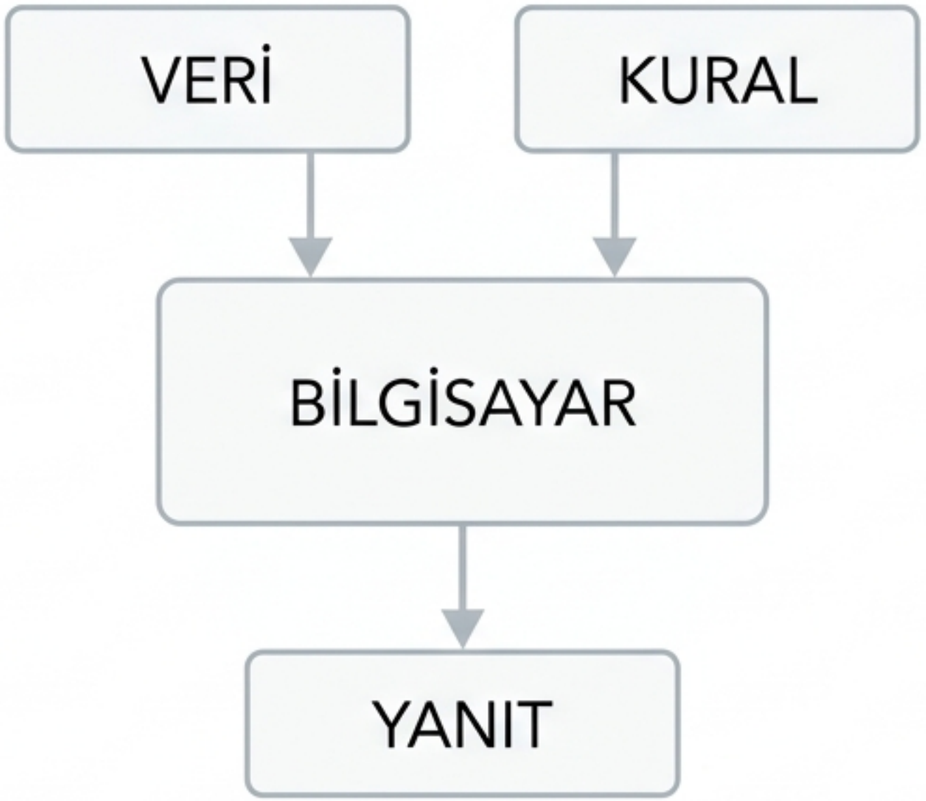


# Makine Öğrenmesi Temelleri ve Algoritmaları

Veriden kurala giden yol: Konseptlerden temel algoritmalara (Karar Ağaçları, KNN, Naive Bayes ve SVM) kapsamlı bir bakış.

Gemini Notebook

## GELENEKSEL PROGRAMLAMA




## MAKİNE ÖĞRENMESİ




Geleneksel sistemlerde kuralları biz yazarız; makine öğrenmesinde ise sistem, mevcut yanıtları ve verileri inceleyerek kuralları (matematiksel modeli) kendisi öğrenir.

Gemini Notebook

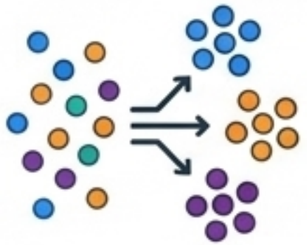



Gemini Notebook




## Gözetimli Öğrenme (Supervised)

Eğitim verileri doğru yanıtlarla etiketlenmiştir. Model, girdi ve hedef (çıktı) arasındaki ilişkiyi öğrenir.



## Gözetimsiz Öğrenme (Unsupervised)

Etiketsiz verilerdeki gizli kalıpları, benzerlikleri ve yapıları (kümeleri) kendi başına keşfeder.



