**MACHINE LEARNING REGRESSION MODEL SELECTION 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.) Problem statement Identification- **Here we need to calculate the Insurance charges. So dependent variable is CHARGES and remaining all fields will be independent variables**

2.) Basic info about the dataset (Total number of rows, columns): **Here we have 1338 rows × 6 columns**

3.) Mention the pre-processing method if you’re doing any (like converting string to number – nominal data): **After preprocessing we are having the**

**following columns** (['age', 'bmi', 'children', 'charges', 'sex\_male', 'smoker\_yes']

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.- **Model created**

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

**Model r2 value documentation :**

**R2 value comparison between various models**

**Multiple Linear Regression:**

|  |  |
| --- | --- |
| **Model** | **R2 value** |
| Multiple Linear regression | 0.80 |

**Support Vector Machine:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SNO** | **C**  **PARAMETER** | **LINEAR**  **R2value** | **POLY**  **R2value** | **RBF**  **R2 value** | **SIGMOD**  **R2 value** |
| 1 | Default | -0.14 | -0.09 | -0.11 | -0.11 |
| 2 | 5 | -0.08 | -0.13 | -0.11 | -1 |
| 3 | 10 | 0.004 | -0.14 | -0.12 | -2 |
| 4 | 500 | 0.62 | -0.12 | -0.14 | -5 |
| 5 | 1000 | 0.75 | -0.09 | -0.14 | -9 |
| 6 | 2000 | 0.74 | -0.01 | -0.15 | -11 |

\*Default C value=1

**DecisionTree:**

|  |  |  |  |
| --- | --- | --- | --- |
| **SN0** | **CRITERION** | **SPLITTER** | **R2value** |
| 1 | Default | Default | 0.66 |
| 2 | Squared\_error | best | 0.64 |
| 3 | Squared\_error | Random | 0.85 |
| 4 | friedman\_mse | best | 0.66 |
| 5 | friedman\_mse | Random | 0.78 |
| 6 | absolute\_error | best | 0.85 |
| 7 | absolute\_error | Random | 0.83 |
| 8 | poisson | best | 0.74 |
| 9 | poisson | Random | 0.80 |

**Random Forest Regression :**

|  |  |
| --- | --- |
| **Model** | **R2 value** |
| RF regression | 0.90 |

6.) Final model: **Models created. Decision tree using Random forest is the final model Since it is having r2 value as 0.90 which is a better value when compared to other models. So it is finalized.**