Regression Assignment

1.Problem Statement:-

Predict insurance charges

2.Data set basic info:-

Total no.of Rows = 1338

Total no.of Columns = 6

3.Pre-processing method:-

Out of six, two columns have nominal data(categorical) data. So, it needs to be converted to numerical data by using "get_dummies()" function

4. Model Development:-

Multiple Linear Regression, Support Vector Machine, Decision Tree and Random Forest algorithms have been used to develop a model with r2_Score.

5.Research values of each algorithm:-

a)Multiple Linear Regression – R Value = 0.7894

b)Support Vector Machine Matrix:-

S.No.	Kernel	C	R Value
1	linear	1	-1.4368
2	linear	10	-113.0486
3	linear	100	-146.1431
4	poly	1	-12266.2045
5	poly	10	-1163348.2386
6	poly	100	-32979013.981
7	rbf	1	-0.0883
8	rbf	10	-0.0804
9	rbf	100	-0.0212
10	sigmoid	1	-0.0734
11	sigmoid	10	0.0131
12	sigmoid	100	-0.5435

c)Decision Tree Matrix:-

S.No.	Criterion	Max Features	Splitter	R Value
1	squared_error	auto	best	-1.3929
2	squared_error	auto	random	-1.8454
3	squared_error	sqrt	best	-0.7769
4	squared_error	sqrt	random	-0.8045
5	squared_error	log2	best	-0.7529
6	squared_error	log2	random	-0.0242
7	friedman_mse	auto	best	-1.3929
8	friedman_mse	auto	random	-0.6949
9	friedman_mse	sqrt	best	-0.5503
10	friedman_mse	sqrt	random	0.4558
11	friedman_mse	log2	best	-0.9323
12	friedman_mse	log2	random	-0.5247

d)Random Forest Matrix:-

S.No.	Criterion	Max Features	N_Estimators	R Value
1	squared_error	auto	10	0.6984
2	squared_error	sqrt	100	0.7652
3	squared_error	log2		0.7469
4	friedman_mse	auto	10	0.7040
5	friedman_mse	sqrt	100	0.7323
6	friedman_mse	log2		0.7148

6.Final Model:-

Multiple Linear Regression has been considered as final model. This model provided highest r2_Score as 0.7894 among all other algorithms.