import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

dataset=pd.read\_csv("Data.csv")

x=dataset.iloc[:,:-1].values

y=dataset.iloc[:,3].values

#missing data

from sklearn.impute import SimpleImputer

imputer = SimpleImputer(missing\_values=np.nan, strategy='mean')

imputer=imputer.fit(x[:,1:3])

x[:,1:3]=imputer.transform(x[:,1:3])

#encoding categorical data

from sklearn.preprocessing import LabelEncoder,OneHotEncoder

from sklearn.compose import ColumnTransformer

labelencoder\_x=LabelEncoder()

x[:,0]=labelencoder\_x.fit\_transform(x[:,0])

ct = ColumnTransformer(

[('one\_hot\_encoder', OneHotEncoder(), [0])], # The column numbers to be transformed (here is [0] but can be [0, 1, 3])

remainder='passthrough' # Leave the rest of the columns untouched

)

x = np.array(ct.fit\_transform(x), dtype=np.float)

labelencoder\_y=LabelEncoder()

y=labelencoder\_x.fit\_transform(y)

#splitting into test and training dataset

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

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=0)

#feature scaling

from sklearn.preprocessing import StandardScaler

sc\_x=StandardScaler()

x\_train=sc\_x.fit\_transform(x\_train)

x\_test=sc\_x.transform(x\_test)