Repo 2

Muhammad Bayu Chandra Adi

A11.2022.14666

Data Mining A11.4506

1. Hitung Entropy dan Gain serta tentukan pohon keputusan yang terbentuk dari contoh kasus keputusan bermain tenis dibawah ini :

OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
Sunny	Hot	High	No	Don't Play
Sunny	Hot	High	Yes	Don't Play
Cloudy	Hot	High	No	Play
Rainy	Mild	High	No	Play
Rainy	Cool	Normal	No	Play
Rainy	Cool	Normal	Yes	Play
Cloudy	Cool	Normal	Yes	Play
Sunny	Mild	High	No	Don't Play
Sunny	Cool	Normal	No	Play
Rainy	Mild	Normal	No	Play
Sunny	Mild	Normal	Yes	Play
Cloudy	Mild	High	Yes	Play
Cloudy	Hot	Normal	No	Play
Rainy	Mild	High	Yes	Don't Play

36

1. Perhitungan Entropy Total

Jumlah data = 14

Play = 9

Don't Play = 5

Entropy(S) = - (9/14) log2(9/14) - (5/14) log2(5/14)

Entropy(S) ≈ 0.940

2. Perhitungan Entropy dan Gain Atribut Outlook

- Sunny: 5 data (2 Play, 3 Don't Play)

Entropy(Sunny) = - (2/5) log2(2/5) - (3/5) log2(3/5) ≈ 0.971

- Cloudy: 4 data (4 Play, 0 Don't Play)

Entropy(Cloudy) = 0

- Rainy: 5 data (3 Play, 2 Don't Play)

Entropy(Rainy) = - (3/5) log2(3/5) - (2/5) log2(2/5) ≈ 0.971

Gain(Outlook) = 0.940 - [(5/14×0.971) + (4/14×0) + (5/14×0.971)]

Gain(Outlook) ≈ 0.246

3. Tentukan Pohon Keputusan

Dari Root **Outlook**, kita memiliki 3 cabang:

1. **Outlook = Cloudy**
    - Cek data: Ada 4 data, semuanya **Play**.
    - **Keputusan:** Langsung menjadi *Leaf Node* **Play**.
2. **Outlook = Sunny**
    - Ada 5 data (Campuran: 2 Play, 3 Don't Play). Kita lihat subset datanya:
        - Sunny, Hot, **High**, No -> Don't Play
        - Sunny, Hot, **High**, Yes -> Don't Play
        - Sunny, Mild, **High**, No -> Don't Play
        - Sunny, Cool, **Normal**, No -> Play
        - Sunny, Mild, **Normal**, Yes -> Play
    - Perhatikan pola atribut **Humidity**:
        - Jika Humidity = **High** -> Semuanya **Don't Play**.
        - Jika Humidity = **Normal** -> Semuanya **Play**.
    - **Keputusan:** Cabang Sunny dipecah oleh atribut **Humidity**.
3. **Outlook = Rainy**
    - Ada 5 data (Campuran: 3 Play, 2 Don't Play). Kita lihat subset datanya:
        - Rainy, Mild, High, **No** -> Play
        - Rainy, Cool, Normal, **No** -> Play
        - Rainy, Cool, Normal, **Yes** -> Play (Catatan: Baris ke-6 di gambar Anda adalah 'Play', di dataset standar seringkali 'No', tapi kita ikuti gambar).
        - Rainy, Mild, Normal, **No** -> Play
        - Rainy, Mild, High, **Yes** -> Don't Play
    - Analisis:
        - Jika **Windy = No** (3 data) -> Semuanya **Play**.
        - Jika **Windy = Yes** (2 data: 1 Play, 1 Don't Play) -> Masih campuran.
    - Biasanya untuk kasus ini, cabang Rainy dipecah oleh atribut **Windy** (karena memisahkan 'No' secara murni).

Gambar Pohon Keputusan