

REGRESSION ASSIGNMENT RESEARCH PAPER

1) PROBLEM STATEMENT

- Machine Learning
- Supervised Learning
- Regression

2) DATASET INFORMATION

This dataset has 1339 Rows and 6 Columns.

3) PRE-PROCESSING METHODS

In this module, I used Dummies and StandardScaler methods.

4) RESEARCH VALUES

1. MULTIPLE LINEAR REGRESSION – $r_score = 0.7894$

2. SUPPORT VECTOR MACHINE:

S.No	Hyper Parameter (r value)	RBF (r value)	Poly (r value)	Sigmoid (r value)
1	C=100	0.3906	0.7508	0.5275
2	C=500	0.6964	0.8593	0.4906
3	C=1000	0.8283	0.8605	0.1437
4	C=1500	0.8530	0.8599	-1.1491
5	C=2000	0.8607	0.8601	-2.5840
6	C=3000	0.8685	0.8600	-6.8861

In the SVM Algorithm, **RBF with Hyperparameter C3000** has the highest $r_score = 0.8605$.

3. DECISION TREE:

S.No	CRITERION	SPLITTER	R VALUE
1	Squared Error	Best	0.6981
2	Squared Error	Random	0.7222
3	Friedman Mse	Best	0.6875
4	Friedman Mse	Random	0.7367
5	Absolute Error	Best	0.6634
6	Absolute Error	Random	0.7140
7	Poisson	Best	0.7289
8	Poisson	Random	0.6185

In a Decision tree algorithm, **Friedman Mse with Random** has the highest r_score = **0.7367**.

4. RANDOM FOREST:

S.No	CRITERION	N_ESTIMATOR	R VALUE
1	Squared Error	10	0.8425
2	Squared Error	100	0.8539
3	Friedman Mse	10	0.8189
4	Friedman Mse	100	0.8554
5	Absolute Error	10	0.8373
6	Absolute Error	100	0.8520
7	Poisson	10	0.8406
8	Poisson	100	0.8563

In Random Forest algorithm, **Poisson with 100 n_estimator** has the highest r_score = **0.8563**.

5) FINAL MODEL

Support Vector Machine is the best algorithm for this model.
Because, among the 4 algorithms, SVM gives the best r_score.