

Takip et

Arkadaşlarını bul

Türkiye Gündemi · Değiştir

#Chapecoense

@Azuth bu konu hakkında Tweetliyor

#DünyaSevilmeyenlerGünü

17,1 B Tweet

#Kweuke

2.837 Tweet

#KalemimDiyorKi

16,7 B Tweet

#AskGöksenin

1.477 Tweet

Silivri Cezaevi'nde

2.029 Tweet

Devlet Bahçeli

17 B Tweet

Gülten Kışanak

2.096 Tweet

# Bilgi Danışmanlığı ve Veri Madenciliği

- "Talep o kadar çok ki firmalar arasında paslaşmak zorunda kalıyoruz"
- Talepler ne olabilir
- Bilgi danışmanı neyi nasıl sunar? Sunmalı?
- Sosyal medya ve büyük veri:  
<https://www.youtube.com/watch?v=z1RXxeZpBM8>