**Requirement Analysis for the Insurance Data set Provided by the client**

**Problem Statement:**

Client Needs to predict the insurance rates for the new customers based on the existing customers data.

**Solution:**

As Client has provided the sufficient existing data set of count nearly 1350 records and also the requirements are clear on the input and output variables. We can build a machine learning model since majority of the inputs and outputs are numeric values.

**Domain Selected** : **Machine Learning** (As majority of the inputs and outputs are numeric values.)

**Learning Method Selected** : **Supervised Learning** (Requirements are clear with input and output variables with sufficient amount of past data provided)

**Machine Learning Model**: **Regression Model** ( As the expected output variable is continuous numerical value)

**Regression Models created and validated as part of Model Development:**

- **Multiple Linear Regression**

- **Support Vector Machine (Works for both classification and regression problems)**

- **Decision Tree** (**Works for both classification and regression problems**)

- **Random Forest** (**Works for both classification and regression problem**)

**Note:- Random Forest comes under the ensemble models**

**Model Evaluation Metric : R2\_SCORE**

**Exploratory Data Analysis: Refer the below file for further investigation on data-set   
Exploratory Data Analysis Document.ipynb**

**Final Model: As SVM model has highest r2\_score on predictions we have finalized SVM-Regressor as a production model for this problem statement**

**Below are the r2\_scores of each model developed and tested**

**- Multiple Linear - Regression --->**  r2\_score = 0.789

- **Support Vector Machine Regressor -->**r2\_score = 0.877

- **Decision Tree - Regressor ----->** r2\_score = 0.745

- **Random Forest Regressor ----->** r2\_score = 0.875

**Application Name : Life insurance rate predictor**