ASSIGNMENT-REGRESSION ALGORITHM

Problem Statement or Requirement

A client's requirement is, he wants to predict the insurance charges based on the several parameters. The Client has provided the dataset of the same. As a data scientist, you must develop a model which will predict the insurance charges.

1) **PROBLEM STATEMENT:**
   i)Predict insurance charges

   ii)Machine Learning

   iii)Supervised Learning

   iv)Regression

2) **BASIC INFO ABOUT DATASET:**
   i)1338 rows × 6 columns

3) **PREPROCESSING METHOD:**

   i) Nominal data

To find following the machine learning regression method using in r2 value

1.Multiple Linear Regression (R2 value): 0.78947
2.Support Vector Machine

S.NO	HYPER PARAMETER	LINEAR	RBF	POLY	SIGMOID
1.	C=10	0.78947	-0.03227	0.03871	0.03930
2.	C=100	0.62887	0.32003	0.61795	0.52761
3.	C=500	0.76310	0.66429	0.82636	0.44460
4.	C=1000	0.76493	0.81020	0.85664	0.28747
5.	C=2000	0.74404	0.85477	0.86055	-0.59395
6	C=3000	0.74142	0.86633	0.85989	-2.12441

The svm regression (r2 value is **0.86633**, hyper parameter=3000, rbf)

3.Decision Tree

S.NO	Criterion	Splitter	Max_feature	R value
1.	Squared Error	best	None	0.69752
2	Squared Error	best	Sqrt	0.63371
3	Squared Error	best	Log2	0.75533
4	Squared Error	random	None	0.72340
5	Squared Error	random	Sqrt	0.71536
6	Squared Error	random	Log2	0.67698
7	Friedman mse	best	None	0.70212
8	Friedman mse	best	Sqrt	0.70848
9	Friedman mse	best	Log2	0.77441
10	Friedman mse	random	None	0.75598
12	Friedman mse	random	sqrt	0.67961
13	Friedman mse	random	Log2	0.63350
14	Absolute Error	best	None	0.65620
15	Absolute Error	best	sqrt	0.66319
16	Absolute Error	best	Log2	0.66319
17	Absolute Error	random	None	0.73143
18	Absolute Error	random	sqrt	0.71801

19	Absolute Error	random	Log2	0.75259
20	Poisson	best	None	0.73250
21	Poisson	best	sqrt	0.68757
22	Poisson	best	Log2	0.72786
23	Poisson	random	None	0.71618
24	Poisson	random	sqrt	0.71719
25	Poisson	random	Log2	0.68476

The Decision Tree (r2 value is 0.75598, criterion=friedman_mse, splitter=random, Max_feature= None)

4.Random Forest

s.no	criterion	N_Estimators	Max_feature	R value
1	Squared_ error	50	None	0.84983
2	Squared _error	50	sqrt	0.86958
3	Squared error	50	Log2	0.86958
4	Squared_erorr	100	None	0.85383
5	Squared_erorr	100	sqrt	0.87102
6	Squared_erorr	100	Log2	0.87102
7	Absolute error	50	None	0.85266
8	Absolute error	50	sqrt	0.87081
9	Absolute error	50	Log2	0.87081
10	Absolute error	100	None	0.85200
11	Absolute error	100	sqrt	0.87106
12	Absolute error	100	Log2	0.87106
13	Friedman_mse	50	None	0.85007
14	Friedman_mse	50	sqrt	0.87024
15	Friedman_mse	50	Log2	0.87024
16	Friedman_mse	100	None	0.85405

17	Friedman_mse	100	sqrt	0.87105
18	Friedman_mse	100	Log2	0.87105
19	Poisson	50	None	0.84910
20	Poisson	50	sqrt	0.86323
21	Poisson	50	Log2	0.86323
22	Poisson	100	None	0.85263
23	Poisson	100	sqrt	0.86801
24	Poisson	100	Log2	0.86801

Random Forest (r2 value is 0.87106, criterion=absolute error, n-estimators=100, max_features= sqrt)

The FINAL MODEL IS RANDOM FOREST

R2 VALUE IS 0.87106