BATCH : **B 150** Data Science
LESSON : **Machine Learning**
DATE : 26.06.2023
SUBJECT : **Introduction**

techproeducation
techproeducation
techproeducation
techproeducation
techproedu

techproeducation.com info@techproeducation.com +1 (917) 768-7466

MACHINE LEARNING INTRO

Makine Öğrenmesi Derslerine Giriş

Peardeck Time !

ML konularında daha önce çalıştım.. Biraz bilgim var..

Overall Table of Contents

General Content

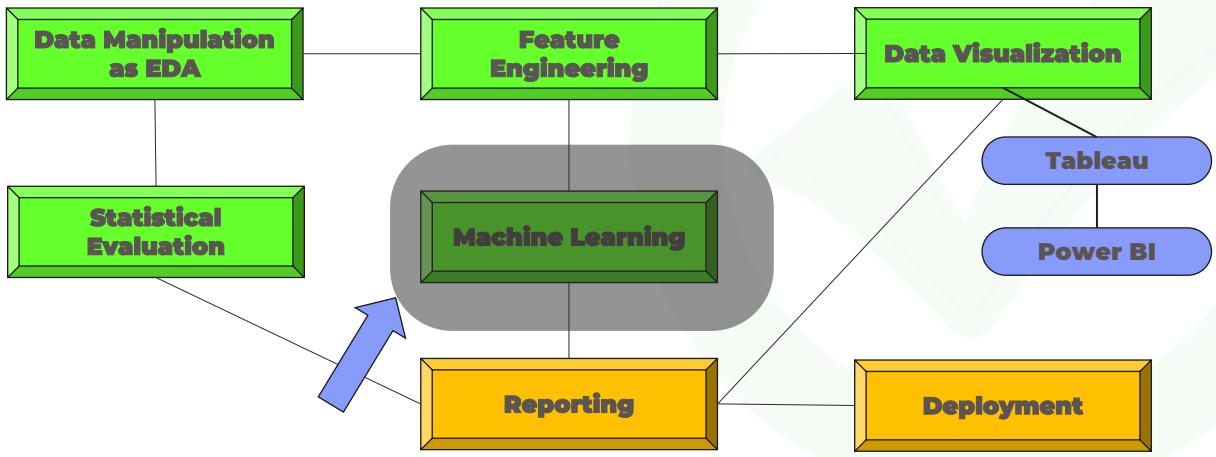
- Data and Machine Learning Introduction
- Supervised Algorithm
- Unsupervised Algorithm
- Capstone Projects

LMS Pre-Class'ta bu dersle ilgili kısma çalıştım

Peardeck Time

ML deyince alakalı aklıma ilk gelenler..

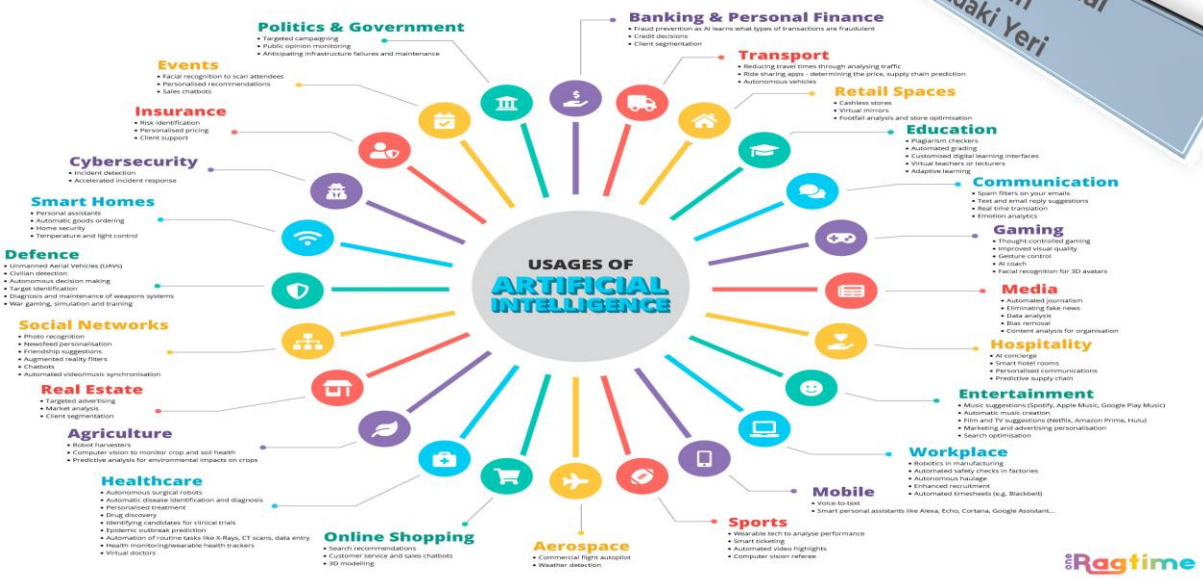
Internship Project Steps



Summary

Yapay Zeka (Artificial Intelligence)'nın Hayatımızdaki Yeri

Machine Learning

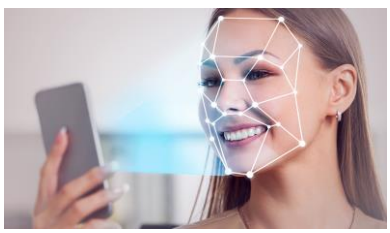


Günlük Yaşamdan Örnekler

Görüntü Tanıma

Siber Güvenlik

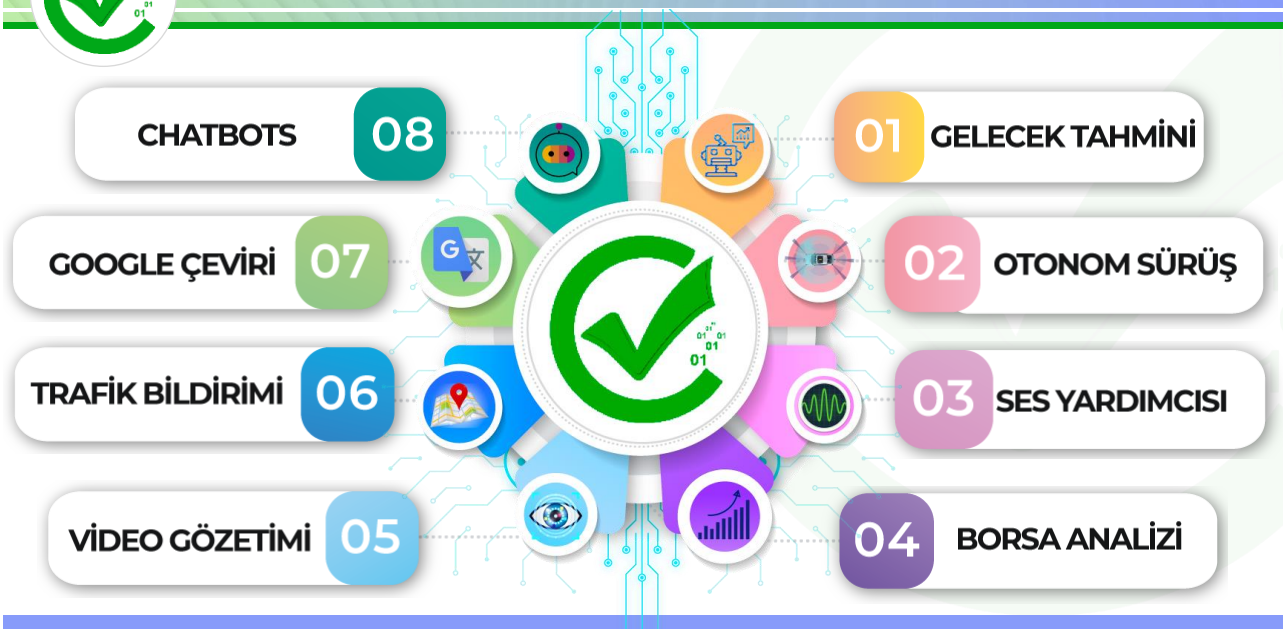
Duygu Analizi



Günlük Yaşamdan Örnekler

Ürün Tavsiyesi | Sosyal Medya | Sağlık

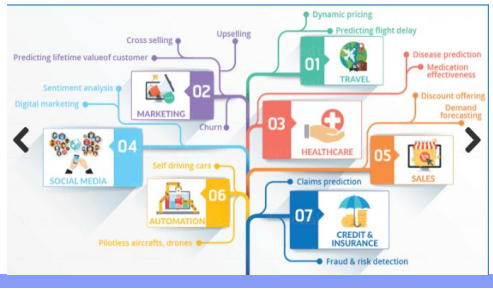
Günlük Yaşamdan Örnekler



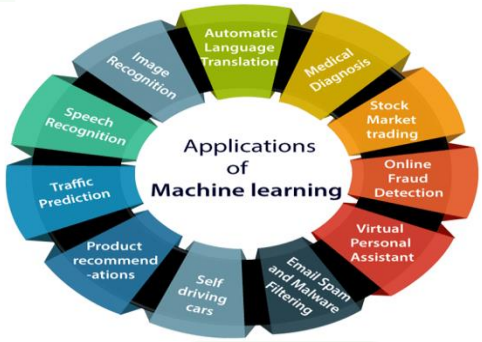
Summary

Önsöz

- ✓ ML dünyada genel durumu
- ✓ Hangi sektörler...
- ✓ Neden Python



Machine Learning applications across industries



ML Area Samples

Machine Learning

Recommendation Engines

Credit Scoring

Customer Churn

Purchasing Trends

New Pricing Models

Predictive Inventory Plan

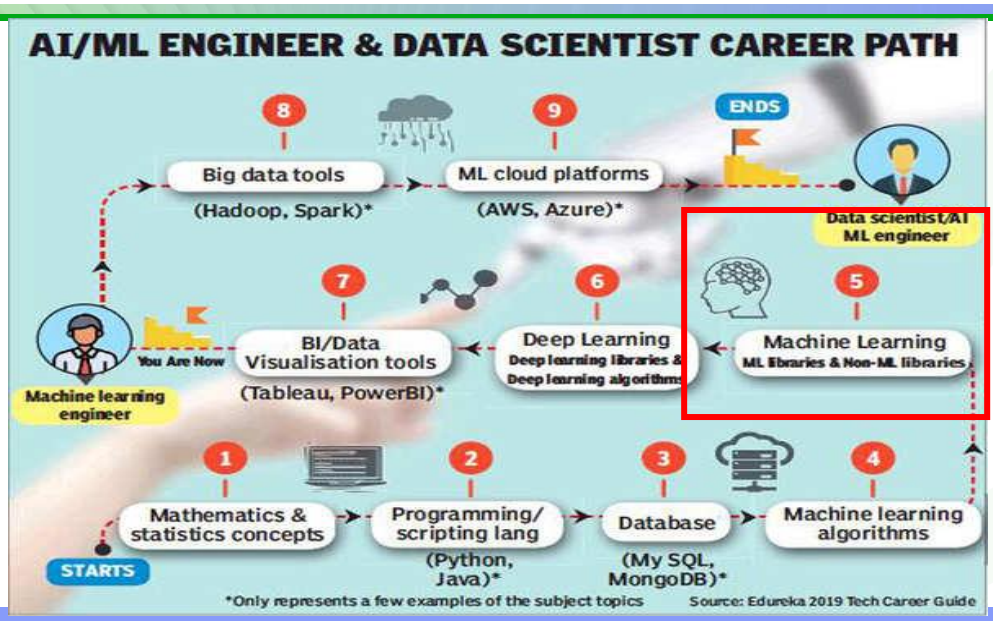
Email Spam Filtering

Material and Stock Estimates

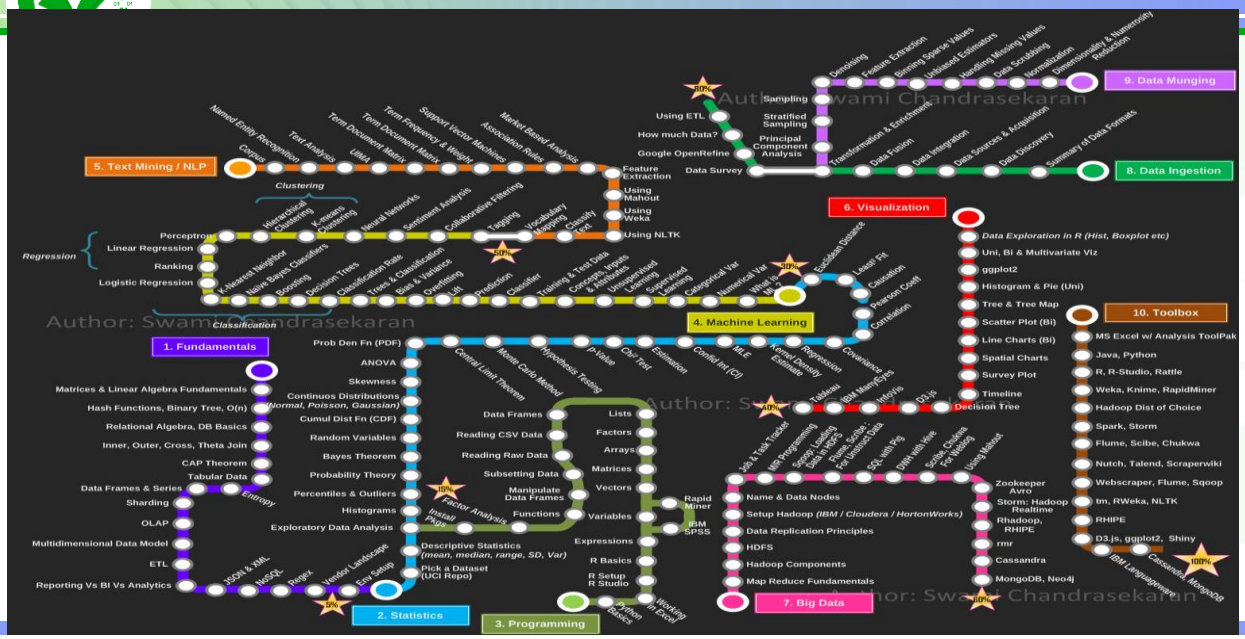
Pattern and Image Recognition

Where are we ? Run Navigation !!!

