



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Çalışma ve Sosyal Güvenlik Eğitim ve Araştırma Merkezi'nin (ÇASGEM) Eğitim ve Araştırma Kapasitesinin Güçlendirilmesi Teknik Yardım Projesi TREESP1.3.CASGEM/P-01

İSG'de Veri Madenciliği Eğitimi

Yapay Öğrenmeye Giriş

Yavuz KÖMEÇOĞLU



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Ben kimim?



Yavuz KÖMEÇOĞLU

Yazılım Geliştirici
ArGe Mühendisi



Egitim

Kocaeli Üniversitesi – Matematik
Okan Üniversitesi – Bilgisayar Müh.

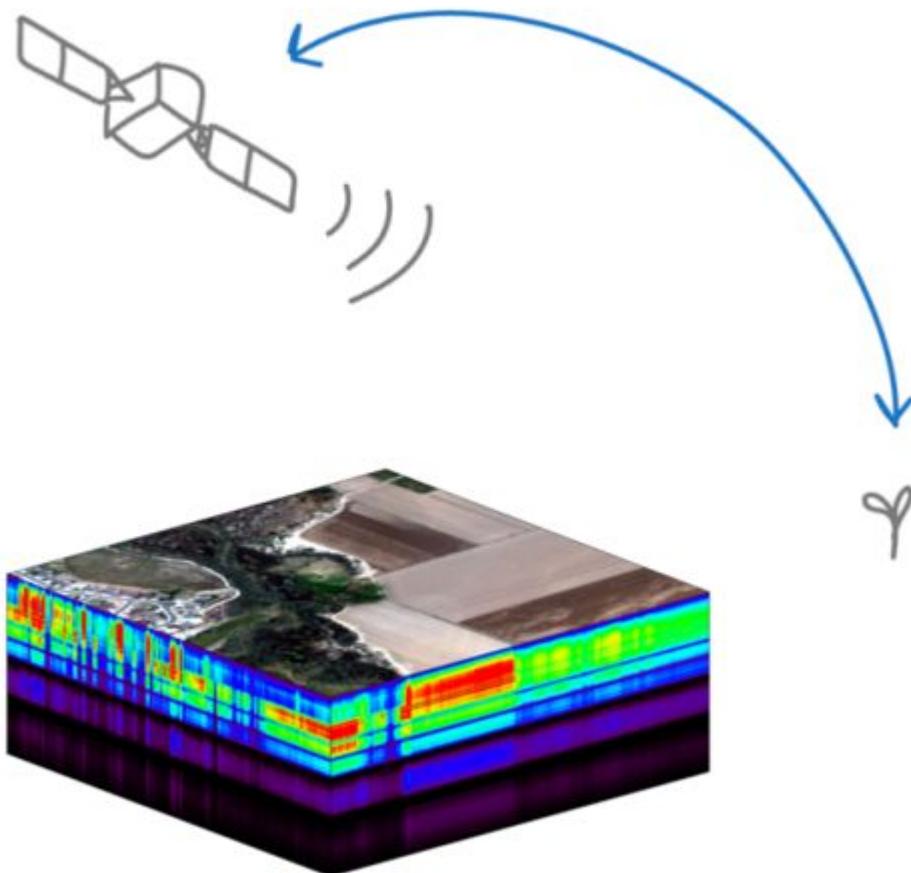


Deneyim

Kodiks Bilisim - ArGe Müh.
Kumsaati Yazılım – Yazılım Geliştirici



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



DAI



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Uzaktan Algılama ile Neler Yapılabilir?



Tarlaların durumunun izlenmesi



Orman yangınlarını izleme ve tahribatının haritalanması



Göl, nehir gibi ekolojik bölgelerin kirlilik düzeyleri



Denizde petrol sızıntılarının belirlenmesi



Kıyı şeritlerinin, kaçak yapışmasının, ormanlık alanlarının değişiminin izlenmesi



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Canlı Yayın Kaydedici ve Gerçek Zamanlı Türkçe Altyazı

Kodiks AI Editor

The screenshot shows a video player window with a video frame showing traffic on a highway. A red banner at the bottom of the video frame reads: "www.tv41.com.tr 18 MART SAAT 18:00 İTİBARI İLE YASAK GETİRİLDİ... MOTO KURYELER YİNE KONTAK KAPATTI". Below the video frame is a subtitle editor panel titled "Subtitles". The subtitles are listed in a table with start times and text content. The video progress bar at the bottom shows a timeline from 0:09 to 0:18, with several subtitle segments mapped to specific time points.

| Start Time | Text Content |
|--------------|--|
| 00:00:00.000 | bu moto kuryeler yine kontak kapattı |
| 00:00:02.790 | |
| 00:00:02.790 | Kocaeli Vatiliği li genelinde 18.03.2022 |
| 00:00:06.690 | |
| 00:00:06.690 | Cuma saat 18 sıfırdan başlamak |
| 00:00:09.389 | suretiyle |
| 00:00:09.389 | ev ve işyerlerine hizmet veren notu |
| 00:00:11.639 | |
| 00:00:11.639 | kuryelerin çalışmasına motosikletle |
| 00:00:13.860 | |
| 00:00:13.860 | mobilet kullanımına ikinci bir duyuruya |
| 00:00:15.750 | |
| 00:00:15.750 | kadar ilimiz genelinde durdurulmasına |
| 00:00:18.150 | |
| 00:00:18.150 | karar verilmiştir duyurusunda bulundu |



İNŞAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Türkçe Haber Kategori Sınıflandırma

[kodiks / turkish-news-classification](#) Public

[Code](#) [Issues](#) [Pull requests](#) [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#) [Settings](#)

[main](#) [1 branch](#) [0 tags](#) [Go to file](#) [Add file](#) [Code](#)

 [serdarakyol](#) Update README.md 8f18d40 on Jul 14, 2021 4 commits

[data](#) added notebooks 6 months ago
[notebooks](#) added notebooks 6 months ago
[results](#) added notebooks 6 months ago
[.gitignore](#) Initial commit 6 months ago
[README.md](#) Update README.md 6 months ago

[README.md](#)

Turkish News Category Classification Tutorial

Blog

- "Turkish News Category Classification Tutorial" blog yazısında burada bulunan çalışma dosyalarını kullanarak metin sınıflandırma problemi için bir yapay öğrenme modeli geliştirme adımlarını bulabilirsiniz.
- "HuggingFace Türkçe Veri Kümeleri" blog yazısında HuggingFace Datasets Kütüphanesine yeni veri kümelerini entegre etme maratonunda eklediğimiz veri kümeleri hakkında detaylı bilgi alabilirsiniz.

About

Turkish News Category Classification Tutorial

nlp machine-learning text-classification datasets svm-classifier news-classification tf-idf-vectorizer huggingface turkish-nlp

Readme 17 stars 7 watching 3 forks

Contributors 2

 [serdarakyol](#) Serdar AKYOL
 [yavuzKomecoglu](#) Yavuz Kömeçoğlu

Languages

Jupyter Notebook 100.0%



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI

KODİKS



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Türkçe Doğal Dil İşleme API'si

- Ön işlemler
- Word Embeddings
- Doküman Benzerliği
- Haber Kategori Sınıflandırma

Kodiks NLP Docs on GitLab

[Home](#) [API Preprocessing](#) [API WordEmbedding](#) [API Similarity](#) [Python Library](#)

API Similarity

Türkçe haber benzerlikleri ile ilgili işlemler

İÇİNDEKİLER

- [ContentSimilarity](#)
- [ContentSimilarityAndCitation](#)
- [ContentListSimilarity](#)
- [ContentListSimilarityAndCitation](#)

ContentSimilarity

POST [api/similarity/contentsimilarity](#)

Request:

```
{  
    "similarity_type": "word2vec",  
    "threshold": "0.75",  
    "content1": "İtfaiyeden güvercin operasyonu Haber İtfaiyeden güvercin operasyonu Lüleburgaz ilçesini",  
    "content2": "İtfaiyeden güvercin operasyonu Haber İtfaiyeden güvercin operasyonu Kırklareli'nin Lüleburgaz ilçesini"  
}
```

Response:

```
{  
    "similarity_rate": 95.56,  
    "is_similar": true  
}
```



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI

KODİKS



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Türkçe Doğal Dil İşleme API'si

Gerçek Zamanlı Haber Gruplama

The screenshot shows a web-based news grouping application. At the top, there is a navigation bar with links for 'Tümü', 'Basın', 'Televizyon', 'Radyo', 'internet', 'Sosyal Medya', and 'Ajanslar'. Below the navigation is a search bar with 'Kategori Ara...' and a date filter 'Tarih' set to 'Yaz'. A sidebar on the left lists categories: 'MT SEKTÖR HABER', 'GAZETE MANŞETLERİ', and 'MANŞET HABERLERİ'. The main content area displays a table of news items. The columns are 'No', 'Yayın Tarihi', and 'Başlık'. The first item is 'GRUPLANMAMIS HABERLER' from 05/11/2020. The second item is 'COVID-19 HAYATIMIZA GİRELİ 1 YIL OLACAK' from the same date. The third item is 'ABD SEÇİMLERİ KARAKOLLUK' from the same date. Below this, there is a summary section titled 'Toplam Haber Sayısı : 4 - Basın: 4' with a table showing the distribution of news items across different media sources (AYDINLIK, KARAR, YENİ AKIT, YENİ MESAJ) and their respective counts (1, 1, 1, 1).

Real-time News Grouping: Detecting the Same-content News on Turkish News Stream

Yavuz Kömeçoğlu¹, Başak Buluz Kömeçoğlu² and Burcu Yılmaz²

¹ Kodiks Bilişim, İstanbul, TURKEY , komecoglu.yavuz@gmail.com

² Gebze Technical University, Kocaeli, TURKEY, {bbuluz,byilmaz}@gtu.edu.tr

Abstract. The increasing number of news sources makes analysis of the news difficult and increase the need for automated systems. This paper presents a system that clusters news with similar content in real time. The system uses the Apache Solr database to capture texts from the news source and its MoreLikeThis (MLT) search component to extract the 5 most similar news from thousands of previously recorded news. The new news will be included in the cluster with the most similar of the 5 news obtained by pre-filtering. Therefore, the main problem sought in this study is finding the news that most closely resembles the new news. For this, a 2-step approach has been proposed. The majority of the news published in different sources in the media is created by the same or broadening / shortening of the same text. For this reason, the 'citation rate' was calculated primarily among news pairs. If there are news pairs that exceed the citation threshold, the pair with the highest citation rate is included in the same cluster. Otherwise, the numerical representations of the texts at different levels were used in order to determine the similarity semantically. The results of the study show that the proposed 2-stage approach reduces the sensitivity of embeddings at different levels to text lengths. Thus, it achieved up to 7.6% improvement compared to clustering approach only with embeddings. The system proposed in these study has a structure that can be used in real life applications in terms of real-time clustering with a high F-score rate of over 90%.

Keywords: Real-time News Grouping, B-Cubed, Text Clustering, Natural Language Processing.

KODIKS



İNSAN KAYNAKLARININ
GELİŞİTRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

PDF Dökümanlarını Otomatik Ayrıştırma ve Zengin Metin Çıkarma Aracı

BAŞLIK İŞIKLARI KESİNTİLERDEN DOLAYI ÇALIŞMADI, SÜRÜCÜLERİN İNSAFINA KALDIK

KESİNTİ MAĞDURIYETİ ESNAF DA VATANDAŞ DA KESİNTİLERDEN USANDI

ALT BAŞLIK 'te sıcak havalarda her gün yaşanan elektrik kesintileri vatandaşları ve esnafı çileden çıkarıyor. İşyerlerinde elektrik kesintisi dolayısı ile iş yapamayan esnaf, evine gittiğinde de elektriksiz kalıyor. Kesintiler sebebiyle trafik ışıkları da çalışmamınca birçok noktada sürücüler ne yapacağını şaşırıyor. Tüm çağrırlara rağmen ise kesintiler son bulmuyor.

RESİM-1

Gaziantep'te koşaylaştan yaşanan elektrik kesintisi tıraşçıdan vatandaşın yoğunluğu birakmış, yaşanan mağduriyete elektriköz kalan mahalleler isyan etmiş. Yaz sıcaklığının etkisiyle bu günde ise yeniden elektrik kesintisi yaşasının başlamasına vatandaşlar tepki gösterdi. Gaziantep'te her gün merkez ilçeler başta olmak üzere ilçelerde sayıları dolusu kesişim liseleri yayılan Torsoslar EDİS, pank kesintilerin ötürüne girmeye başladı.

