**Problem Identification:**

Stage 1:

Domain Selection – Machine Learning

Reason: Requirement output is Numeric only

Stage 2:

Learning Selection – Supervised Learning

Reason: Clear requirement and data set input and output well defined.

Stage 3:

Supervised Learning – Regression

Reason: Requirement expected output the numeric value.

**Data Set Details:**

Use the panda module **‘shape’** attribute find the rows and columns details.

Number of Row : 6

Number of Column : 1338

**Pre-Processing the Data Set:**

In the input dataset, some of the columns contain categorical data. Therefore, we need to convert them to numeric data using the nominal 'One Hot Encoding' algorithm.

**Multi** **Linear Regression:**

R² Score values is: 0.7894790349867009

**Support Vector Machine:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **Hyper Parameter** | **Linear R² Value** | **RBR R² Value** | **Poly R² Value** | **Sigmoid R² Value** |
| 1 | C=1.0 | -0.010102665 | -0.083382386 | -0.075699656 | -0.075429243 |
| 2 | C=10 | 0.462468414 | -0.032273294 | 0.038716223 | 0.039307144 |
| 3 | C=100 | 0.628879286 | 0.320031783 | 0.617956962 | 0.527610355 |
| 4 | C=500 | 0.763105805 | 0.664298465 | 0.826368354 | 0.444606103 |
| 5 | C=1000 | 0.764931174 | 0.810206485 | 0.856648768 | 0.287470695 |

The SVM prediction uses a Poly kernel and C=1000. The R² value is 0.8566.

**Decision Tree Regression:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Hyper Parameter** | **splitter= best R²** | **splitter= random R²** |
| 1 | squared\_error | 0.691828288 | 0.633761209 |
| 2 | friedman\_mse | 0.701261502 | 0.771053222 |
| 3 | absolute\_error | 0.669913656 | 0.72477199 |
| 4 | Poisson | 0.726750813 | 0.680477697 |

The Decision tree prediction uses a criterion of 'Friedman\_mse' and a splitter set to 'random'. The R² value is 0.7710.

**Random Forest Regression:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.No** | **n\_estimators** | **squared\_error** | **friedman\_mse** | **absolute\_error** | **Poisson** |
| 1 | 100 | 0.848807706 | 0.852324013 | 0.854778956 | 0.853805422 |
| 2 | 200 | 0.854486062 | 0.853637168 | 0.852425236 | 0.854331393 |
| 3 | 300 | 0.853333235 | 0.855604103 | 0.852918364 | 0.852869132 |
| 4 | 400 | 0.855111429 | 0.855006219 | 0.855277306 | 0.854338332 |
| 5 | 600 | 0.855737523 | 0.853385465 | 0.855237048 | 0.854948928 |
| 6 | 800 | 0.855671249 | 0.855728861 | 0.855606813 | 0.855226635 |
| 7 | 1000 | 0.85572907 | 0.854337787 | 0.855367749 | 0.854705117 |

The random forest prediction uses a criterion of ‘**squared\_error**’and n\_estimators set to **600**. The R² value is 0.85573."

**Final Model:**

We have predicted the final model using a Support Vector Machine algorithm with the criterion set to ' Poly ' and C set to 1000. The R² value is 0.85664. Because this model's R² value is close to the maximum value of 1