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

Machine Learning

Supervised Learning

Regression

2.) Tell basic info about the dataset (Total number of rows, columns=1338 rows × 6 columns)

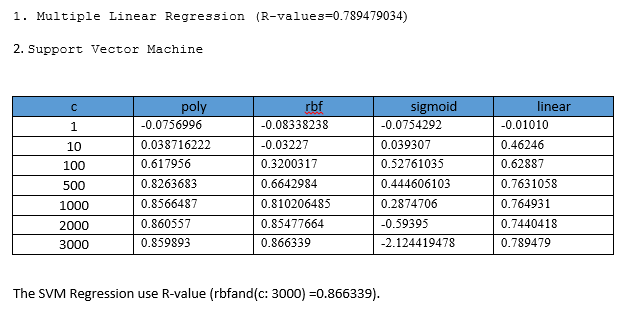
 Total **number of rows& columns**: [independent=1338 rows × 5 columns][ dependent= 1338 rows × 1 columns].

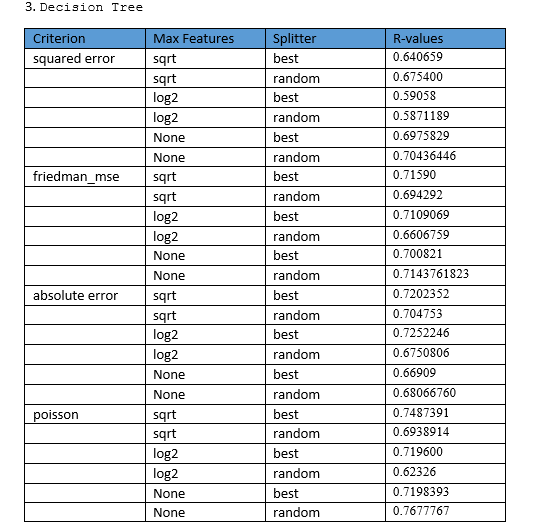
3.) Mention the pre-processing method if you’re doing any (like converting string to number – nominal data)

Convert categorical data (e.g., gender, smoker) into numerical format using techniques like one-hot encoding or label encoding.

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.

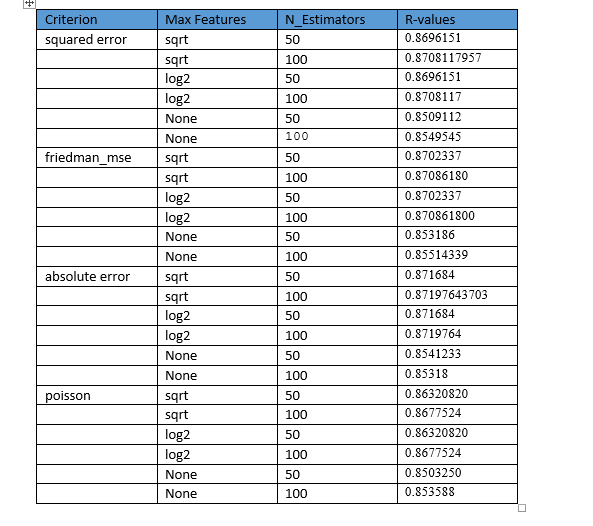
5.) All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

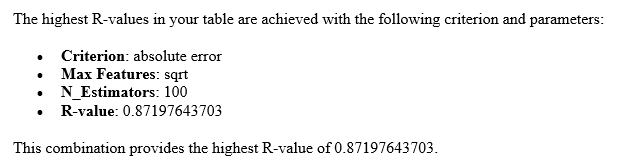












6.) Mention your final model, justify why u have chosen the same.

#### Final Model and Justification

The **Random Forest Regression** model with the following parameters:

* **Criterion**: Absolute Error
* **Max Features**: sqrt
* **N\_Estimators**: 100

This model was chosen because it achieved the highest R2 Score of 0.87197643703 among all the models tested.