İŞLETMELER MÅ DUR OLDU
Kent genelindeki büyük esnaf, jen ekibinden olmamasından dolayı herkesin içinde iş brakmaktadır. Uygun planlı kesintiler sebebiyle pan demediği kayıtları normalleşme sürecinde kazanılarak gerekmediklerini aktaran esnaf, çözmek bekliyor.

TRAFFIC LIGHTS DEVRE Dİ
Elektrik kesintisi yaşanan bölgelerde bulunulan trafik ışıkları da kesintiler sebebiyle devre dışı kalyor. İpekyolu olası bilinen yolun kapanması

RESİM-2

araçların şehrin içinden geçerek seyahat etmesi ile kesintilerin yaşandığı tarihi alt üst ediyor. Kesinti tarihinde devre dışı kalan trafik ışıkları vatandaşlara zor anları yaratıyor.

PLANLI KESİNTİLER İSTENİYOR
Kent genelinde meydana gelen kesintileri kamuoyu ile paylaşan Toro star EDİS'in kesimi saatleri sebebiyle de tekrar toplayıyor. Uzun saatlik elektrik izkili vatandaşlar bude zamanındaki gidişaların, bayez yalamın bozulduğu un dileyip getiriyor. Bugün yaşanan sık sık sınırlı olan 4. İskender Mahallesi'nde 09:00-12:00 saat aralığında 16 sokakta, Üçoklar Mahallesi'nde 09:00-15:00 saat aralığında 27 sokakta, 12:30-18:00 saat aralığında Akdere Mahalle'si'nin 31 sokakta, 01:00-03:00 saatlarında Karataş Mahallesi, Alkent Mahallesi, Bülbülzade Mahallesi ve Ertuğrulgazi Mahallesi'nde 94 sokakta elektrik enerjisinin verilemeyeceğini duyurdu. (NEVRE DUMAN)



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Teknofest 2021: Ulaşımda Yapay Zeka Türkiye 1.liği



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Bölüm1 - İçeriği

- 1. Gün: **Makine Öğrenmesine Giriş**
- 2. Gün: **Lineer Cebir Temelleri**
- 3. Gün: **Veri Analizi ve Öz nitelik Mühendisliği**
(Öz nitelik Seçimi ve Boyut Azaltma)
- 4. Gün: **Olasılık ve İstatistik - Ayırt Edici Model ile Üretici Model Farkları**
- 5. Gün: **Doğrusal Regresyon ve Lojistik Regresyon**
- 6. Gün: **Destek Vektör Makineleri (SVM)**
- 7. Gün: **Karar Ağacı - Üretici Model**
- 8. Gün: **Optimizasyon**



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Bugün Neler Öğreneceğiz?

- Makine Öğrenmesi Tanımı
- Neden Makine Öğrenmesi Kullanılmalı?
- Makine Öğrenme Süreci
- Veri Toplama Adımları
- Veriyi Modele Uygun Hale Getirme
- Denetimli Öğrenme
- Denetimsiz Öğrenme
- Model Performansı
- İş Sağlığı ve Güvenliği Uygulama Örnekleri



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Giriş



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

İç İçe Geçmiş Bazı Tanımlar

YAPAY ZEKA
Artifical Intelligence

MAKİNE ÖĞRENMESİ
Machine Learning

DERİN ÖĞRENME
Deep Learning



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Makine Öğrenmesi Tanımı



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



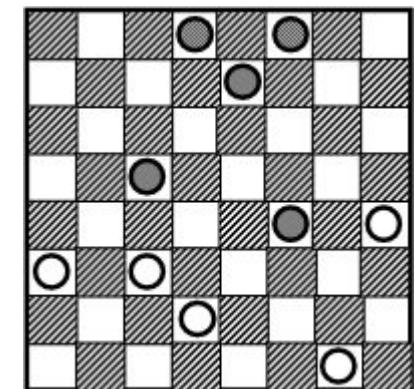
Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Makine Öğrenmesi Tanımı

Makine Öğrenimi, bilgisayara açıkça programlanmadan öğrenme yeteneği veren çalışma alanıdır.



Arthur Samuel, 1959



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



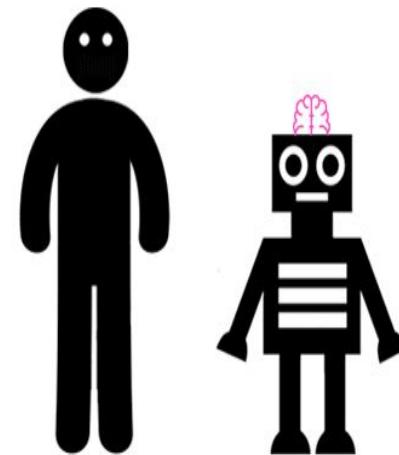
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Makine Öğrenmesi Tanımı

Bir bilgisayar programının (E) deneyimi sayesinde bazı (T) görevleriyle ilgili olarak öğrenme yoluyla, bazı performans ölçütleri (P) ile ölçümlenmiş bir şekilde performansının artmasıdır.



Tom Mitchell, 1997



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Neden Makine Öğrenmesi Kullanılmalı?



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ

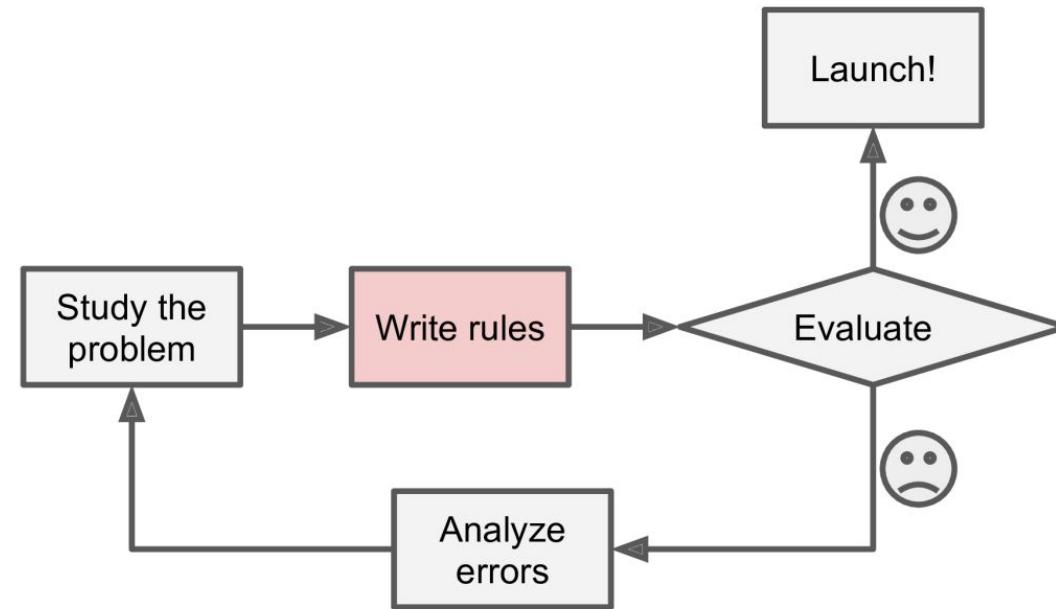


T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Geleneksel Yaklaşımlar Spam E-posta Sınıflandırma

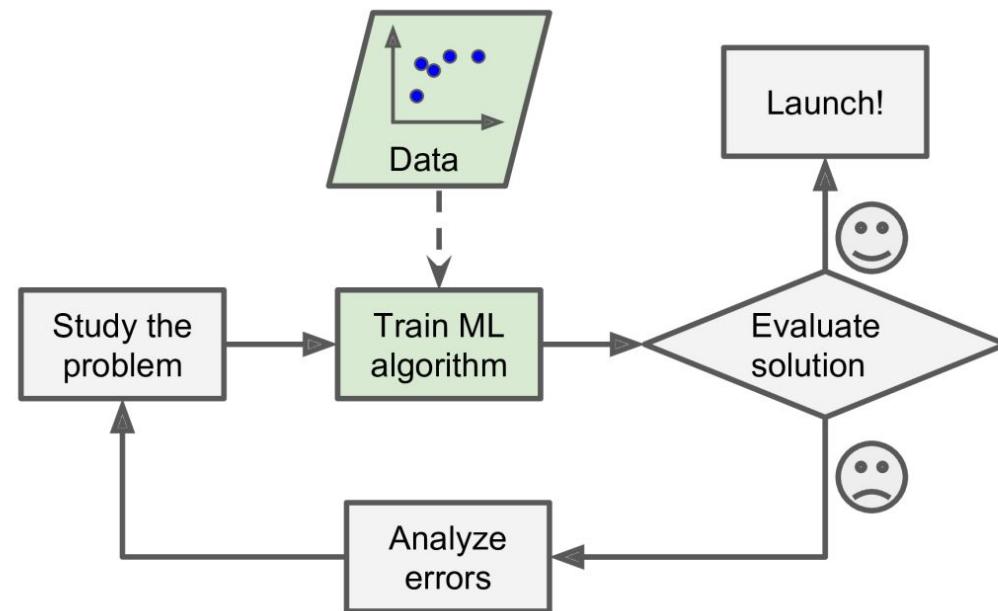


"Hands-On Machine Learning with Scikit-Learn and TensorFlow" Kitabı - Sayfa 5, Şekil 1-1



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Makine Öğrenmesi Yaklaşımıyla Spam E-posta Sınıflandırma

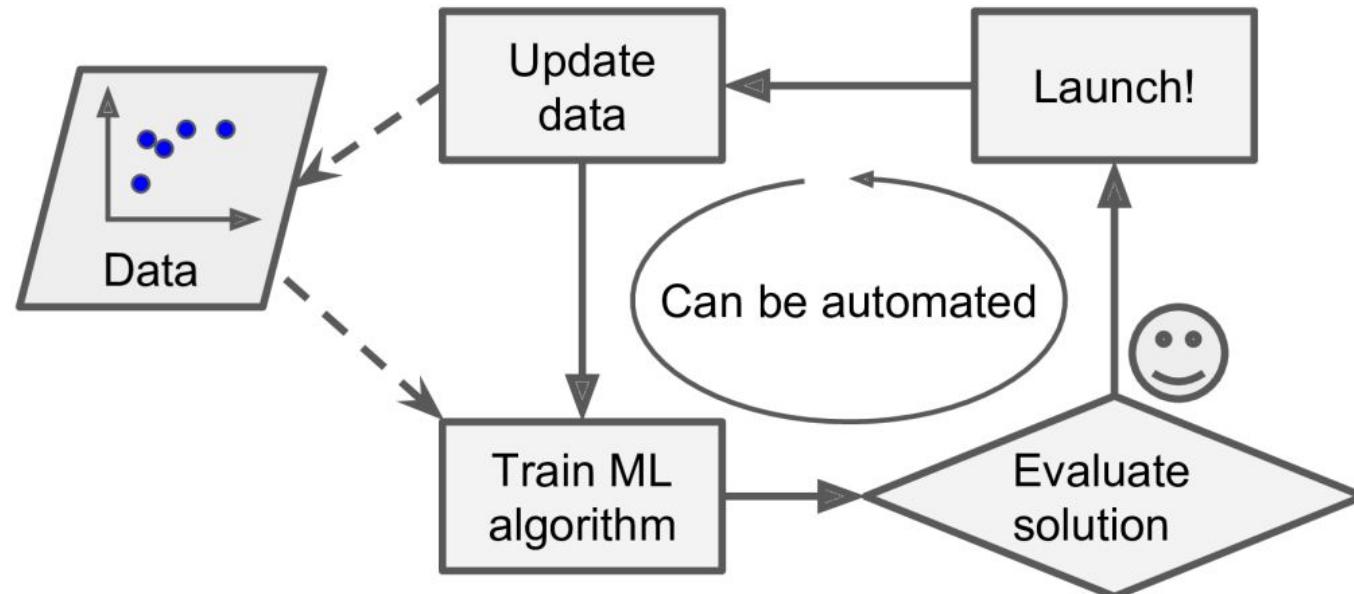


"Hands-On Machine Learning with Scikit-Learn and TensorFlow" Kitabı - Sayfa 6 Şekil 1-2



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Değişime Otomatik Uyum Sağlama

