Primer klasifikacije

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
from sklearn import model_selection
from sklearn import preprocessing
from sklearn import metrics
```

In [3]:
```
import xgboost
from sklearn import tree
from sklearn import ensemble
```

Pravimo XGBoost i AdaBoost klasifikatore na skupu diabetes.

In [4]:
```
data = pd.read_csv('diabetes.csv')
```

In [5]:
```
y = data['Outcome']
X = data.drop(columns=['Outcome'], axis=1)
```

In [6]:
```
X_train, X_test, y_train, y_test = model_selection.train_test_split(
    X, y, test_size=0.3, stratify=y, random_state = 7)
```

In [7]:
```
scaler = preprocessing.StandardScaler()
scaler.fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)
```

In [8]:
```
model_xgboost = xgboost.XGBClassifier(n_estimators=100, max_depth=5)
model_adaboost = ensemble.AdaBoostClassifier(
    base_estimator=tree.DecisionTreeClassifier(max_depth=5),
    n_estimators=100, random_state=7)
```

In [9]:
```
model_xgboost.fit(X_train, y_train);
```

In [10]:
```
model_adaboost.fit(X_train, y_train);
```

In [11]:
```
y_predicted1 = model_xgboost.predict(X_test)
```

In [12]:
```
metrics.accuracy_score(y_test, y_predicted1)
```

Out[12]:
```
0.7359307359307359
```

In [13]:
```
metrics.f1_score(y_test, y_predicted1)
```

Out[13]:
```
0.6013071895424836
```

In [14]:
```
y_predicted2 = model_adaboost.predict(X_test)
```

In [15]:
```
metrics.accuracy_score(y_test, y_predicted2)
```

Out[15]:
```
0.7532467532467533
```

In [16]:
```
metrics.f1_score(y_test, y_predicted2)
```

Out[16]:
```
0.6174496644295301
```

U ovom slučaju je AdaBoost dao bolje rezultate. Pokušajmo sa 500 stabala.

In [17]:
```
model_xgboost = xgboost.XGBClassifier(n_estimators=500, max_depth=5)
model_adaboost = ensemble.AdaBoostClassifier(
    base_estimator=tree.DecisionTreeClassifier(max_depth=5), n_estimators=500, random_st
```

In [18]:
```
model_xgboost.fit(X_train, y_train);
```

In [19]:
```
model_adaboost.fit(X_train, y_train);
```

In [20]:
```
y_predicted1 = model_xgboost.predict(X_test)
```

In [21]:
```
metrics.accuracy_score(y_test, y_predicted1)
```

Out[21]:
```
0.7186147186147186
```

In [22]:
```
metrics.f1_score(y_test, y_predicted1)
```

Out[22]:
```
0.5695364238410596
```

In [23]:
```
y_predicted2 = model_adaboost.predict(X_test)
```

In [24]:
```
metrics.accuracy_score(y_test, y_predicted2)
```

Out[24]:
```
0.7445887445887446
```

In [25]:
```
metrics.f1_score(y_test, y_predicted2)
```

Out[25]:
```
0.6040268456375839
```

Oba modela su malo lošija nego sa 100 stabala, ali je AdaBoost i dalje bolji.