

# Örüntü Tanıma

Emir Öztürk

Oğuz Kırat

# Giriş

- Sınıflandırma
- Kümeleme
- Sınıflandırma – Kümeleme
- Sınıflandırma Algoritmaları
  - Karar ağaçları
  - K-NN
  - SVM

# Sınıflandırma

- Girdi verisine özelliklere göre etiket atama
- Öğrenme sırasında etiketli veri verilir
- Eğitim aşamasında model veri ile etiketler arasındaki ilişkiyi öğrenir
- Duygu analizi
- Paket analizi
- Medikal analiz

# Kümeleme

- Etiketsiz veriler ile çalışılır
- Verilerin tespitinden çok gruplandırılması gerçekleştirilir
- Denetimsiz öğrenmedir
- Müşteri gruplama
- Anomali tespiti
- Veri yönlendirme

# Sınıflandırma - Kümeleme

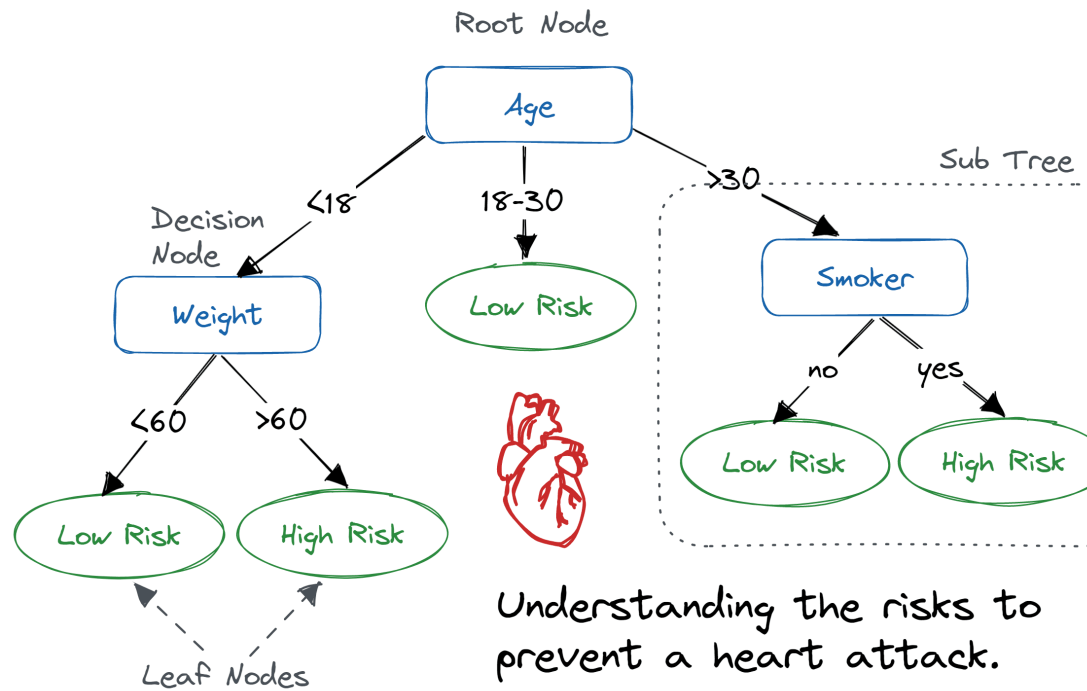
- Sınıflandırmada model hedef sınıfları bilir
- Kümelemede yakın olan küme tespiti yapılır
- Kümeleme sınıflandırmanın ön aşaması olarak kullanılabilir
  - Sonradan etiketleme için avantaj sağlar
  - Semisupervised learning
- Sınıflandırmadaki gibi doğru – yanlış ölçümü bulunmaz

# Sınıflandırma algoritmaları

- Karar ağaçları
- KNN
- SVM

# Karar Ağaçları

- Eldeki veriye göre bir koşul ağacı çıkartılır.
- Verinin belirli değerlerine göre yönlendirme yapılır.



# Karar Ağaçları

- ID3
- C4.5
- CART
- Random Forest



# ID3

- Ross Quinlan
- Yalnızca kategorik verilerde çalışır
- Entropi (ya da bilgi kazanımı) hesabı yapılır
- En düşük entropiye sahip değer karar düğümü olarak seçilir

## C4.5

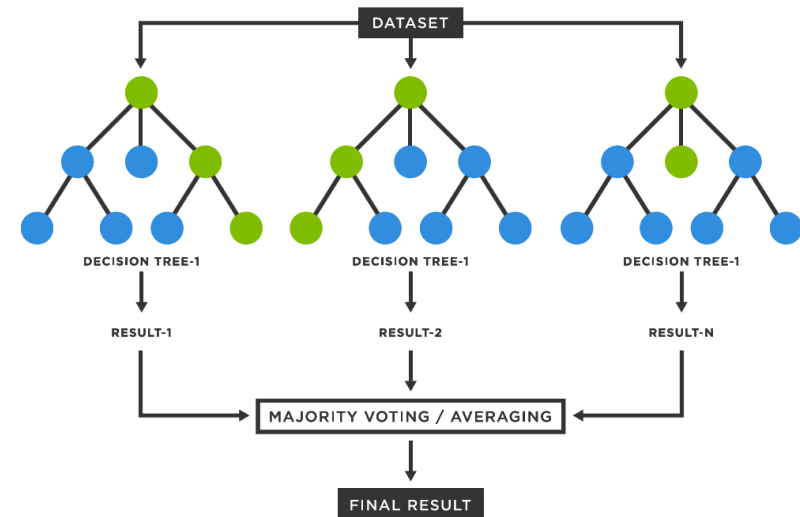
- Ross Quinlan
- Entropi hesabı yerine kazanımın oranı hesaplanır
- Bias olmasının engellenmesi amacı ile bu işlem seçilmiştir
- Sürekli verilerde verileri belirli noktalardan bölerek çalışabilir
- Olmayan değerleri olasılık dağılımları ile tespit eder
- Ağaç küçültme (pruning) ile overfit engellenir

# CART

- Breiman
- Classification and regression trees
- Düğümlerin bölünmesi için “Gini impurity” hesaplanır
- Tüm bölme ihtimalleri denenir. “Impurity” değerinin minimum olduğu bölünme seçilir
- Seçilen bir değer yanlıs sınıflandırılma değeridir

# Random Forest

- Birden fazla karar ağacının karar vermesi
- Eğitim verisetinin farklı parçaları ile farklı ağaçlar oluşturulur
- Test verisinde her ağaçtan gelen sonuç ortalaması alınır



# K-NN

- Lazy learner
  - Verisetinden öğrenilmiş bir model bulunmaz
  - Her istekte veriseti üzerinden yeniden hesaplama gerçekleştirilir
- Bir veri noktasının sınıflandırılması için basitçe k adet en benzer örneğe bakılır
- Benzerlik bir “uzaklık metriği” ile ölçülür
  - Öklit uzaklığı
  - Manhattan uzaklığı
  - Kosinüs benzerliği

# SVM

- N boyutlu bir uzayda optimum hyperplane'i bulmayı amaçlar
- Bu hyperplane oluşturulduğunda buna en yakın noktalar support vector olarak adlandırılır.
- Support vector'ler margin hesaplamada kullanılır.
- Margin'in maksimize edilmesi başarıyı arttırmak için amaçlanır
- Verilerin üst boyutlara çıkarılabilmesi için kernel'lar kullanılır