## Pekiştirmeli Öğrenme (Reinforcement)

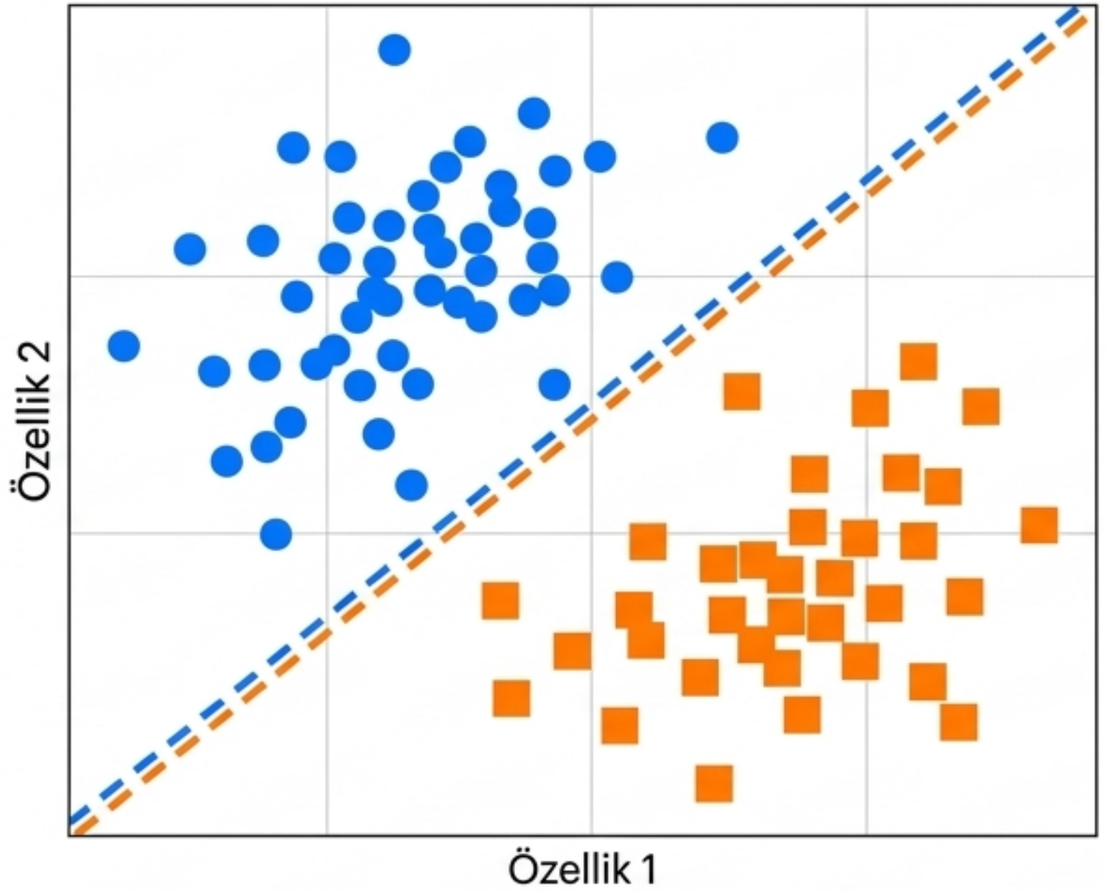
Bir ajanın ortamda deneme yanılma (ödül/ceza) yoluyla en iyi stratejiyi (politikayı) öğrenmesidir.



