**ASSIGNMENT-REGRESSION ALGORITHM**

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:**

**Stage1**: Since the dataset is in numerical form we are going to use **Machine learning Algorithm** for prediction.

**Stage2:** The requirements from the client are very clear having the dataset with clear input and output values. Hence it falls under **Supervised Learning**

**Stage3:** The output is in numerical form. So that we are going to use **Regression Algorithm**

**2. Basic Information**

The dataset contains 1338 rows × 6 columns

**3.Data-Pre Processing**

The dataset contains two columns of categorical data(**Nominal data**).Therefore it has to be converted into numerical values using **One Hot Encoding Method** and then moved to further calculation

**4.Hyper Tuning Parameters**

**1 .Multiple Linear Regression**

|  |  |
| --- | --- |
| **Standardisation** | **R2 Value** |
| yes | 0.7894790349867009 |

**2. Support Vector Machine**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Standardisation** | **Kernel** | **C** | **R2 Value** |
| 1 | No | linear | 10 | -0.0016176324886458815 |
| 2 | No | linear | 100 | 0.5432818196692782 |
| 3 | No | linear | 1000 | 0.634036931263208 |
| 4 | Standardized | linear | 1000 | 0.7649311738596382 |
| 5 | **Standardized** | **poly** | **1000** | **0.8566487675946569** |
| 6 | Standardized | poly | 100 | 0.6179569624059795 |
| 7 | Standardized | rbf | 100 | 0.3200317832050831 |
| 8 | Standardized | rbf | 1000 | 0.8102064851758545 |
| 9 | Standardized | sigmoid | 1000 | 0.28747069486976207 |
| 10 | Standardized | sigmoid | 100 | 0.5276103546510404 |
| 11 | Standardized | sigmoid | 200 | 0.5455532067863539 |

**3. Decision Tree**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Standardisation** | **Criterion** | **Splitter** | **R2 Value** |
| 1 | No | friedman\_mse | random | 0.6738959768023087 |
| 2 | Standardized | friedman\_mse | random | 0.7080677820542198 |
| 3 | Standardized | friedman\_mse | best | 0.6858772546514931 |

**4. Random Forest**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Standardisation** | **N\_estimators** | **Criterion** | **R2 Value** |
| 1 | Standardized | 50 | - | 0.850961172329626 |
| 2 | Standardized | 100 | - | 0.855040880868016 |
| 3 | Standardized | 100 | Friedman\_mse | 0.8550986344127547 |
| 4 | Standardized | 1000 | Friedman\_mse | 0.8548388977480252 |
| 5 | Standardized | 10 |  | 0.8337914810384686 |
| 6 | Standardized | 100 | Friedman\_mse | 0.8540051307771396 |

**5.Final Model**

**Support Vector Machine** algorithm model gives the maximum prediction value of **85.6%** of accuracy while checking with standardization technique as well as with hyper-tuned parameters.