**Assignment – Regression Algorithm**

**Problem Statement:**

Customer wants to predict the insurance charges based on several parameters.

**Problem Identification:**

Machine Learning - Supervised – Regression

**Basic Info about dataset:**

The dataset consist of 6 Columns and 1338 Rows. In this we are going to consider 6 columns for predicting the output,

**Input data:** Age, Sex, bmi, No of Child, Smoker

**Output data:** Insurance Charges

This dataset consist of two nominal data columns, here we have to convert string to numbers for creating best model

**Machine Learning Regression**

**Phase 01: R Value**

1. **Multiple Linear Regression**

**R Value** = 0.7894

1. **Support Vector Machine (SVM)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S NO** | **HYPER PARAMETER** | **LINEAR**  **(R Value)** | **RBF**  **(R Value)** | **POLY**  **(R Value)** | **SIGMOID**  **(R Value)** |
| 1 | C10 | -0.9050 | -0.0804 | -7374 | -0.0894 |
| 2 | C100 | -9.1717 | -0.0212 | -1868 | -0.4782 |
| 3 | C1000 | -3.0566 | -0.0804 | -4897 | -9.7875 |

1. **Decision Tree**

|  |  |  |  |
| --- | --- | --- | --- |
| **S NO** | **CRITERION** | **SPLITTER** | **R VALUE** |
| 1 | Squared\_error | best | 0.6816 |
| 2 | Squared\_error | random | 0.6648 |
| 3 | poisson | best | 0.6679 |
| 4 | poisson | random | 0.6467 |
| 5 | Friedman\_mse | best | 0.6877 |
| 6 | Friedman\_mse | Random | 0.7151 |

1. **Random Forest**

|  |  |  |  |
| --- | --- | --- | --- |
| **S No** | **n estimators** | **Criterion** | **R VALUE** |
| 1 | 50 | squared\_error | 0.8498 |
| 2 | 100 | squared\_error | 0.8539 |
| 3 | 50 | absolute\_error | 0.8529 |
| 4 | 100 | absolute\_error | 0.8521 |
| 5 | 50 | friedman\_mse | 8.8499 |
| 6 | **100** | **friedman\_mse** | **0.8540** |
| 7 | 50 | poisson | 0.8279 |
| 8 | 100 | poisson | 0.8332 |

**Best Model :**

**Random Forest – 0.8540 (Using criterion – friedman\_mse & n\_estimator -100)**