## Üretken Yapay Zeka (Generative)

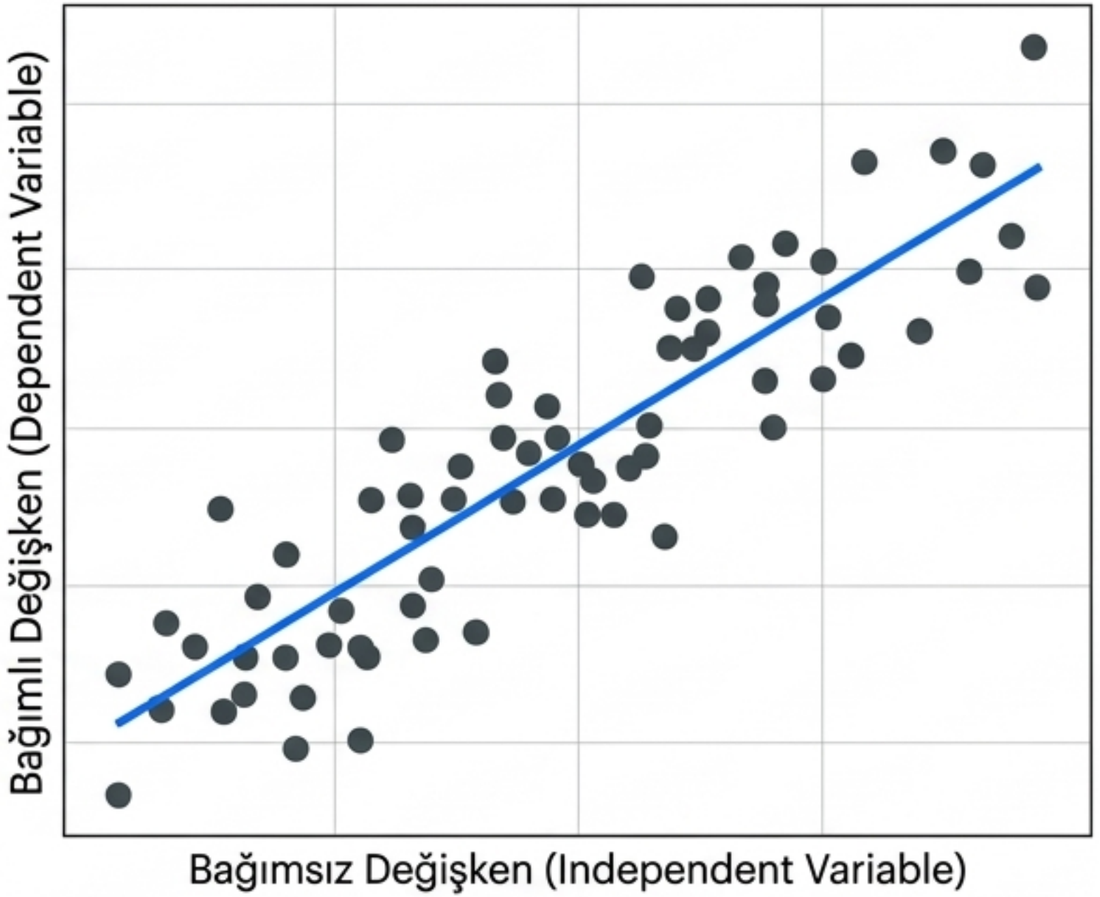
Giriş verilerinden (metin, resim, ses) öğrenerek tamamen yeni ve benzer içerikler üreten model sınıfıdır.

