ID3 Algoritması

* Entropi değerleri hesaplayıp oluşan entropi sonucu elde edilen bilgi kazançları hesaplayıp karar ağacı oluşturmayı amaçlayan bir algoritma türüdür.

Karar Ağacı Nedir?

* Veriyi sınıflandırmak veya tahmin yapmak için kullanılan, bir dizi karar kuralına dayalı yapısal bir modeldir. Karar ağacında önemli olan terimler:
* **Düğümler (Nodes):** Karar ağacının herhangi bir seviyesinde yer alan ve bir testin yapıldığı noktadır. Bir ağacın içinde, her iç düğüm (yani kök düğüm hariç) veri kümesini belirli bir özelliğe göre böler.
* **Dallar (Branches):** Bir düğümdeki testin sonucunu temsil eder ve test sonucuna göre yönlendiren bir bağlantıdır.

“Yaş> 30 mu?” bir düğümü temsil ederken ‘Evet’ ve ‘Hayır’ gibi cevaplar da dalları temsil eder.

* **Kök Düğüm (Root Node):** Karar ağacının en üst düzeyindeki düğümdür. Burada, veriye ait en önemli özellik veya kriter test edilir.

Entropi Nedir?

* Bir veri kümesinin karışıklığını veya düzensizliğini ölçen bir kavramdır. Temelde veri kümesinin ne kadar "belirsiz" veya "düzensiz" olduğunu gösterir. Yani verideki rastgeleliğini ölçer. Eğer bir sınıfa ait ise sınıf içindeki düzensizlik azalırsa veri daha düzenli olur ve entropi düşük olur.  
  + **Düşük Entropi**: Veri daha düzenlidir ve belirli sınıflara ayrılmak daha kolaydır. Örneğin, veri kümesinde çoğunlukla bir sınıf varsa, o zaman entropi düşer.
  + **Yüksek Entropi**: Veri daha dağılmış ve belirsizdir. Veri kümesindeki sınıflar eşit dağıldığında entropi yüksek olur, çünkü sınıflar arasındaki ayrım net değildir.

metin, yazı tipi, beyaz, tipografi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**(Pi) = Durumun olasılığı**

**Bu olasılığı hem 2 tabanlı log içinde de değeri bu olasılık ile çarpılır.**

Uygulama Üzerinden Örnek

metin, ekran görüntüsü, sayı, numara, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

[1] Hava Durumuna Göre Futbol Oynayıp Oynayamama Durumu

* Bu veri setinde “Play” sütunu karar sütunudur.
* Bu sütuna göre ana entropiyi hesaplamalıyız.
  + Veri setinde: 5 adet hayır + 9 adet evet

Buna göre hesaplanmasında :

yazı tipi, çizgi, beyaz, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

Oyun Entropisi = 0.940285

* Ben uygulamanın devamında 0.9402 olarak kabul edeceğim. –
* Şimdi diğer sütunlara göre entropi hesabı yapılır.  
  + Outlook: - Hava Durumu-
    - Sunny -Güneşli- :
      * 5 adet veri güneşlidir. Bu verilerin üçü ‘Hayır’, ikisi ‘Evet’ olduğu için entropi hesabı yapılır.
      * yazı tipi, beyaz, çizgi, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * E(Sunny|Outlook) = 0.9709
    - Overcast -Bulutlu-:
      * 4 adet veri bulutludur. Bu verilerin dördü de ‘Evet’ olduğu için entropi hesaplanır.
      * yazı tipi, beyaz, diyagram, çizgi içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * E(Overcast|Outlook) = 0
      * Bu formüle döküldüğünde sonuç tanımsız çıkacağı için iki adet kabul edeceğimiz durum vardır.

! Eğer tüm üyeler aynı sınıf üyesi ise Entropi SIFIR.

! Eğer eşit dağılımda sınıf üyeleri var ise Entropi BİR.

* + - Rainy -Yağmurlu-:
      * 5 adet veri yağmurludur. Bu verilerin 3 adeti ‘Evet’ ikisi de ‘Hayır’dır.

yazı tipi, beyaz, çizgi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

* + - * E(Rainy|Outlook) = 0.9709

Elimizde tüm veri sınıflarına göre entropileri hesaplandıktan sonra bu kategori için bilgi kazancı hesaplanır.

Entropi(Outlook) = yazı tipi, metin, beyaz, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu = 0.6935

~ 14 adet veriden beşi güneşli, dördü bulutlu, beşi yağmurlu ve bunların entropileri ile olasılıklarının çarpımı ile bulunuyor.