Machine Learning



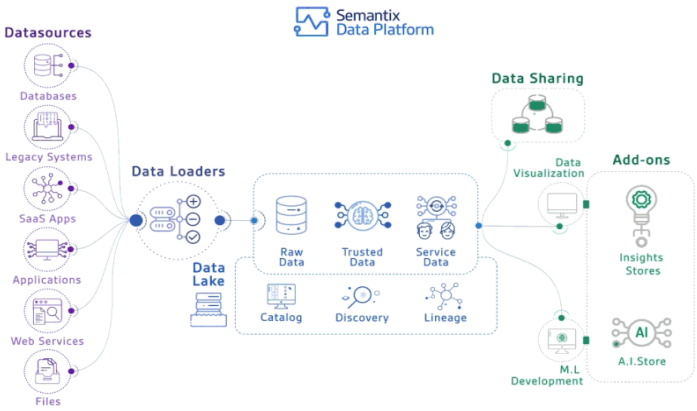
More Focus... Run Navigation..



ML Temel Kavramlar

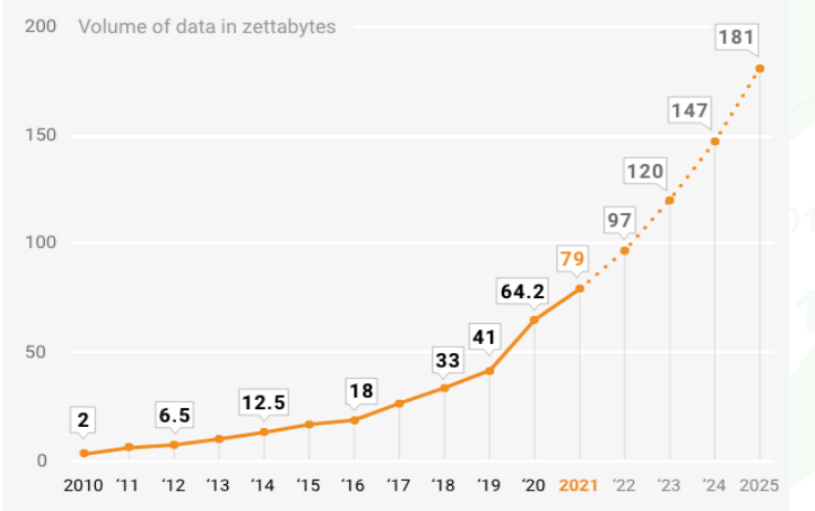
Data (Veri) nedir ?

- Veri yüzyılımızın petrolüdür
- Big Data Kavramı
- Data Science Kavramı



DATA HACMİ (Yıllık Seyir)

- 2021 de **79** *zettabytes* data üretildi.
- 2022 de **94** zettabyte data üretildi.
- 2025 e kadar **150 zettabytes** den daha fazla data **analiz edilmesi** gerekecek.
- Bütün datanın yaklaşık %90'ı replicate, %10'u orijinal-yeni data (*statista.com*)
- 2010-20 arası data etkileşimi %5.000 artmış. (*forbes.com*)
- İnternette üretilen günlük datanın %80-90 ı yapılandırılmamış data. (*cio.com*)
- 1 zettabyte = 1 trilyon GB



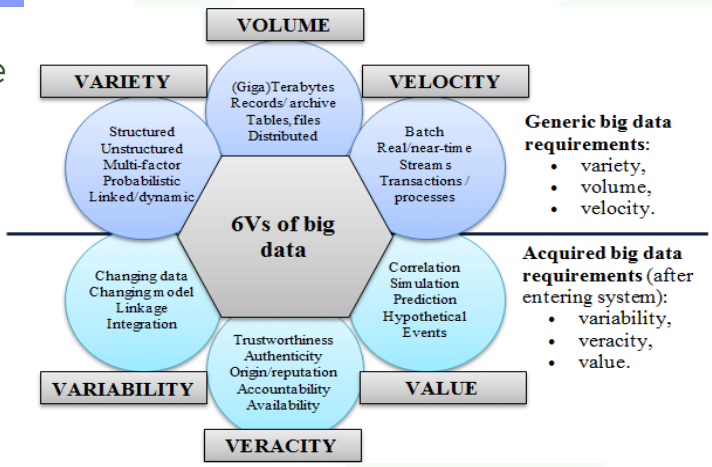
Kaynak: statista.com

ML Temel Kavramlar

6V formülü

Bir veri setini büyük yapan nedir? Cevap 6V formülü ile açıklanmaktadır

- Volume (Hacim)
- Velocity (Hız)
- Variety (Çeşitlilik)
- Veracity (Doğruluk)
- Variability (Değişkenlik)
- Validty (Geçerlilik)

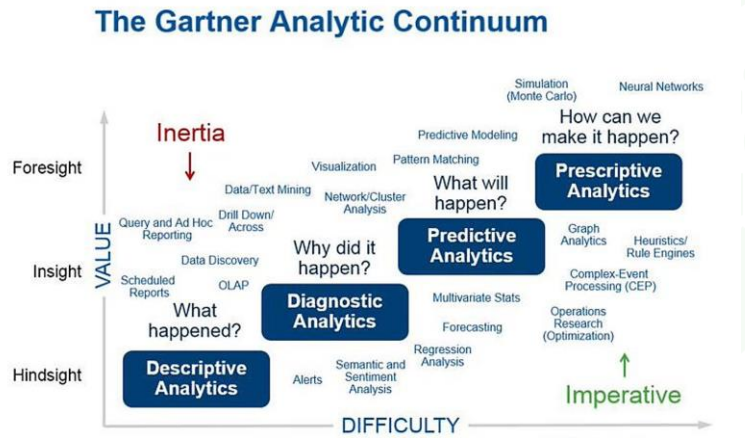


ML Temel Kavramlar

Bilgi Üretme

Datadan değerli bilgi nasıl üretilir ?

- Insight oluşturma
- Data Analitiği türleri

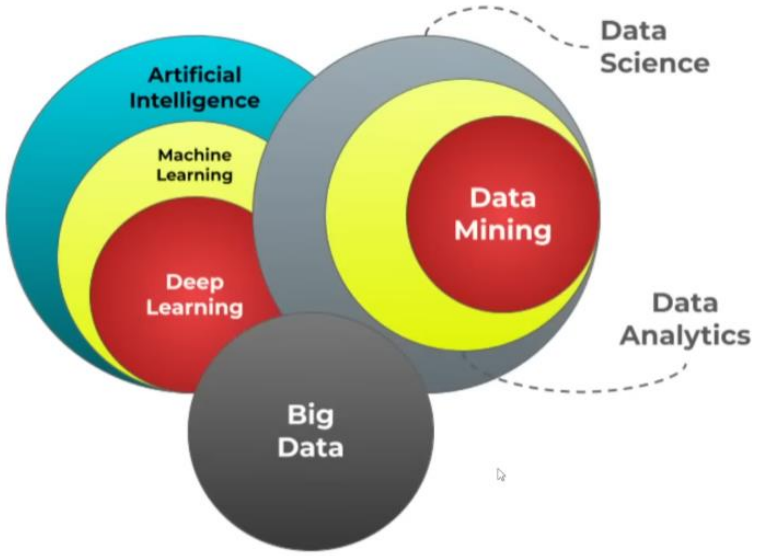


ML Temel Kavramlar

Datadan değerli bilgi nasıl üretilir?

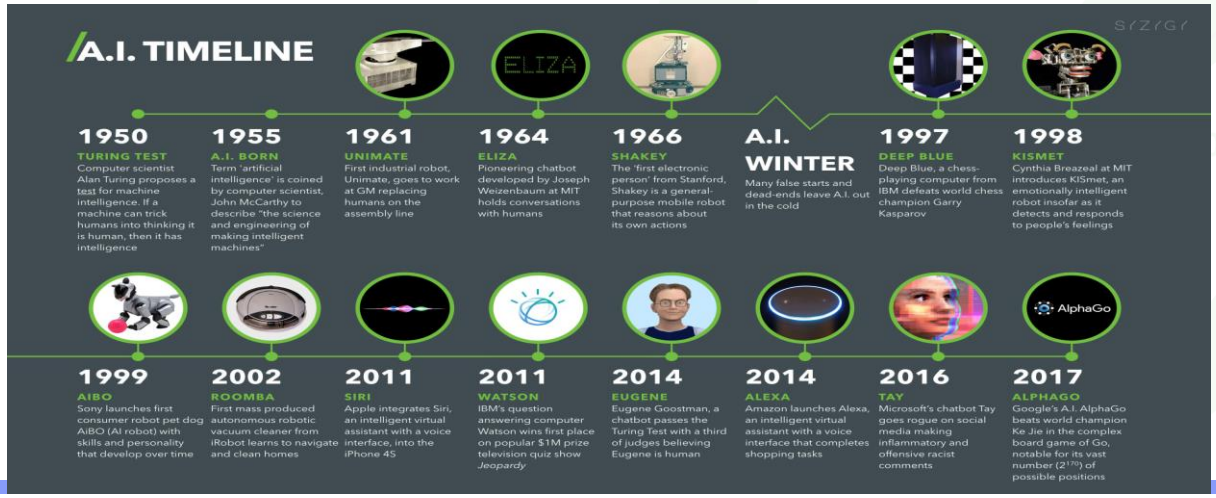
- Data Analitiği türleri

Veri Bize Ne Anlatır ?		
Analitiğin Türü	**Cevaplanan Soru**	**Örnek**
Tanımlayıcı	Ne oldu ?	Geçen sene ne kadar ürün sattık ?
Teşhise Yönelik	Neden Oldu ?	Geçen Sene neden sadece A ürününü satabildik ?
Tahmine Yönelik	Ne Olacak ?	Eğer A ürünü için reklam kampanyası düzenelrsek ne kadar satarız ?
Kural Oluşturucu	Nasıl Yapmalı ?	A ürününden 100.000 adet daha satmak için nasıl bir strateji uygulamalıyız ?

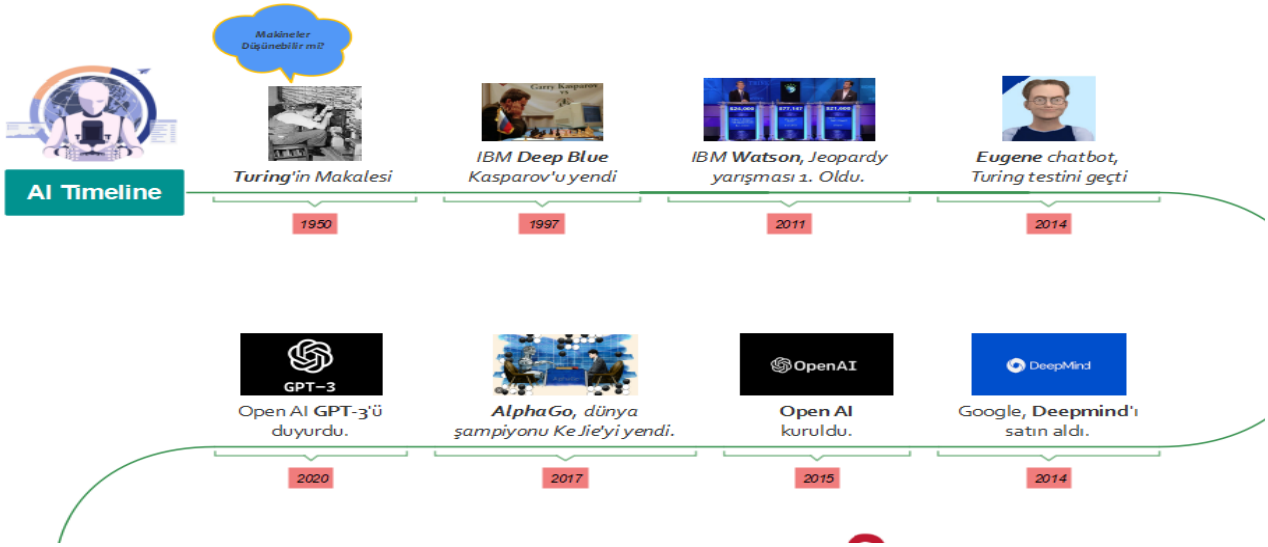


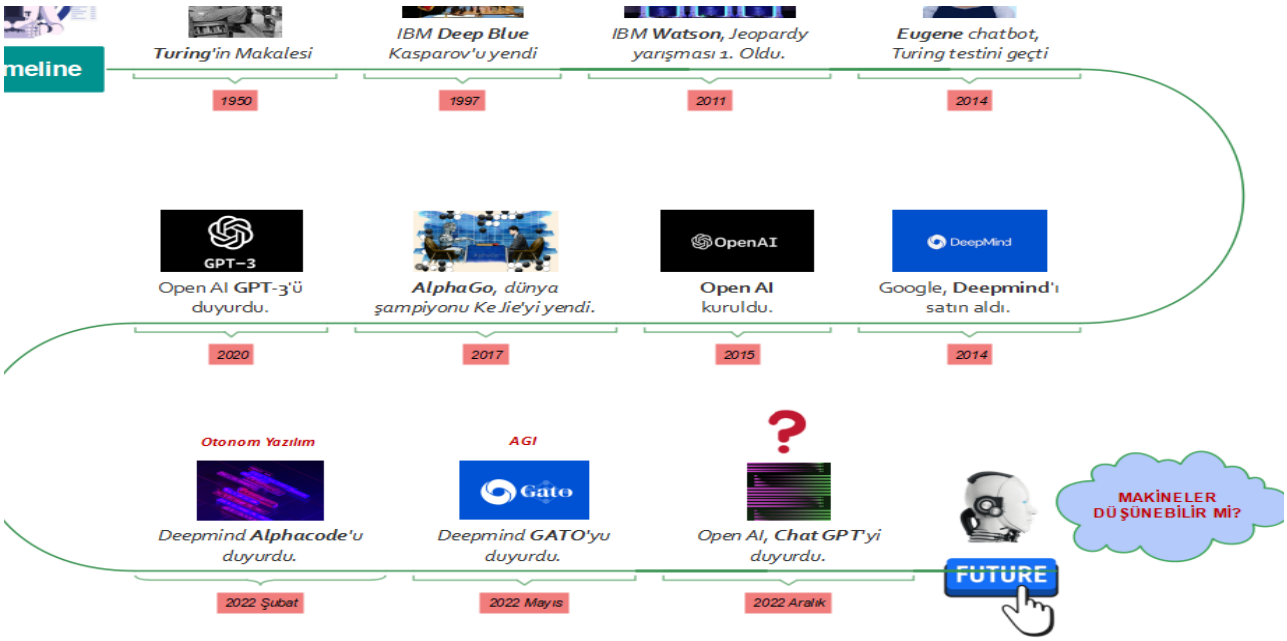
AI (Yapay Zeka) ve Bileşenleri

Tanım ve History

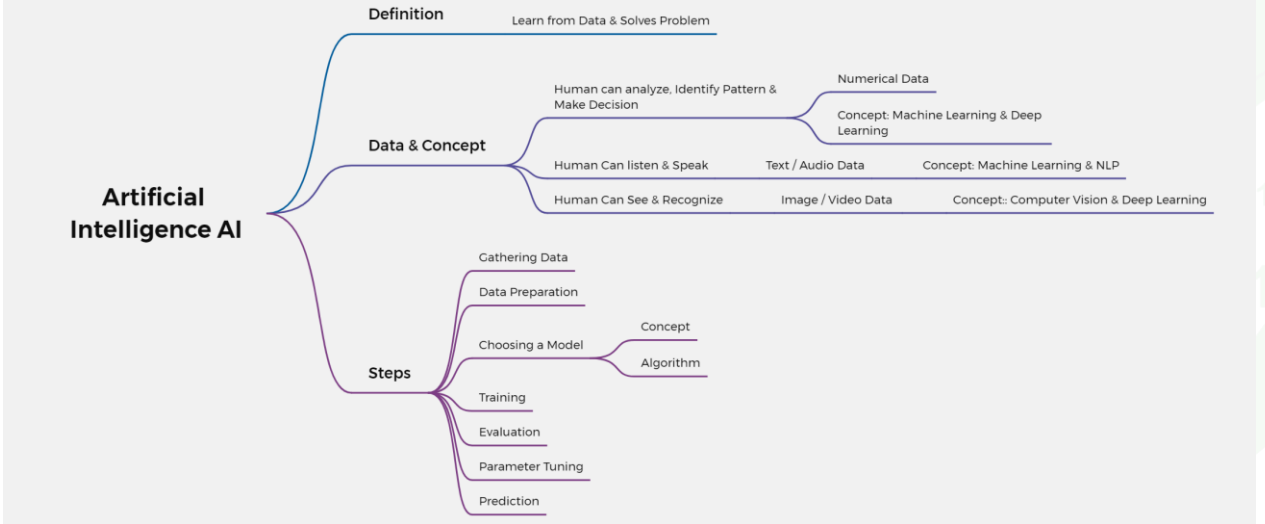


Yapay Zeka Timeline

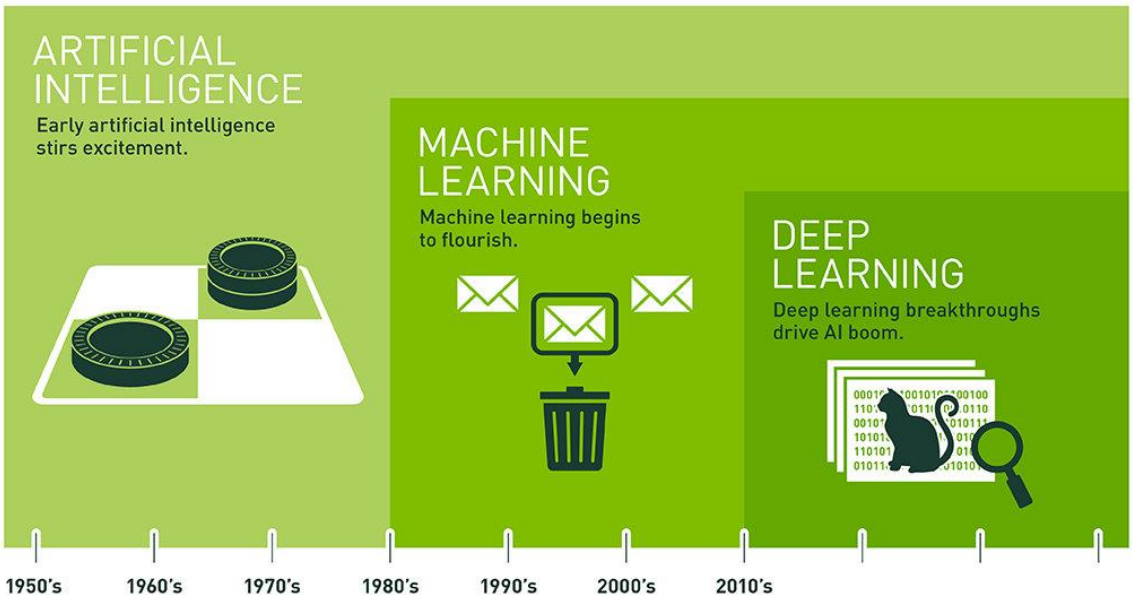



meline
Turing'in Makalesi
1950
IBM Deep Blue Kasparov'u yendi
1997
IBM Watson, Jeopardy yarışması 1. Oldu.
2011
Eugene chatbot, Turing testini geçti
2014
GPT-3
Open AI GPT-3'ü duyurdu.
2020
AlphaGo, dünya şampiyonu Ke Jie'yi yendi.
2017
OpenAI
Open AI kuruldu.
2015
DeepMind
Google, Deepmind'ı satın aldı.
2014
Otonom Yazılım
Deepmind Alphacode'u duyurdu.
2022 Şubat
AGI
Gato
Deepmind GATO'yu duyurdu.
2022 Mayıs
Open AI, Chat GPT'yi duyurdu.
2022 Aralık
FUTURE
MAKİNELER DÜŞÜNEBİLİR Mİ?

AI Mind Map

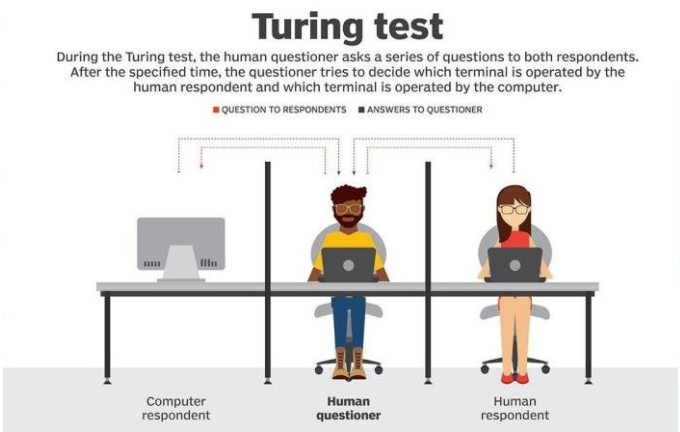

Artificial Intelligence AI
Definition
Learn from Data & Solves Problem
Data & Concept
Human can analyze, Identify Pattern & Make Decision
Numerical Data
Concept: Machine Learning & Deep Learning
Human Can listen & Speak
Text / Audio Data
Concept: Machine Learning & NLP
Human Can See & Recognize
Image / Video Data
Concept:: Computer Vision & Deep Learning
Steps
Gathering Data
Data Preparation
Choosing a Model
Concept
Algorithm
Training
Evaluation
Parameter Tuning
Prediction

