Nama : Margareta Valencia

NIM : A11.2022.14704

Kelompok : A11.4509

**SOAL KUIS 1 PERTEMUAN 7**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **NO** | **OUTLOOK** | **TEMPERATURE** | **HUMIDITY** | **WINDY** | **PLAY** |
| 1 | Sunny | Hot | High | No | Don't Play |
| 2 | Sunny | Hot | High | Yes | Don't Play |
| 3 | Cloudy | Hot | High | No | Play |
| 4 | Rainy | Mild | High | No | Play |
| 5 | Rainy | Cool | Normal | No | Play |
| 6 | Rainy | Cool | Normal | Yes | Play |
| 7 | Cloudy | Cool | Normal | Yes | Play |
| 8 | Sunny | Mild | High | No | Don't Play |
| 9 | Sunny | Cool | Normal | No | Play |
| 10 | Rainy | Mild | Normal | No | Play |
| 11 | Sunny | Mild | Normal | Yes | Play |
| 12 | Cloudy | Mild | High | Yes | Play |
| 13 | Cloudy | Hot | Normal | No | Play |
| 14 | Rainy | Mild | High | Yes | Don't Play |

**JAWAB**

Rumus

1. Entropy
2. Gain

**Langkah 1**

Menghitung jumlah kasus, jumlah kasus untuk keputusan **Yes** dan **No**, serta Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut **OUTLOOK, TEMPERATURE, HUMIDITY, dan WINDY.**

Detail perhitungan :

Entropy(Total)Gain (Total, Outlook)

1. Entropy (cloudy)
2. Entropy (rainy)
3. Entropy (sunny)

Gain (Total, Temp)

1. Entropy (cool)
2. Entropy (hot)
3. Entropy (mild)

Gain(Total, Humidity)

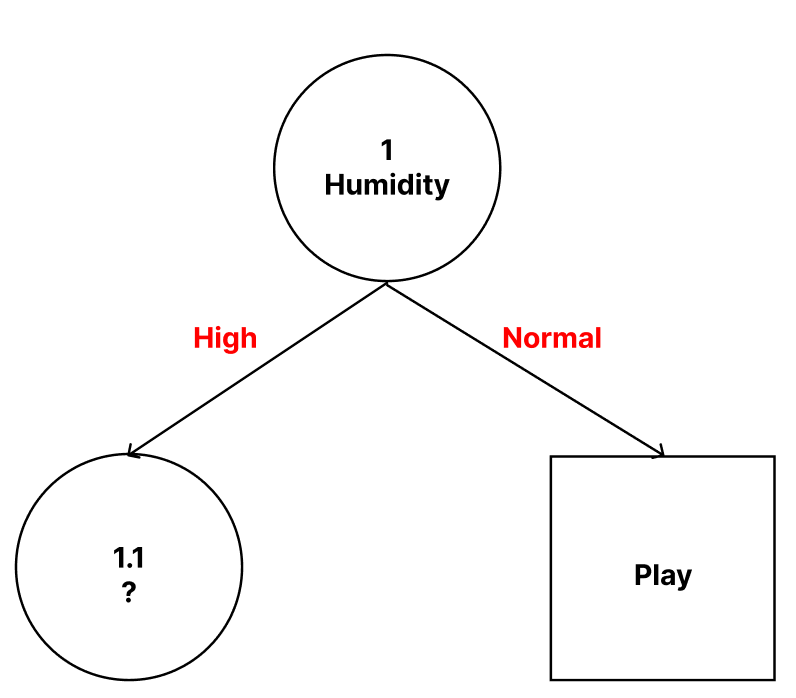
1. Entropy (high)
2. Entropy (normal)

Gain (Total, Windy)

