

In [1]:

```
import pandas as pd
from pandas import DataFrame
df_tennis = DataFrame.from_csv(r"C:\Users\Student\Downloads\data set\weather.csv")
df_tennis
```

C:\Users\Student\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: FutureWarning: from_csv is deprecated. Please use read_csv(...) instead. Note that some of the default arguments are different, so please refer to the documentation for from_csv when changing your function calls

This is separate from the ipykernel package so we can avoid doing imports until

Out[1]:

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
3	yes	hot	high	weak	overcast
4	yes	mild	high	weak	rainy
5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
8	no	mild	high	weak	sunny
9	yes	cool	normal	weak	sunny
10	yes	mild	normal	weak	rainy
11	yes	mild	normal	strong	sunny
12	yes	mild	high	strong	overcast
13	yes	hot	normal	weak	overcast
14	no	mild	high	strong	rainy

In [2]:

```
def entropy(probs):
    import math
    return sum( [-prob*math.log(prob, 2) for prob in probs] )

def entropy_of_list(a_list):
    from collections import Counter
    cnt = Counter(x for x in a_list)
    print("No and Yes Classes:",a_list.name,cnt)
    num_instances = len(a_list)*1.0
    probs = [x / num_instances for x in cnt.values()]
    return entropy(probs) # Call Entropy:
total_entropy = entropy_of_list(df_tennis['PlayTennis'])
print("Entropy of given PlayTennis Data Set:",total_entropy)
```

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})
 Entropy of given PlayTennis Data Set: 0.9402859586706309

In [3]:

```
def information_gain(df, split_attribute_name, target_attribute_name, trace=0):
    df_split = df.groupby(split_attribute_name)
    for name,group in df_split:
        print(name)
        print(group)
```

In [4]:

```
def information_gain(df, split_attribute_name, target_attribute_name, trace=0):
    df_split = df.groupby(split_attribute_name)
    for name,group in df_split:
        print(name)
        print(group)
    nobs = len(df.index) * 1.0
    df_agg_ent = df_split.agg({target_attribute_name : [entropy_of_list, lambda x: len(x)/nobs] })[target_attribute_name]
    df_agg_ent.columns = ['Entropy', 'PropObservations']
    new_entropy = sum(df_agg_ent['Entropy'] * df_agg_ent['PropObservations'] )
    old_entropy = entropy_of_list(df[target_attribute_name])
    return old_entropy - new_entropy
```

In [5]:

```
print('Info-gain for Outlook is :'+str( information_gain(df_tennis, 'Outlook', 'PlayTennis')), "\n")
print('\n Info-gain for Humidity is: ' + str( information_gain(df_tennis, 'Humidity', 'PlayTennis')), "\n")
print('\n Info-gain for Wind is:' + str( information_gain(df_tennis, 'Wind', 'PlayTennis')), "\n")
print('\n Info-gain for Temperature is:' + str( information_gain(df_tennis , 'Temperature', 'PlayTennis')), "\n")
```

overcast

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
3	yes	hot	high	weak	overcast
7	yes	cool	normal	strong	overcast
12	yes	mild	high	strong	overcast
13	yes	hot	normal	weak	overcast

rainy

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
4	yes	mild	high	weak	rainy
5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
10	yes	mild	normal	weak	rainy
14	no	mild	high	strong	rainy

sunny

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
8	no	mild	high	weak	sunny
9	yes	cool	normal	weak	sunny
11	yes	mild	normal	strong	sunny

No and Yes Classes: PlayTennis Counter({'yes': 4})

No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 2})

No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

Info-gain for Outlook is :0.2467498197744391

high

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
3	yes	hot	high	weak	overcast
4	yes	mild	high	weak	rainy
8	no	mild	high	weak	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

normal

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
9	yes	cool	normal	weak	sunny
10	yes	mild	normal	weak	rainy
11	yes	mild	normal	strong	sunny
13	yes	hot	normal	weak	overcast

No and Yes Classes: PlayTennis Counter({'no': 4, 'yes': 3})

No and Yes Classes: PlayTennis Counter({'yes': 6, 'no': 1})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

Info-gain for Humidity is: 0.15183550136234136

strong

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
2	no	hot	high	strong	sunny
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast

11	yes	mild	normal	strong	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

weak

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					

1	no	hot	high	weak	sunny
3	yes	hot	high	weak	overcast
4	yes	mild	high	weak	rainy
5	yes	cool	normal	weak	rainy
8	no	mild	high	weak	sunny
9	yes	cool	normal	weak	sunny
10	yes	mild	normal	weak	rainy
13	yes	hot	normal	weak	overcast