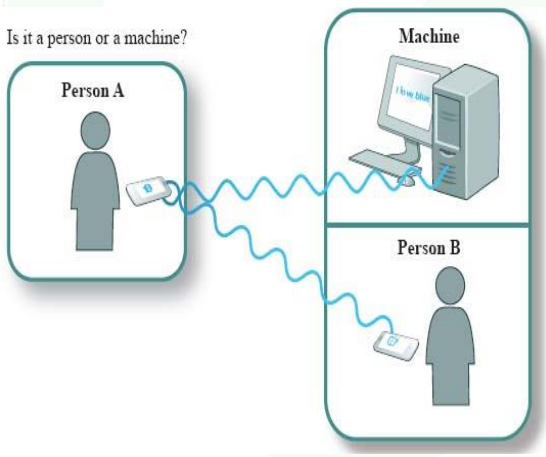
AI (Yapay Zeka) History



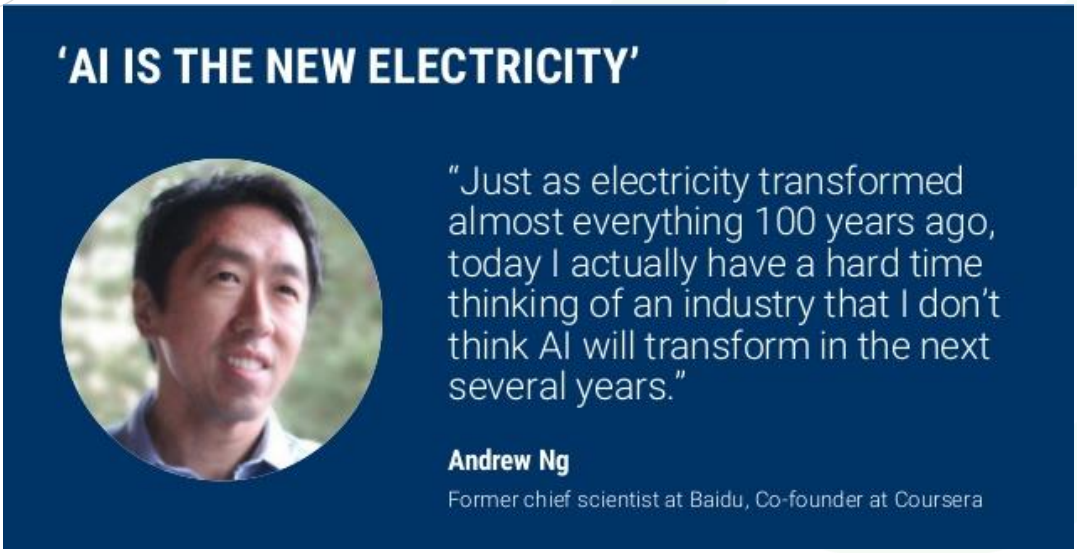
AI (Yapay Zeka) ve Bileşenleri

Turing Testi



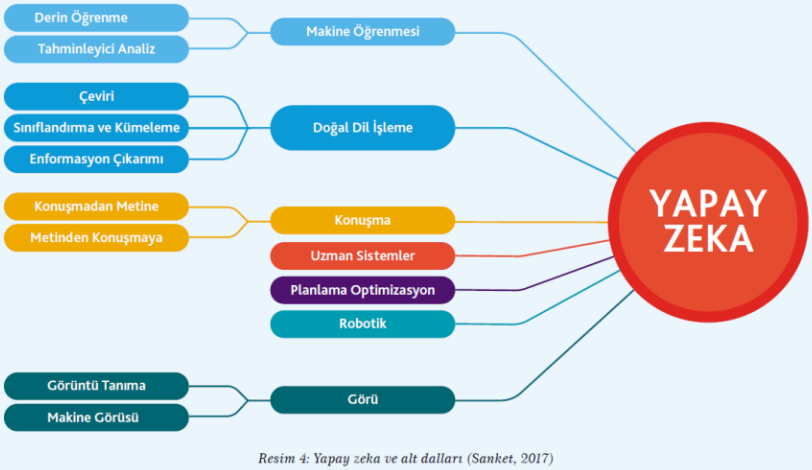


AI and Future



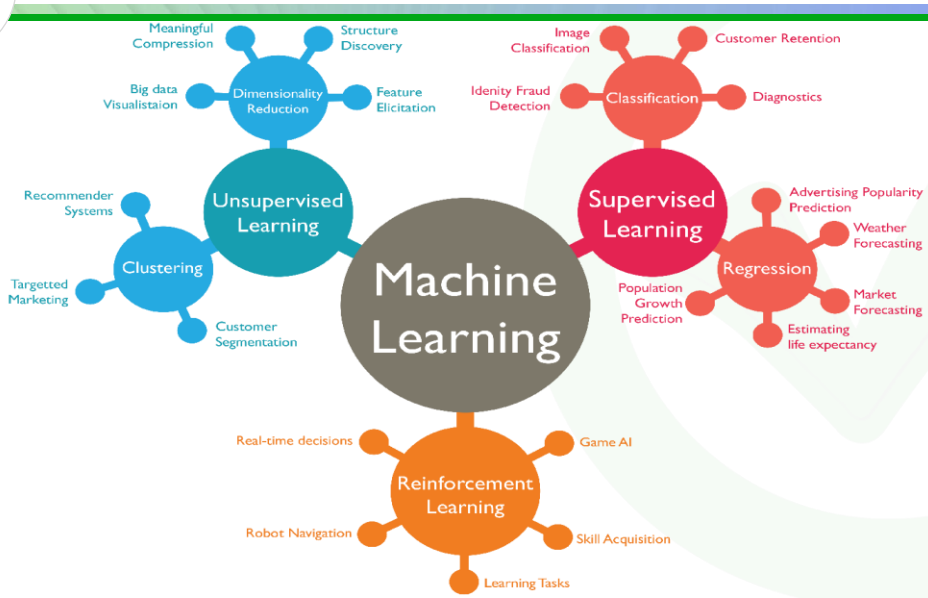
AI (Yapay Zeka) ve Bileşenleri

AI Çalışma Alanları



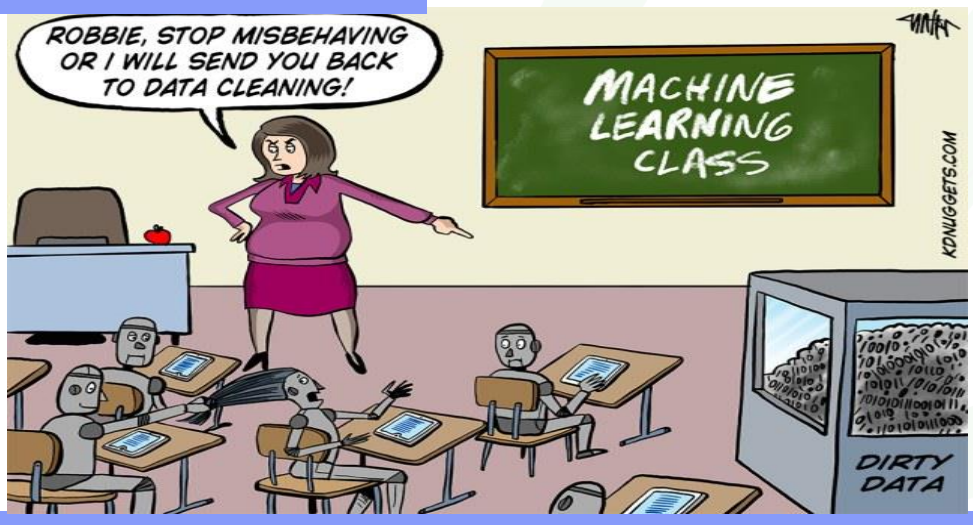
Resim 4: Yapay zeka ve alt dalları (Sanket, 2017)

ML TYPES



Machine Learning ve Türleri

Tanım



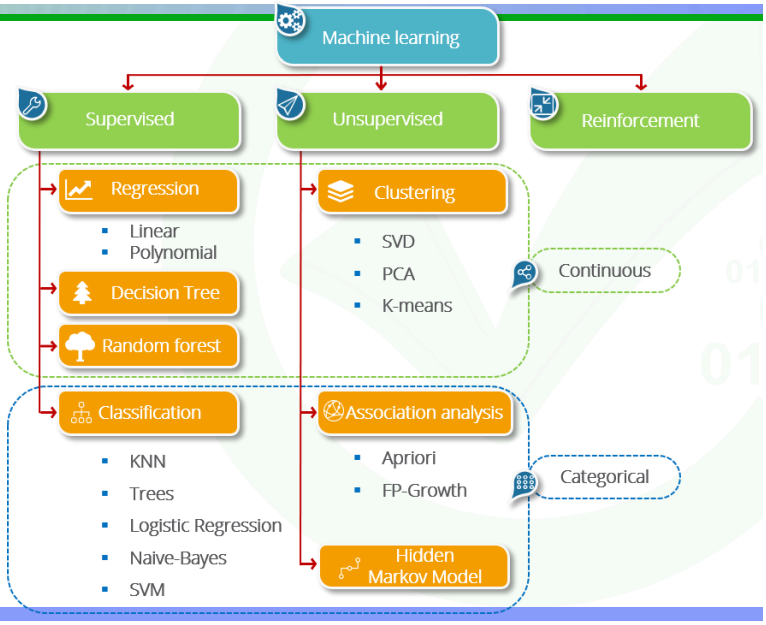
Machine Learning ve Türleri

Bir Örnek

Machine Learning ve Türleri

Türleri

- Türleri
  - ✓ Supervised
  - ✓ Unsupervised
  - ✓ Reinforcement
- Labeled - Unlabeled

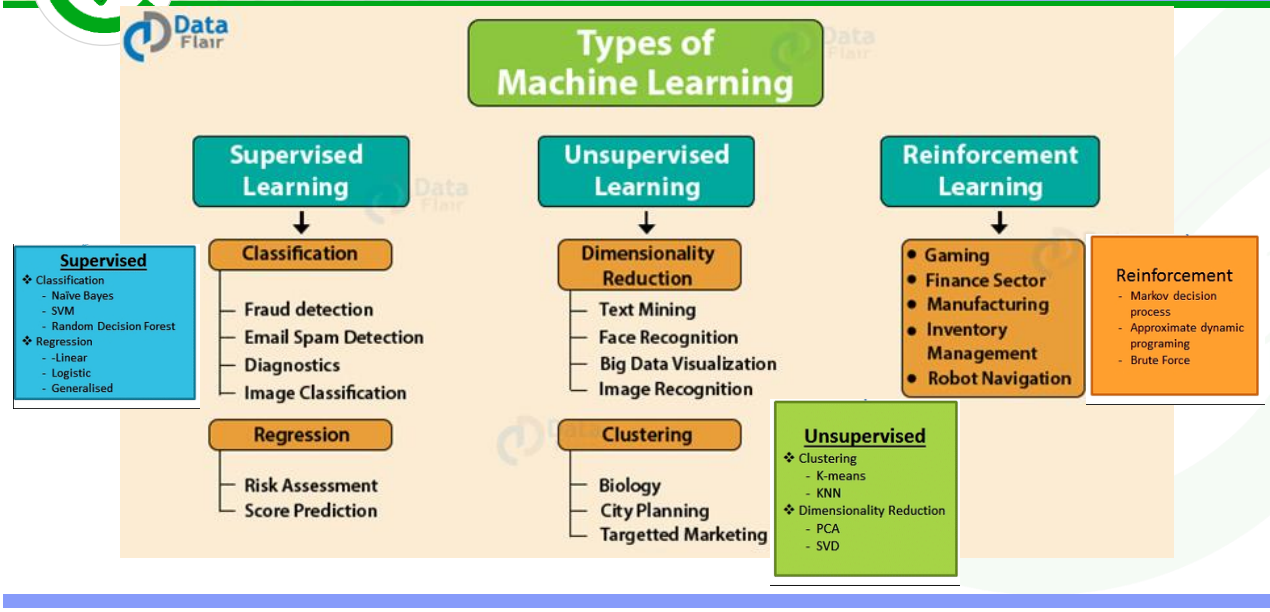


Machine Learning ve Türleri

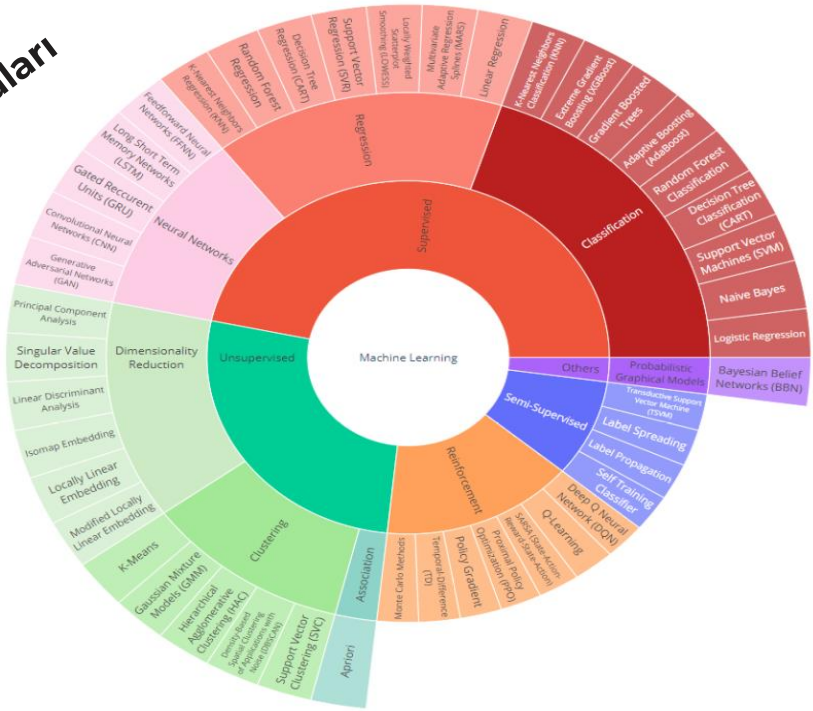
ML Algoritmasının Türleri

Denetimli Öğrenme	Denetimsiz Öğrenme	Pekiştirmeli Öğrenme
• Regresyon • Sınıflandırma	• Kümeleme • Boyut Azaltma • Birliktelik Kuralları	

