**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** | **GAMMA** | **LINEAR**  **(R Value)** | **RBF**  **(R Value)** | **POLY**  **(R Value)** | **SIGMOID**  **(R Value)** |
| 1 | C10 | Scale | -0.9050 | -0.0804 | -7374 | -0.0894 |
| 2 | C10 | auto | -0.1795 | -0.1021 | -1.021 | -0.1019 |
| 3 | C100 | Scale | -9.1717 | -0.0212 | -1868 | -0.4782 |
| 4 | C100 | auto | -37.194 | -0.1024 | -6073 | -0.1017 |
| 5 | C1000 | Scale | -3.0566 | -0.0804 | -4897 | -9.7875 |
| 6 | C1000 | auto | -1564 | -0.1055 | -5562 | -0.0931 |

1. **Decision Tree**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S NO** | **CRITERION** | **MAX FEATURES** | **SPLITTER** | **R VALUE** |
| 1 | Squared\_error | auto | Best | 0.707 |
| 2 | Squared\_error | sqrt | Best | 0.404 |
| 3 | Squared\_error | log2 | Best | 0.744 |
| 4 | Squared\_error | auto | Random | 0.710 |
| 5 | Squared\_error | sqrt | Random | 0.565 |
| 6 | Squared\_error | log2 | Random | 0.699 |
| 7 | Poisson | auto | Best | 0.680 |
| 8 | Poisson | sqrt | Best | 0.673 |
| 9 | Poisson | log2 | Best | 0.600 |
| 10 | Poisson | auto | Random | 0.638 |
| 11 | Poisson | sqrt | Random | 0.617 |
| 12 | Poisson | log2 | Random | 0.603 |
| 13 | Friedman\_mse | auto | Best | 0.692 |
| 14 | Friedman\_mse | sqrt | Best | 0.616 |
| 15 | Friedman\_mse | log2 | Best | 0.740 |
| 16 | Friedman\_mse | auto | Random | 0.748 |
| 17 | Friedman\_mse | sqrt | Random | 0.723 |
| 18 | Friedman\_mse | log2 | Random | 0.657 |

1. **Random Forest**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S No** | **n estimators** | **Max\_features** | **Criterion** | **R VALUE** |
| 1 | 50 | none | squared\_error | 0.8498 |
| 2 | 50 | sqrt | squared\_error | 0.8694 |
| 3 | 50 | log2 | squared\_error | 0.8694 |
| 4 | 100 | none | squared\_error | 0.8539 |
| 5 | 100 | sqrt | squared\_error | 0.8709 |
| 6 | 100 | log2 | squared\_error | 0.8709 |
| 7 | 50 | none | absolute\_error | 0.8529 |
| 8 | **50** | **sqrt** | **absolute\_error** | **0.8721** |
| 9 | **50** | **log2** | **absolute\_error** | **0.8721** |
| 10 | 100 | none | absolute\_error | 0.8521 |
| 11 | 100 | sqrt | absolute\_error | 0.8717 |
| 12 | 100 | log2 | absolute\_error | 0.8717 |
| 13 | 50 | none | friedman\_mse | 0.8499 |
| 14 | 50 | sqrt | friedman\_mse | 0.8700 |
| 15 | 50 | log2 | friedman\_mse | 0.8700 |
| 16 | 100 | none | friedman\_mse | 0.8540 |
| 17 | 100 | sqrt | friedman\_mse | 0.8709 |
| 18 | 100 | log2 | friedman\_mse | 0.8709 |
| 19 | 50 | none | Poisson | 0.8279 |
| 20 | 50 | sqrt | Poisson | 0.8287 |
| 21 | 50 | log2 | Poisson | 0.8287 |
| 22 | 100 | none | Poisson | 0.8332 |
| 23 | 100 | sqrt | Poisson | 0.8293 |
| 24 | 100 | log2 | Poisson | 0.8293 |

**Best Model :**

**Random Forest – 0.8721 (Using criterion – absolute\_error, max\_feature – sqrt (or) log2 & n\_estimator - 50)**