**Assignment**

# Problem statement/Identification of Problem in stages:

Stage 1: Machine Learning

Stage 2: Supervised Learning

Stage 3: Regression

# Basic info of Dataset:

Total No. of ROWS: 1338 rows

Total No. of COLUMNS: 6 columns

# Pre-processing of Dataset:

In Dataset some inputs were present in categorical form. So for converting it into Numerical value here Nominal Data is used. In that One-Hot Encoding method is used.

**1.Multiple Linear Regression** (R\_Score value)=**0.78948**

**2.Support Vector Machine – REGRESSION = 0.86634**

**Without Standardisation(R\_Score value) = -0.08338**

**Standardised:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SI.No** | **HYPER PARAMETER** | **LINEAR** | **rbf** | **Poly** | **sigmoid** |
| **1** | **C=1.0** | -0.0101 | -0.0834 | -0.0757 | -0.0754 |
| **2** | **C=10.0** | 0.4625 | -0.0323 | 0.0387 | 0.0393 |
| **3** | **C=100.0** | 0.6289 | 0.3200 | 0.6180 | 0.5276 |
| **4** | **C=500.0** | 0.7631 | 0.6643 | 0.8264 | 0.4446 |
| **5** | **C=1000.0** | 0.7649 | 0.8102 | 0.8566 | 0.2875 |
| **6** | **C=2000.0** | 0.7440 | 0.8548 | 0.8606 | -0.5940 |
| **7** | **C=3000.0** | 0.7414 | 0.8663 | 0.8599 | -2.1244 |

SVM Regression using hyper tuning parameter with **C**=**3000** in **rbf** has **0.8663 Highest Accuracy.**

**3. Decision TREE– REGRESSION = *0.76936***

**Without passing any Arguments(R\_Score value) = 0.70804**

**Standardised:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SI.No** | **criterion** | **splitter** | **max\_features** | **r2\_score** |
| **1** | ***squared\_error*** | ***best*** | ***None*** | 0.68408 |
| **2** | ***squared\_error*** | ***Random*** | ***None*** | 0.79495 |
| **3** | ***squared\_error*** | ***Best*** | ***Sqrt*** | 0.57367 |
| **4** | ***squared\_error*** | ***Random*** | ***Sqrt*** | 0.70649 |
| **5** | ***squared\_error*** | ***Best*** | ***log2*** | 0.65933 |
| **6** | ***squared\_error*** | ***random*** | ***log2*** | 0.65196 |
| **7** | ***friedman\_mse*** | ***best*** | ***None*** | 0.69270 |
| **8** | ***friedman\_mse*** | ***Random*** | ***None*** | 0.69480 |
| **9** | ***friedman\_mse*** | ***Best*** | ***Sqrt*** | 0.64866 |
| **10** | ***friedman\_mse*** | ***Random*** | ***Sqrt*** | 0.65249 |
| **11** | ***friedman\_mse*** | ***Best*** | ***log2*** | 0.67241 |
| **12** | ***friedman\_mse*** | ***random*** | ***log2*** | 0.69696 |
| **13** | ***absolute\_error*** | ***best*** | ***None*** | 0.69782 |
| **14** | ***absolute\_error*** | ***Random*** | ***None*** | 0.73120 |
| **15** | ***absolute\_error*** | ***Best*** | ***Sqrt*** | 0.75919 |
| **16** | ***absolute\_error*** | ***Random*** | ***Sqrt*** | 0.72237 |
| **17** | ***absolute\_error*** | ***Best*** | ***log2*** | 0.76936 |
| **18** | ***absolute\_error*** | ***random*** | ***log2*** | 0.55106 |
| **19** | ***Poisson*** | ***best*** | ***None*** | 0.73063 |
| **20** | ***Poisson*** | ***Random*** | ***None*** | 0.73191 |
| **21** | ***Poisson*** | ***Best*** | ***Sqrt*** | 0.72247 |
| **22** | ***Poisson*** | ***Random*** | ***Sqrt*** | 0.67312 |
| **23** | ***Poisson*** | ***Best*** | ***log2*** | 0.69654 |
| **24** | ***poisson*** | ***random*** | ***log2*** | 0.68842 |

Decision Tree Regression using hyper tuning parameter with ***absolute\_error, Best, Log2 has highest Accuracy of about 0.76936***

**4. RANDOM FOREST = *0.*87711**

**Without passing any Arguments (R\_Score value) = 0.84685**

**Standardised:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SI.No** | **criterion** | **N**\_**estimator** | **max\_features** | **r2\_score** |
| **1** | ***squared\_error*** | ***50*** | ***None*** | 0.84951 |
| **2** | ***squared\_error*** | ***100*** | ***None*** | 0.85720 |
| **3** | ***squared\_error*** | ***50*** | ***Sqrt*** | 0.86884 |
| **4** | ***squared\_error*** | ***100*** | ***Sqrt*** | 0.87128 |
| **5** | ***squared\_error*** | ***50*** | ***log2*** | 0.86415 |
| **6** | ***squared\_error*** | ***100*** | ***log2*** | 0.86734 |
| **7** | ***friedman\_mse*** | ***50*** | ***None*** | 0.85319 |
| **8** | ***friedman\_mse*** | ***100*** | ***None*** | 0.84900 |
| **9** | ***friedman\_mse*** | ***50*** | ***Sqrt*** | 0.87192 |
| **10** | ***friedman\_mse*** | ***100*** | ***Sqrt*** | 0.86825 |
| **11** | ***friedman\_mse*** | ***50*** | ***log2*** | 0.87156 |
| **12** | ***friedman\_mse*** | ***100*** | ***log2*** | 0.86794 |
| **13** | ***absolute\_error*** | ***50*** | ***None*** | 0.85374 |
| **14** | ***absolute\_error*** | ***100*** | ***None*** | 0.85537 |
| **15** | ***absolute\_error*** | ***50*** | ***Sqrt*** | 0.86559 |
| **16** | ***absolute\_error*** | ***100*** | ***Sqrt*** | 0.87139 |
| **17** | ***absolute\_error*** | ***50*** | ***log2*** | 0.87408 |
| **18** | ***absolute\_error*** | ***100*** | ***log2*** | 0.87210 |
| **19** | ***Poisson*** | ***50*** | ***None*** | 0.85242 |
| **20** | ***Poisson*** | ***100*** | ***None*** | 0.85255 |
| **21** | ***Poisson*** | ***50*** | ***Sqrt*** | 0.87711 |
| **22** | ***Poisson*** | ***100*** | ***Sqrt*** | 0.87135 |
| **23** | ***Poisson*** | ***50*** | ***log2*** | 0.86998 |
| **24** | ***poisson*** | ***100*** | ***log2*** | 0.86990 |

Random Forest Regression using hyper tuning parameter with ***Poisson, 50, Sqrt has highest Accuracy of about 0.*87711**

# From all the models using different parameters passed in it, came to an conclusion that

# **Random Forest Regression has highest Accuracy of about 87%**