

Problem Statement:

A client's requirement is, He wants to redirect the insurance charges based on the several parameters. The clients have provided dataset the same. As a data scientist you must develop the same model.

SL.No.	CRITERION	MAX FEATURES	SPLITER	R2_VALUE
1	friedman_mse	auto	best	0.702
2.	friedman_mse	auto	random	0.661
3.	friedman_mse	sqrt	best	0.738
4.	friedman_mse	sqrt	random	0.652
5.	friedman_mse	log2	best	0.727
6	friedman_mse	log2	random	0.636
7.	absolute_error(mae)	auto	best	0.679
8.	mae	auto	random	0.696
9.	mae	sqrt	best	0.659
10.	mae	sqrt	random	0.699
11.	mae	log2	best	0.661
12.	mae	log2	random	0.678
13.	squared_error(mse)	auto	best	0.688
14.	mse	auto	random	0.709
15.	mse	sqrt	best	0.730
16.	Mse	sqrt	random	0.612
17	Mse	log2	best	0.716
18.	Mse	log2	random	0.732

The above Decision Tree algorithm shows r2 scores maximum value = 0.738

The Multiple Linear Regression shows r2 scores = 0.789

The SVR algorithm is not fit for this dataset. Because of R2 Score is= - 0.111

The Random Forest Algorithm shows r2 scores maximum value =

SL.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R2_VALE
-------	-----------	--------------	--------------	---------

1	mse	auto	10	0.842
2	mse	auto	50	0.850
3	mse	auto	100	0.853
4	Mse	sqrt	10	0.858
5	mse	sqrt	50	0.867
6	mse	sqrt	100	0.874
7	mse	Log2	10	0.833
8	mse	Log2	50	0.869
9	mse	Log2	100	0.870
10	mae	auto	10	0.847
11	mae	auto	50	0.853
12	mae	auto	100	0.856
13	mae	sqrt	10	0.855
14	mae	sqrt	50	0.875
15	mae	sqrt	100	0.877
16	mae	Log2	10	0.859
17	mae	Log2	50	0.873
18	mae	Log2	100	0.871

The final machine learning best method of Regression :

Random Forest R^2 Score (mae, sqrt,100) = 0.877