**SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

**Department of Electronics and Telecommunication Engineering**

**Subject: Machine Learning Program: B.Tech/MBA.Tech**

**Sem: III/V ACAY: 2020-21**

**EXPERIMENT NO. 5**

**Aim:**

1. To be able to perform multiple linear regression using sklearn and statsmodels.
2. To be able to interpret the results obtained from multiple linear regression.

**Software:**  PYTHON.

**Prerequisite:**

|  |  |
| --- | --- |
| Sr. No | Concepts |
| 1. | Knowledge of multiple linear regression |

**Outcome:**

After successful completion of this experiment students will be able to:

1. Implement multiple linear regression by using sklearn package statsmodels.
2. Interpret the results obtained from different models and choose the best model for the given data set.

**Theory:**

* The data which we will be using for our multiple **linear regression** example is in a .csv file called: ‘MLR\_data.csv.
* For loading the data use the command:

df=pd.read\_csv(MLR\_data.csv')

**TO BE COMPLETED BY STUDENTS**

* Students must upload the soft copy of the program in the given format.

|  |  |
| --- | --- |
| Name of the Experiment: Multiple Linear Regression | |
| Roll No.: B032 | Name: NAMAN GARG |
| Program: B Tech CS (B div) | Semester: V |
| Date of Performance: 21/8/2020 | Date of Submission: 21/8/2020 |

**Step 1: Importing the Relevant Libraries**

### ****Step 2: Loading the Data****

### ****Step 3: Visualizing the Data Frame****

### ****Step 4: Exploring the Data****

### ****Step 5: Plot the scatter plots of IV vs DV and IV vs IV****

### Step 6: Fit a multiple linear regression model using sklearn

### ****Step 7: Get the model statistics using statsmodel****

### ****Step 8: Interpret the results and implement other models****

### ****Step 9: Choose the best possible model****

### 

### 

### 

### 

### 

### 

### 

### 

### 

**Observation & Learning**

The R-squared value is the best for the first model, but the P value is not statistically significant. the one with interest rate (X1) has a better standard error and the one with unemployment rate (X2) has a better P value. all the models are performing well but considering both the independent variables for the model training is redundant. In between X1 and X2, the model using X2 can be considered better because of the P-values of constant and slope.