import pandas as pd

import matplotlib.pyplot as plt

from sklearn import tree

from sklearn.datasets import load\_breast\_cancer

data = load\_breast\_cancer()

dataset = pd.DataFrame(data=data['data'], columns=data['feature\_names'])

dataset

#splitting the dataset

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

x = dataset.copy()

y = data['target']

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.33, random\_state=0)

#training the decision tree model on the training set

from sklearn.tree import DecisionTreeClassifier

clf = DecisionTreeClassifier()

clf = clf.fit(x\_train, y\_train)

#fetching parameters

clf.get\_params()

#relevant parameters that define the approach of the tree

#criterion : {gini, entropy}

#splitter : {best, random} (best : choose the better of the 2 criteria, random : pick any one)

#max\_features : use a fix no. of features out of all

#random\_state : will result in the same tree

#class\_weight : one of the classes might be imp than the other

#ccp\_alpha : complexity parameter used for minimal-cost complexity pruning. Subtree with the lasgest cost complexity smaller

# than ccp alpha will be chosen

y\_pred = clf.predict(x\_test)

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

cm = confusion\_matrix(y\_test, y\_pred)

cm

accuracy\_score(y\_test, y\_pred)

print(classification\_report(y\_test, y\_pred, target\_names=['malignant', 'benign']))

#fetching all the features

feature\_names = x.columns

feature\_names

#fetch the imp of all these features and print in descending order

feature\_imp = pd.DataFrame(clf.feature\_importances\_, index=feature\_names).sort\_values(0, ascending=False)

feature\_imp

#fetch all the features with weigths>0

features = list(feature\_imp[feature\_imp[0]>0].index)

features

feature\_imp.head(10).plot(kind='bar')

fig = plt.figure(figsize=(25,20))

\_ = tree.plot\_tree(clf,

feature\_names = feature\_names,

class\_names = {0:'Malignant', 1:"Benign"},

filled=True,

fontsize=12

)