**Regression Assignment**

Problem Statement or Requirement:

A client’s requirement is, he wants to predict the insurance charges based on the several parameters.

The Client has provided the dataset of the same.As a data scientist, you must develop a model which will predict the insurance charges.

1.) Identify your problem statement:

Ans: Client would like to predict the insurance charges. The insurance charges will be the dependent parameter.

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

Ans: The total number of rows is 1338

The total number of columns is 6

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

Ans: Converted ordinal data (sex and smoker) to numerical data as phyton algorithm cannot interpret categorical 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.

Ans: The good model is based on Random Forest, which has a R2 score of 0.86 or 86% with the following as input parameter.

*n\_estimators=100, criterion='friedman\_mse', max\_features= None, random\_state=0. R2=0.86*

5.) All the research values (r2\_score of the models) should be documented.

Ans: PDF uploaded in Github

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

Ans: The finalized model is Random Forest, based on the fact that it has the highest R2 score of 86%.