

# ML Intro



# Akıl defteri



- MI çekici ama her problem çivi değil
  - Kural yazmanın zor olduğu problemler
    - If-else ile Netflix
    - Bazukayla sinek
  - Sihirli değnek? Her ML projesi başarılı olmak zorunda değil
- ML türleri (3)
- [ML yaklaşımlar \(4\)](#)
- [ML Yaşam Döngüsü ve Template](#)
- [Öğrenme ve Öğrenme problemleri](#)
- Train test ayrimı ve süreç önemli
- [Cross Validation, Öğrenme Problemlerinin çözümleri](#)
- Kirli Veri → Kötü Model
  - FE sanattır
  - FE'i iyi olan kazanır
- Veri Tipleri (Nümerik, Kategorik(nominal & ordinal))
- Beyin cimnastiği (Netflixtte Veri, feature, label?)
-

# Bias Variance (ve Noise)



- Overfitting ve Underfitting tespiti
- Overfitting ve Underfitting tespiti  
çözüm önerileri
- Ders sonrasında
- <https://medium.com/analytics-vidhya/three-steps-in-case-of-imbalanced-data-and-close-look-at-the-splitter-classes-8b73628a25e6>
-



# Splitting Dataset

- Terminolojik kargaşa
  - Train – test (test=holdout=validation)
  - Train - validation - test (test= holdout, 1-fold cv)
  - Train - test with cv (cv => validation, test=holdout)
  - Train – holdout
  - Mini Özeti:
    - Train → öğren
    - Validation / CV → karar ver / seçim yap
    - Test → hüküm ver
    - Nested CV → “hakem bile tarafsız olsun” yaklaşımı
    - (Holdout: Ayırma biçimi: test veya validation)
  - [machinelearningmastery tanımları](#)
- Which validation and when?
  - Small data: nested CV.
  - Medium data: nested CV if you can afford it. Use k-fold CV otherwise.
  - Big Data(and/or short deadline): Use a Single train-validation-test split (no CV at all). Or train-test (no validation either)
- CV: genelleme simülatörü
  - Aynı model + parametre → overfit mi?
  - Farklı model / parametre → hangisi daha iyi genelleniyor?