Gemini Notebook




## Sınıflandırma (Classification)

Bir öğenin belirli bir kategoriye veya sınıfa ait olma olasılığını tahmin eder (örn. E-posta Spam mi? Görüntü kedi mi köpek mi?). Ayrık/kategorik çıktılar üretir.




## Regresyon (Regression)

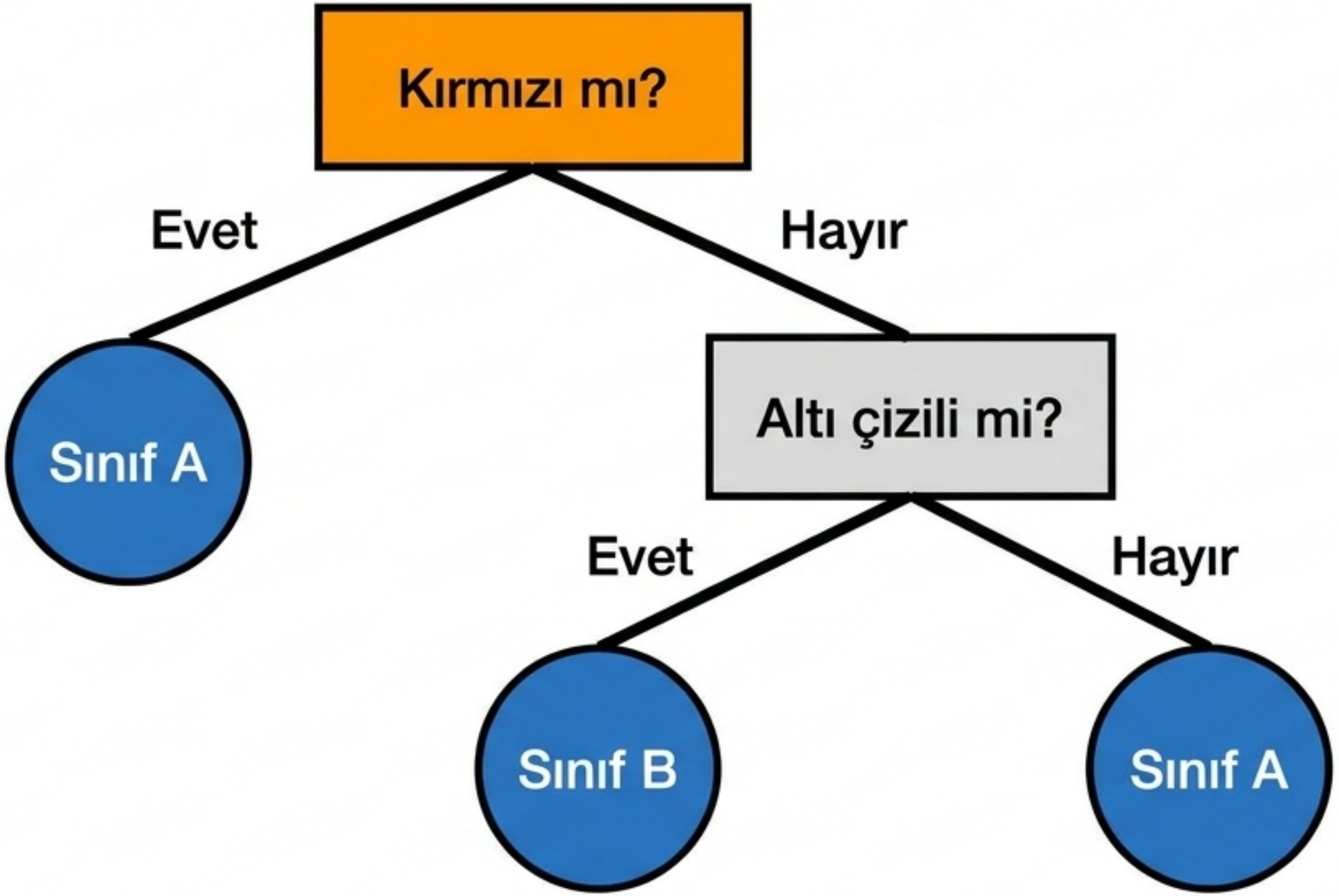
Sürekli ve sayısal değerleri tahmin eder (örn. Gelecekteki ev fiyatı, yağmur miktarı, yolculuk süresi). Nicel değişkenler arası ilişkileri belirler.

