

Entropi = Bilgi transfer oranının logaritmik ölçüsüdür.

= ID3 Algoritması için Entropi Hesabı =

Gain (kazanım)

$$\text{Gain}(x_i) = H(Y) - \sum_{j=1}^{k_i} P(x_{ij}) H(Y/x_{ij}) \quad i=1,2,3 \dots$$

$\text{Gain}(x_i) \rightarrow i$. bağımsız değişkenin katsayı değeri

$H(Y) \rightarrow$ Bağımlı değişkene ait entropi değeri

$k_i \rightarrow i$. bağımsız değişkenin katsayı (düzey/sınıf) sayısı

$P(x_{ij}) \rightarrow i$. bağımsız değişkenin j . durumunun ortaya çıkma olasılığı

$m \rightarrow$ bağımsız değişken sayısı

§ Kazanım değeri büyük olan bağımsız değişken (sınıflık) bağımlı değişken üzerinde en fazla belirleyiciliğe sahiptir ve dallandırma için seçilir.

§ Y değişkeninin k ^{tem} düzeyi için entropi değeri aşağıdaki gibi hesaplanır.

$$H(Y) = H(P_1, P_2, P_3 \dots P_n) = - \sum_{j=1}^k (P_j \log_2 (1/P_j))$$

$P_j \rightarrow j$. düzeyin ortaya çıkma olasılığı

$$P_1 = \frac{9 \rightarrow \text{yes sayısı}}{14 \rightarrow \text{mcg sayısı}} \quad \frac{5 \rightarrow \text{no sayısı}}{14 \rightarrow \text{mcg sayısı}}$$

$$H(Y) = H\left(\frac{9}{14}, \frac{5}{14}\right) = \frac{9}{14} \log_2 \frac{14}{9} + \frac{5}{14} \log_2 \frac{14}{5} = 0,410 + 0,530 = 0,940$$

Genel entropi

§ Hava, sıcaklık, nem ve rüzgar için entropiler aşağıdaki gibi hesaplanır.

$H_{\text{Hava}}(x_1)$ değişkeni için (Güneşli Bulutlu ve Yağmurlu o.ö 3 durum var)

$$G(x_1) = H(Y) = - \sum_{j=1}^k P(x_{ij}) H(Y/x_{ij})$$

$$H(Y/x_{11}) = H\left(\frac{2}{5}, \frac{3}{5}\right) = \frac{2}{5} \log_2 \frac{5}{2} + \frac{3}{5} \log_2 \frac{5}{3} = 0,971 \quad x_{11} \text{ Hava Güneşli}$$

$$H(Y/x_{12}) = H\left(\frac{4}{4}, \frac{0}{4}\right) = \frac{4}{4} \log_2 \frac{4}{4} + 0 = 0 \quad x_{12} \text{ Hava Bulutlu}$$

↳ toplam bulutlu gün

$$H(Y/X_{13}) = H\left(\frac{3}{5} \cdot \frac{2}{5}\right) = \frac{3}{5} \log_2 \frac{5}{3} + \frac{2}{5} \log_2 \frac{5}{2} = 0,971 \quad X_{13} \text{ Hava Yağmurlu}$$

Hava değişkeni için kazanç

$$\text{Gain}(X_1) = 0,940 = \frac{\overset{\text{Toplam güreşli}}{5}}{\underset{\text{Toplam maç}}{14}} (0,971) + \frac{4}{14} (0) + \frac{5}{14} (0,971) = 0,940 - 0,694 = 0,246 \quad \text{Hava}$$

= Gain (Kazanım) Soruuları =

$$X_1 (\text{Hava}) \rightarrow 0,246$$

$$X_2 (\text{Sıcaklık}) \rightarrow 0,029$$

$$X_3 (\text{Mem}) \rightarrow 0,152$$

$$X_4 (\text{Rüzgar}) \rightarrow 0,048$$

Herbir özdeşliğin çıktı üzerindeki kazanımı hesaplandıktan sonra en yüksek kazanım değerine sahip özdeşlik "Kök Düzümü" (Ağacın gövdesi) olur. Buna "En bilgilendirici özdeşlik" denir.

Ağacın kalan dalları gövde içerisinde kalan özdeşlikler için ayrı ayrı entropi hesabı yapılmaya başlanır.

Dallanma artık iletilenmediği zaman yaprak bulunmuş olur; ve bu duruma "Saflik" (Pure) adı verilir.

Bir karar ağacı temel olarak dallar ile birbirine bağlanan kök düğüm, iç düğüm ve yaprak düğümlerinden oluşur.

= ID3 Algoritmasında Ağac Oluşumu =

- 1- Genel entropi değeri hesaplanır.
 - 2- Herbir bağımsız değişken için kazanç değeri hesaplanır.
 - 3- En büyük kazanım değerine sahip bağımsız değişkenin (X) düzeylerine göre dallanmayı gerçekleştirir.
 - 4- Yeni oluşan ve "Saf" olmayan düğümlerin herbiri için veriyi indirgeyerek Adım 1'e dön, tüm düğümler "Saf" olduğunda dur.
- § ID3 Algoritmasını kullanabilmek için veriler kategorik olmalıdır.
- § Gain birimi "bits" dir.