Entropi(Play) – Entropi (Outlook) = 0.9402 – 0.6935

**Outlook – Hava Durumu için Bilgi Kazancı = 0.2467**

* + Temperature-Isı-
    - Hot-Sıcak-:
      * 4 adet veri sıcaktır. İkisi ‘Hayır’ ve ikisi ‘Evet’ sınıfına üyedir.
      * Eşit sayıda sınıfa karşılık geldiği için entropisi BİR’dir.
    - Mild-Ilık-:
      * 6 adet veri ılık sınıfına aittir. Bu sınıf üyelerin ikisi ‘Hayır’ dördü ‘Evet’ karar sınıfı üyesidir.
      * yazı tipi, çizgi, beyaz, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * Entropi(Mild|Temperature) = 0.9182
    - Cold-Soğuk-:
      * 4 adet veri soğuk sınıfındadır. Biri ‘Hayır’ , üçü ‘Evet’ sınıfına karşılık gelmiştir.
      * yazı tipi, beyaz, çizgi, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * Entropi(Cold|Temperature) = 0.8112

Tüm sınıflara göre entropileri hesaplandıktan sonra bilgi kazancı hesaplanabilir.

Entropi (Temperature) = yazı tipi, metin, beyaz, tipografi içeren bir resim

Açıklama otomatik olarak oluşturuldu = 0.911

Entropi (Play)- Entropi (Temperature) = 0.9402 – 0.911 = 0.0292

**Temperature – Sıcaklık – için Bilgi Kazancı = 0.0292**

* + Humidity – Nem –:
    - High-Yüksek-:
      * 7 adet veri nemi yüksek olarak işaretlenmiştir. Dördü ‘Hayır’, üçü ‘Evet’ durumundadır.
      * yazı tipi, çizgi, beyaz, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * Entropi(High|Humidty) = 0.9852
    - Normal-Normal-:
      * 7 adet veri nemi normaldir. Biri ‘Hayır’ iken altısı ‘Evet’tir.
      * yazı tipi, beyaz, çizgi, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * Entropi(Normal|Humidty) = 0.5916

Tüm sınıflara göre entropileri hesaplandıktan sonra bilgi kazancı hesaplanabilir.

Entropi (Humidity) = yazı tipi, metin, beyaz, hat sanatı, kaligrafi içeren bir resim

Açıklama otomatik olarak oluşturuldu = 0.7884

Entropi (Play) – Entropi (Humidty) = 0.9402 – 0.7884 =0.1518

**Humidty – Nem- için Bilgi Kazancı = 0.1518**

* + Windy – Rüzgâr Durumu-:
    - False-Yok-:
      * 8 adet veri rüzgârsız durumdur. İkisi ‘Hayır’, altısı ‘Evet’ sınıfına üyedir.
      * yazı tipi, beyaz, çizgi, diyagram içeren bir resim

        Açıklama otomatik olarak oluşturuldu
      * Entropi (False|Windy) = 0.8112
    - True-Var-:
      * 6 adet veri rüzgarlı durumu karşılar. Üçü ‘Hayır’, üçü ‘Evet’tir.
      * Eşit üyelere sahip sınıflara üyeler olduğu için Entropi Birdir.

Tüm sınıflara göre entropileri hesaplandıktan sonra bilgi kazancı hesaplanabilir.

Entropi (Windy) = yazı tipi, metin, sayı, numara, beyaz içeren bir resim

Açıklama otomatik olarak oluşturuldu = 0.8921

Entropi (Play) – Entropi (Windy) = 0.9402 – 0.8921 = 0.048

**Windy-Rüzgar- için Bilgi Kazancı = 0.048**

**Tüm bilgi kazançlar hesaplandıktan sonra en büyük bilgi kazancına sahip olan sütun düğüm olur.**

**Hava = 0.2467**

**Sıcaklık = 0.0292**

**Nem = 0.1518**

**Rüzgâr = 0.048**

**Bu bilgi kazançlarına göre ilk düğüm ‘Hava’ sütunudur.**

Buna göre oluşan karar ağacının ilk düğüm hava durumudur. Bu çıkan sonuca göre veriler ayrılır.

Rainy

YES

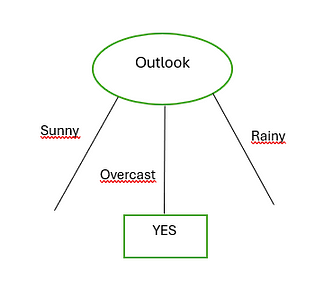
Overcast

Sunny



Overcast - Rainy - Sunny durumlarına göre veriler ayrıştırılır.

Eğer oluşan veri kümelerinde karar sınıfı eğer aynı üyeye sahip ise tekrar bir işlem yapılmadan o sonuca göre karar sınıfı üyeliği verilebilir.



Karar sınıflarında dağılım olan “rainy” ve “sunny” tabloları için önce karar sınıfı entropisi hesaplanır, ardından da tekrar seçilmeyen sütunlar için entropiler hesaplanıp bilgi kazançları bulunur. Bulunan bilgi kazancının en yüksek değerine göre sıradaki düğüm her dal için belirlenir. Farklı olmak zorunda değil aynı düğüme de gidebilirler.