1. Predict Insurance Charges
2. Dataset Total Number of Rows and Columns **1338 Rows x 6 Columns**
3. The Pre-processing method will depend on the dataset.it involves handling missing values encoding categorical variables , scaling numerical features and splitting the dataset into training and testing sets.

Once the dataset is available, I can determine the pre-processing steps.

1. To develop a model for predicting insurance charges various machine learning algorithms can be explored.It is used to Muliple Linear Regression, Decision tree, Random Forest algorithms used the R2 value.
2. Once the dataset is provided and the models are trained,I will document the R2 scores of each model. This can be done the tabular format.

6.Final model is based on R2 scores obtained, the final model will be selected.

**1.MULTIPLE LINEAR REGRESSION**

**R2 Value** is = **0.78651**

**2.SUPPORT VECTOR REGRESSION**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.NO** | **HYPER PARAMETER** | **LINEAR** | **RBF** | **POLY** | **SIGMOID** |
| 1 | C=100 | 0.62887 | 0.32003 | 0.61795 | 0.52761 |
| 2 | C=500 | 0.7631 | 0.66429 | 0.82636 | 0.4446 |

The SVM RegressIon use R2 value (**POLY**) and hyper parameter C=500 is = **0.82636**

**3.DECISION TREE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.NO** | **CRITERION** | **SPLITTER** | **R2 VALUE** | **SPLITTER** | **R2 VALUE** |
| 1 | squared\_error | best | 0.71625 | random | 0.67742 |
| 2 | friedman\_mse | best | 0.71138 | random | 0.73534 |
| 3 | absolute\_error | best | 0.64001 | random | 0.68667 |
| 4 | poisson | best | 0.72825 | random | 0.73068 |

The Decision Tree Regression use R2 value (**Criterion,Splitter**) friedman\_mse and Random is = **0.73534**

**4.RANDOM FOREST**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **CRITERION** | **n\_estimators** | **random\_state** | **R2 VALUE** |
| 1 | squared\_error | 50 | 0 | 0.8496 |
| 2 | absolute\_error | 50 | 0 | 0.85364 |
| 3 | friedman\_mse | 50 | 0 | 0.8497 |
| 4 | poisson | 50 | 0 | 0.849333 |

The Random Forest Regression use R2 value (**Criterion,n\_estimators,random\_state**) absolute\_error,50 and 0 = **0.85364**

**The Final Machine Learning best method of Regression:**

**Random Forest R2 value** (**Criterion,n\_estimators,random\_state**) = **0.85364**