Ensemble methods – Decision trees

In the example, a person will try to decide if he/she should go to a comedy show or not.

Luckily our example person has registered every time there was a comedy show in town, and registered some information about the comedian, and also registered if he/she went or not.

1. Import libraries

```
import pandas as pd
import pandas
from sklearn import tree
from sklearn.tree import c
import matplotlib.pyplot as plt
```

2. Import data

```
df = pd.read_csv("../data/comedian.csv")

print(df)
```

3. Data pre treatment

```
#change string into numerical

d = {'UK': 0, 'USA': 1, 'N': 2}
df['Nationality'] = df['Nationality'].map(d)
d = {'YES': 1, 'NO': 0}
df['Go'] = df['Go'].map(d)

print(df)
```

4. Preparing data for training

```
#X is the feature columns, y is the target column:

features = ['Age', 'Experience', 'Rank', 'Nationality']

X = df[features]
y = df['Go']

print(X)
print(y)
```

5. Training and ploting the tree

```
dtree = DecisionTreeClassifier()

dtree = dtree.fit(X, y)

tree.plot_tree(dtree, feature_names=features)
```

6. What happends if we change the criterion for the Split? Try with these for tests Criterion = {“gini”, “entropy”, “log_loss”}, default=”gini”