import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

import tensorflow as tf

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import confusion\_matrix, accuracy\_score, classification\_report

from sklearn.metrics import recall\_score

from mlxtend.plotting import plot\_confusion\_matrix

df = pd.read\_csv('breast-cancer.csv')

df.head()

df

df.diagnosis.value\_counts()

df.diagnosis.value\_counts().plot(kind="bar", color=['blue', 'red'])

df.info()

df.isna().sum()

## Iybacivanje kolone ID

df = df.drop(columns=['id'])

df.describe().T

corr\_matrix = df.corr()

fix, ax = plt.subplots(figsize=(25,20))

ax = sns.heatmap(corr\_matrix,

annot = True,

linewidth=0.5,

fmt=".2f",

cmap='YlGnBu')

df.drop(columns=['perimeter\_worst', 'area\_worst', 'area\_mean', "perimeter\_mean"], inplace=True)

df

X = df.drop(columns='diagnosis')

y = df.diagnosis

X

Y

for i in range(len(y)):

if (y[i] == 'M'):

y[i] = 1

else:

y[i] = 0

y

## Podela podataka u trening i test

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

print('Num of train instances:', len(X\_train))

print('Num of test instances:', len(X\_test))

test\_df = pd.DataFrame(y\_test)

test\_df.value\_counts().plot(kind="bar", color=['blue', 'red'])

sc = StandardScaler()

X\_train = sc.fit\_transform(X\_train)

X\_test = sc.transform(X\_test)