

VERİ BİLİMİ AKADEMİSİ

Kasım 2025





Neden Ücretsiz?

Ikigai

A JAPANESE CONCEPT MEANING "A REASON FOR BEING"





Tanışma

Ben

- 2000, İTÜ Endüstri Müh,
- 2004-...: Akbank, KVO
- 2018: Python'la tanılaşma
- 2018/19: Udemy/Coursera
- 2020/21, İleri Analitik Programı, Accelera
- 2021/22, SÜ Data Analytics
- Eğitmenlik Deneyimi
 - Udemy, Excel VBA
 - Banka içi münferit eğitimler

Siz

- Yaş: 20 - 42
- K/E: ~%50
- 6 öğrenci (+/-)
- 2 hoca
- Neden?
 - Kariyerlerini veri bilimi veya yazılım alanında ilerletmek
 - Mevcut bilgilerini güncellemek
 - Yeni beceriler kazanmak
 - Profesyonel rehberlik almak

Beklentiler



Benim Beklediklerim

- Devamlılık
 - Özel durumları önden bildirme
 - Benden kaynaklı ders olmaması durumları
 - 8 Aralık Pzt
 - Yaz tatili ?
 -?
- Aktif katılım
- Ödev ve Proje teslimleri
 - Manuel/Random kontrol
 - Web'e yükleme

Kurs Sonunda?

- Sertifika?
- Portföy (Github & Kaggle)
- İşin mantığını anlamış olmanız
 - Low Code / No Code Tools
- **Stairway to heaven?**

Süreç ve Anlatım tarzı



- %20-30 Teori (*), %70-80 Pratik
- %70-80 ben, %20-30 siz anlatacaksınız
- Canlı kod yazımı + kısmen copy-paste
- Betimlemeli çizimler
- Dallanıp budaklanma
- Teams Kanallar
- Ekran Kaydı?



Başlıyoruz...



