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

Insurance company provided the dataset which contains their customers information such as sex, age, children's details, smoking habits and the corresponding insurance charge they collected. Based on the history we need to build a model which can predict the charges.

Below are the details about dataset and modelling.

1. Machine learning works because the input is numerical
2. This comes under supervised learning, the input and the output is defined. And the data is complete.
3. Regression model – the expected output is a numerical value

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

1. Total number of inputs 5

2. Expected output 1

3. Number of rows : 1338

4. number of columns 6

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

1. we need to do pd.get\_dummies(dataset,drop\_first = True) because the smoking parameter is in string. And Sex is male or female.

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.)

Random forest Regressor is chosen as best model with below mentioined parameters.

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

This gives r score value as 0.873 which is between 0 and 1 and nearly 1 compared to others

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S. No | Criterion | N\_estimators | Max features | R score |
| **11** | **Absolute\_error** | **100** | **Sqrt** | **0.873** |
| **4** | **squared\_error** | **100** | **Log2** | **0.873** |