# playtennis

July 4, 2023

GROUP K - Data Mining Étudiants

[22]:
```
import pandas as pd
```

[23]:
```
import graphviz
```

[25]:
```
from sklearn.model_selection import train_test_split
```

[26]:
```
from sklearn.tree import DecisionTreeClassifier, export_graphviz
```

[5]:
```
from sklearn.metrics import accuracy_score
```

[5]:
```

```

### 0.0.1 playtennis data loading

[6]:
```
data = pd.read_csv('data/playtennis.csv')
```

[6]:
```

```

### 0.0.2 Splitting the data into features & target

[7]:
```
X = data.drop('PlayTennis', axis=1)
```

[8]:
```
y = data['PlayTennis']
```

[8]:
```

```

### 0.0.3 categorical vars to num convertion while avoidin' multicollinearity

[9]:
```
data_encoded = pd.get_dummies(data, drop_first=True)
```

[9]:
```

```

### 0.0.4 Split the data into features and target

[10]:
```
X = data_encoded.drop('PlayTennis_Yes', axis=1)
```

[11]:
```
y = data_encoded['PlayTennis_Yes']
```

### 0.0.5 data splitting(training & tests sets)

[12]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)
```

[12]:

### 0.0.6 now creating a D.T classifier

[13]:
```
classifier = DecisionTreeClassifier()
```

[13]:

### 0.0.7 Classifier training based on training data

[14]:
```
classifier.fit(X_train, y_train)
```

[14]:
```
DecisionTreeClassifier()
```

[15]:
```
y_pred = classifier.predict(X_test)
```

[15]:

### 0.0.8 classifier accuracy score

[16]:
```
accuracy = accuracy_score(y_test, y_pred)
```

[17]:
```
print('Accuracy:', accuracy)
```

```
Accuracy: 0.6666666666666666
```

[17]:

### 0.0.9 Visualize the decision tree

[27]:
```
dot_data = export_graphviz(classifier, out_file=None, feature_names=X.columns,␣
↪class_names=['No', 'Yes'], filled=True)
```

[28]:
```
graph = graphviz.Source(dot_data)
```

[29]:
```
graph.render('decision_tree', format='png')
```

[29]:
```
'decision_tree.png'
```