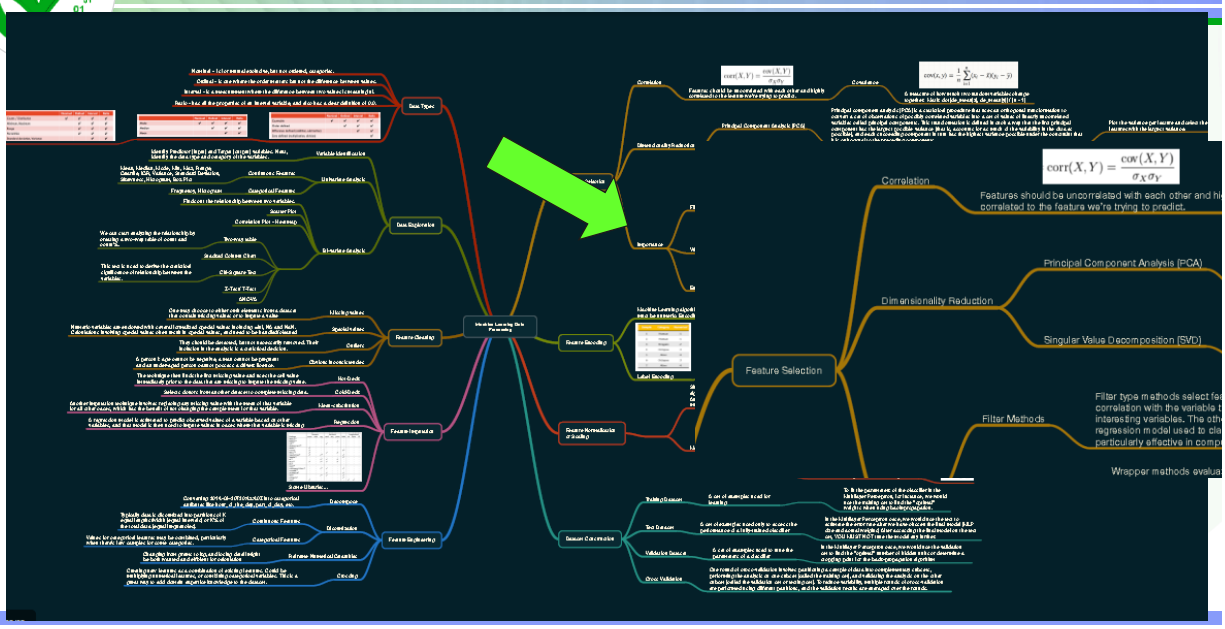
ML Algoritmaları

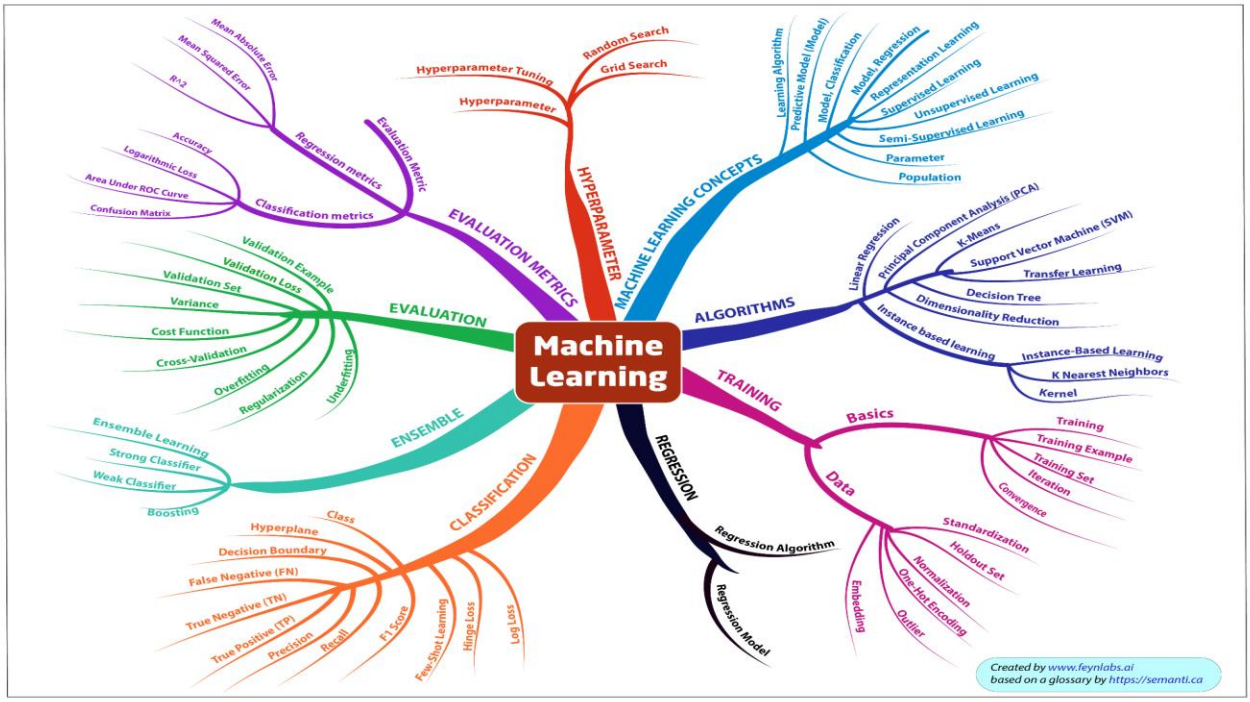


ML Algoritmaları



Mind Map of ML



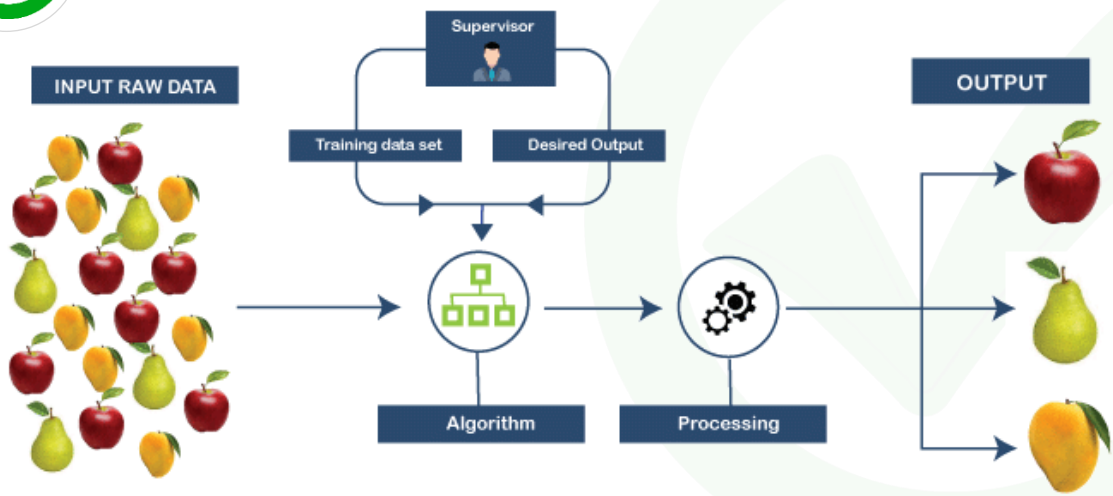


Everything is clear ?

SUPERVISED LEARNING

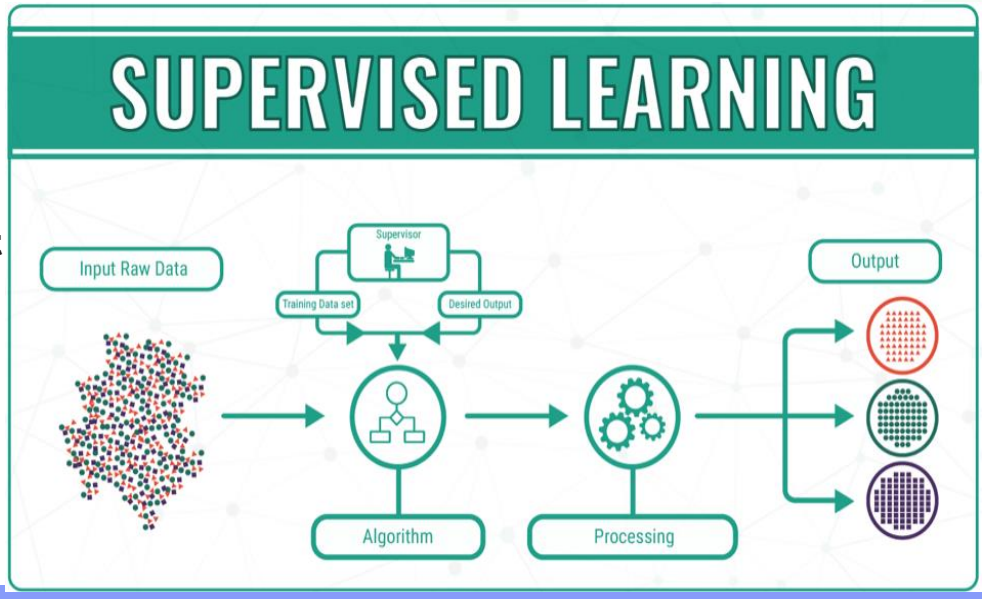
Denetimli Öğrenme

Supervised Learning



Supervised Learning

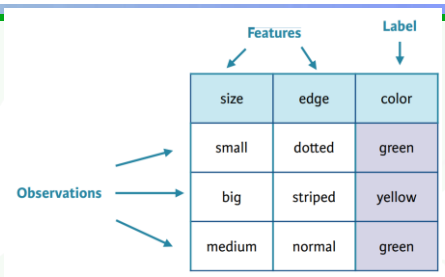
- ✓ **Features**
- ✓ **Target**
- ✓ **Independent variable**
- ✓ **Dependent variable**

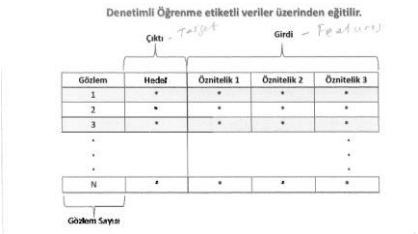


Supervised Learning

Feature-Label

- Features (Independent variable)- X
- Label (target- dependent var.) - Y



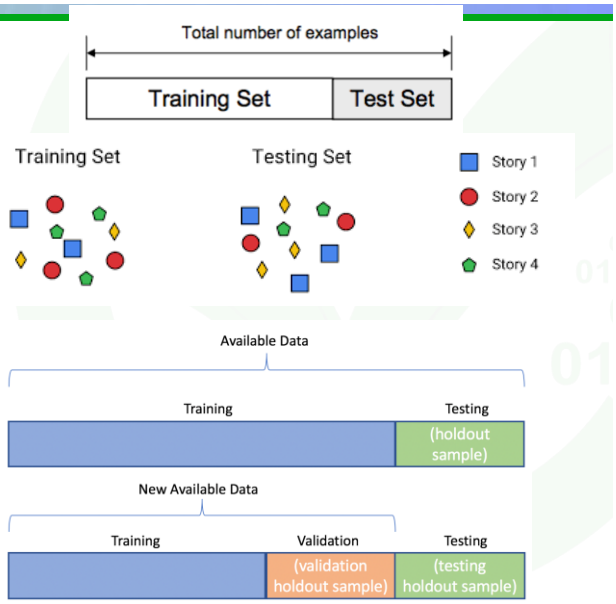


Features					Label
Position	Experience	Skill	Country	City	Salary ($)
Developer	0	1	USA	New York	103100
Developer	1	1	USA	New York	104900
Developer	2	1	USA	New York	106800
Developer	3	1	USA	New York	108700
Developer	4	1	USA	New York	110400
Developer	5	1	USA	New York	112300
Developer	6	1	USA	New York	114200
Developer	7	1	USA	New York	116100
Developer	8	1	USA	New York	117800
Developer	9	1	USA	New York	119700
Developer	10	1	USA	New York	121600