1. Entropy (no)
2. Entropy (yes)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Node 1** |  | **jml kasus(S)** | **No (S1)** | **Yes (S2)** | **Entropy** | **Gain** |
| **total** |  | 14 | 4 | 10 | 0,863120569 |  |
| **outlook** |  |  |  |  |  | 0,258521037 |
|  | cloudy | 4 | 0 | 4 | 0 |  |
|  | rainy | 5 | 1 | 4 | 0,721928095 |  |
|  | sunny | 5 | 3 | 2 | 0,970950594 |  |
| **temp** |  |  |  |  |  | 0,183850925 |
|  | cool | 4 | 0 | 4 | 0 |  |
|  | hot | 4 | 2 | 2 | 1 |  |
|  | mild | 6 | 2 | 4 | 0,918295834 |  |
| **humidity** |  |  |  |  |  | 0,370506501 |
|  | high | 7 | 4 | 3 | 0,985228136 |  |
|  | normal | 7 | 0 | 7 | 0 |  |
| **windy** |  |  |  |  |  | 0,005977713 |
|  | No | 8 | 2 | 6 | 0,811278124 |  |
|  | Yes | 6 | 4 | 2 | 0,91829583 |  |

Dari hasil perhitungan diketahui bahwa atribut dengan gain tertinggi adalah **HUMIDITY** dengan nilai **0,370506501**. Oleh karena itu, HUMIDITY menjadi node akar dari pohon keputusan. HUMIDITY memiliki dua nilai atribut, yaitu **HIGH** dan **NORMAL.** Untuk cabang dengan atribut HUMIDITY NORMAL, dapat langsung diklasifikasikan sebagai leaf node dengan keputusan "Yes" (Play), karena seluruh instance dengan nilai HUMIDITY NORMAL konsisten menghasilkan keputusan yang sama. Sebaliknya, cabang dengan atribut HUMIDITY HIGH memerlukan analisis lebih lanjut. Hal ini menunjukkan bahwa data dengan nilai HUMIDITY HIGH memiliki variasi yang perlu diteliti lebih dalam sebelum dapat ditentukan keputusannya secara pasti.



**Langkah 2**

Menghitung jumlah kasus, jumlah kasus untuk keputusan **Yes**, jumlah kasus untuk keputusan **No**. Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut **OUTLOOK, TEMPERATURE** dan **WINDY**, yang dapat menjadi node akar dari nilai atribut **HIGH**.

Detail perhitungan :

Entropy(S)Gain(S, Outlook)