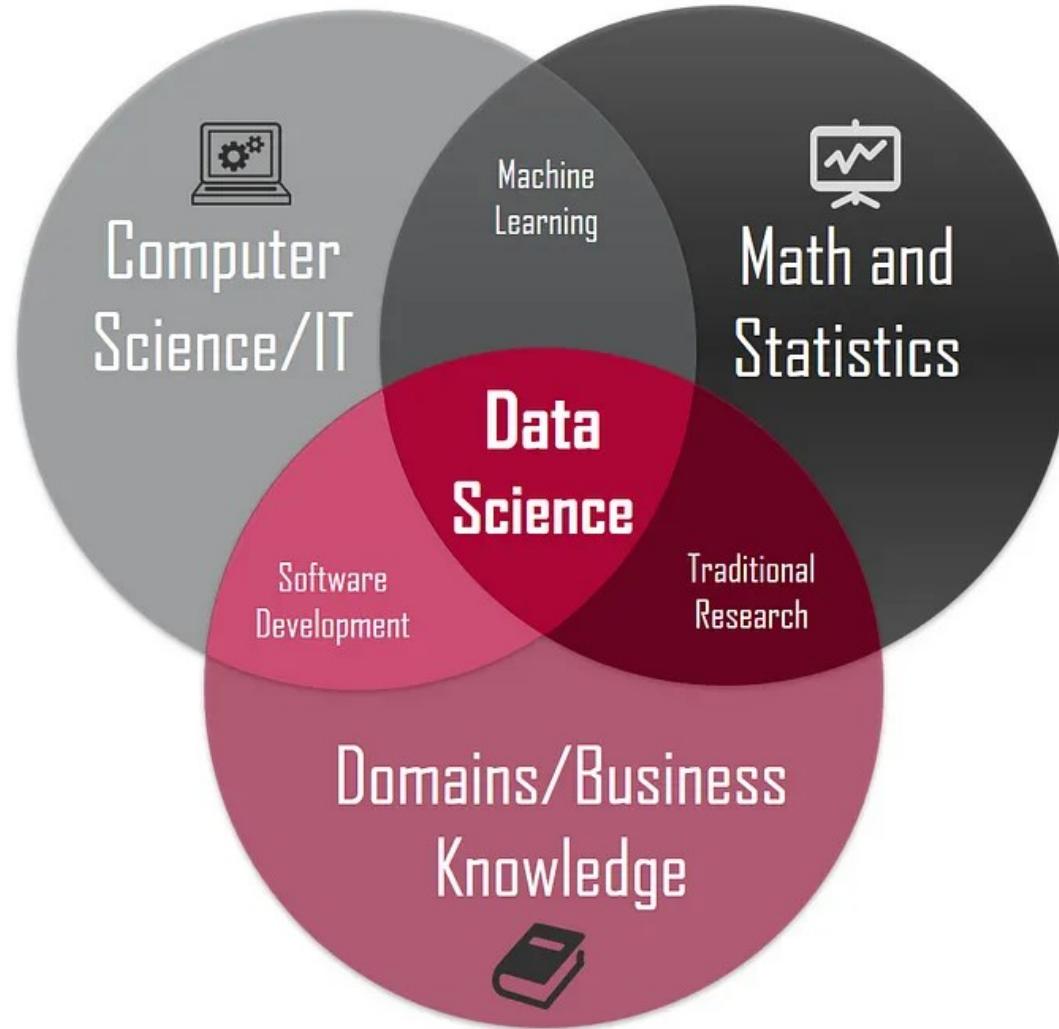
- ML Tani
- ML vs D

Machine I

Machine I

"Field of stud
computers the a
without bein
program"

-Arthur San



data science

# of bathrooms	newly renovated	price (1000\$)
2	N	115
3	N	150
1	N	210
3	Y	280
4	N	355
5	Y	440

ies with 3 bedrooms are more expensive
omes with 2 bedrooms of a similar size.
ly renovated homes have a 15% premium.

Andrew Ng



ChatGPT ne diyor

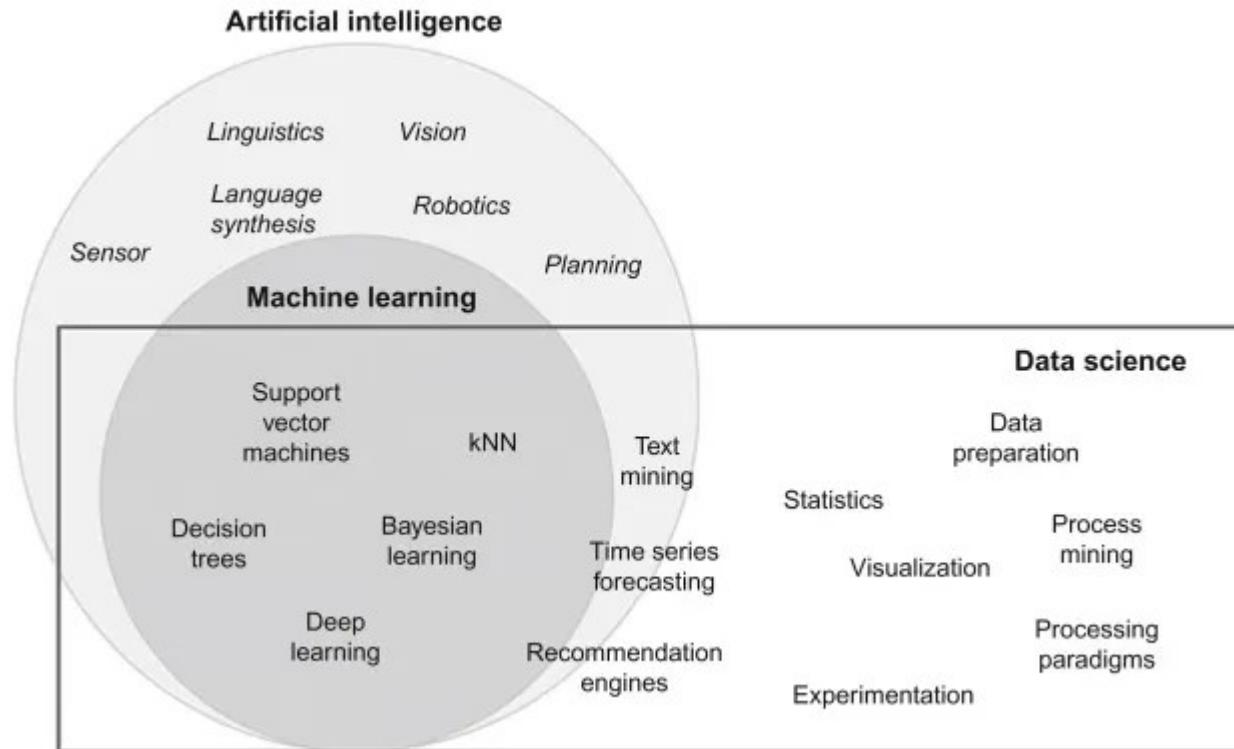
imi

edebilir miyim?" (*)

Amaç: Veride

Soru: "Bu veri

- Veri toplama (Sensor)
- Veri temizleme
- Veri analizi
- **İstatistiksel n**
- Görselleştirme
- Raporlama
- İş kararları üre



* Supervised Learning örneği



ML Kullanım Alanları

1. Tahminleme (Forecasting)

- Satış tahmini
- Stok/lojistik talep planlama
- Enerji tüketimi tahmini
- Hava durumu tahmini
- Konut fiyat tahmini
- Sigorta prim hesaplama

