Cheatography

Python - Decision Tree & Random Forest Cheat Sheet

by Pitbull (aggialavura) via cheatography.com/83764/cs/19971/

TO START

```
# IMPORT DATA LIBRARIES
import pandas as pd
import numpy as np
# IMPORT VIS LIBRARIES
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
# IMPORT MODELLING LIBRARIES
from sklearn.model_selection import train_tes-
t_split
# libraries for decision trees
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classific ation_rep-
ort,confusion_matrix
# libraries for random forest
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classific ation_rep-
ort,confusion_matrix
```

PRELIMINARY OPERATIONS

df = pd.read_csv('data.csv')	import data
sns.pairplot(df,hue='col')	pairplot
df.info()	check info df
df.describe()	check stats df
df.head()	check head df

TRAIN MODEL - DECISION TREES

SPLIT DATASET

X = df[['col1','col2',etc.]]	create df features
y = df['col']	create df var to predict
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)	split df in train and test df

FIT THE MODEL

tree = DecisionTreeClassifier()	instatiate model
tree.fit(X_train, y_train)	train/fit the model

MAKE PREDICTIONS

pred = tree.predict(X_test)	make predictions

TRAIN MODEL - DECISION TREES (cont)

EVAUATE MODEL

```
print(classification_report(y_test,pred))
print(confusion_matrix(y_test,pred))
```

TRAIN MODEL - RANDOM FOREST

SPLIT DATASET

X = df[['col1','col2',etc.]]	create df features
y = df['col']	create df var to predict
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3)	split df in train and test df

FIT THE MODEL

rfc = RandomForestClassifier (n_estimators=200)*	instatiate model
rfc.fit(X_train, y_train)	train/fit the model

MAKE PREDICTIONS

rfc_pred = rfc.predict(X_test)	make predictions

EVAUATE MODEL

```
print(confusion_matrix(y_test,rfc_pred))
print(classification_report(y_test,rfc_pred))
```

n_estimators: number of trees to be used in the forest.