ASSIGNMENT REGRESSION ALGORITHM

INSURANCE PREMIUM CHARGES REPORT

I] PROBLEM STATEMENT.

➔ **Stage 1**: Machine Learning since its output is in Number.

➔ **Stage 2:** It's supervised learning since I/O and Requirement is clear.

➔ **Stage 3:** It comes under Regression since it contains continuous Numbers.

II] INFORMATION ABOUT THE DATASET.

➔ **Total No of Rows**: 6 Total No of Columns: 1340.

➔ **Column Data**: Totally **2** integers, Float and Objects present in the Data.

III] THE PRE-PROCESSING METHOD USED IN THE DATASET.

➔ **One Hot Encoder**: It's used to convert Nominal Data such as "**Sex and Smoker**" to Numerical data.

V] THE RESEARCH OF THE R2 VALUES IN DIFFERENT ALGORITHMS.

➔ **Multi Linear Regression**: 0.7894

➔ **Support Vector Machine**:

S.NO	HYPER PARAMETER	LINEAR	RBF	POLY	SIGMOID
1	C10	-0.0016	-0.0819	-0.0931	-0.0907
2	C100	0.543	-0.1248	-0.0097	-0.1181
3	C500	0.6270	-0.1246	-0.0820	-0.4562
4	C1000	0.6340	-0.1174	-0.0555	-1.6659
5	C2000	0.6893	-0.1077	-0.0027	-5.6164
6	C3000	0.7590	-0.0962	0.0489	-12.019
7	C7000	0.7629	-0.5030	0.2360	-59.941

➔ **Decision Tree**:

S.NO	CRITERION	MAX_FEATURES	SPLITTER	R VALUE

1	squared_error	None	Best	0.6911
2	squared_error	None	Random	0.7300
3	squared_error	Sqrt	Best	0.7054
4	squared_error	Sqrt	Random	**0.7511**
5	squared_error	Log2	Best	0.7341
6	squared_error	Log2	Random	0.6490
7	MAE	None	Best	0.6719
8	MAE	None	Random	0.7256
9	MAE	Sqrt	Best	0.7067
10	MAE	Sqrt	Random	0.6384
11	MAE	Log2	Best	0.7028
12	MAE	Log2	Random	0.6754
13	friedman_mse	None	Best	0.6780
14	friedman_mse	None	Random	0.6923
15	friedman_mse	Sqrt	Best	0.7127
16	friedman_mse	Sqrt	Random	0.6318
17	friedman_mse	Log2	Best	0.6324
18	friedman_mse	Log2	Random	0.6725

➔ **Random Forest**:

S.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R VALUE
1	squared_error	None	10	0.8446
2	squared_error	None	100	0.8583
3	squared_error	Sqrt	10	0.8370
4	squared_error	Sqrt	100	0.8698
5	squared_error	Log2	10	0.8578
6	squared_error	Log2	100	0.8704
7	MAE	None	10	0.8475
8	MAE	None	100	0.8553
9	MAE	Sqrt	10	0.8388
10	MAE	Sqrt	100	0.8724
11	MAE	Log2	10	0.8548
12	MAE	Log2	100	**0.8739**

IV] THE FINAL MODEL.

→ I PREFER TO CHOOSE RANDOM FOREST WHICH HAVE A HIGHER ACCURACY