|  |
| --- |
| **A Project Report on**    **Sale Prediction (Linear Regression)**      Submitted in partial fulfillment of award of    **BACHELOR OF TECHNOLOGY**    Degree in  Computer Science & Engineering    By  **Gaurang Gupta**  **1808210053**  **(**2020-2021)    SUPERVISOR  **Miss Prachi Aggarwal**  **Assistant Professor**  **Dept. of Computer Science & Engineering**      **Moradabad Institute of Technology**  **Dept. of Computer Science & Engineering**  **Moradabad (U.P.)** |

**Project Objective**

**We have to predict the salary of an employee given how many years of experience they have.**

# Step 1: Load the Dataset We will be using the pandas data frame. Here X is the independent variable which is the “Years of Experience” and y is the dependent variable which is the “Salary”

# Step 2: Split dataset into training set and test set

Next we have to split the dataset into training and testing. We will use the training dataset for training the model and then check the performance of the model on the test dataset.

For this we will use the train\_test\_split method from library model\_selection  
We are providing a test\_size of 1/3 which means test set will contain 10 observations and training set will contain 20 observations

# Step 3: Fit Simple Linear Regression model to training set

We will be using the LinearRegression class from the library sklearn.linear\_model. First we create an object of the LinearRegression class and call the fit method passing the X\_train and y\_train.

# Step 4: Predict the test set

Using the regressor we trained in the previous step, we will now use it to predict the results of the test set and compare the predicted values with the actual values.

# Step 5 — Visualizing the training set

We will plot the actual data points of training set — X\_train and y\_train

**plt.scatter(X\_train, y\_train, color = ‘red’)**

Next we’ll plot the regression line — which is the predicted values for the X\_train

**plt.plot(X\_train, regressor.predict(X\_train), color=’blue’)**

# Step 6 — Visualizing the test set

we will plot the actual data points of training set — X\_test and y\_test

**plt.scatter(X\_test, y\_test, color = ‘red’)**

Next we’ll plot the regression line

**plt.plot(X\_train, regressor.predict(X\_train), color=’blue’)**

# Step 7 — Make new predictions

We can also make brand new predictions for data points that do not exist in the dataset. Like for a person with 15 years experience