#Jinit Joshi

#importing all libraries

Import pandas as pd

Import numpy as np

Import matplotlib.pyplot as plt

# %matplotlib inline

#reading dataset from remote link

url = http://bit.ly/w-data

student\_data = pd.read\_csv(url)

print(“Data imported successfully”)

#checking the first 10 rows of the imported dataset Student\_data.head(10)

#checking the dataset shape i.e num of rows and cols Student\_data.shape

#describe the datset in terms of mean,min,max,count etc. Student\_data.describe()

“””## Data Visulaisation”””

# Plotting the distribution of scores

Student\_data.plot(x=’Hours’, y=’Scores’, style=’om’) Plt.title(‘Hours vs Percentage’)

Plt.xlabel(‘Hours Studied’)

Plt.ylabel(‘Percentage Score’)

Plt.show()

“””## Preparing the data

Divide the data into the attributes and the output/label for those attributes.Here the “score” is the output and “hours” studied by the attribute.

“””

X = student\_data.iloc[:, :-1].values

Y = student\_data.iloc[:, 1].values

“””Now we split the dataset into training and test sets using scikit-learn train\_test\_split() method. Here we split the datset in 80/20 ratio.

“””

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) “””Training the algorithm using Linear Regression”””

From sklearn.linear\_model import LinearRegression

Regressor = LinearRegression()

Regressor.fit(X\_train, y\_train)

Print(“Training completed.”)

# Plotting the regression line

Line = regressor.coef\_\*X+regressor.intercept\_

# Plotting for the test data

Plt.scatter(X, y,c=’r’)

#plt.scatter()

Plt.plot(X,line,’m’)

Plt.show()

“””## Making Prediction”””

Print(X\_test) # Testing data – In Hours

Y\_pred = regressor.predict(X\_test) # Predicting the scores

# Comparing Actual vs Predicted

Df = pd.DataFrame({‘Actual’: y\_test, ‘Predicted’: y\_pred})

Df

“””## What will be predicted score if a student studies for 9.25 hrs/ day? “””

Hours = np.array(9.25)

Hours\_new = hours.reshape(-1,1)

Main\_pred = regressor.predict(hours\_new)

Print(“For a student studying 9.25 hours a day the predicted score is : {}”.format(main\_pred[0])) “””Error Calculation”””

From sklearn import metrics

Print(‘Mean Absolute Error:’, metrics.mean\_absolute\_error(y\_test, y\_pred)) Print(‘Mean Squared Error:’,metrics.mean\_squared\_error(y\_test, y\_pred))