Gemini Notebook

# Makine Öğrenimi Çalışma Prensibi



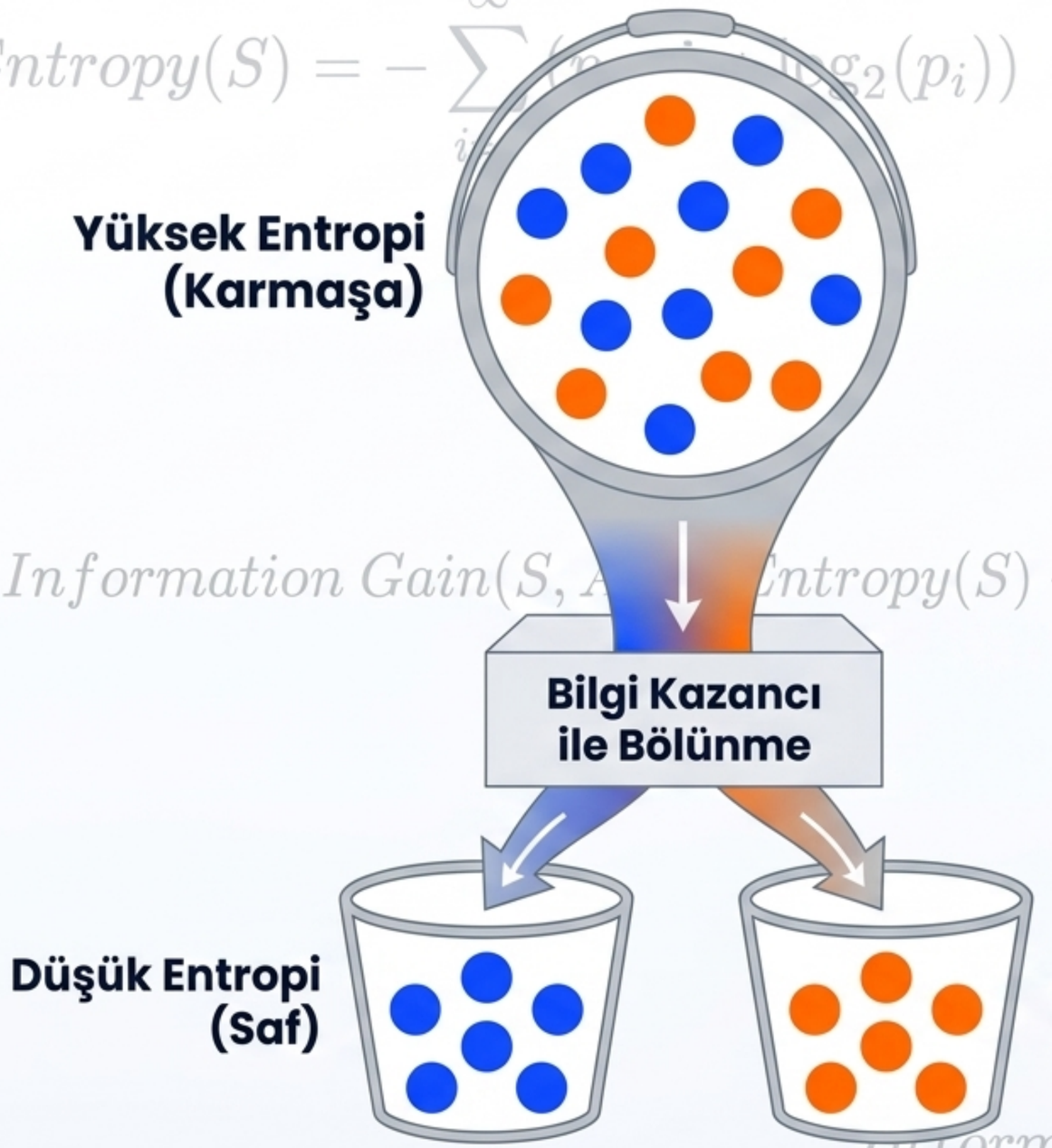

Gemini Notebook




## Karar Ağaçları (Decision Trees)

Veriyi belirli kurallara (özniteliklere) göre ardışık olarak bölen, hem sınıflandırma hem regresyon yapabilen, yorumlanması en kolay gözetimli öğrenme algoritmasıdır.

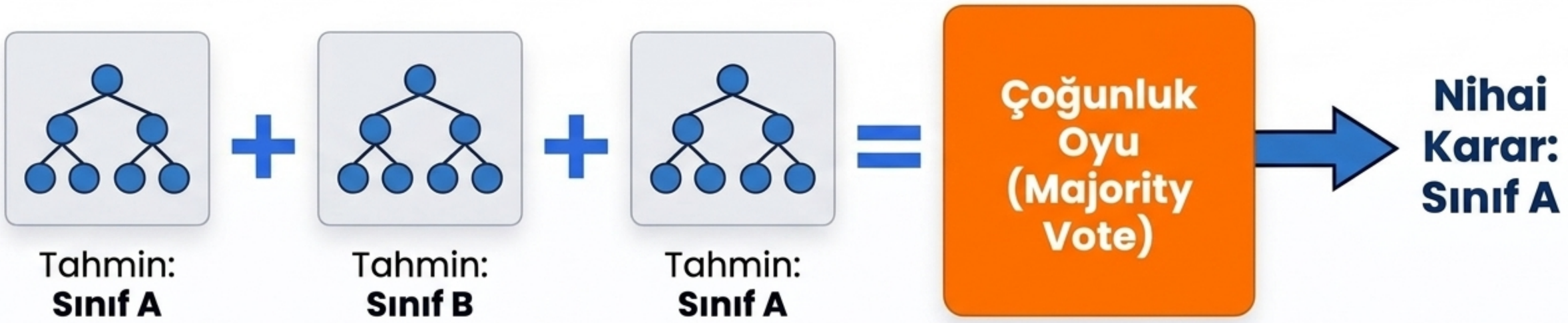
Gemini Notebook




**Entropi:**
Veri setindeki düzensizliğin veya **belirsizliğin** ölçüsüdür. Sistem, entropiyi sıfıra (tamamen saf gruplar) indirmeye çalışır.

**Bilgi Kazancı (Information Gain):**
Bir özelliğe göre bölünme yapıldığında entropideki azalma miktarıdır. Karar ağacı her adımda en yüksek bilgi kazancını sağlayan özelliği seçerek dallanır.

