In [2]:
```
import pandas as pd
import numpy as np

data = {
    'Outlook': ['Sunny', 'Sunny', 'Overcast', 'Rain', 'Rain', 'Rain', 'Overcast', 'Sunny', 'Sunny', 'Rain', 'Sunny', 'Overcast
    'Temperature': ['Hot', 'Hot', 'Hot', 'Mild', 'Cool', 'Cool', 'Cool', 'Mild', 'Cool', 'Mild', 'Mild', 'Mild', 'Hot', 'Mild'
    'Humidity': ['High', 'High', 'High', 'High', 'Normal', 'Normal', 'Normal', 'High', 'Normal', 'Normal', 'Normal', 'High', '
    'Wind': ['Weak', 'Strong', 'Weak', 'Weak', 'Weak', 'Strong', 'Strong', 'Weak', 'Weak', 'Weak', 'Strong', 'Strong', 'Weak',
    'PlayTennis': ['No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'Yes', 'No']
}

df = pd.DataFrame(data)

def gini_index(groups, classes):

    n_instances = float(sum([len(group) for group in groups]))
    gini = 0.0
    for group in groups:
        size = float(len(group))
        if size == 0:
            continue
        score = 0.0
        for class_val in classes:
            p = (group.count(class_val)) / size
            score += p * p
        gini += (1.0 - score) * (size / n_instances)
    return gini


def test_split(attribute, value, dataset):
    left, right = list(), list()
    for i in range(len(dataset)):
        if dataset[attribute].iloc[i] == value:
            left.append(dataset['PlayTennis'].iloc[i])
        else:
            right.append(dataset['PlayTennis'].iloc[i])
    return left, right
```

```
def get_best_split(dataset):
    class_values = list(set(dataset['PlayTennis']))
    best_index, best_value, best_score, best_groups = None, None, float('inf'), None
    for attribute in dataset.columns[:-1]:
        for value in dataset[attribute].unique():
            groups = test_split(attribute, value, dataset)
            gini = gini_index(groups, class_values)
            if gini < best_score:
                best_index, best_value, best_score, best_groups = attribute, value, gini, groups
    return {'index': best_index, 'value': best_value, 'groups': best_groups, 'gini': best_score}


best_split = get_best_split(df)
print("Aryan Chugh 21BCE3118\n\n\n")
best_split
```

```
Aryan Chugh 21BCE3118
```

Out[2]:
```
{'index': 'Outlook',
 'value': 'Overcast',
 'groups': (['Yes', 'Yes', 'Yes', 'Yes'],
  ['No', 'No', 'Yes', 'Yes', 'No', 'No', 'Yes', 'Yes', 'Yes', 'No']),
 'gini': 0.35714285714285715}
```

In [ ]: