

## **REGRESSION ASSIGNMENT**

### **1. Identify your problem statement**

Since the dataset is related to numbers, Hence i have been selected Machine learning .since we have both input and output, chosen Supervised machine Learning.  
Then its prediction about Numerical value so that chosen Regression.

### **2.) Tell basic info about the dataset (Total number of rows, columns)**

In the given dataset we have 1338 Rows and 6 columns

### **3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)**

we have two categorical column which is nominal data present in the dataset.  
we have chosen One hot encoding as its Nominal data.

### **4.) Develop a good model with r2\_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.**

r2\_score is 0.78 for Multilinear regression

r2\_score is 0.84 for svm

r2\_score is 0.73 for Decision tree

r2\_score is 0.85 for random forest

### **5.) All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.)**

#### **Model=SVM Regressor**

R2 value is 0.84 by using kernel = rbf , C=3000

<b>kernel</b>	<b>C=10 R value</b>	<b>C=100 R value</b>	<b>C=1000 R value</b>	<b>C=2000 R value</b>	<b>C=3000 R value</b>
Linear	0.46	0.63	0.76	-0.88	0.74
Poly	0.39	0.61	0.84	0.84	0.81
rbf	-0.03	0.31	0.81	0.79	0.84
Sigmoid	0.40	-0.53	0.28	-0.56	-0.34

#### **Model=Decision tree Regressor**

R2 value is 0.73 by using Criterion=friedman mse and splitter =random

<b>Criterion</b>	<b>Splitter (best) R value</b>	<b>Splitter (Random) R value</b>
Squared error	0.63	0.69
Friedman_mse	0.67	0.73
Absolutr_error	0.71	0.71
Poly	0.71	0.63

**Model = Random Forest**

R<sup>2</sup> value is 0.85 without any parameter , Got same value after using parameter also.

**6.) Mention your final model, justify why u have chosen the same.**

I have chosen the **Random Forest Regressor** as the final model because it achieved the highest R<sup>2</sup> value (0.85) compared to the other models.