Gemini Notebook

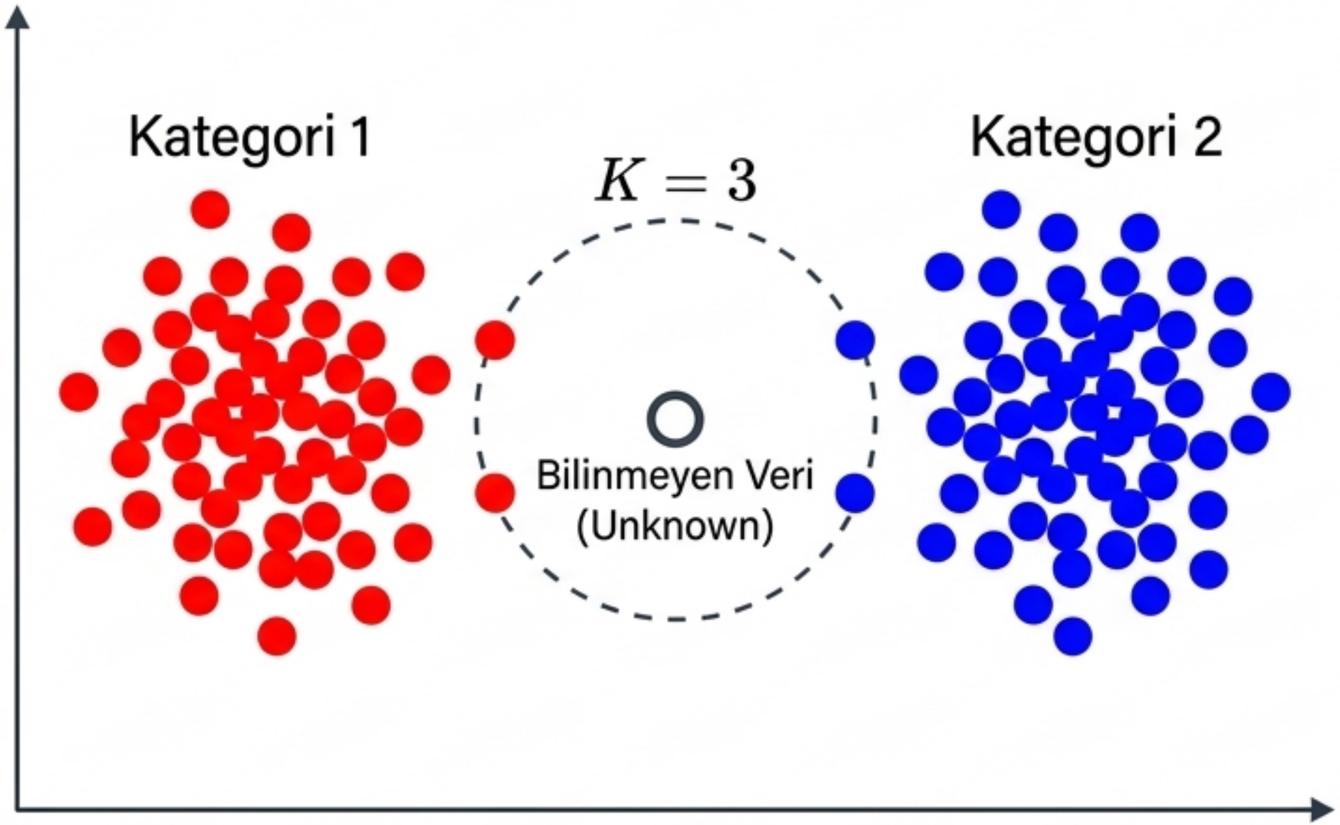



# Rassal Orman Modeli (Random Forest)

Tek bir karar ağacının ezberleme (overfitting) riskini ortadan kaldıran 'Topluluk (Ensemble)' modelidir.

**Torbalama (Bagging):** Algoritma, verinin farklı alt kümelerini ve rastgele özelliklerini kullanarak yüzlerce farklı karar ağacı oluşturur. Nihai karar, sınıflandırmada çoğunluk oyu, regresyonda ise ortalama alınarak verilir.