Supervised Learning

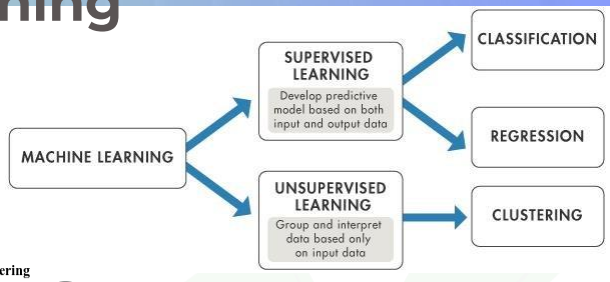
Datayı Bölme

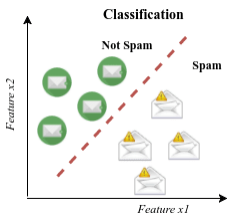
- Data Split
  - Training
  - Test

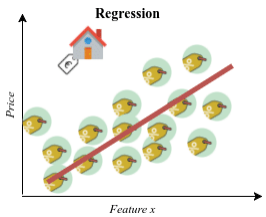


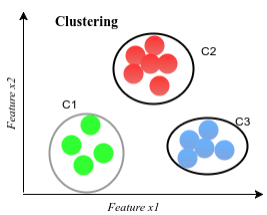
Supervised Learning

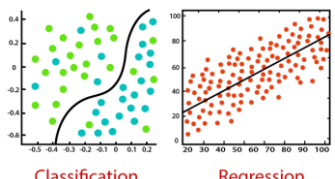
- **Regression**
- **Classification**

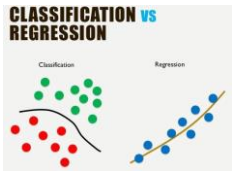


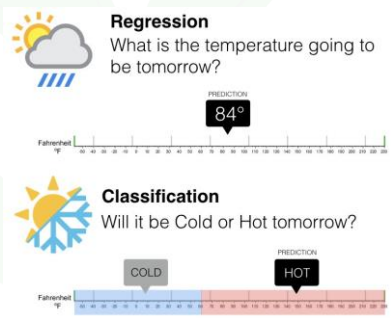


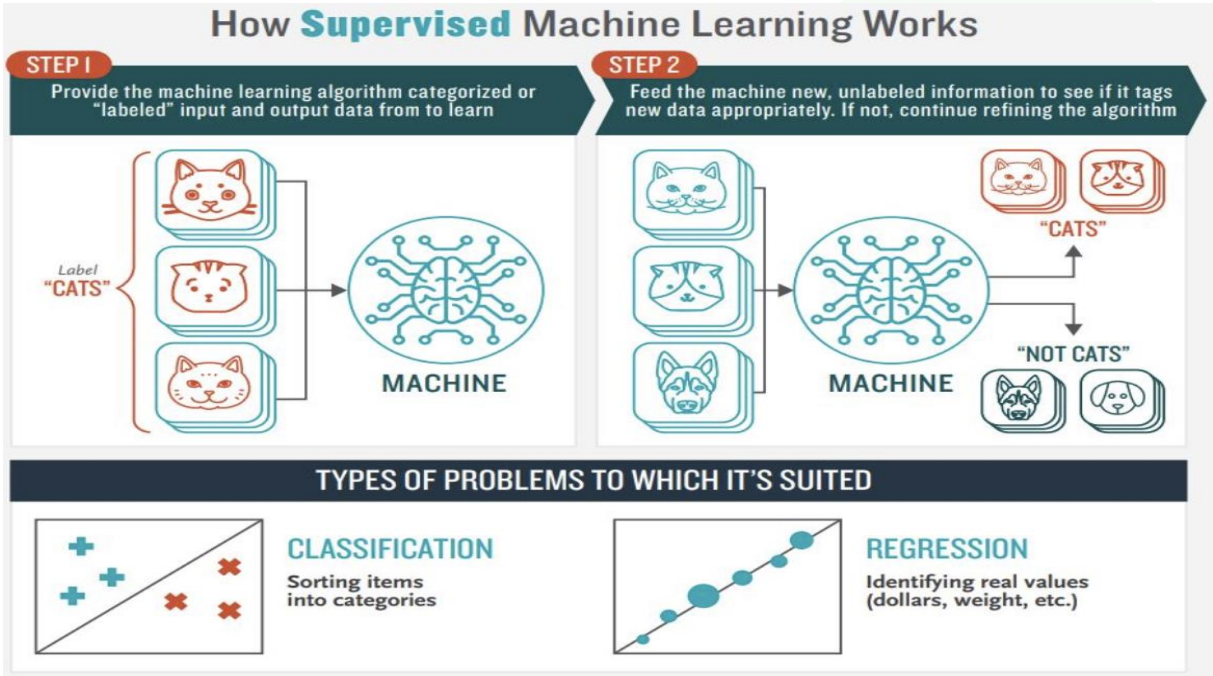


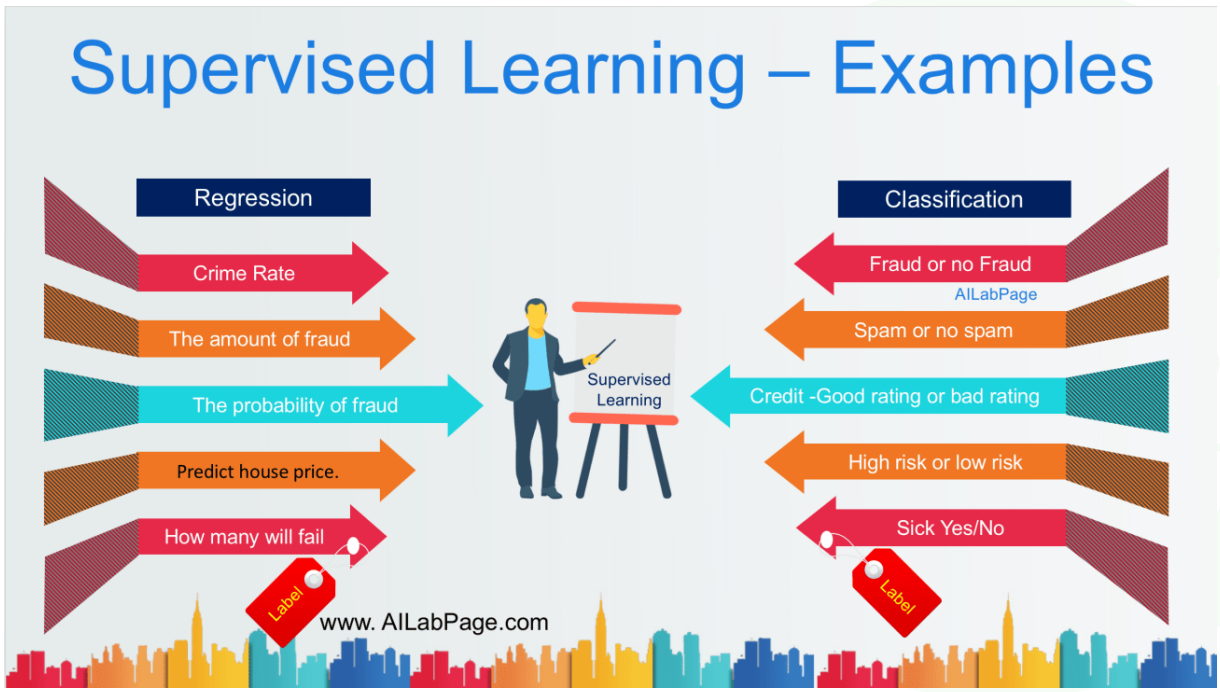








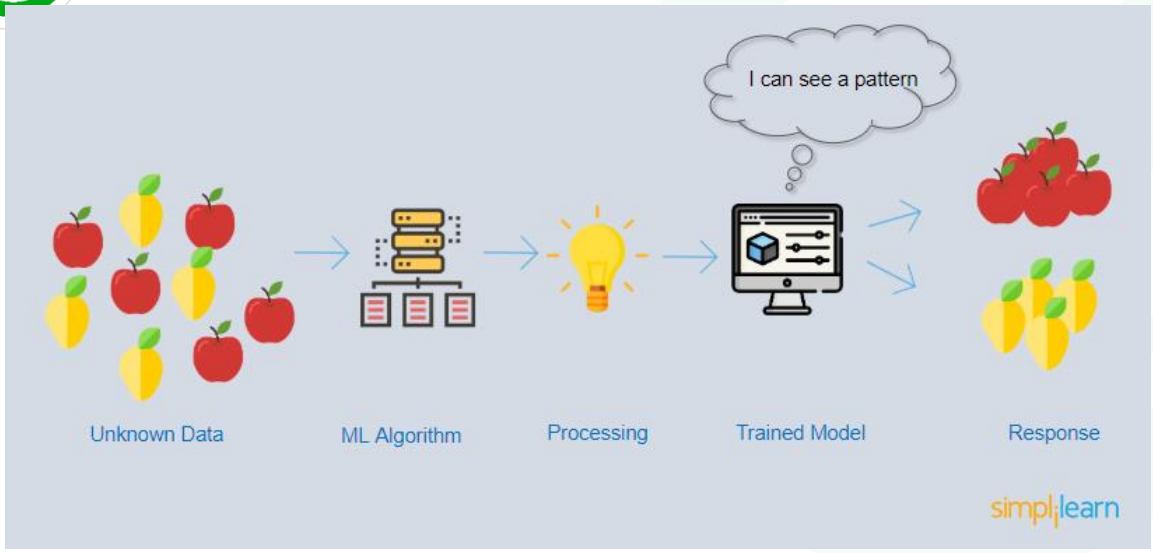

How Supervised Machine Learning Works
STEP 1
Provide the machine learning algorithm categorized or "labeled" input and output data from to learn
STEP 2
Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm
Label "CATS"
MACHINE
"CATS"
"NOT CATS"
MACHINE
TYPES OF PROBLEMS TO WHICH IT'S SUITED
CLASSIFICATION
Sorting items into categories
REGRESSION
Identifying real values (dollars, weight, etc.)


Supervised Learning – Examples
Regression
Crime Rate
The amount of fraud
The probability of fraud
Predict house price.
How many will fail
Label
Classification
Fraud or no Fraud
AILabPage
Spam or no spam
Credit -Good rating or bad rating
High risk or low risk
Sick Yes/No
Label
Supervised Learning
www. AILabPage.com

UNSUPERVISED LEARNING

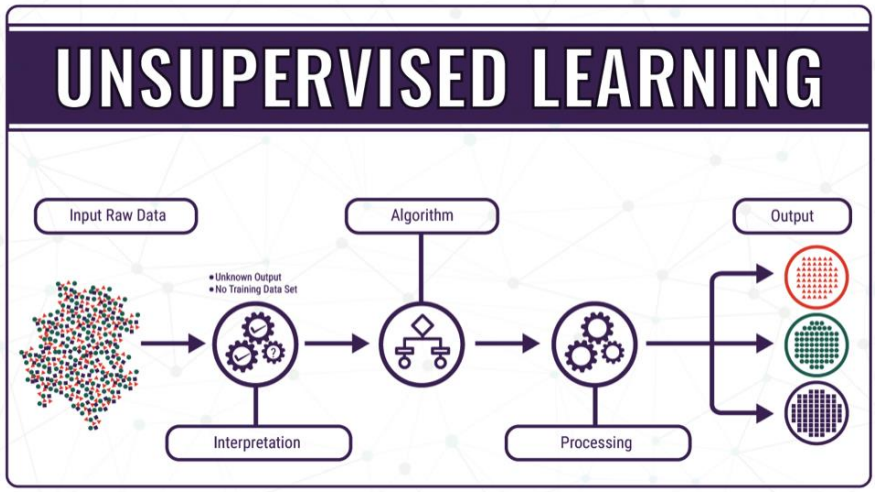
Denetimsiz Öğrenme

Unsupervised Learning



Unsupervised Learning

- ✓Unlabeled data
- ✓Insight

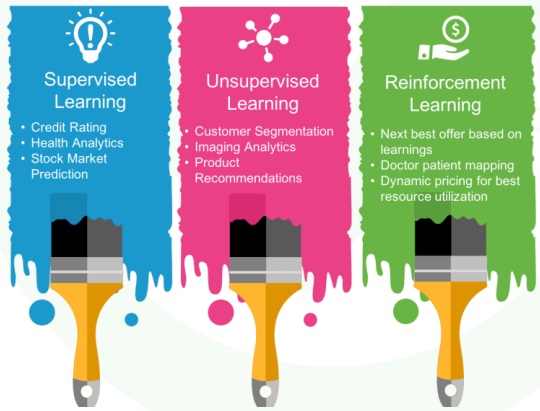


Unsupervised Learning

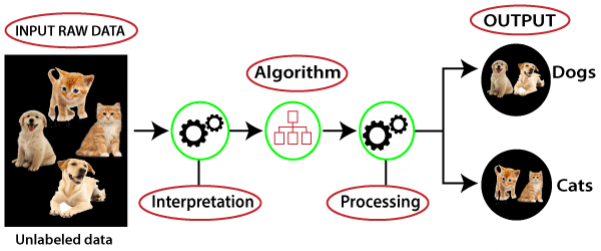
Denetimsiz Öğrenme etiketsiz veriler üzerinden eğitilir.