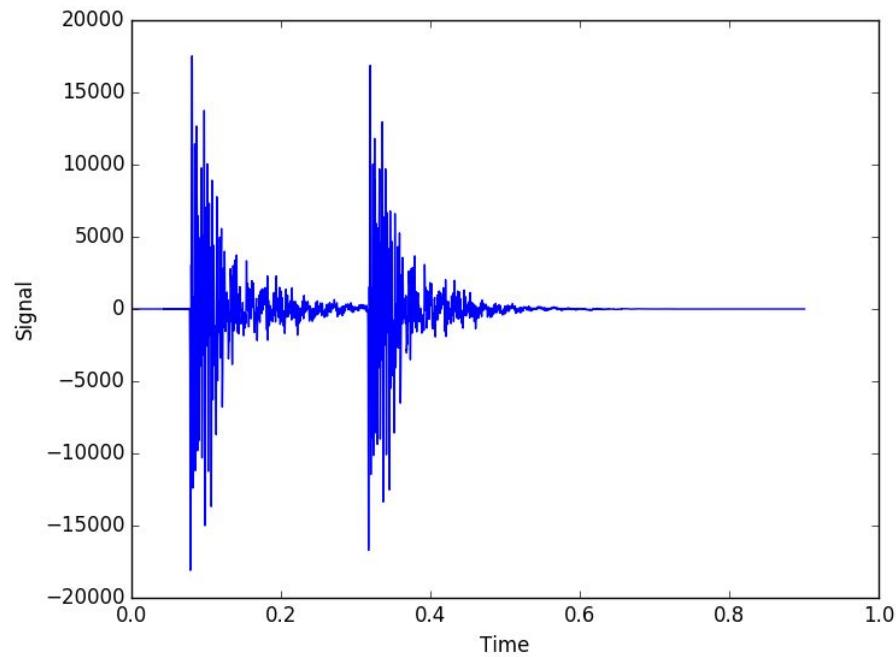


"Hands-On Machine Learning with Scikit-Learn and TensorFlow" Kitabı - Sayfa 6 Şekil 1-3



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Konuşma Tanıma



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



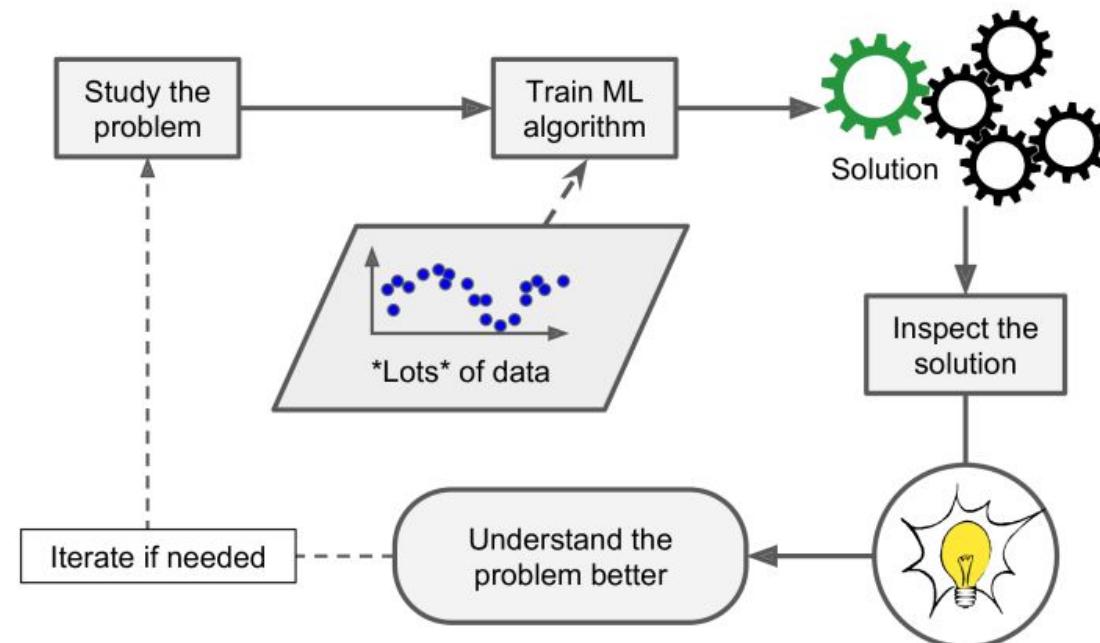
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Makine Öğrenimi, insanların öğrenmesine yardımcı olabilir



"Hands-On Machine Learning with Scikit-Learn and TensorFlow" Kitabı - Sayfa 7 Şekil 1-4



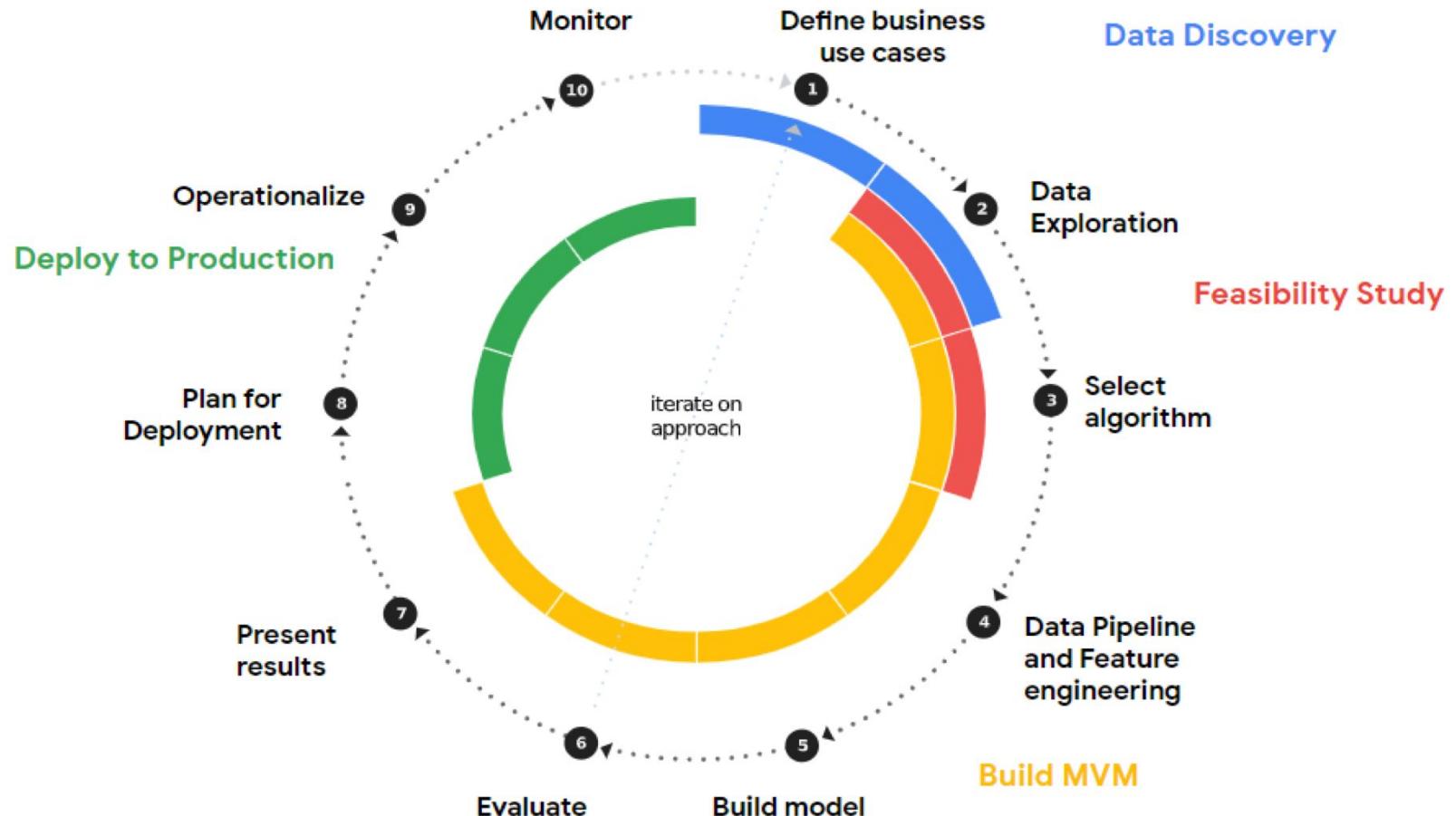
Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Özetle ne zaman/hangi durumlarda makine öğrenmesi?

- Geleneksel yöntemlerle bulunan çözümlerde çok fazla elle ayarlama veya uzun kural listeleri gerektirdiğinde
- Geleneksel bir yaklaşım kullanarak hiçbir şekilde iyi bir çözüm bulunamayan karmaşık problemlerde
- Yeni verilere uyum sağlamayı gerektiren değişken ortamlarda
- Büyük miktarda veriler ve karmaşık sorular hakkında bir öngörü elde etme ihtiyacı duyulduğunda



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ

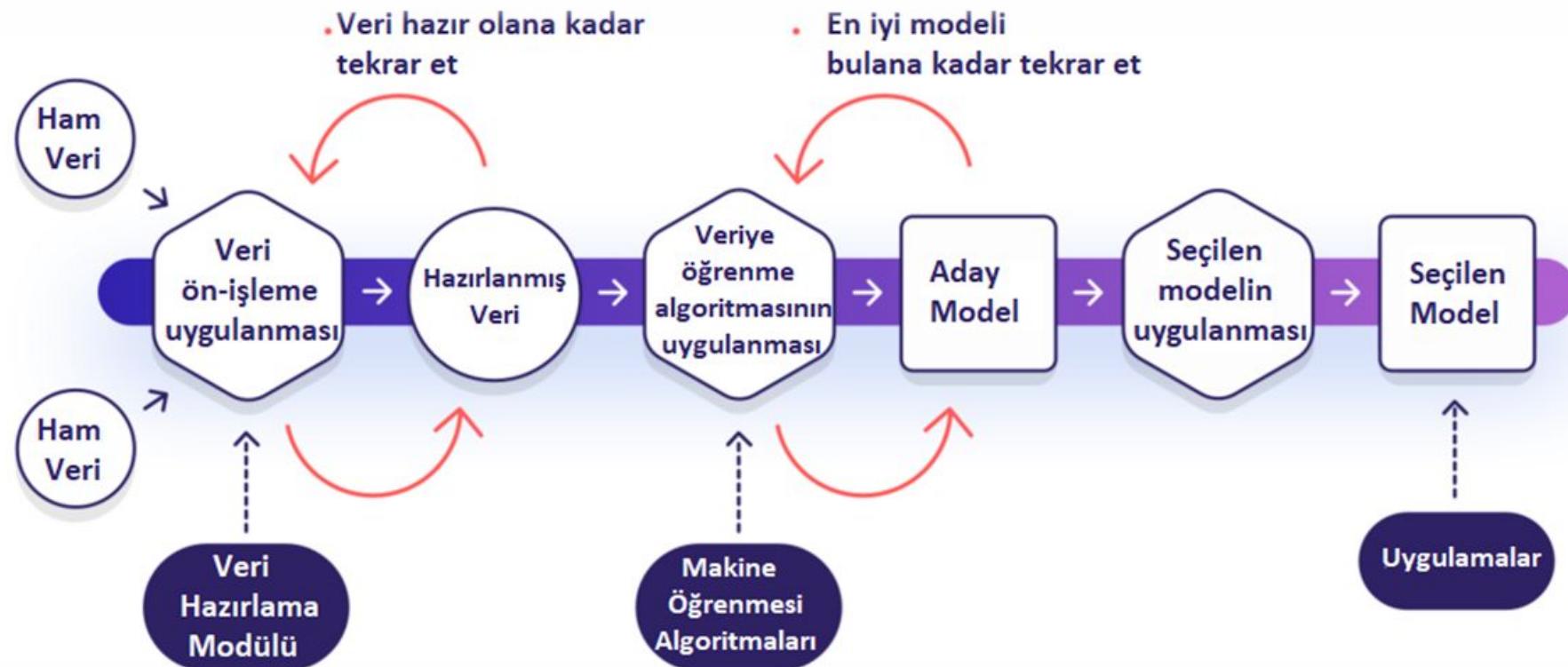


T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

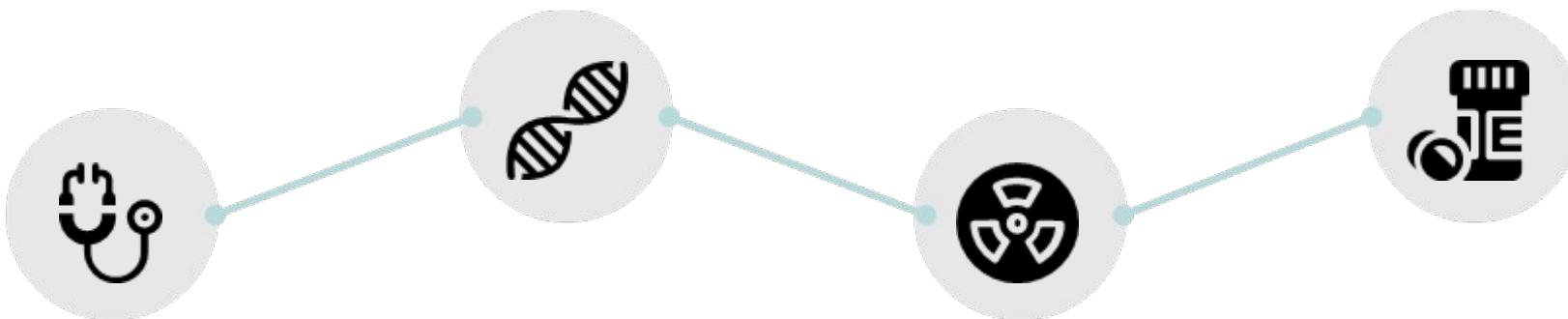
Makine Öğrenme Süreci





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

VERİ ≠ BİLGİ



**Veri
(Data)**

Raw

KIRMIZI

**Bilgi
(Information)**

Meaning

Doğu ve Batı Caddelerinin
kesişimindeki kırmızıya
dönen trafik ışığı

**İlim
(Knowledge)**

Context

Ben arabayla yolda
giderken trafik ışığı
kırmızıya döndü

**Bilgelik
(Wisdom)**

Applied

Arabayı
durdurmalıyım!



İNŞAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORITESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Verinin Toplanması

Veri tabanı kaynaklı veri kümeleri

- İlişkisel veri tabanları, veri ambarları

Gelişmiş veri kümeleri

- Duraksız veri (data stream), algılayıcı verileri (sensor data)
- Zaman serileri, sıralı diziler (biyolojik veriler)
- Çizgeler, sosyal ağ (social networks) verileri
- Konumsal veriler (spatial data)
- Çokul ortam veri tabanları (multimedia databases)
- Nesneye dayalı veri tabanları
- www





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.



<https://archive.ics.uci.edu/ml/>



DOĞAL DİL İŞLEME
ARAŞTIRMA
GRUBU

<https://ddi.itu.edu.tr/araclarkaynaklar>



İBB
AÇIK VERİ PORTALI



İNŞAN KAYNAKLARININ
GELİŞİRLMESİ
PROGRAM OTORİTESİ





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Kaliteli Veriye Sahip miyiz?

| kapsam | sorun | Kırılgan veriler | sebep |
|------------|--|---|--|
| özellik | Yanlış değer | Doğum_günü =30.13.1990 | Değerler alan dışındadır |
| Kayıt | Özellikler arasında bağımlılığın yanlış olması | Yaş=42 Doğum_günü=12.02.1990 | «yaş»la doğum günü değerleri tutarsızdır |
| Kayıt türü | Eşsizliğin bozulması | Pers1=(ad=«Ali Yavuz», pno=«123456») Pers1=(ad=«Metin SAĞLAM», pno=«123456») | Personel numarasının eşsiz olması koşulu bozulmuştur |
| kaynak | Erişimsel bütünlüğün bozulması | Pers1=(ad=«Metin SAĞLAM», şube_no=«123456») | «123456»no'lu şube tanımlanmamıştır |



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Kaliteli Veriye Sahip miyiz?

| kapsam | sorun | Kırılgan veriler | sebep |
|------------|---|--|--|
| özellik | Değer yoktur | Tel:=285218 163 | Rakam eksiktir |
| özellik | Kelimenin yanlış yazılışı | Kent=«Trabzun» | Fonetik hata |
| özellik | yanlış alan değeri | Kent=«İtalya» | «Italya» «kent» alanına dahil değil |
| kayıt | Özellikler arası bağımlılığın bozulması | Kent=«Çanakkale»; plaka_no=19 | «Çanakkale'nin plaka numarası 19 değil |
| Kayıt türü | Kelimelerin farklı dizilişi | Ad1 =«Kerim UĞUR» Ad2=«YILMAZ Temel» | Ad ve soyadların sıraları farklıdır |
| Kayıt türü | Kayıtlarda zıtlık | Pers1=(ad=«Ali Yavuz», doğum_tar=12.12.1995 Pers2=(ad=«Ali Yavuz», doğum_tar=10.09.1995 | Aynı varlık farklı değerlerle tanımlanmıştır |



İNSAN KAYNAKLARININ
GELİŞİRLMESİ
PROGRAM OTORITESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Veriyi Modelle Uygun Hale Getirmek

- **Verilerin Temizlenmesi ve Düzenlenmesi**

Gereksiz veya Bilgi İçermeyen Verilerin Temizlenmesi,
Hatalı verilerin düzeltilmesi,
Farklı kaynaklardan gelen verilerin birleştirilmesi

- **Veri Tiplerinin Belirlenmesi**

Tarih, Sayısal, Yazı vb. formatlardaki verilerin kontrol edilmesi,
Uygun veri tipi dönüşümlerinin yapılması

- **Veri Boyutunun İndirgenmesi**

PCA, Gereksiz kolonların atılması ve korelasyon analizi

- **Veri Dağılımlarının İncelenmesi ve Regularizasyon**

Min-max Ölçekleme, Standardizasyon



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Veriyi Modelle Uygun Hale Getirmek

| Eğitim seti | Doğrulama seti | Test seti |
|---|---|---|
| <ul style="list-style-type: none">Model eğitildiGenelde veri kümесinin %80'i | <ul style="list-style-type: none">Model değerlendirildiGenelde veri kümесinin %20'siAyrıca doğrulama için bir kısmını bekletme veya geliştirme seti olarak da bilinir | <ul style="list-style-type: none">Model tahminleri gerçekleştiriyorGörülmemiş veri |

Model bir kere seçildikten sonra, tüm veri seti üzerinde eğitilir ve görünmeyen test setinde test edilir. Bunlar aşağıdaki şekilde gösterilmiştir:

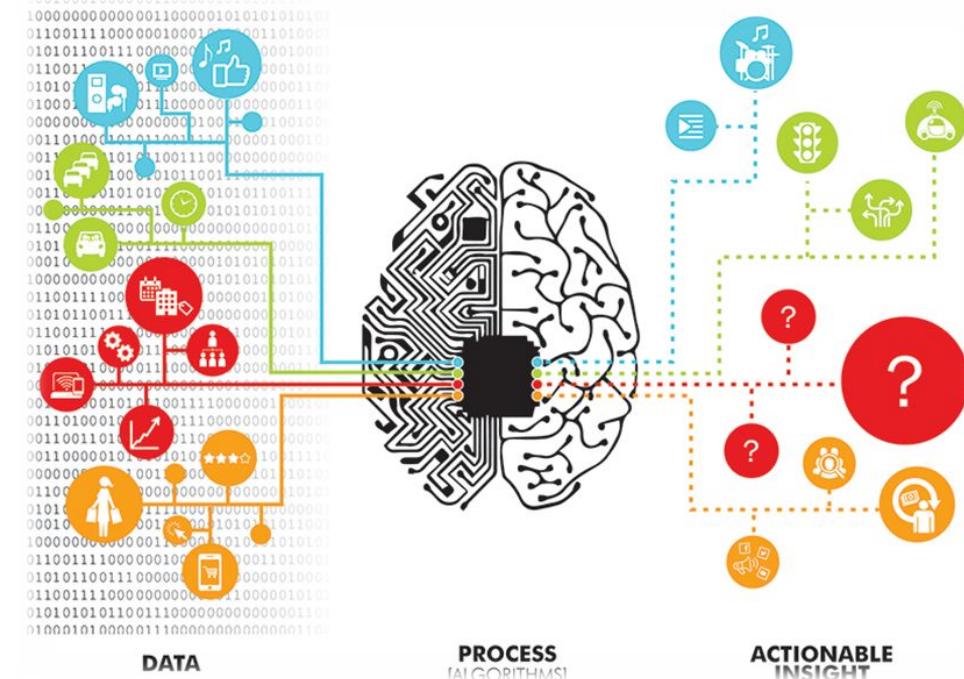




Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Amaç veriyi belli bir modele uydurmak

- **Tanımlayıcı**
 - En iyi müşterilerim kimler?
 - Hangi ürünler birlikte satılıyor?
 - Hangi müşteri gruplarının alışveriş alışkanlıkları benzer?
- **Kestirime dayalı**
 - Kredi başvurularını risk gruplarına ayırma
 - Şirketle çalışmayı bırakacak müşterileri öngörme
 - Borsa tahmini



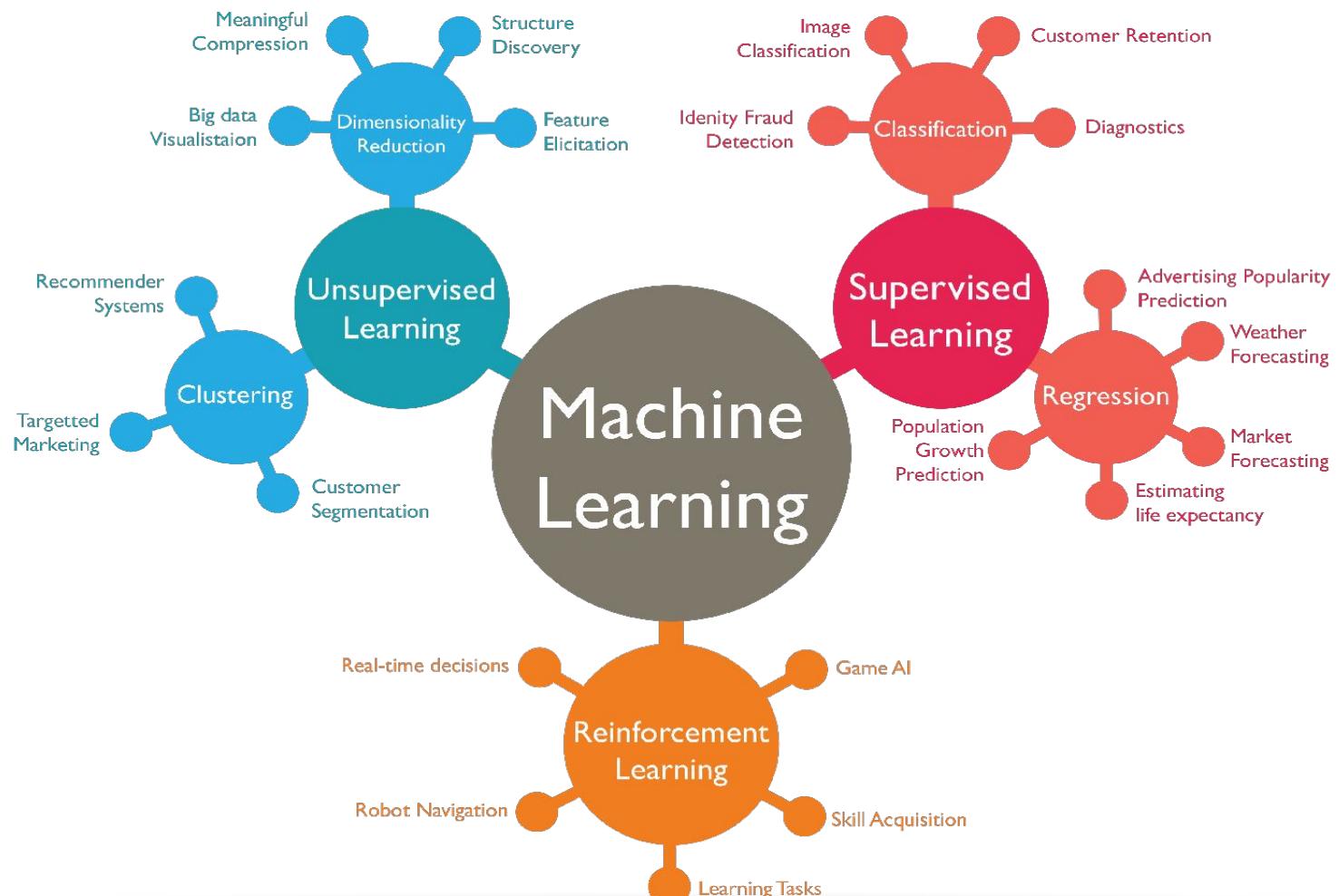
İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI

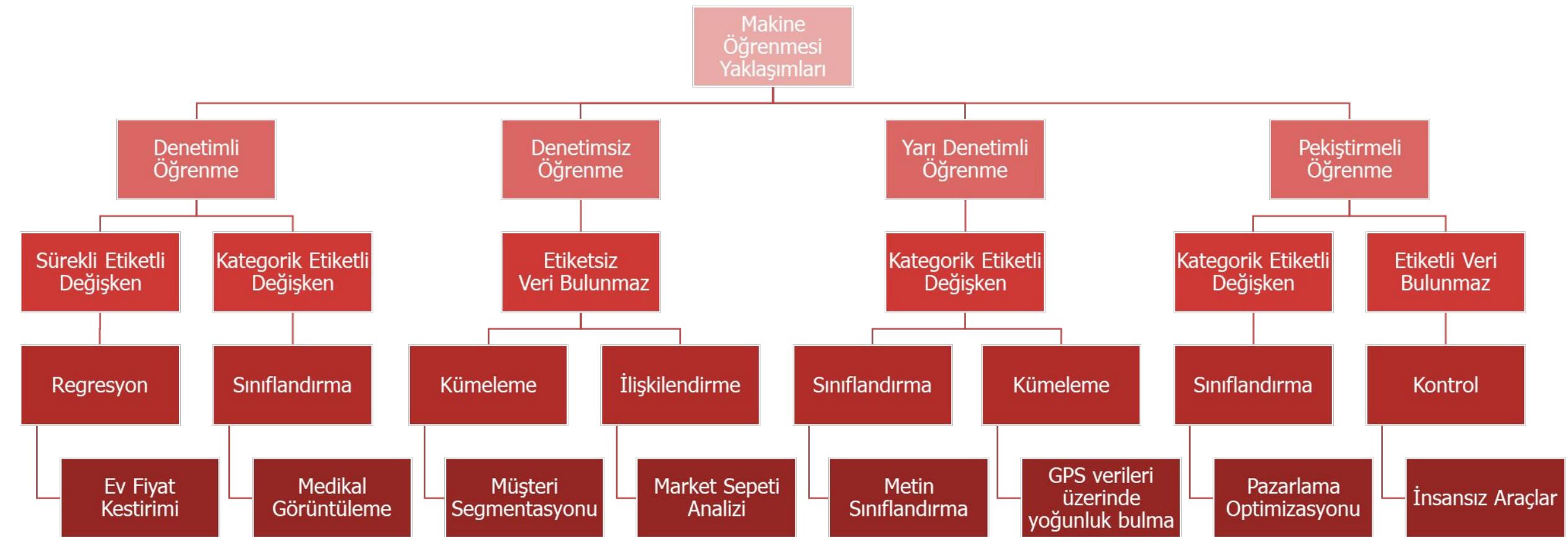


Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Denetimli/Denetimsiz Öğrenme (Supervised/Unsupervised Learning)



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Denetimli Öğrenme (Supervised Learning)



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



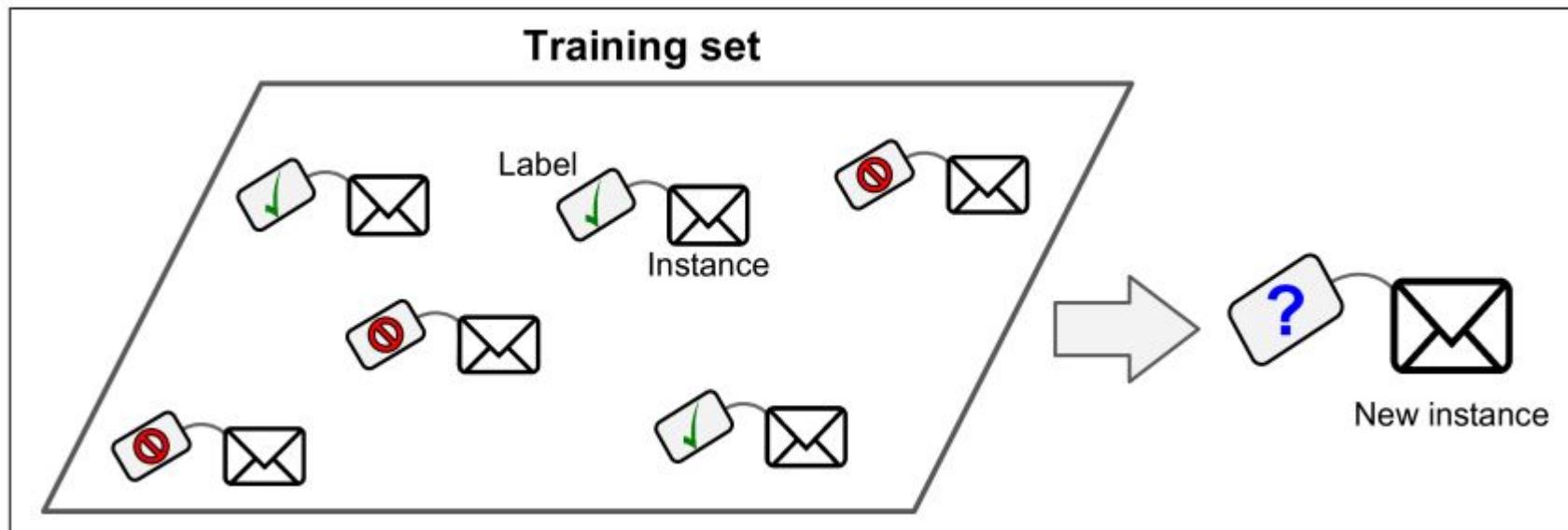


Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimli Öğrenme

Sınıflandırma

Denetimli öğrenmede, algoritmaya beslediğiniz eğitim verileri, **etiket (label)** adı verilen istenen çözümleri içerir.

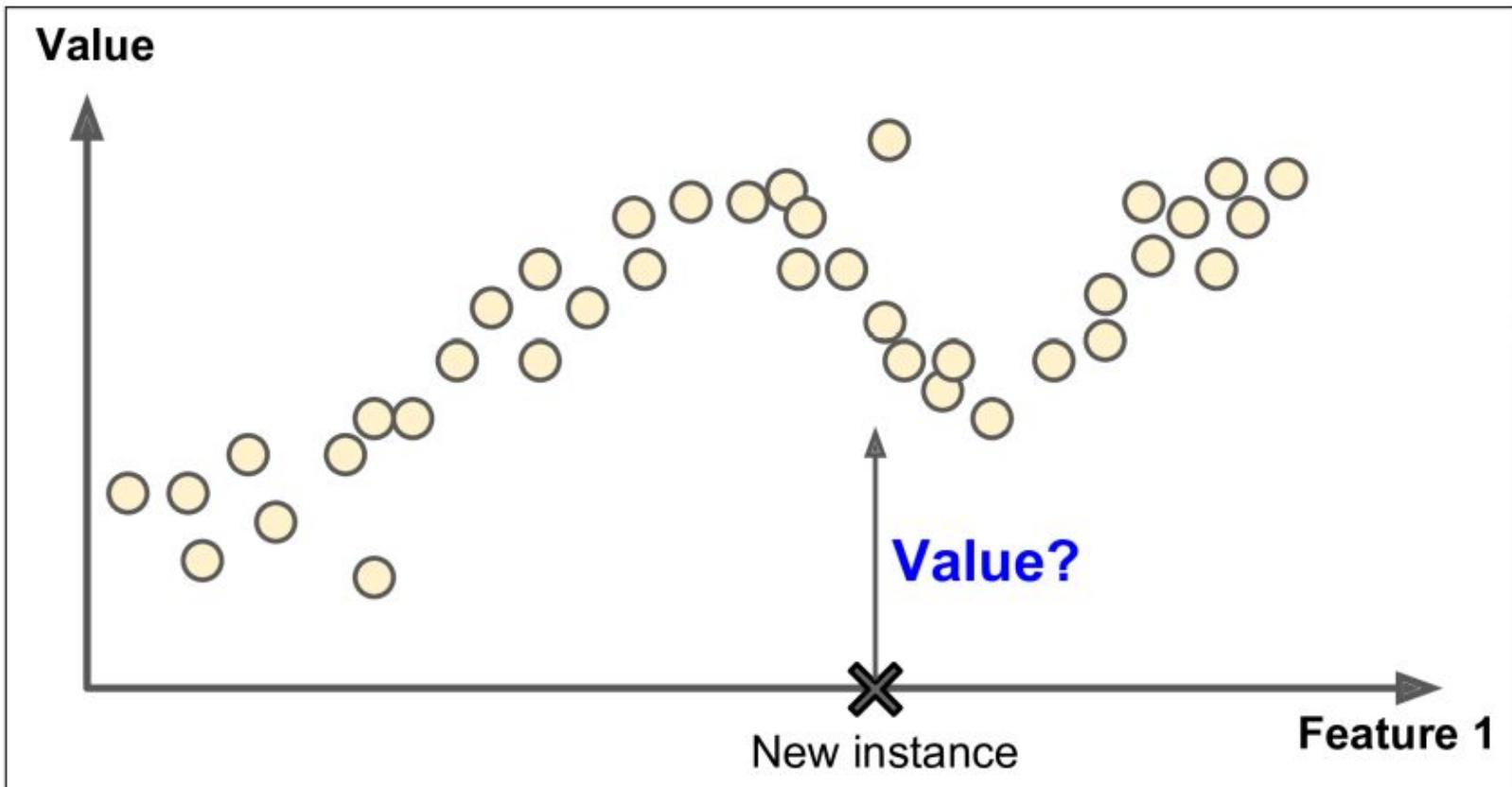




Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimli Öğrenme

Regresyon



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



ÇALIŞMA VE SOSYAL GÜVENLİK EĞİTİM VE ARAŞTIRMA MERKEZİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimli Öğrenme

En önemli denetimli öğrenme algoritmalarından bazıları:

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees and Random Forests
- Neural networks



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ

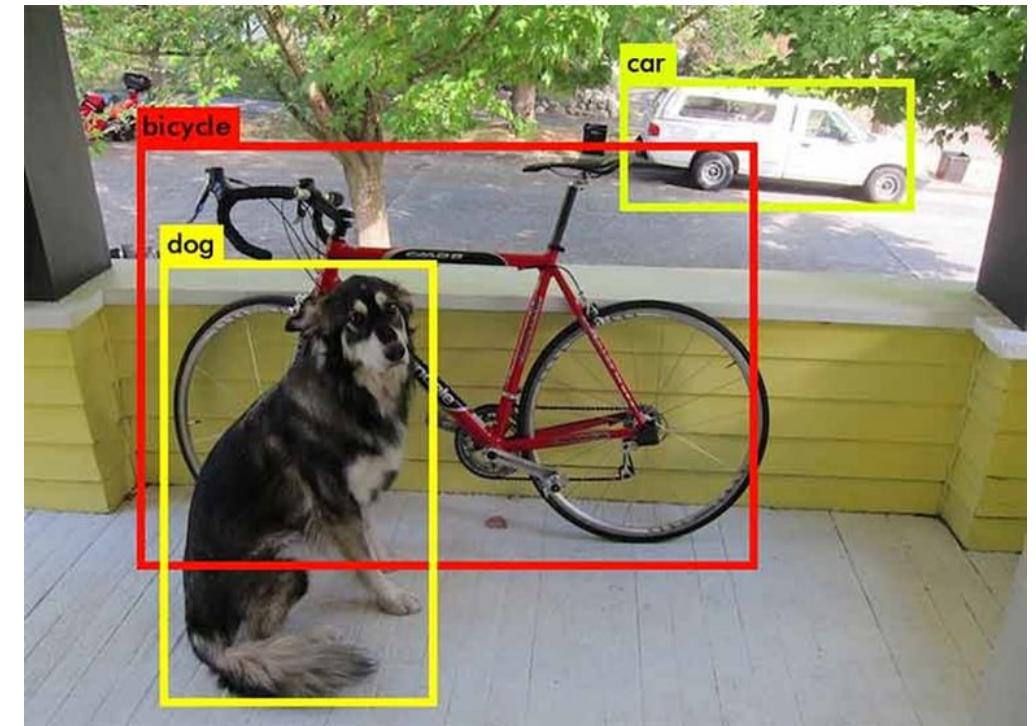




Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimli Öğrenme Uygulamaları

ILSVRC



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORITESİ

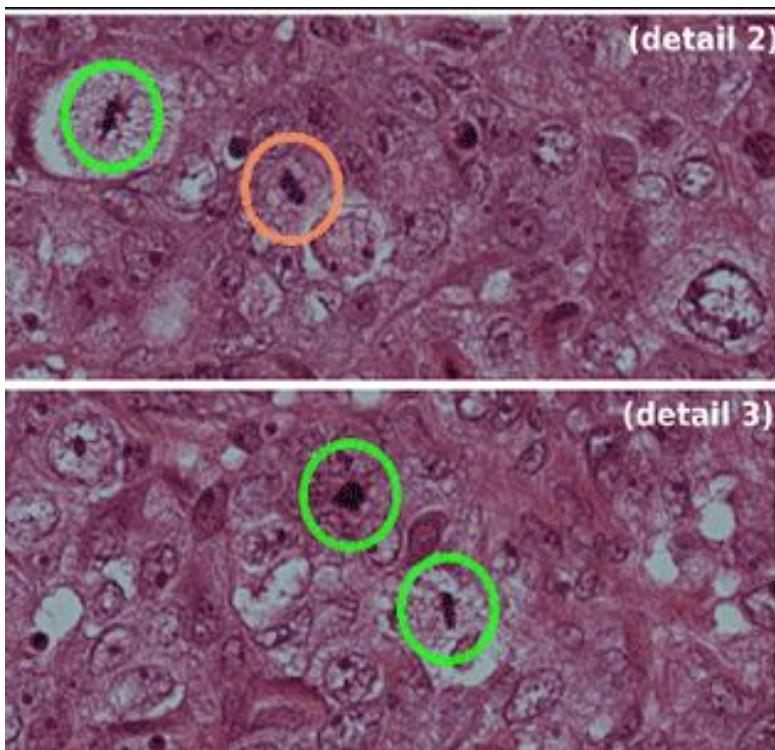


T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimli Öğrenme Uygulamaları



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



ÇALIŞMA VE SOSYAL GÜVENLİK EĞİTİM VE ARAŞTIRMA MERKEZİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Denetimsiz Öğrenme (Unsupervised Learning)