Gemini Notebook

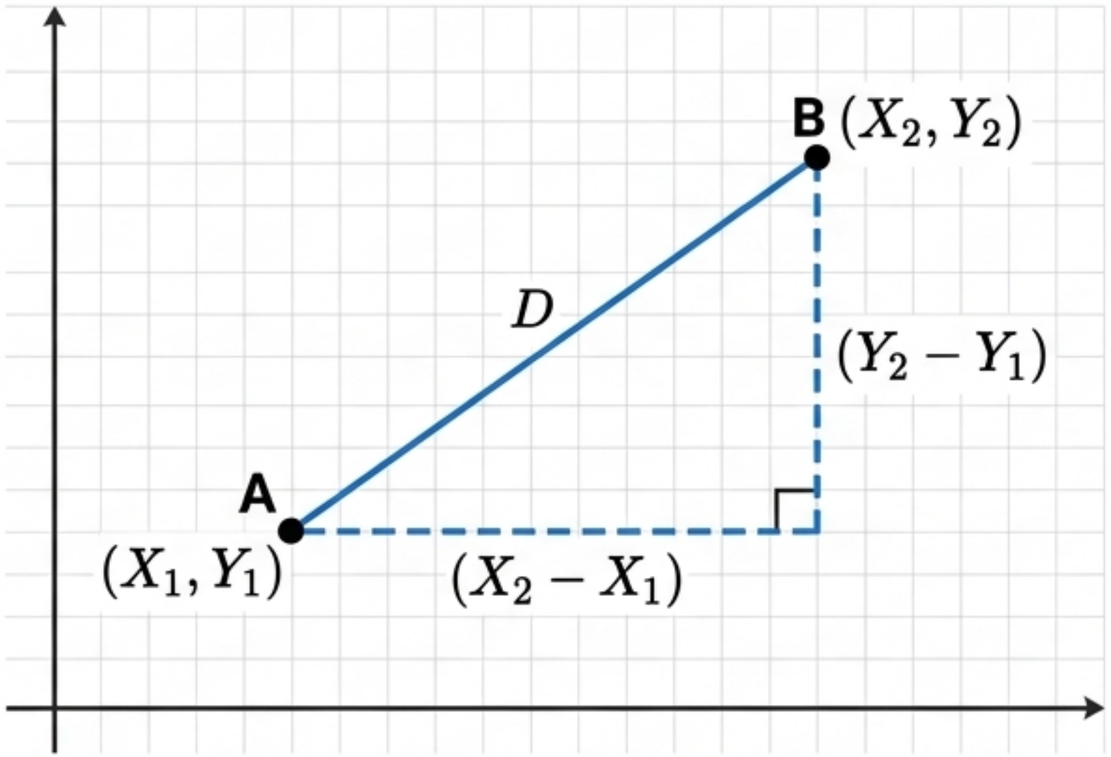



# K-En Yakın Komşu (K-Nearest Neighbors - KNN)

‘Bana arkadaşını söyle, sana kim olduğunu söyleyeyim.” KNN ‘tembel (lazy)’ bir algoritmadır; eğitim aşamasında bir model oluşturup kurallar çıkarmaz, sadece veriyi ezberler. Yeni bir veri geldiğinde, ona mesafe olarak en yakın K adet komşusuna bakar ve sınıflandırmayı yapar.

Gemini Notebook

# Uzaklık Ölçümü




$$D = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}.$$

Benzerlik, Öklid (Euclidean) veya Manhattan gibi uzaklık formülleriyle anında hesaplanır.

## $K$ Parametresinin Seçimi

- $K$ parametresi sonuca doğrudan etki eder.

- $K$ çok küçükse model gürültüye duyarlı olur (aşırı uyum/overfitting).

- $K$ çok büyükse ana prensip bozulur ve kararlar çarpıklaşır.

Gemini Notebook

$$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$$

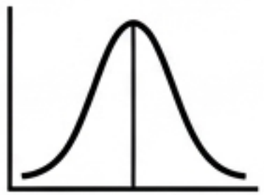
$P(Y|X)$ (Posterior): $X$ verisi görüldüğünde $Y$ sınıfının gerçekleşme olasılığı **(Asıl aradığımız sonuç).**
$P(X|Y)$ (Likelihood): $Y$ sınıfı verildiğinde $X$ verisinin gözlemlenme olasılığı.
$P(Y)$ (Prior): $Y$ sınıfının önceden bilinen yalın olasılığı.

Naive Bayes, geçmiş deneyimlere dayanarak gelecekteki olayların olasılığını hesaplayan matematiksel bir yaklaşımdır.

Gemini Notebook

## Neden ‘Saf’ (Naive)?

Algoritma, veri setindeki tüm özniteliklerin birbirinden tamamen bağımsız olduğunu varsayar (gerçekte nadiren böyle olsa da inanılmaz hızlı ve başarılı çalışır).



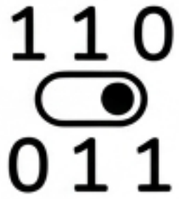
### Gaussian

Sürekli ve normal dağılım gösteren verilerde kullanılır.



