**Assignment for Insurance project**

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.

Solution:

After analyse the customers data, we have to predict the insurance charges.

1.) Identify your problem statement

- We have clear set of data's with numbers so will take ML

Domain Selection- Machine Learning

- We have clear requirements to predict the insurance charges

Learning selection- Supervised learning

- The requirements for the prediction is numeric value.

Departments – Regression

2.) Tell basic info about the dataset

   Columns – 6
   Rows - 1338

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

Since the values are in words we converting words to number using One Hot Encoding by 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.

1 Multiple Linear Regression

```
R Score = 0.789479034986701
```

| S No | Hyper parameter | Linear | RBF(Non Linear) | Poly | Sigmoid |
|---|---|---|---|---|---|
| | **Support Vector Machine R_Score** | | | | |
| 1 | C=0.1 | −0.080959968 | −0.08970665 | −0.088 | −0.088 |
| 2 | C=10 | 0.628879286 | −0.08970665 | 0.0387 | 0.0393 |
| 3 | C=100 | 0.628879286 | 0.320031783 | 0.618 | 0.5276 |
| 4 | C=1000 | 0.764931174 | 0.810206491 | 0.8566 | 0.2875 |

| Decision Tree | | | |
|---|---|---|---|
| Sno | Crterion | Splitter | R Value |
| 1 | mae | best | 0.6900954 |
| 2 | mae | random | 0.745554277 |
| 3 | friedman_mse | best | 0.696553742 |
| 4 | friedman_mse | random | 0.697402953 |

| Random forest ensamble | | |
|---|---|---|
| S No | estimator | R Value |
| 1 | 50 | 0.849637541 |
| 2 | 100 | 0.853703664 |

The highest R score we have got is Random forest ensemble model but still we have received below score value.