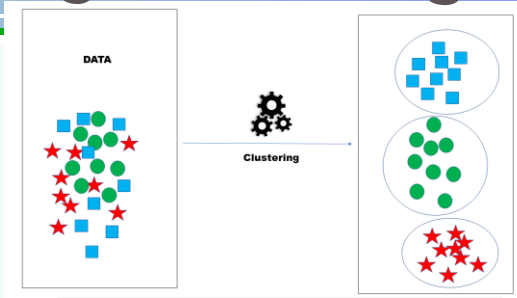
	Çıkış	Girdi		
Gözlem	Hedef	Öznitelik 1	Öznitelik 2	Öznitelik 3
1	*	*	*	*
2	*	*	*	*
3	*	*	*	*
.				.
.				.
.				.
N	*	*	*	*

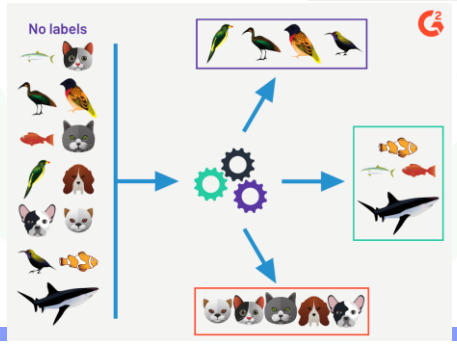
Gözlem Sayısı



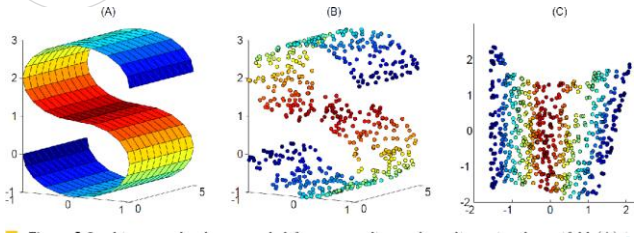
Unsupervised Learning - Clustering

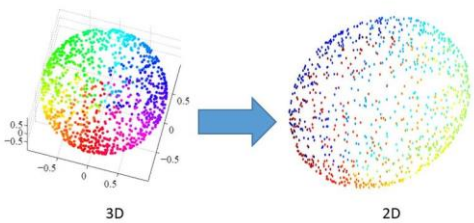


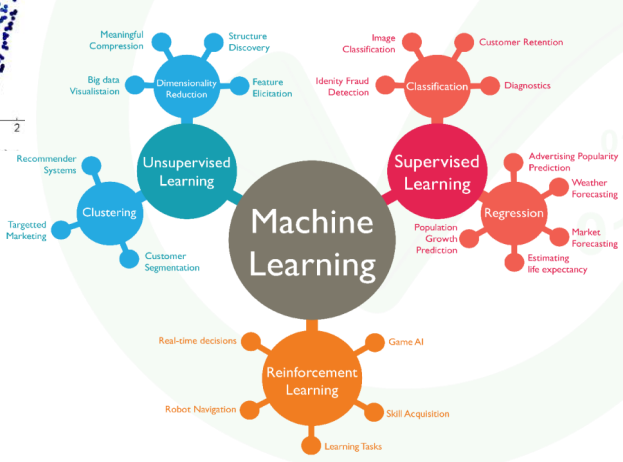




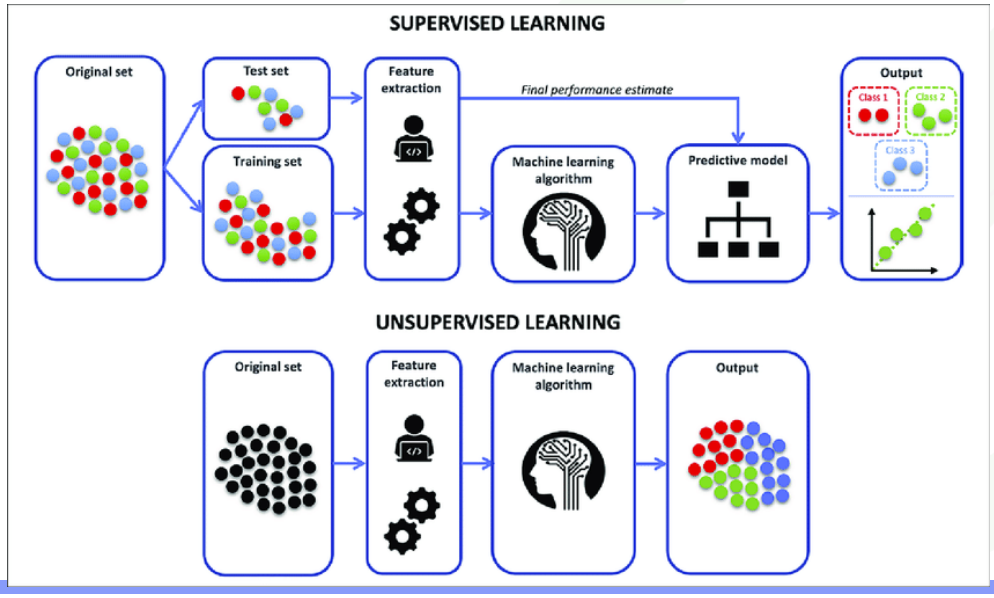
Dimensional Reduction







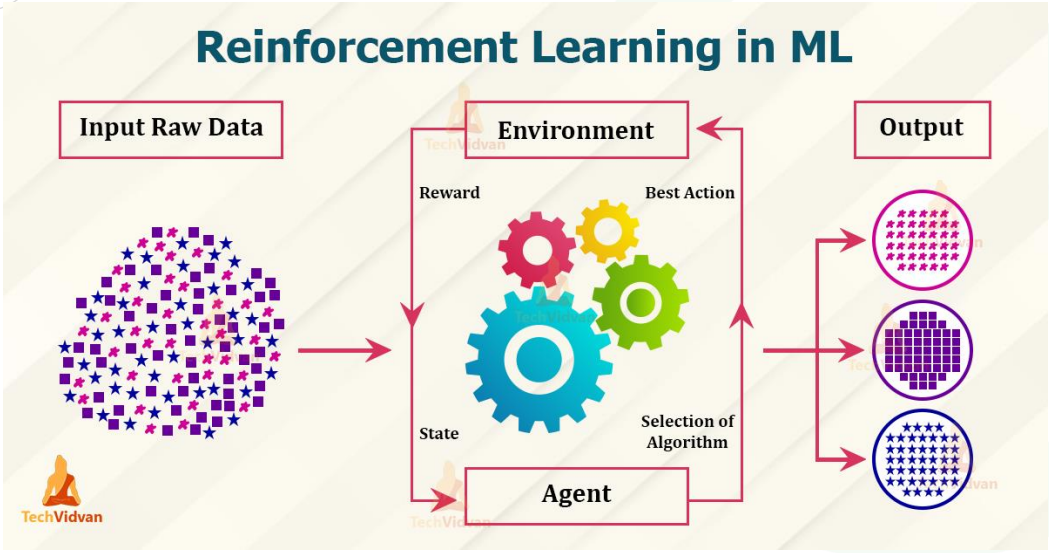
Supervised vs. Unsupervised Learning



REINFORCEMENT LEARNING

Reinforcement Learning

Reinforcement Learning



MACHINE LEARNING STEPS

Makine Öğrenmesi Aşamaları

Machine Learning Steps

ML Aşamaları

1. Data Toplama (Data Collection)
   - ML ilk amacı
   - Data, makine öğrenmesiyle çalışan yapay zekanın gıdasıdır
   - Veri toplama sırasında karşılaşılan en yaygın sorunlar
     - Hatalı data
     - Data yapısı
     - Eksik data
     - Data dengesizliği
     - Yanlı Data

Machine Learning Steps

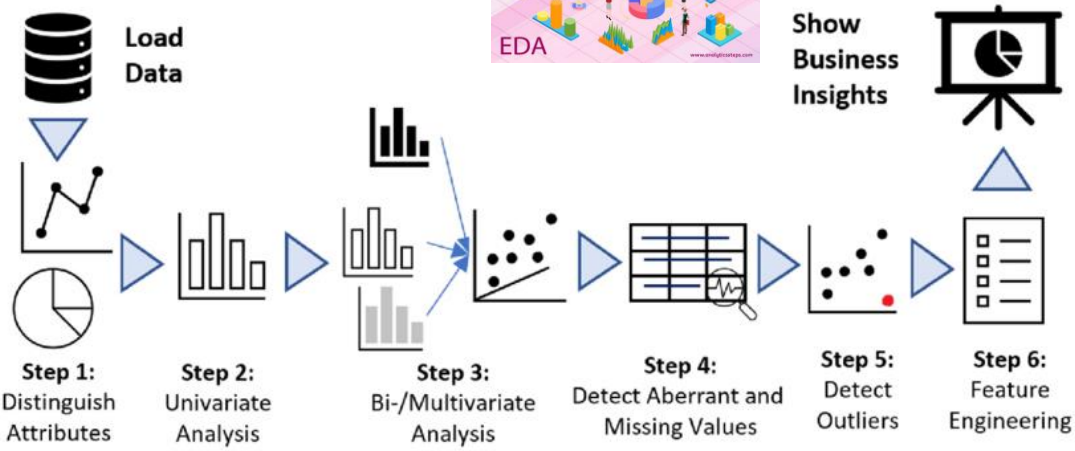
ML Aşamaları

2. Data Hazırlama (DATA PREPARATION)

- EDA (Exploratory Data Analysis)
- Data Ön işleme (Data preprocessing)
- Data Seti bölümleme (Data split)

Machine Learning Steps

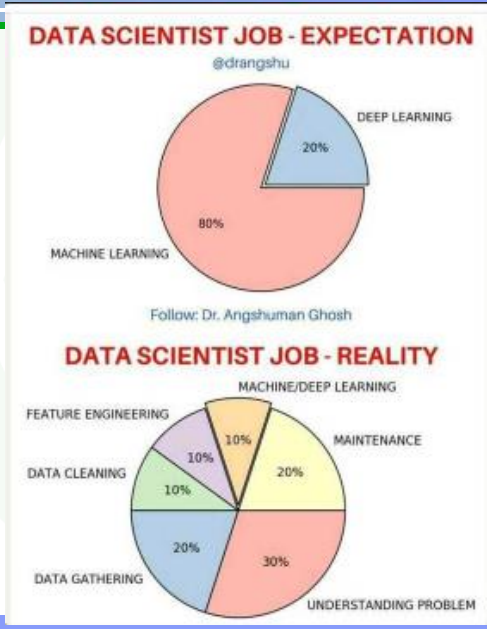
EDA (Exploratory Data Analysis)



Machine Learning Steps

Data Preprocessing

- Verilerin iyi modellenmesi bu adımın sağlıklı şekilde yürütülmesine bağlıdır. Veri Ön-işleme adımı (Data Preprocessing) makine öğrenmesi sürecinde en fazla zaman alan kısımdır. **Bir çalışmaya göre, veri bilimciler, zamanlarının %60'ını verileri temizlemek ve düzenlemek için harcamaktadırlar**.