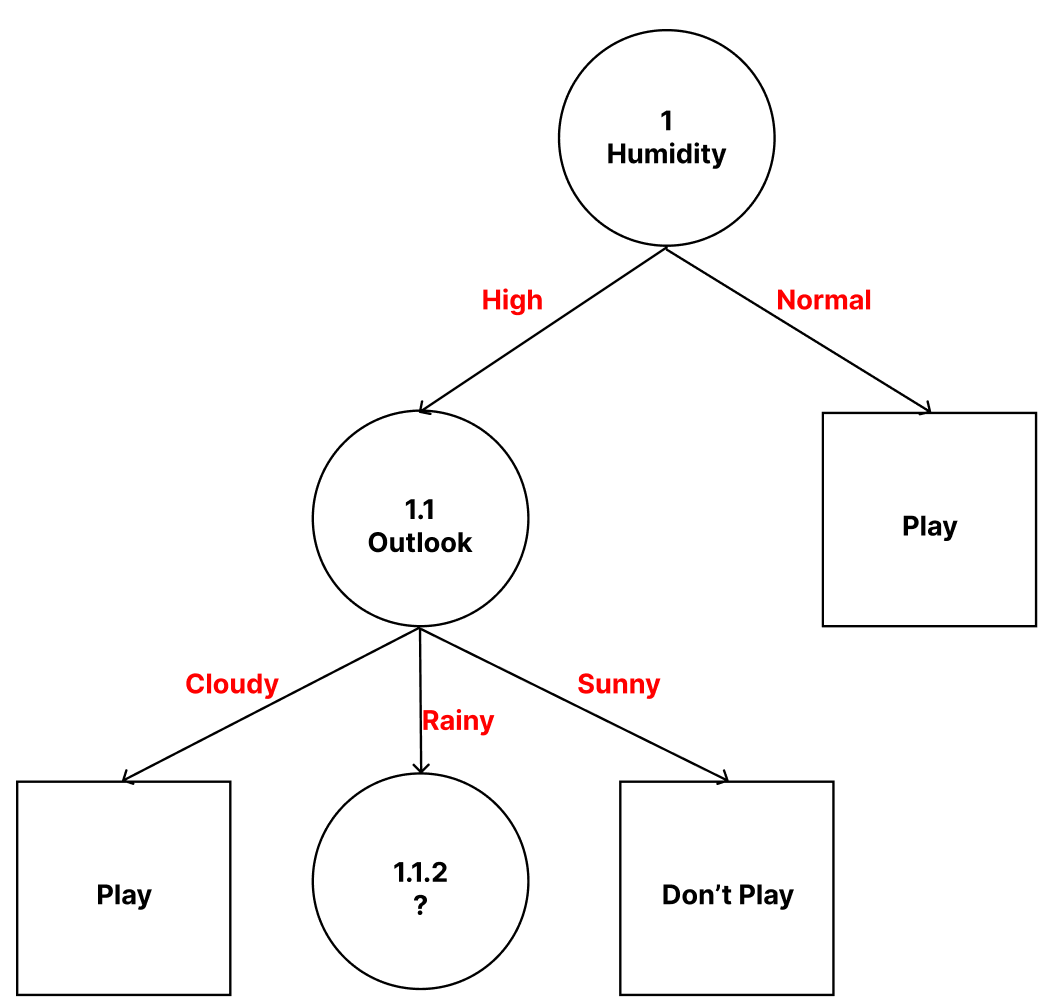
Gain(S, Temp)

Gain(S, Windy)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Node 1.1** |  | **jml kasus(S)** | **No (S1)** | **Yes (S2)** | **Entropy** | **Gain** |
| **humidity (high)** |  | 7 | 4 | 3 | 0,98522814 |  |
| **outlook** |  |  |  |  |  | 0,69951385 |
|  | cloudy | 2 | 0 | 2 | 0 |  |
|  | rainy | 2 | 1 | 1 | 1 |  |
|  | sunny | 3 | 3 | 0 | 0 |  |
| **temp** |  |  |  |  |  | 0,02024421 |
|  | cool | 0 | 0 | 0 | 0 |  |
|  | hot | 3 | 2 | 1 | 0,91829583 |  |
|  | mild | 4 | 2 | 2 | 1 |  |
| **windy** |  |  |  |  |  | 0,02024421 |
|  | No | 4 | 2 | 2 | 1 |  |
|  | Yes | 3 | 2 | 1 | 0,91829583 |  |

OUTLOOK dipilih sebagai node cabang dari nilai HUMIDITY = High karena Gain tertingginya sebesar **0.69951385**. OUTLOOK memiliki tiga nilai:

* CLOUDY : Semua kasus di cabang ini diklasifikasikan sebagai Play.
* SUNNY : Semua kasus di cabang ini diklasifikasikan sebagai Don't Play.
* RAINY : Masih ada variasi antara Play dan Don't Play, sehingga perlu dilakukan perhitungan lebih lanjut untuk atribut lain agar bisa mengklasifikasikan kasus di cabang ini.



**Langkah 3**

Menghitung jumlah kasus, jumlah kasus untuk keputusan Yes, jumlah kasus untuk keputusan No. Entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut TEMPERATURE dan WINDY, yang dapat menjadi node cabang dari nilai atribut RAINY.

