

## Makine Öğrenmesinde Entropi Ne İşe Yarar?

Entropi, bir veri kümesinin belirsizliğinin ölçüsü olarak tanımlanır. Bir sistemin entropisi ne kadar yüksekse, o sistem o kadar düzensizdir ve gelecekteki durumları hakkında bilgi edinmek o kadar zordur.

Makine Öğrenmesi ve özellikle karar ağaçları gibi algoritmalarda veri kümesinin bölünmesi ile elde edilen belirsizlik azalmasını yani veri kümesindeki belirsizliği ne kadar azalttığını ölçer.

Formülü: 
$$H(X) = - \sum p(x) \log_2 p(x)$$

## Entropi ve Bilgi Kazancının Kullanım Alanları:

### Makine Öğrenmesi ve Karar Ağaçları:

Örnek: Bir karar ağacı oluştururken, her düğümde hangi özelliğin kullanılacağına karar vermek için bilgi kazancı kullanılır. Özelliklerin entropisi hesaplanır ve en yüksek bilgi kazancına sahip olan özellik seçilir. Bu süreç, ağacın her bir bölünmesinde belirsizliği en çok azaltan özelliğin seçilmesini sağlar.

Örnek Senaryo: Bir müşteri veri tabanını kullanarak, müşterilerin belirli bir ürünü satın alıp almayacağını tahmin etmek istiyoruz. Veri kümesinde yaş, gelir, eğitim seviyesi gibi özellikler bulunuyor. Karar ağacı algoritması, her düğümde bu özelliklerin bilgi kazancını hesaplayarak, müşterilerin satın alma kararını en iyi açıklayan özelliği seçer ve ağacı buna göre dallandırır.

### Doğal Dil İşleme (NLP):

Örnek: Metin sınıflandırma veya duygu analizi gibi görevlerde, kelimelerin bilgi kazancını hesaplayarak hangi kelimelerin daha ayırt edici olduğunu belirlemek için kullanılır.

Örnek Senaryo: Bir film inceleme veri setinde, incelemelerin olumlu veya olumsuz olduğunu belirlemek istiyoruz. Bu incelemelerde geçen kelimelerin entropisi ve bilgi kazancı hesaplanarak, incelemelerin sınıflandırılmasında en etkili olan kelimeler belirlenir.

### Genetik ve Biyoinformatik:

Örnek: Genetik veri analizinde, belirli genlerin veya genetik işaretçilerin hastalıklarla ilişkisini belirlemek için bilgi kazancı kullanılabilir.

Örnek Senaryo: Bir genetik araştırmada, belirli genlerin bir hastalıkla ilişkisini incelemek için, her genin entropisi hesaplanır ve hastalıkla en yüksek bilgi kazancına sahip olan genler belirlenir. Bu, araştırmacıların hastalıkla ilişkili genleri daha hızlı ve doğru bir şekilde tespit etmelerini sağlar.

#### Görüntü İşleme:

Örnek: Görüntü sınıflandırma ve nesne tanıma gibi görevlerde, piksellerin veya özelliklerin entropisi ve bilgi kazancı kullanılarak en ayırt edici özellikler belirlenir.

Örnek Senaryo: Bir görüntü veri setinde, belirli nesneleri tanımak için, görüntülerdeki farklı piksellerin veya bölgelerin bilgi kazancı hesaplanır. Bu, nesneleri tanımak için en bilgilendirici özelliklerin belirlenmesine yardımcı olur.

#### Finansal Analiz:

Örnek: Piyasa analizlerinde, belirli göstergelerin (örneğin, hisse senedi fiyatları, işlem hacmi) bilgi kazancı hesaplanarak, hangi göstergelerin piyasa hareketlerini en iyi açıkladığı belirlenir.

Örnek Senaryo: Bir hisse senedi fiyat tahmin modelinde, farklı ekonomik göstergelerin (örneğin, faiz oranları, enflasyon oranları) bilgi kazancı hesaplanarak, hisse senedi fiyatlarını en doğru şekilde tahmin eden göstergeler seçilir.