No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 3})

No and Yes Classes: PlayTennis Counter({'yes': 6, 'no': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

Info-gain for Wind is:0.04812703040826927

cool

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					

5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
9	yes	cool	normal	weak	sunny

hot

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					

1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
3	yes	hot	high	weak	overcast
13	yes	hot	normal	weak	overcast

mild

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					

4	yes	mild	high	weak	rainy
8	no	mild	high	weak	sunny
10	yes	mild	normal	weak	rainy
11	yes	mild	normal	strong	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 1})

No and Yes Classes: PlayTennis Counter({'no': 2, 'yes': 2})

No and Yes Classes: PlayTennis Counter({'yes': 4, 'no': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

Info-gain for Temperature is:0.029222565658954647

In [6]:

```
def id3(df, target_attribute_name, attribute_names, default_class=None):
    from collections import Counter
    cnt = Counter(x for x in df[target_attribute_name])
    if len(cnt) == 1:
        return next(iter(cnt))
    elif df.empty or (not attribute_names):
        return default_class
    else:
        gainz = [information_gain(df, attr, target_attribute_name) for attr in attribute_names]
        index_of_max = gainz.index(max(gainz))
        best_attr = attribute_names[index_of_max]
        tree = {best_attr: {}}
        remaining_attribute_names = [i for i in attribute_names if i != best_attr]
        for attr_val, data_subset in df.groupby(best_attr):
            subtree = id3(data_subset,
                           target_attribute_name,
                           remaining_attribute_names,
                           default_class)
            tree[best_attr][attr_val] = subtree
        return tree
```

In [7]:

```
attribute_names = list(df_tennis.columns)
print("List of Attributes:", attribute_names)
attribute_names.remove('PlayTennis')
print("Predicting Attributes:", attribute_names)
```

List of Attributes: ['PlayTennis', 'Temperature', 'Humidity', 'Wind', 'Outlook']

Predicting Attributes: ['Temperature', 'Humidity', 'Wind', 'Outlook']

In [8]:

```
from pprint import pprint
tree = id3(df_tennis, 'PlayTennis', attribute_names)
print("\n\nThe Resultant Decision Tree is :\n")
pprint(tree)
```

cool

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
9	yes	cool	normal	weak	sunny

hot

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
3	yes	hot	high	weak	overcast
13	yes	hot	normal	weak	overcast

mild

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
4	yes	mild	high	weak	rainy
8	no	mild	high	weak	sunny
10	yes	mild	normal	weak	rainy
11	yes	mild	normal	strong	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 1})

No and Yes Classes: PlayTennis Counter({'no': 2, 'yes': 2})

No and Yes Classes: PlayTennis Counter({'yes': 4, 'no': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

high

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
1	no	hot	high	weak	sunny
2	no	hot	high	strong	sunny
3	yes	hot	high	weak	overcast
4	yes	mild	high	weak	rainy
8	no	mild	high	weak	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

normal

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
5	yes	cool	normal	weak	rainy
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
9	yes	cool	normal	weak	sunny
10	yes	mild	normal	weak	rainy
11	yes	mild	normal	strong	sunny
13	yes	hot	normal	weak	overcast

No and Yes Classes: PlayTennis Counter({'no': 4, 'yes': 3})

No and Yes Classes: PlayTennis Counter({'yes': 6, 'no': 1})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

strong

	PlayTennis	Temperature	Humidity	Wind	Outlook
id					
2	no	hot	high	strong	sunny
6	no	cool	normal	strong	rainy
7	yes	cool	normal	strong	overcast
11	yes	mild	normal	strong	sunny
12	yes	mild	high	strong	overcast
14	no	mild	high	strong	rainy

weak

	PlayTennis	Temperature	Humidity	Wind	Outlook
--	------------	-------------	----------	------	---------


```

id
1      no      hot      high  weak  sunny
3      yes     hot      high  weak  overcast
4      yes     mild     high  weak  rainy
5      yes     cool     normal weak  rainy
8      no      mild     high  weak  sunny
9      yes     cool     normal weak  sunny
10     yes     mild     normal weak  rainy
13     yes     hot      normal weak  overcast

```

No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 3})

No and Yes Classes: PlayTennis Counter({'yes': 6, 'no': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

overcast

```

      PlayTennis Temperature Humidity    Wind  Outlook
id
3      yes      hot      high    weak  overcast
7      yes      cool     normal strong  overcast
12     yes      mild     high    strong overcast
13     yes      hot      normal   weak  overcast