2. Sınıflandırma

- E-mail spam tespiti
- Sahtekârlık (fraud) tespiti
- Müşteri churn tahmini
- Hastalık teşhisleri (ör. diyabet tahmini)

3. Tavsiye Sistemleri

- Netflix/YouTube önerileri
- Spotify kişisel listeleri
- E-ticaret ürün önerileri

4. Zaman Serisi ve Anomali Tespiti

- Bankacılıkta şüpheli işlem yakalama
- Sunucu/metrik monitoring
- IoT sensör verilerinde arıza erken uyarı

5. Müşteri Segmentasyonu

- Pazarlama segmentleri
- Kullanıcı kümeleri

6. Doğal Dil İşleme (NLP)

- Chatbot'lar
- Sentiment analizi
- Otomatik metin özetleme
- Dil çeviri
- LLM+

9. Reinforcement Learning

- Oyun botları
- Robot kontrolü
- Self-driving car karar mekanizmaları

7. Ses İşleme

- Komut tanıma ("Hey Siri")
- Gürültü azaltma
- Konuşmacı tanıma

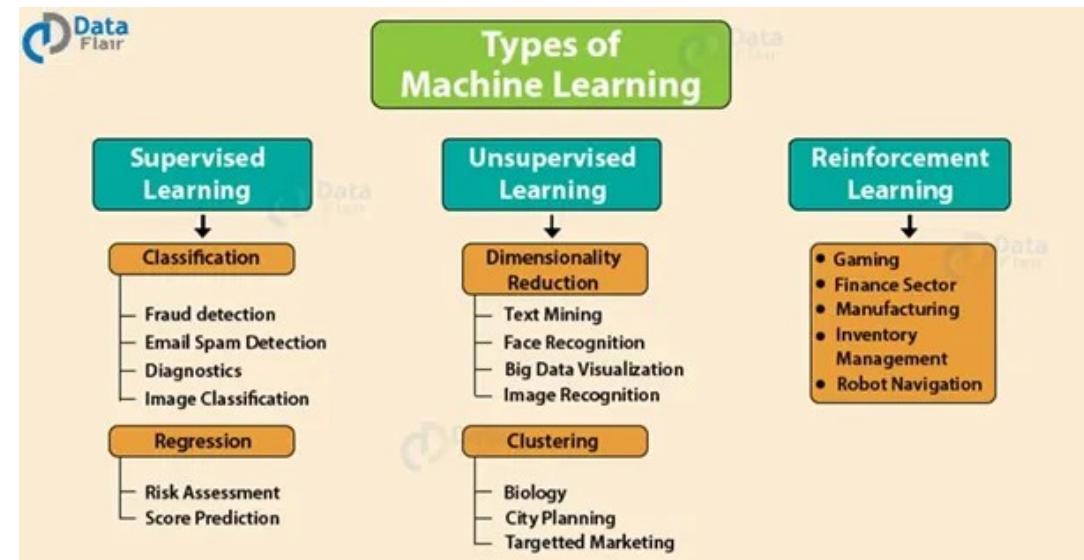
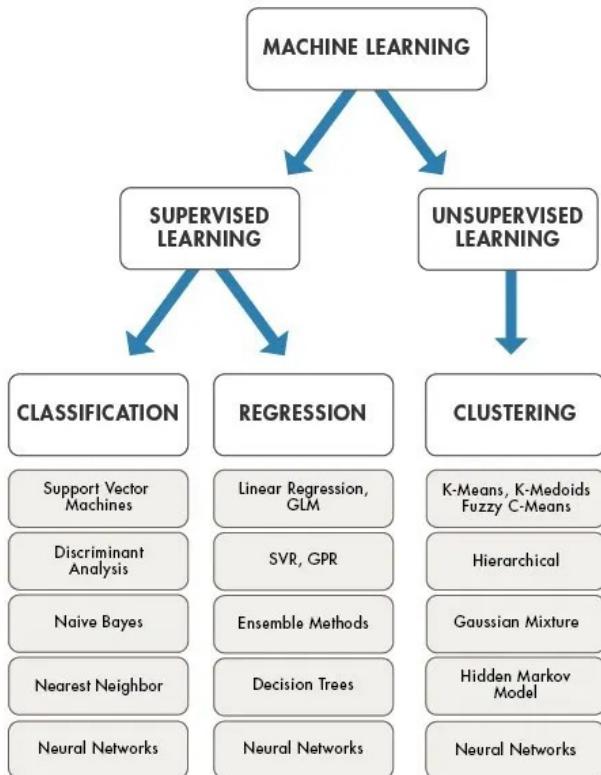
Bunların
nesi yapay
zeka?

8. Görüntü İşleme

- Yüz tanıma
- Nesne tanıma (self-driving araçlar dahil)
- Tıbbi görüntü analizi (MR/BT tanıları)



ML Türleri ve Farklı(*) Gösterimler



* Farklı: Hem “çeşitli” anlamında hem de “uyumsuz” anlamında. Zamana göre de değişebiliyor

ML Türleri ve Farklı Gösterimler – Devam (ChatGPT)



1. Supervised Learning

1.1. Regression (Sürekli çıktı) (Ör: Evin fiyatı kaç olmalı)

Linear Regression / Polynomial Regression

Ridge / Lasso / ElasticNet

Decision Tree Regression

Random Forest Regression

Gradient Boosting Regression (XGBoost, LightGBM, CatBoost)

SVR (Support Vector Regression)

k-NN Regression

Neural Network Regression

1.2.2. Multi-class Classification (Mutlu, Öfkeli, Nötr)

Softmax Regression

One-vs-Rest / One-vs-One modeller

Tree-based yöntemler

Deep Learning cnn'leri (ör: görüntü sınıflandırma)

1.2.1 ve 1.2.2- → İkisi de One-Label classificationdır

1.2.3. Multi-label Classification

Binary Relevance

Classifier Chains

ML-kNN

1.2. Classification (Sınıflandırma) (Kategorik çıktı)

1.2.1. Binary Classification (Ör: Spam vs Not Spam v.s)

Logistic Regression

SVM (Linear / RBF)

Decision Tree Classifier

Random Forest Classifier

Gradient Boosting Classifiers

k-NN Classifier

Naïve Bayes

Neural Networks

Linear Discriminant Analysis (LDA)

2. Unsupervised Learning

2.1. Clustering (Kümeleme)

2.2. Dimensionality Reduction (Boyut İndirgeme)

2.3. Association Rule Learning (İlişki Kuralları)

2.4. Density Estimation (Yoğunluk Tahmini)

2.5. Anomaly Detection (Anomali Tespiti – Etiketsiz)

Bilim ve Sanat el ele



ML'in Bilim (Science) Yönü

1 Modelin matematiği

- Lojistik regresyonda sigmoid fonksiyonunun neden olasılık verdiği
- Cross-entropy loss'un neden daha iyi optimize edildiği
- Random Forest'in neden overfit'e daha dayanıklı olduğu
- Gradient descent'in nasıl çalıştığı ve neden lokal minimuma takılabildiği

2 İstatistiksel temeller

- Bias-variance tradeoff
- Central limit theorem'in neden işe yaradığı
- Overfitting / Underfitting analizi
- Model performansının istatistiksel güvenilirliği

3 Deney tasarımı

- A/B testlerinin istatistiksel gücü
- Veri bölme stratejilerinin (train/test/validation) bilimsel gerekçesi
- Modelin genelleme kabiliyetinin ölçülmesi

4 Optimizasyon

- Learning rate planlaması
- Regularization teknikleri (L1/L2, dropout)
- Loss surface analizleri

ML'in Sanat (Art) Yönü

1 Feature engineering sanatı

- FE ile yeni featurelar oluşturmak
- "Bu kolon logarithmic scale olursa model kendine gelir" diyebilmek
- Kategorik veriyi nasıl encode etmenin gerektiğini sezmek
- "Bu feature işe yaramaz gibi görünüyor ama model için değerli" diyebilmek

2 Model seçme sezgisi

- "Bu problem için XGBoost daha iyi gider" hissiyatı
- "Bu veriye neural network çok fazla" diyebilmek

3 Veri temizleme ustalığı

- Veride gizli bir pattern'ı fark etmek
- Eksik veriyi doğru yöntemle doldurmak (impute etmek bir sanattır)
- Outlier'ın anlamlı mı yoksa gerçekten çöp mü olduğunu hissetmek

4 Hyperparameter tuning sezgisi

- Grid Search yerine Random Search tercih etmek
- "Bu modelde erken durdurma yapmazsam kaçar" hissi

5 Sonuçları yorumlama

- Feature importance'in hikâyesini çıkarmak
- İş tarafına anlatırken teknik detayı hikayeyleştirmek: "Modeliniz şöyle düşünüyor..."

6 GörSEL sezgi

- Confusion matrix'in nerede garip durduğunu sezmek
- ROC curve şekillerinden model karakteri çıkarmak

Algoritma Cehennemi





• Kaynaklar

- Hands-on Machine Learning with Scikit-Learn, Keras & Tensorflow, Aurelien Geron
- An Introduction to Statistical Learning, Gareth James, Daniela Witten , Trevor Hastie, Robert Tibshirani
- Günceli takip edin