



```
In [1]: import pandas as pd
```

```
In [2]: data = pd.read_csv("Weather-D.csv")
print(data)
```

	Outlook	Temperature	Humidity	Windy	Play Football
0	Sunny	Hot	High	Weak	No
1	Sunny	Hot	High	Strong	No
2	Overcast	Hot	High	Weak	Yes
3	Rainy	Mild	High	Weak	Yes
4	Rainy	Cool	Normal	Weak	Yes
5	Rainy	Cool	Normal	Strong	No
6	Overcast	Cool	Normal	Strong	Yes
7	Sunny	Mild	High	Weak	No
8	Sunny	Cool	Normal	Weak	Yes
9	Rainy	Mild	Normal	Weak	Yes
10	Sunny	Mild	Normal	Strong	Yes
11	Overcast	Mild	High	Strong	Yes
12	Overcast	Hot	Normal	Weak	Yes
13	Rainy	Mild	High	Strong	No

```
In [3]: def gini_index(pos, neg):
        total = pos + neg
        if total == 0:
            return 0
        p_pos = pos/total
        p_neg = neg/total
        return 1 - (p_pos**2 + p_neg**2)
```

```
In [14]: total_yes = len(data[data["Play Football"] == "Yes"])
total_no = len(data[data["Play Football"] == "No"])
total = total_yes + total_no
print(total, total_yes, total_no)
```

14 9 5

```
In [15]: total_gini = gini_index(total_yes, total_no)
print(f" {total_gini:.4f}")
```

0.4592

```
In [19]: features = ["Outlook", "Temperature", "Windy", "Humidity"]
gini_results = {}

for feature in features:
    feature_values = data[feature].unique()
    feature_gini = 0
    print(f"\nCalculating Gini Gain for feature: {feature}")
    for value in feature_values:
        subset = data[data[feature] == value]
        pos = len(subset[subset["Play Football"] == "Yes"])
        neg = len(subset[subset["Play Football"] == "No"])
        gini = gini_index(pos, neg)
        weight = (pos+neg)/total
```

```

        feature_gini += weight * gini
        print(f" {value} -> Yes: {pos}, No: {neg}, Gini: {gini:.4f}")
    gini_gain = total_gini - feature_gini
    gini_results[feature] = gini_gain
    print(f" Gini gain for {feature}: {gini_gain:.4f}")

```

Calculating Gini Gain for feature: Outlook

```

Sunny -> Yes: 2, No: 3, Gini: 0.4800
Overcast -> Yes: 4, No: 0, Gini: 0.0000
Rainy -> Yes: 3, No: 2, Gini: 0.4800
Gini gain for Outlook: 0.1163

```

Calculating Gini Gain for feature: Temperature

```

Hot -> Yes: 2, No: 2, Gini: 0.5000
Mild -> Yes: 4, No: 2, Gini: 0.4444
Cool -> Yes: 3, No: 1, Gini: 0.3750
Gini gain for Temperature: 0.0187

```

Calculating Gini Gain for feature: Windy

```

Weak -> Yes: 6, No: 2, Gini: 0.3750
Strong -> Yes: 3, No: 3, Gini: 0.5000
Gini gain for Windy: 0.0306

```

Calculating Gini Gain for feature: Humidity

```

High -> Yes: 3, No: 4, Gini: 0.4898
Normal -> Yes: 6, No: 1, Gini: 0.2449
Gini gain for Humidity: 0.0918

```

```

In [21]: best_feature = max(gini_results, key=gini_results.get)
          print("\n----Summary----")
          for feature, gini_gain in gini_results.items():
              print(f"{feature} : {gini_gain:.4f}")
          print(f"\n Root node should be {best_feature} (Highest Gini gain: {gini_result

```

----Summary----

```

Outlook : 0.1163
Temperature : 0.0187
Windy : 0.0306
Humidity : 0.0918

```

Root node should be Outlook (Highest Gini gain: 0.1163)