### Multinomial

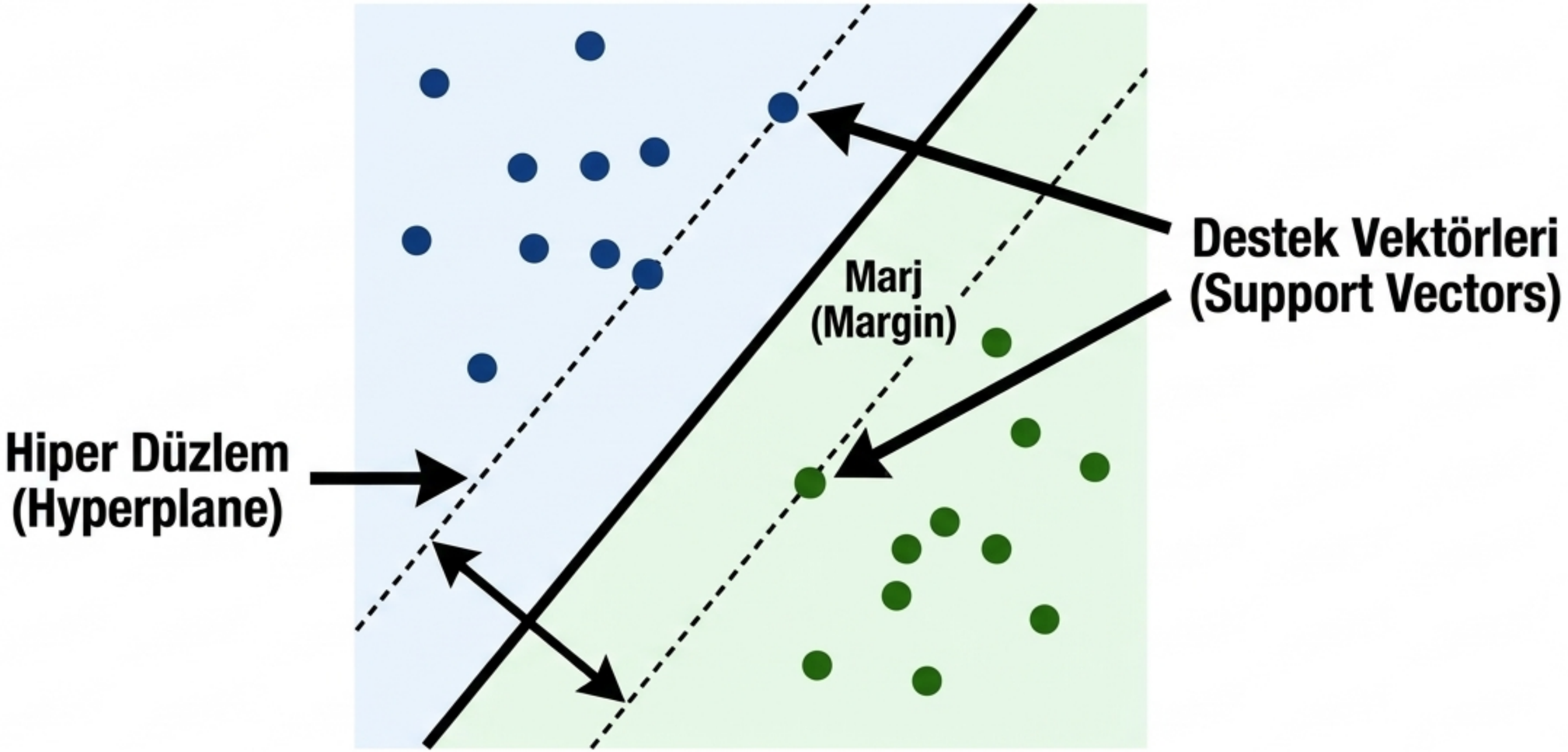
Metin sınıflandırma (Spam tespiti, duygu analizi) ve kelime frekansları gibi kesikli verilerde kullanılır.




### Bernoulli

İkili (var/yok) veriler için uygundur.

Gemini Notebook




# Destek Vektör Makineleri (SVM)

Sınıflar arasındaki en geniş ‘sokağı’ (marjı) çizmeyi hedefler. Sınır çizgisini **(hiper düzlem)** belirleyen şeyler, çizgiye en yakın olan veri noktalarıdır; bunlara ‘**destek vektörleri**’ denir.

Gemini Notebook

# Algoritma Karşılaştırma Matrisi

| Algoritma | Çalışma Mantığı | Avantajı | En İyi Kullanım Alanı |
|---|---|---|---|
| **Karar Ağacı** | Kural tabanlı dallanma | Yorumlanması insan mantığına en yakın olanıdır | Net kuralların çıkarılması gereken durumlar |
| **Random Forest** | Çoklu ağaç oylaması (Torbalama) | Yüksek doğruluk, ezberleme (overfitting) riski düşük | Karmaşık ve büyük veri setleri |
| **KNN** | Komşulara olan uzaklık (Tembel Öğrenme) | Model eğitimi gerektirmez, hemen uygulanır | Benzerliğe dayalı öneri sistemleri |
| **Naive Bayes** | Bağımsızlık varsayımı ile Olasılık | Çok hızlı çalışır ve az veri ile bile eğitilebilir | Metin sınıflandırma (Spam, Duygu analizi) ve çok sınıflı problemler |

Gemini Notebook