**Step 1: Importing the libraries**

import numpy as np

import pandas as pd

**Step 2: Importing dataset**

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

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

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

**Step 3: Handling the missing data**

from sklearn.preprocessing import Imputer

imputer = Imputer(missing\_values = "NaN", strategy = "mean", axis = 0)

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

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

**Step 4: Encoding categorical data**

from sklearn.preprocessing import LabelEncoder, OneHotEncoder

labelencoder\_X = LabelEncoder()

X[ : , 0] = labelencoder\_X.fit\_transform(X[ : , 0])

**Creating a dummy variable**

onehotencoder = OneHotEncoder(categorical\_features = [0])

X = onehotencoder.fit\_transform(X).toarray()

labelencoder\_Y = LabelEncoder()

Y = labelencoder\_Y.fit\_transform(Y)

**Step 5: Splitting the datasets into training sets and Test sets**

from sklearn.cross\_validation import train\_test\_split

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split( X , Y , test\_size = 0.2, random\_state = 0)

**Step 6: Feature Scaling**

from sklearn.preprocessing import StandardScaler

sc\_X = StandardScaler()

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

X\_test = sc\_X.fit\_transform(X\_test)

**Done**