```

rainy

```

      PlayTennis Temperature Humidity    Wind Outlook
id
4      yes      mild     high    weak  rainy
5      yes      cool     normal   weak  rainy
6      no       cool     normal strong  rainy
10     yes      mild     normal   weak  rainy
14     no       mild     high    strong  rainy

```

sunny

```

      PlayTennis Temperature Humidity    Wind Outlook
id
1      no      hot      high    weak  sunny
2      no      hot      high    strong sunny
8      no      mild     high    weak  sunny
9      yes     cool     normal   weak  sunny
11     yes     mild     normal strong  sunny

```

No and Yes Classes: PlayTennis Counter({'yes': 4})

No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 2})

No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 2})

No and Yes Classes: PlayTennis Counter({'yes': 9, 'no': 5})

cool

```

      PlayTennis Temperature Humidity    Wind Outlook
id
5      yes      cool     normal   weak  rainy
6      no       cool     normal strong  rainy

```

mild

```

      PlayTennis Temperature Humidity    Wind Outlook
id
4      yes      mild     high    weak  rainy
10     yes      mild     normal   weak  rainy
14     no       mild     high    strong  rainy

```

No and Yes Classes: PlayTennis Counter({'yes': 1, 'no': 1})

No and Yes Classes: PlayTennis Counter({'yes': 2, 'no': 1})

No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 2})

high

```

      PlayTennis Temperature Humidity    Wind Outlook
id
4      yes      mild     high    weak  rainy
14     no       mild     high    strong  rainy

```

normal

```

      PlayTennis Temperature Humidity    Wind Outlook
id

```

```

5         yes         cool    normal    weak    rainy
6         no          cool    normal    strong   rainy
10        yes         mild    normal    weak    rainy
No and Yes Classes: PlayTennis Counter({'yes': 1, 'no': 1})
No and Yes Classes: PlayTennis Counter({'yes': 2, 'no': 1})
No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 2})
strong
    PlayTennis Temperature Humidity    Wind Outlook
id
6         no          cool    normal    strong   rainy
14        no          mild     high    strong   rainy
weak
    PlayTennis Temperature Humidity    Wind Outlook
id
4         yes         mild     high    weak    rainy
5         yes         cool    normal    weak    rainy
10        yes         mild    normal    weak    rainy
No and Yes Classes: PlayTennis Counter({'no': 2})
No and Yes Classes: PlayTennis Counter({'yes': 3})
No and Yes Classes: PlayTennis Counter({'yes': 3, 'no': 2})
cool
    PlayTennis Temperature Humidity    Wind Outlook
id
9         yes         cool    normal    weak    sunny
hot
    PlayTennis Temperature Humidity    Wind Outlook
id
1         no          hot      high    weak    sunny
2         no          hot      high    strong   sunny
mild
    PlayTennis Temperature Humidity    Wind Outlook
id
8         no          mild     high    weak    sunny
11        yes         mild     normal  strong   sunny
No and Yes Classes: PlayTennis Counter({'yes': 1})
No and Yes Classes: PlayTennis Counter({'no': 2})
No and Yes Classes: PlayTennis Counter({'no': 1, 'yes': 1})
No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 2})
high
    PlayTennis Temperature Humidity    Wind Outlook
id
1         no          hot      high    weak    sunny
2         no          hot      high    strong   sunny
8         no          mild     high    weak    sunny
normal
    PlayTennis Temperature Humidity    Wind Outlook
id
9         yes         cool    normal    weak    sunny
11        yes         mild     normal  strong   sunny
No and Yes Classes: PlayTennis Counter({'no': 3})
No and Yes Classes: PlayTennis Counter({'yes': 2})
No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 2})
strong
    PlayTennis Temperature Humidity    Wind Outlook
id
2         no          hot      high    strong   sunny
11        yes         mild     normal  strong   sunny
weak
    PlayTennis Temperature Humidity    Wind Outlook
id
1         no          hot      high    weak    sunny

```

```
8      no      mild    high  weak  sunny
9      yes      cool   normal weak  sunny
No and Yes Classes: PlayTennis Counter({'no': 1, 'yes': 1})
No and Yes Classes: PlayTennis Counter({'no': 2, 'yes': 1})
No and Yes Classes: PlayTennis Counter({'no': 3, 'yes': 2})
```

The Resultant Decision Tree is :

```
{'Outlook': {'overcast': 'yes',
             'rainy': {'Wind': {'strong': 'no', 'weak': 'yes'}},
             'sunny': {'Humidity': {'high': 'no', 'normal': 'yes'}}}}
```