

Nama : Muhammad Alif al Husain

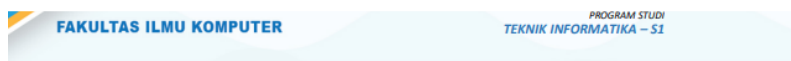
NIM : A11.2022.14718

Hitung Entropy Entropy Dataset Diabetes.

Terdapat dua class di kolom Outcome

- 1 = (Positif Diabetes)
- 0 = (Negatif Diabetes)

Rumus Entropy



Konsep Entropy [2]

- Untuk perhitungan nilai Entropy sbb :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

- Keterangan :
 - S : himpunan kasus.
 - A : fitur.
 - n : jumlah partisi S.
 - p_i : proporsi dari S_i terhadap S

Disini saya menggunakan sample 19 data dari dataset Diabetes.csv saya yang dimana :

Output 1 = 15 data

Output 0 = 14 data

Dengan Menggunakan Rumus Konsep Entropy didapatkan $p_1 = 5/19$ dan $p_0 = 14/19$

Maka Entropy dataset Utama 0.831.

Konsep Gain [2]

- Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
- Untuk menghitung *gain* digunakan rumus :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

- Keterangan :
 - S : himpunan kasus
 - A : atribut
 - n : jumlah partisi atribut A
 - $|S_i|$: jumlah kasus pada partisi ke-i
 - $|S|$: jumlah kasus dalam S

Maka didapatkan nilai Gain dari attribute pregnancies adalah 0.286.