Detail perhitungannya :

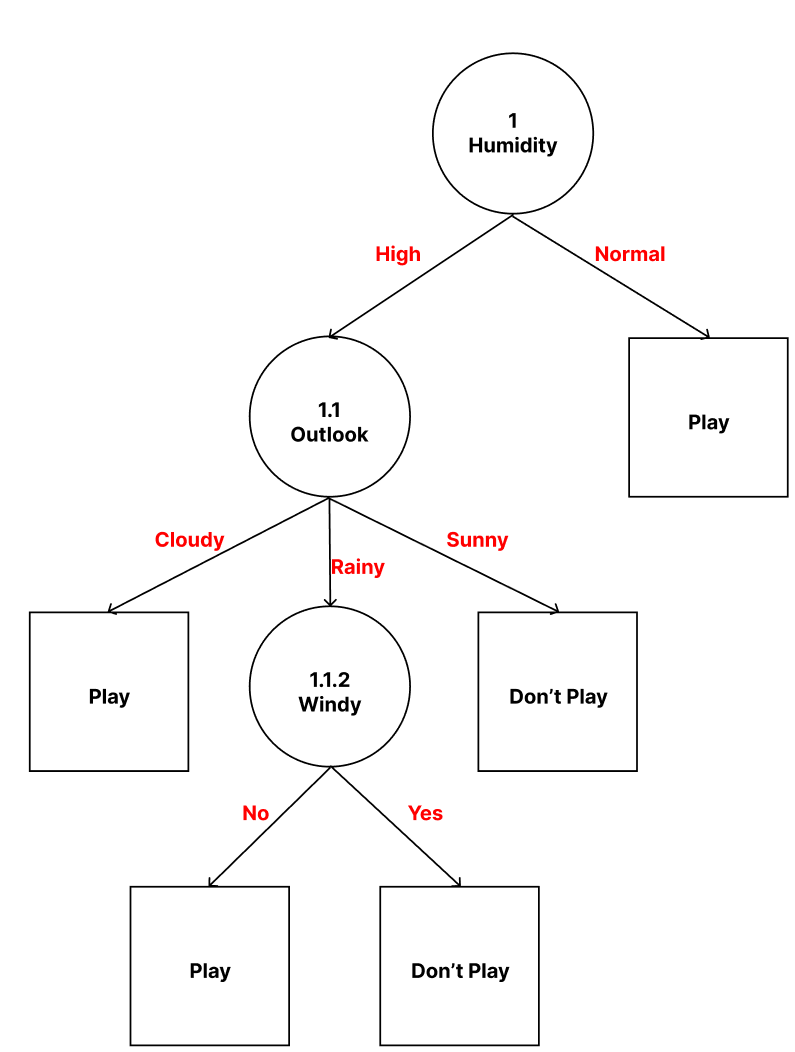
Entropy(S)

Gain(S, Temp)

Gain(S, Windy)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Node 1.1.2** |  | **jml kasus(S)** | **No (S1)** | **Yes (S2)** | **Entropy** | **Gain** |
| **humidity high and outlook rainy** |  | 2 | 1 | 1 | 1 |  |
| **temp** |  |  |  |  |  | 0 |
|  | cool | 0 | 0 | 0 | 0 |  |
|  | hot | 0 | 0 | 0 | 0 |  |
|  | mild | 2 | 1 | 1 | 1 |  |
| **windy** |  |  |  |  |  | 1 |
|  | No | 1 | 0 | 1 | 0 |  |
|  | Yes | 1 | 1 | 0 | 0 |  |

Atribut **WINDY** dengan Gain tertinggi (1) menjadi node untuk OUTLOOK = RAINY. Jika WINDY = No, semua kasus adalah Play, dan jika WINDY = Yes, maka semua kasus adalah Don't Play. Karena kedua nilai sudah mengklasifikasikan kasus, maka tidak perlu perhitungan lebih lanjut.



**SOAL KUIS 2 PERTEMUAN 7**

Link : <https://github.com/margaretavalen/Data-Mining/tree/main/Tugas7>