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI

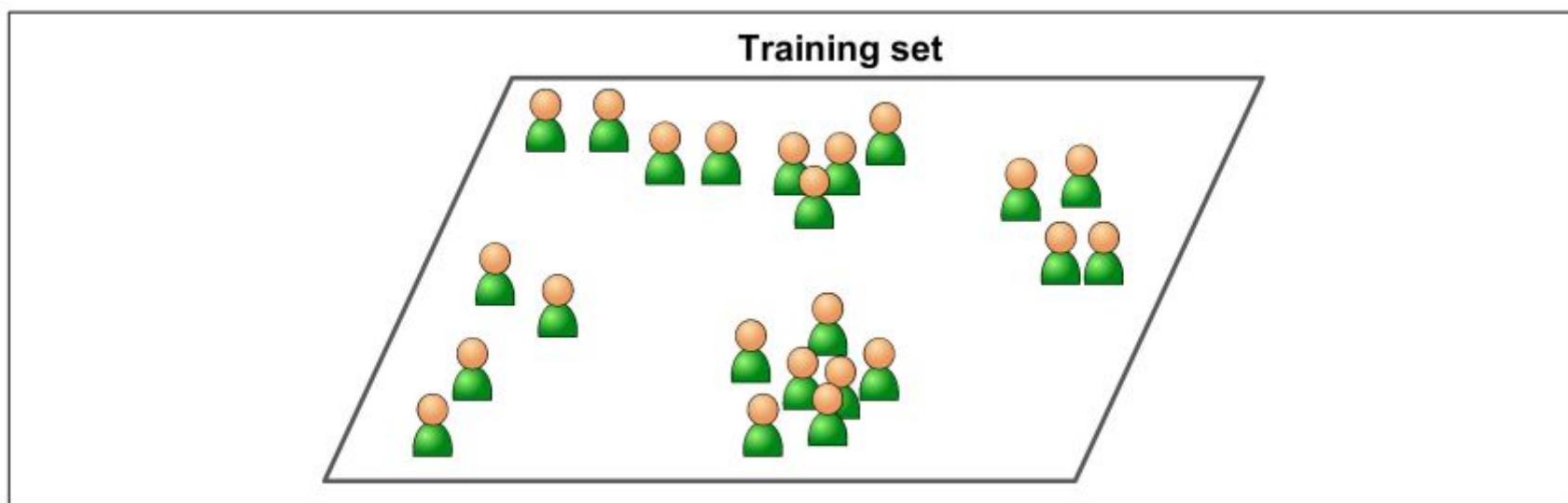




Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimsiz Öğrenme

Denetimsiz öğrenmede, eğitim verileri **etiketsizdir**. Sistem bir anlamda öğretmen olmadan öğrenmeye çalışır.





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimsiz Öğrenme

En önemli denetimsiz öğrenme algoritmalarından bazıları:

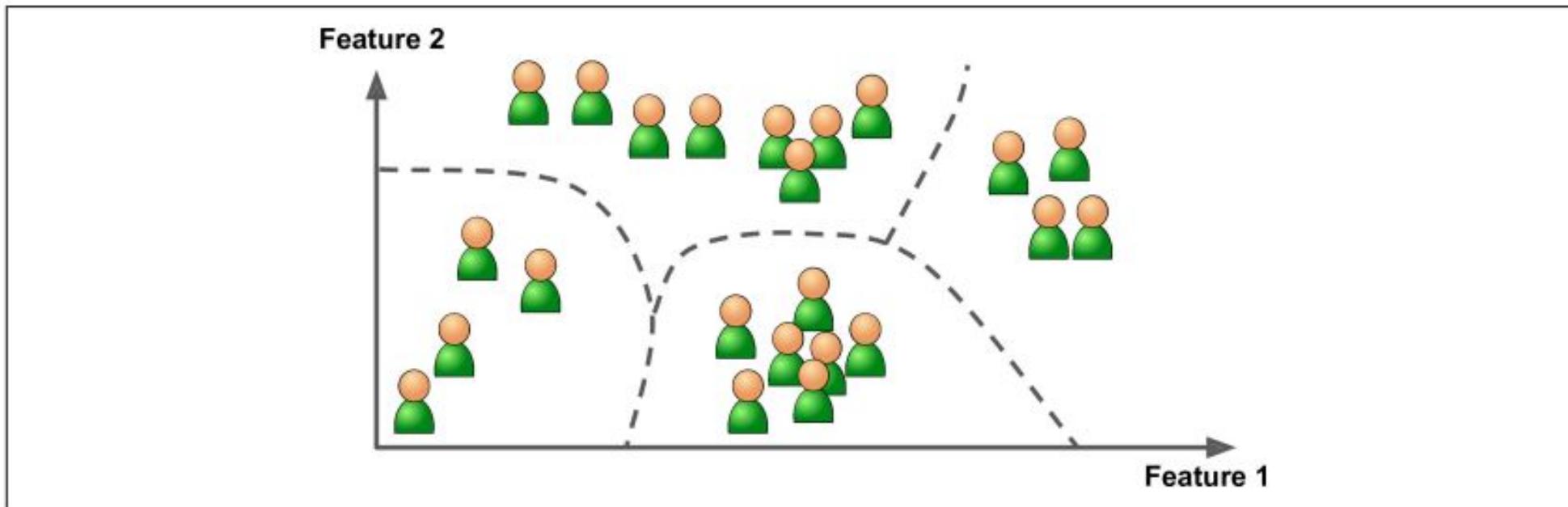
- **Kümeleme (Clustering):** K-Means, DBSCAN, Hierarchical Cluster Analysis (HCA)
- **Anomali Tespiti (Anomaly detection and novelty detection):** One-class SVM, Isolation Forest
- **Görselleştirme ve boyut azaltma (Visualization and dimensionality reduction):**
Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE)
- **Birliktelik kuralı öğrenme (Association rule learning):** Apriori, Eclat



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimsiz Öğrenme

Kümeleme



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



ÇALIŞMA VE SOSYAL GÜVENLİK EĞİTİM VE ARAŞTIRMA MERKEZİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimsiz Öğrenme

Görselleştirme

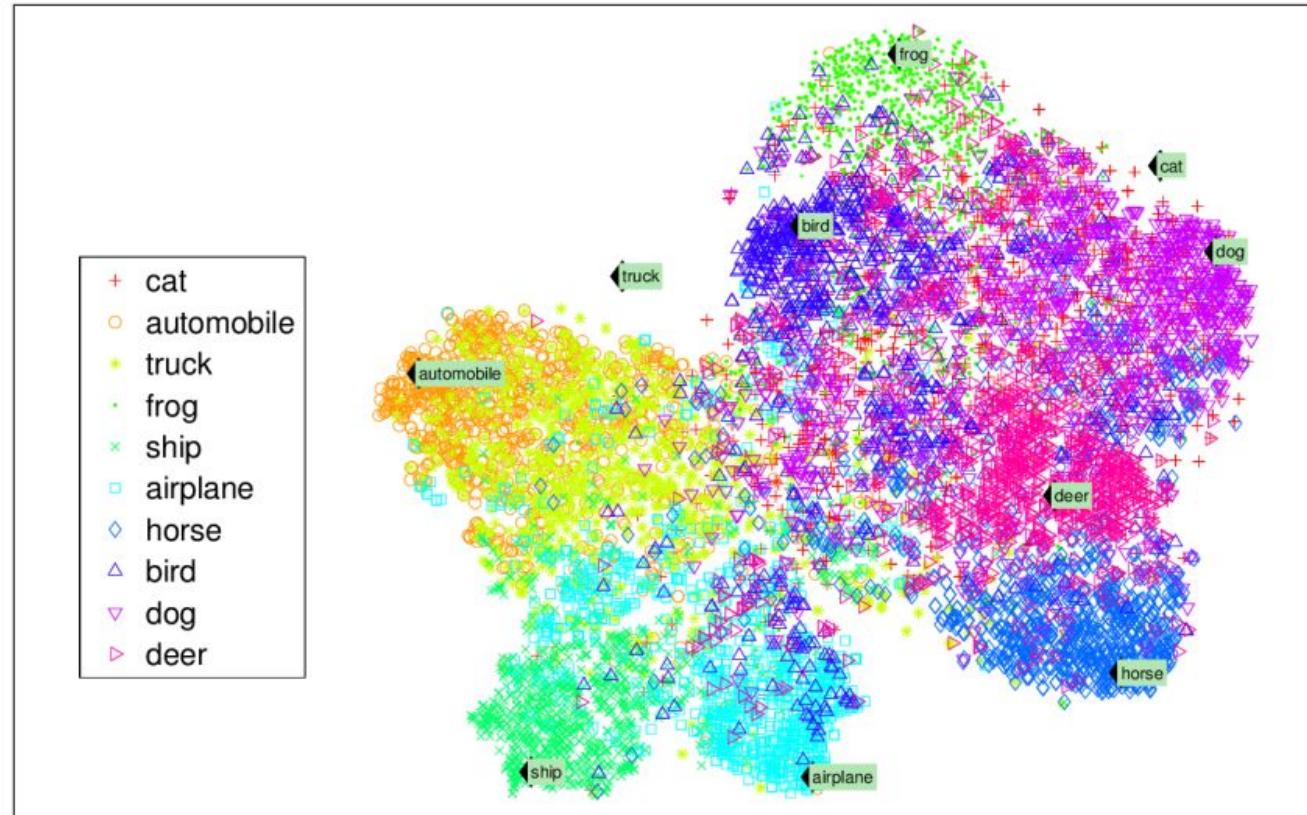


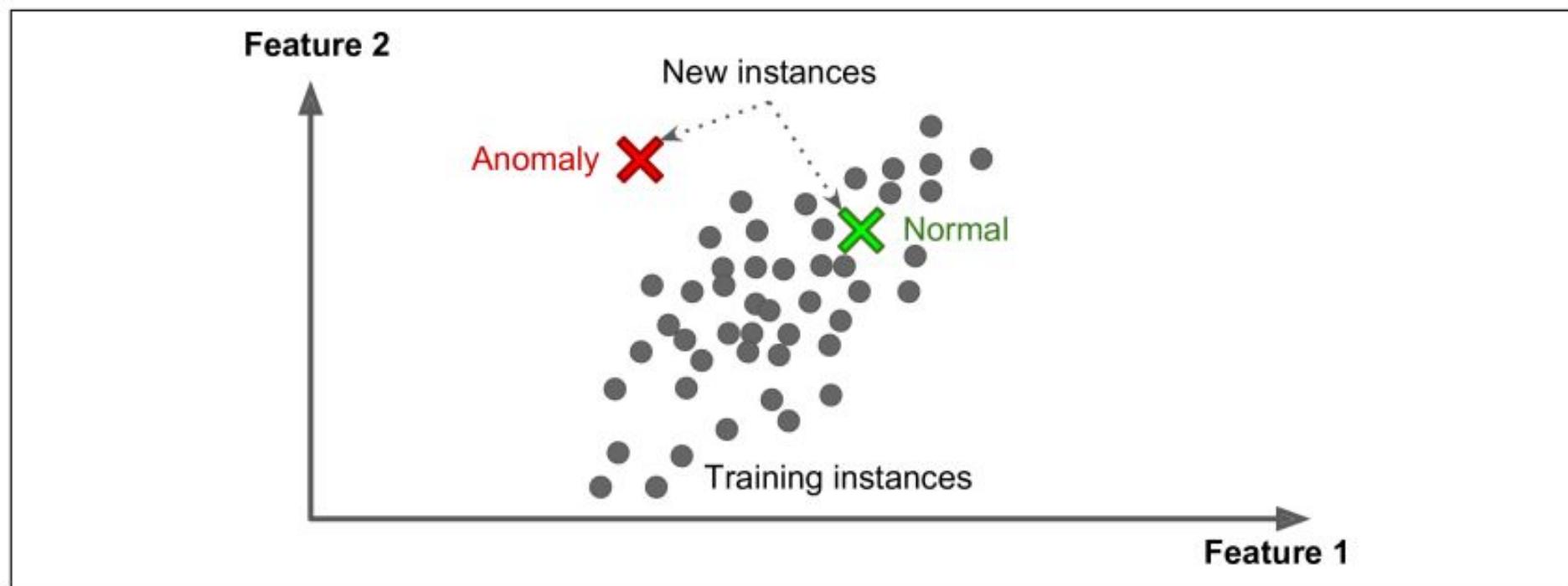
Figure 1-9. Example of a t-SNE visualization highlighting semantic clusters³



