1)PROBLEM STATEMENT IDENTIFICATION:

Step-1: Machine Learning

Step-2: Supervised learning

Step-3: Supervised-Regression

2)Total no of Rows-1339 ; Total no of Columns-6

3)Pre-Processing method: Nominal data method-One Hot Encoding method

4) MLR:- R^2 Value=0.7894790

SVM:-

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SNO | HYPER  PARAMETER | LINEAR  R-Value | RBF  R-Value | SIGMOID  R-Value | POLY  R-Value |
| 1 | C0.1 | -0.08095996 | -0.0890745 | -0.08826991 | -0.08830237 |
| 2 | C10 | 0.462468414 | -0.0322732 | 0.039307143 | 0.038716222 |
| 3 | C1000 | 0.764931173 | 0.81020648 | 0.287470694 | 0.856648767 |
| 4 | C2000 | 0.744041830 | 0.8547766 | -0.59395097 | 0.860557927 |
| 5 | C3000 | 0.741423659 | 0.86633939 | -2.12441947 | 0.859893008 |
| 6 | C4000 | 0.741419880 | 0.87174078 | -5.5103335 | 0.860004953 |

* The SVM Regression R^2value (RBF and hyperparameter C=4000)=0.87174078

DECISION TREE MODEL:-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.NO | CRITERION | SPLITTER | MAX-FETURES | R-SCORE |
| 1 | Squared\_error | best | sqrt | 0.6869230267185223 |
| 2 | Squared\_error | best | Log2 | 0.7471257821177038 |
| 3 | Squared\_error | best | none | 0.7148950565398933 |
| 4 | Squared\_error | random | sqrt | 0.7142263278051157 |
| 5 | Squared\_error | random | Log2 | 0.7144983777755 |
| 6 | Squared\_error | random | none | 0.6912733088690999 |
| 7 | Friedman\_mse | best | sqrt | 0.7218069779405167 |
| 8 | Friedman\_mse | best | Log2 | 0.6977557687829694 |
| 9 | Friedman\_mse | best | none | 0.6948880597155574 |
| 10 | Friedman\_mse | random | sqrt | 0.6947177012541464 |
| 11 | Friedman\_mse | random | Log2 | 0.6883950220347903 |
| 12 | Friedman\_mse | random | none | 0.7143453312281522 |
| 13 | Absolute error | best | sqrt | 0.7023684287476444 |
| 14 | Absolute error | best | Log2 | 0.8007148050626391 |
| 15 | Absolute error | best | none | 0.6977749738237873 |
| 16 | Absolute error | random | sqrt | 0.6794294251590102 |
| 17 | Absolute error | random | Log2 | 0.6563833825860289 |
| 18 | Absolute error | random | none | 0.7564863007662027 |
| 19 | Poisson | best | sqrt | 0.4988677509484797 |
| 20 | Poisson | best | Log2 | 0.6584386874059316 |
| 21 | Poisson | best | none | 0.7056210521632514 |
| 22 | Poisson | random | sqrt | 0.5548878008478138 |
| 23 | Poisson | random | Log2 | 0.7205173872214099 |
| 24 | Poisson | random | none | 0.7123924623514243 |

* In the Decision Tree Regression use R^2 value as (criterion='absolute\_error',splitter='best', max\_features=Log2)= 0.80071480, is the best model.

RANDOM FOREST:-

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S:NO | CRITERION | N-Estimators | MAX-FEATURES | R^2 SCORE |
| 1 | Squared\_error | 50 | None | 0.856119053508626 |
| 2 | Squared\_error | 100 | None | 0.854039046826665 |
| 3 | Squared\_error | 50 | sqrt | 0.866766154217111 |
| 4 | Squared\_error | 100 | Sqrt | 0.870857688319025 |
| 5 | Squared\_error | 50 | Log2 | 0.8674929312562272 |
| 6 | Squared\_error | 100 | Log2 | 0.8714752230065679 |
| 7 | Friedman-mse | 50 | None | 0.8492195911865059 |
| 8 | Friedman-mse | 100 | None | 0.8515132011979629 |
| 9 | Friedman-mse | 50 | Sqrt | 0.865796357440343 |
| 10 | Friedman-mse | 100 | Sqrt | 0.8715319251042227 |
| 11 | Friedman-mse | 50 | Log2 | 0.8666822654338064 |
| 12 | Friedman-mse | 100 | Log2 | 0.8690370029659661 |
| 13 | Absolute-error | 50 | None | 0.8547055693110439 |
| 14 | Absolute-error | 100 | None | 0.8568091365219767 |
| 15 | Absolute-error | 50 | Sqrt | 0.8711583460603914 |
| 16 | Absolute-error | 100 | Sqrt | 0.8715034572612606 |
| 17 | Absolute-error | 50 | Log2 | 0.867962160618138 |
| 18 | Absolute-error | 100 | Log2 | 0.8719890281832289 |
| 19 | Poisson | 50 | None | 0.8526726249938772 |
| 20 | Poisson | 100 | None | 0.8521342502179741 |
| 21 | Poisson | 50 | Sqrt | 0.8676223279719134 |
| 22 | Poisson | 100 | Sqrt | 0.8697522775786619 |
| 23 | Poisson | 50 | Log2 | 0.8685082450317689 |
| 24 | Poisson | 100 | Log2 | 0.8730626737515996 |

* The Random Forest Regression R^2value (Criteron-Poisson;N-estimator=100;Max-features-Log2)==0. 8730626737

FINAL OUTPUT:

\*The best Final model is- Random forest, Using the hyperparameters:(Criterion-Poisson;N-estimator-100;Max-features-Log2,)its R^2 score value is 0.873062,So this is the final highest value when comparing to all other algorithms.

\*So I’m Deploying for the Best model (i.e) -Random forest method