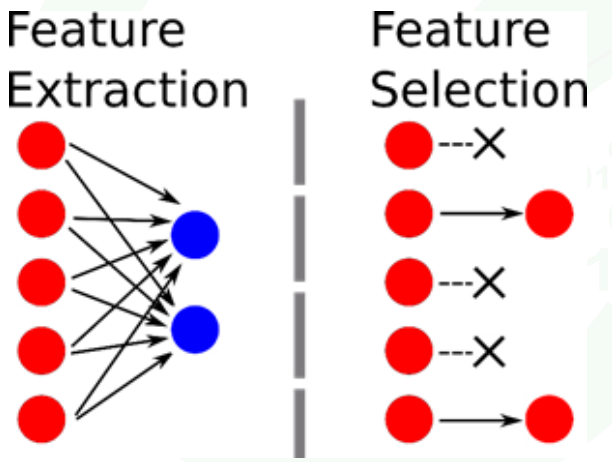


Machine Learning Steps

Datayı Hazırlama

DATA PREPARATION (Veri ön işleme) birkaç teknik adıma dayanır:

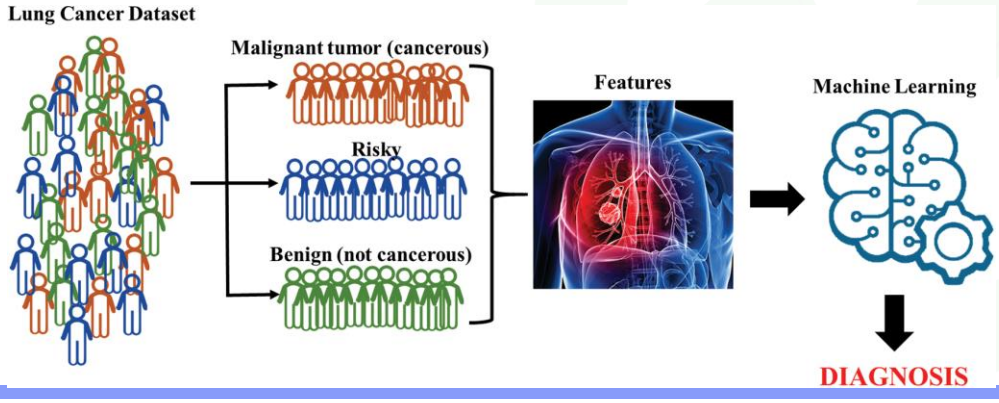
- **Data temizleme**
- **Data yapılandırma**
- **Eksik data tamamlama**
- **Oversampling**
- **Data Entegrasyonu**
- **Data normalizasyonu**
- **Feature Selection**
- **Feature Extraction**
- **Data Split**



Machine Learning Steps

3. Model Oluşturma (Modelling)

- Training (*Algoritmanın Eğitilmesi)*
- Testing

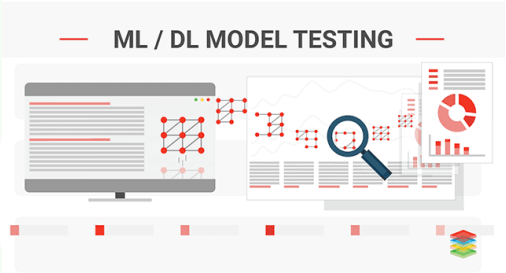


Machine Learning Steps

3. Model Oluşturma (Modelling)

- Training
- Testing (*Modelin test edilmesi)*

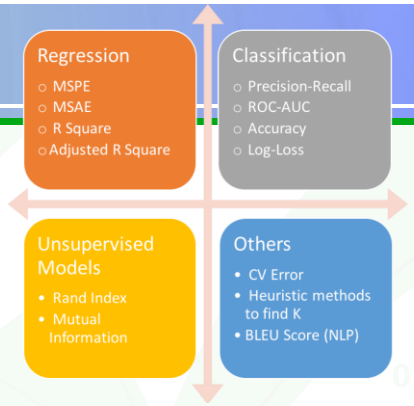
- Eğitim ve test aşamaları askerlik sürecine benzetilebilir. Acemi birliğinde eğitim alan asker daha sonra usta birliğine yollanır. Burada tatbikatlara ve hatta komutanları nezaretinde geri operasyonlara katılır. Sadece çok az bir kısmı başarılı ve istekliyse profesyonel asker olarak kariyerine devam edebilir. **Modelin cepheye sürülmesi sadece eğitim performansına değil, bundan daha çok test performansına bağlıdır.**



Machine Learning Steps

4. Model Değerlendirme (Model Evaluation)

- Modelin tahmin performansı nasıl ölçülür?
- Performans ölçümü (Error Metrics)
- Cross-Validation (Çapraz doğrulama

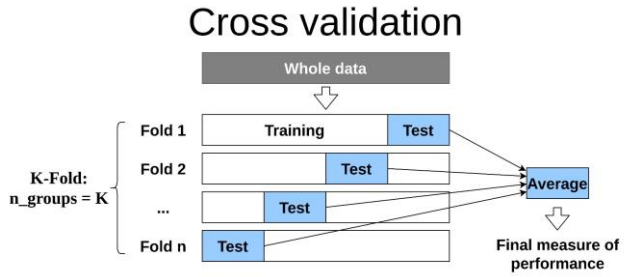


Regresyon	Sınıflandırma	Kümeleme
• MSE	• Doğruluk	• ARI
• RMSE	• Kesinlik	• MIS
• AMSE	• Duyarlılık	• V-measure
• R2	• F1	• Elbow
• Düzeltilmiş R2	• ROC-AUC	
	• Kesinlik-Duyarlılık Grafiği	

Machine Learning Steps

4. Model Değerlendirme (Model Evaluation)

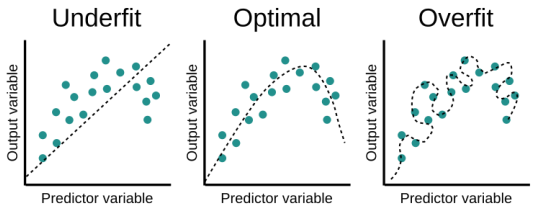
- Cross-Validation (Çapraz doğrulama)

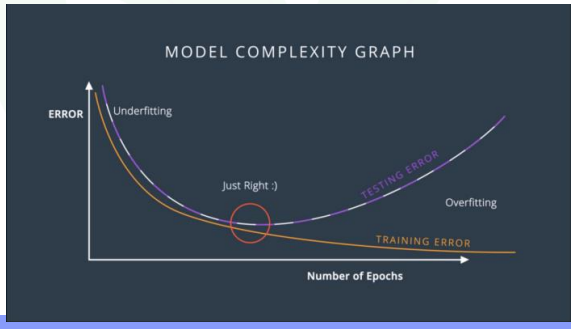


Machine Learning Steps

4. Model Değerlendirme (Model Evaluation)

- Overfitting (Ezberleme) (Aşırı Öğrenme) (Abartılı taklit)
- Training set ☺ Test Set ☹

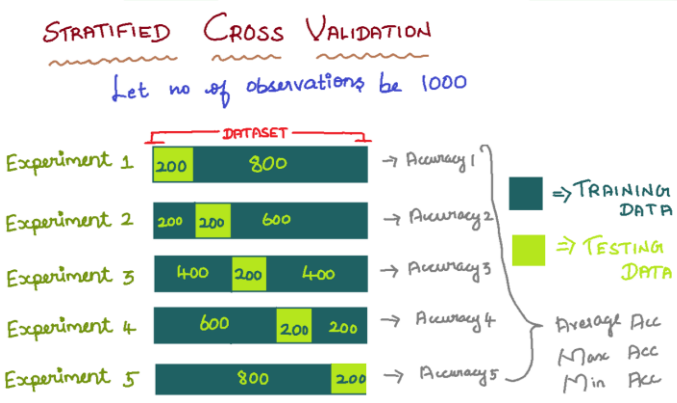




Machine Learning Steps

4. Model Değerlendirme (Model Evaluation)

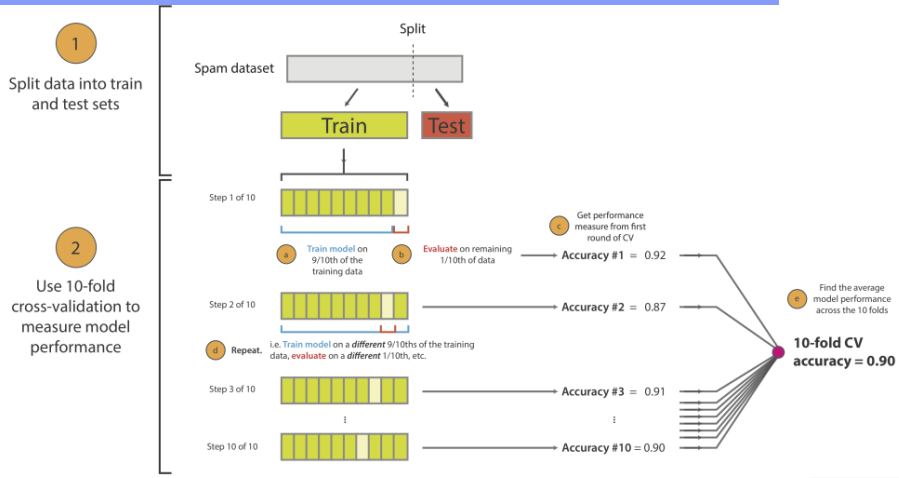
- Overfitting var, ne yapabiliriz ??
  - Cross-validation



Cross validation

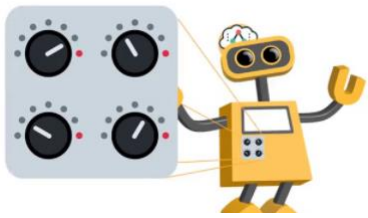
Machine Learning Steps

4. Model Değerlendirme (Model Evaluation)



Machine Learning Steps

5. Hiper Parametre İnce ayar (HyperParameter Tuning)



$$RSS_{LASSO} = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2 + L1\sum_{j=1}^{p}|\beta_j|$$

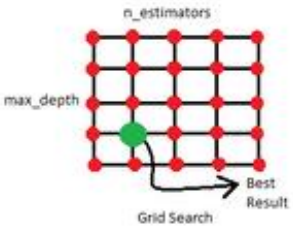
Machine Learning Steps

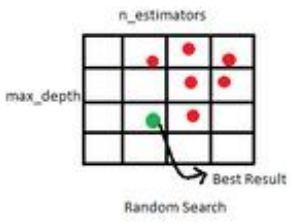
5. Hiper Parametre İnce ayar (HyperParameter Tuning)

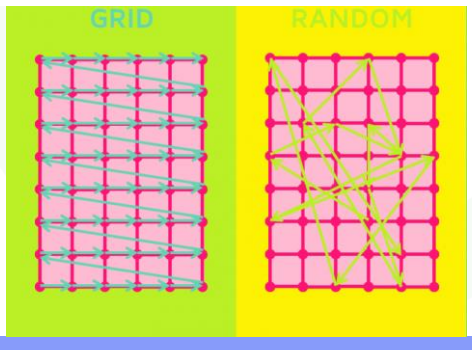
Grid Search

(hiperparametrelerin optimum değerlerini hesaplamaya çalışan bir ayarlama tekniği)

(belirli bir model için en uygun hiperparametreleri yapılandırmak üzere verileri tarama işlemi)

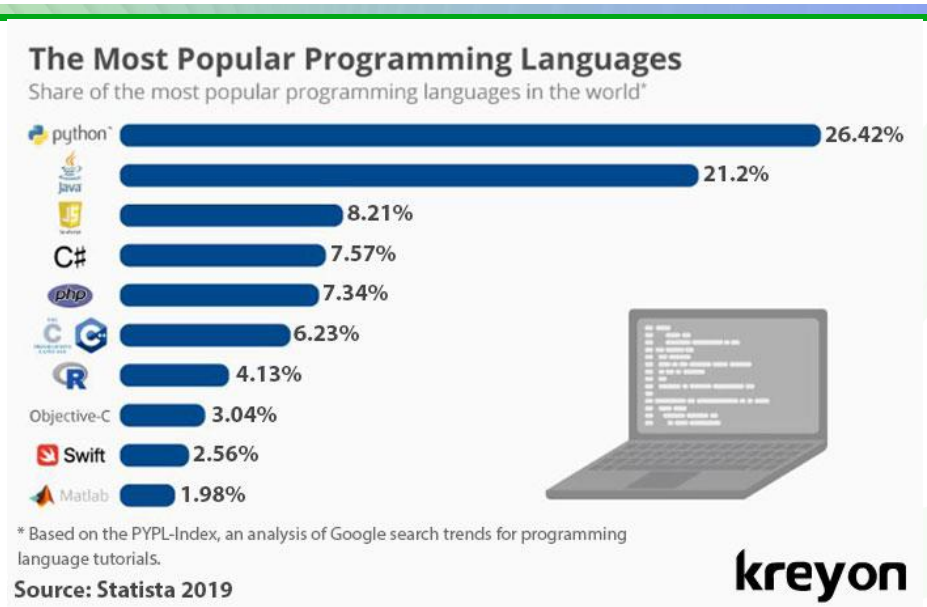






Neden Python

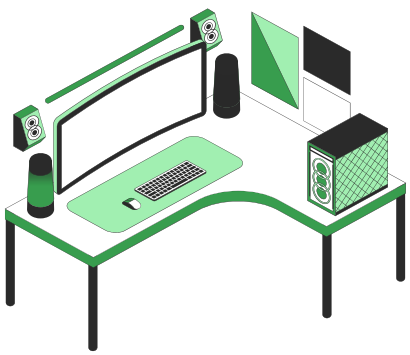
Neden Python ?



Bu ders benim için verimli geçti

Everything is clear ?

FINISH



Do you have any questions?

Send it to us! We hope you learned something new.