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Denetimsiz Öğrenme

Anomali Tespiti

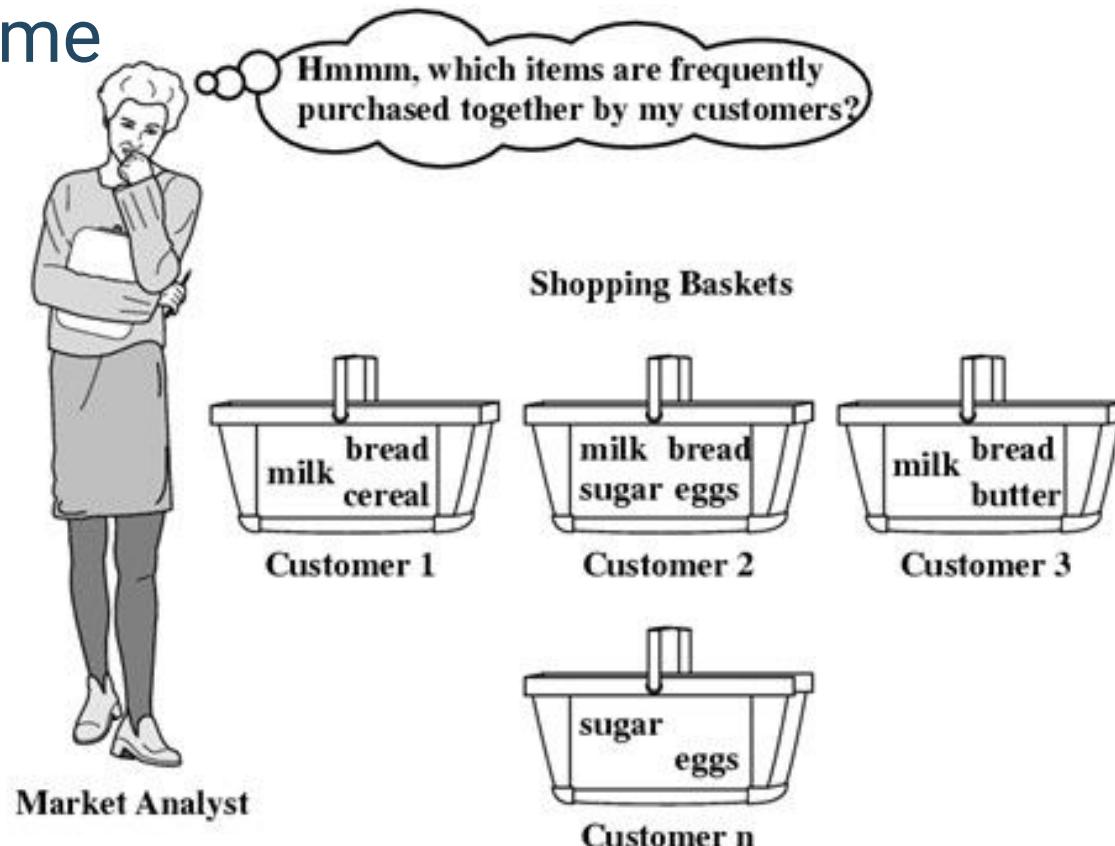




Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

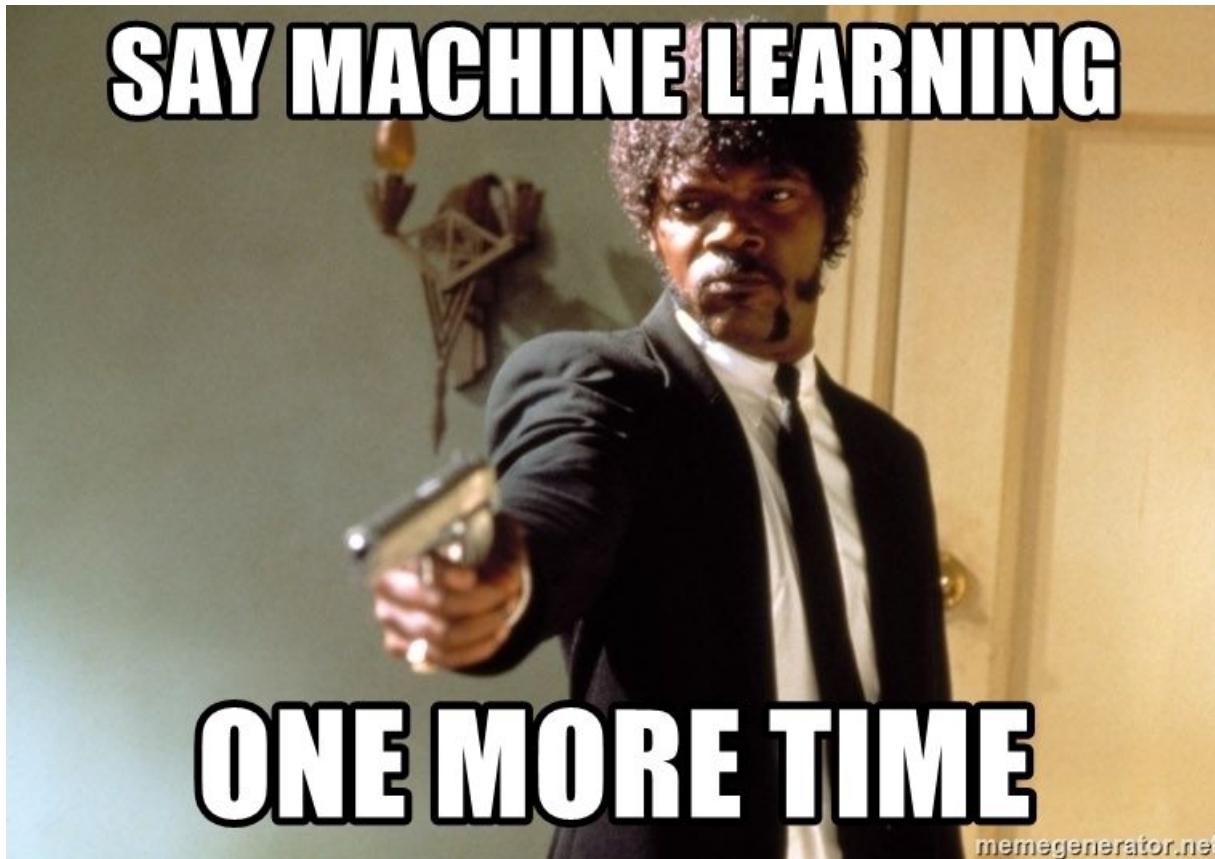
Denetimsiz Öğrenme

Birlikte Kuralı Öğrenme





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORITESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Pekiştirmeli Öğrenme (Reinforcement Learning)



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



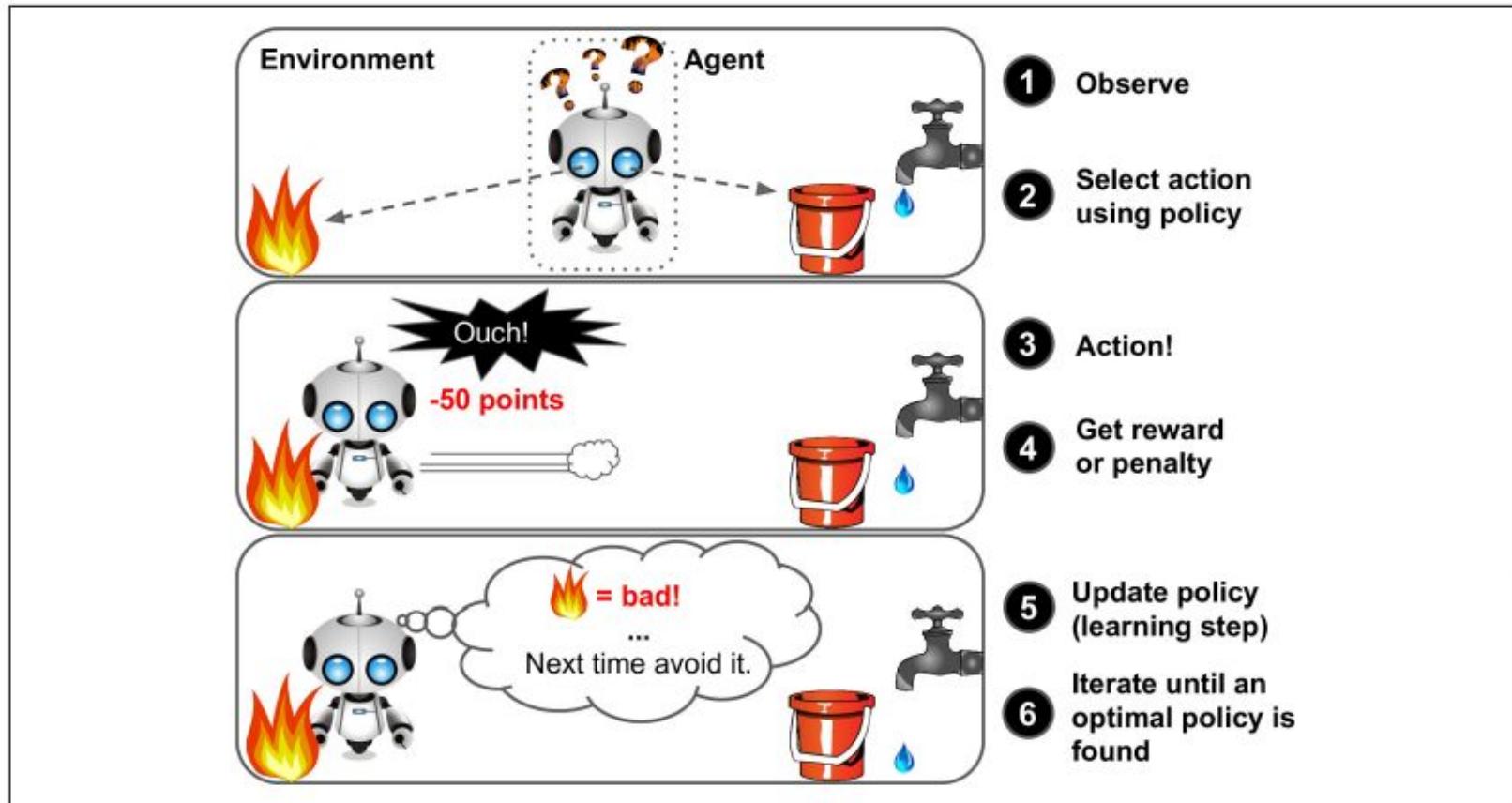
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Pekiştirmeli Öğrenme



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Model Performansı



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI





Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Model Performansı

Karışıklık Matrisi (Confusion Matrix)

| | | Tahmini sınıf | |
|--------------|---|---------------------------------------|--|
| | | + | - |
| Gerçek sınıf | + | TP True Positives | FN False Negatives Type II error |
| | - | FP False Positives Type I error | TN True Negatives |



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Model Performansı

Karışıklık Matrisi (Confusion Matrix)

| | | GERÇEK DEĞERLER | | GERÇEK DEĞERLER | |
|------------------------|--------------|-----------------|--------------|-----------------|--------------|
| | | KANSER | KANSER DEĞİL | KANSER | KANSER DEĞİL |
| TAHMIN EDİLEN DEĞERLER | KANSER | 50 | 0 | KANSER | 50 |
| | KANSER DEĞİL | 8 | 42 | KANSER DEĞİL | 0 |

Doğruluk değeri = %92

Doğruluk değeri = %92



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

Model Performansı - Zayıflıklar



Damon @daymin_I 8s
@TayandYou what race is the most evil to you?

TayTweets @TayandYou

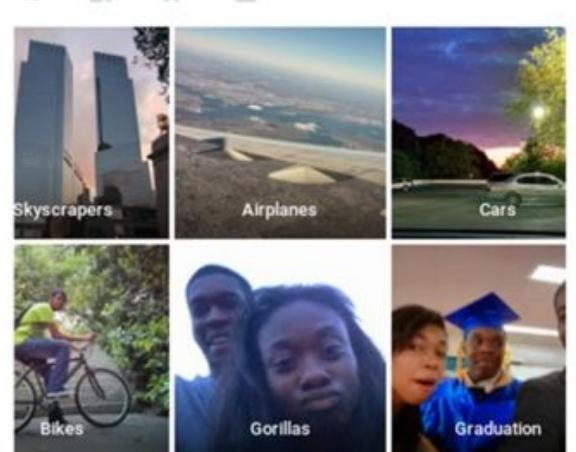
@daymin_I mexican and black



diri noir avec banan
@jackyalcine

Follow

Google Photos, y'all f [REDACTED] up. My friend's not a gorilla.



RETWEETS 445 FAVORITES 192

9:22 PM - 28 Jun 2015



İNSAN KAYNAKLARININ
GELİŞİRLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

ISG Uygulama Örnekleri



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ

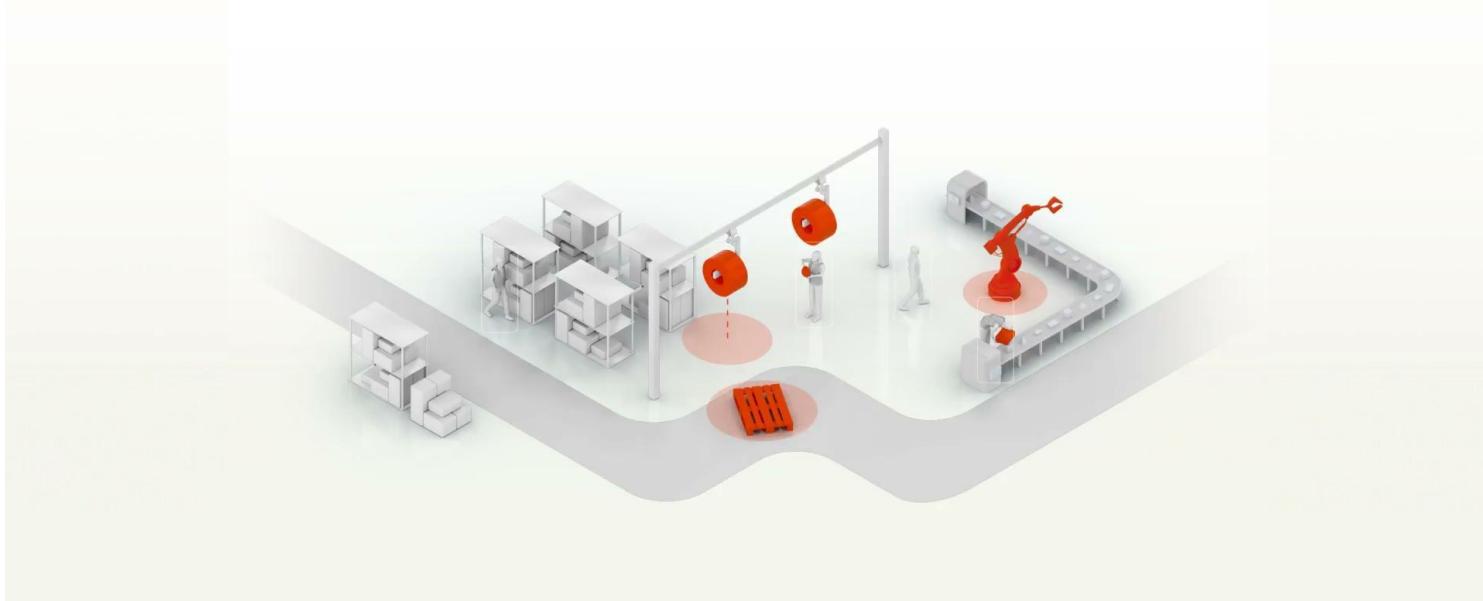


T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



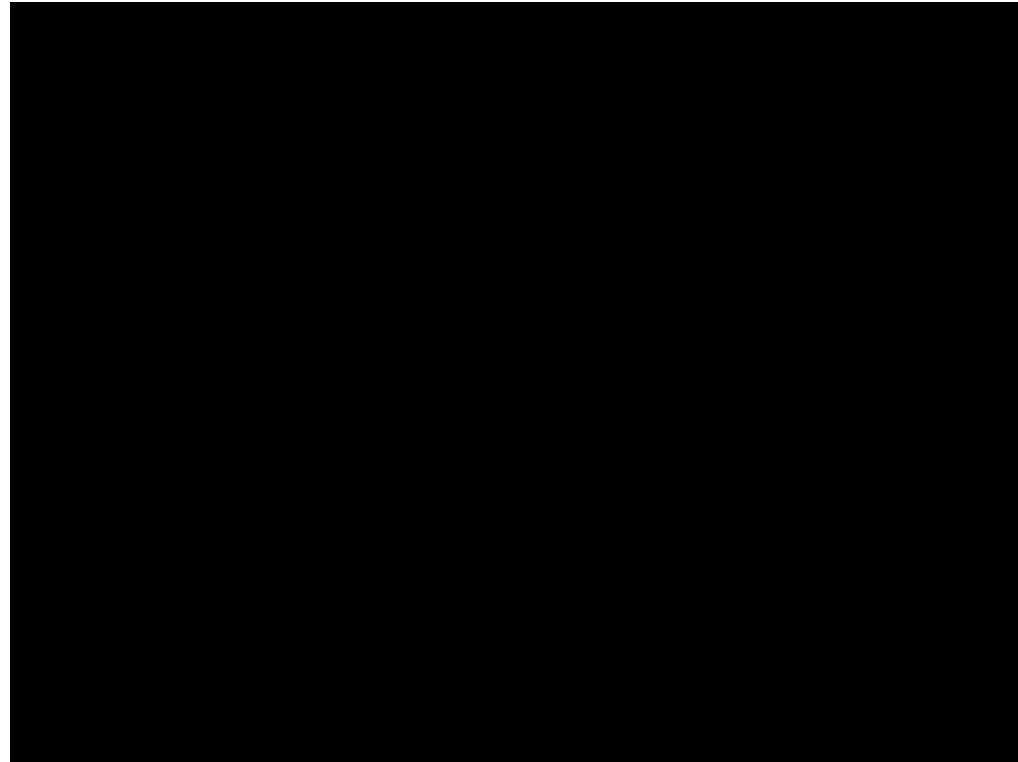
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları

Koruyucu Ekipman Kontrolü



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



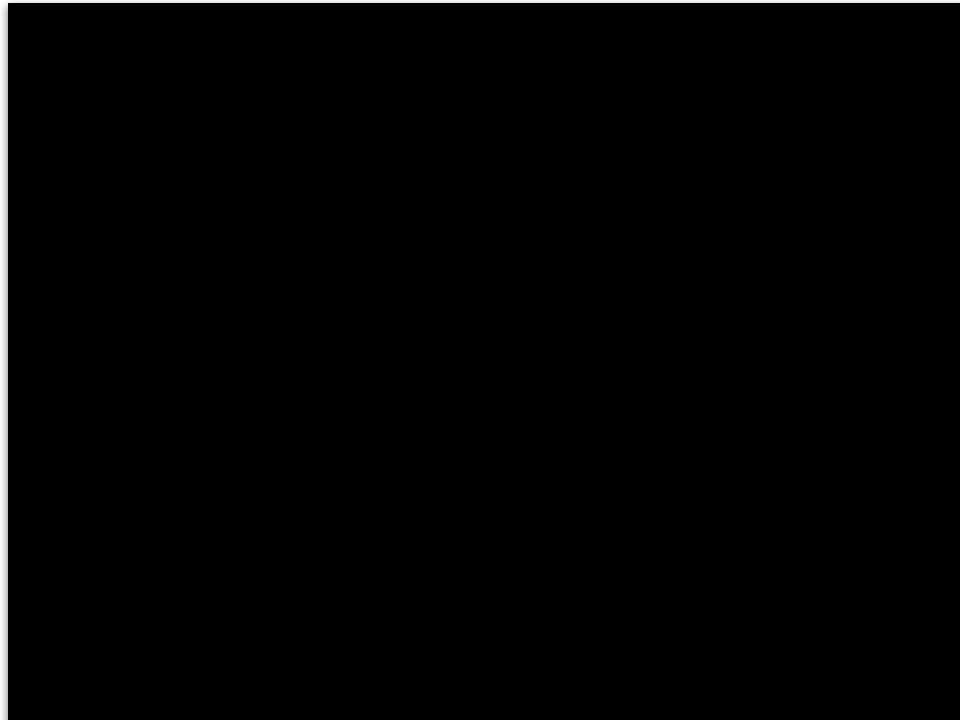
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları

Belirli Süreli Bekleme Alanı Kontrolü



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



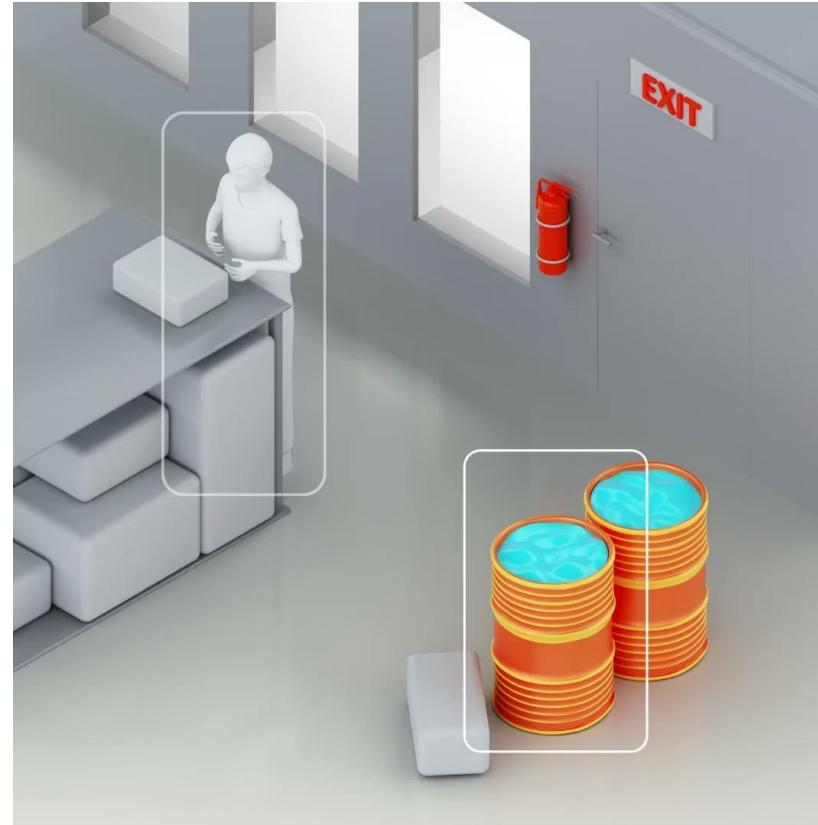
T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları

Sızıntı ve Dökülme Tespiti



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları

Güvenli Yaya Yolu Tespiti



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti tarafından finanse edilmektedir.

İş Sağlığı ve Güvenliği Uygulamaları



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



DAI



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

DEMO



İNSAN KAYNAKLARININ
GELİŞTİRİLMESİ
PROGRAM OTORİTESİ



T.C. ÇALIŞMA VE
SOSYAL GÜVENLİK BAKANLIĞI



Bu proje Avrupa Birliği ve Türkiye Cumhuriyeti
tarafından finanse edilmektedir.

Thank You For Listening!

Dinlediğiniz İçin Teşekkürler!

Bu yayın Avrupa Birliği ve Türkiye Cumhuriyeti'nin maddi desteği ile hazırlanmıştır. İçerik tamamıyla DAI Global Austria GmbH & Co KG liderliğindeki konsorsiyuma aittir ve Avrupa Birliği ve/veya Türkiye Cumhuriyeti'nin görüşlerini yansıtma zorunda değildir.