

Problem statement

1. 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.

2. Total number of rows= 1338, Total number of columns=1
3. To Find following the machine learning regression method using r^2 value for many algorithms.

1. MULTIPLE LINEAR REGRESSION (R^2 VALUE= 0.785)

2. SUPPORT VECTOR MACHINE:

S.NO	HYPER PARAMETER	LINEAR (r value)	RBF (NON LINEAR) (r value)	POLY (r value)	SIGMOID (r value)
1	No changes	-0.010	-0.0833	-0.075	-0.075
2	C0.01	-0.088	-0.0896	-0.0895	-0.089
3	C0.0001	-0.0897	-0.089	-0.0897	-0.089
4	C0.001	-0.0895	-0.0897	-0.0896	-0.089
5	C500	0.7631	0.6642	0.826	0.446
6	C1000	0.7649	0.810	0.8566	0.287
7	C2000	0.7440	0.854	0.8605	-0.593
8	C3000	0.7414	0.866	0.8598	-2.127

The SVM Regression use R^2 value (Non-Linear and hyper parameter(C=3000)) =0.866

3.DECISION TREE:

S.NO	CRITERION	MAX FEATURES	SPLITTER	R VALUE
1	Squared_error	Auto	best	0.7098
2	Squared_error	Auto	random	0.697
3	Squared_error	Sqrt	Best	0.695
4	Squared_error	Sqrt	Random	0.7398

5	Squared_error	Log2	Best	0.7318
6	Squared_error	Log2	random	0.625
7	friedman_mse	Auto	random	0.7408
8	friedman_mse	Auto	best	0.7053
9	friedman_mse	Sqrt	random	0.6652
10	friedman_mse	Sqrt	best	0.6659
11	friedman_mse	Log2	random	0.7131
12	friedman_mse	Log2	best	0.7474
13	poisson	auto	random	0.7073
14	poisson	auto	best	0.7170
15	poisson	sqrt	random	0.678
16	poisson	sqrt	best	0.5417
17	poisson	Log2	random	0.6155
18	poisson	Log2	best	0.7282
19	absolute_error	auto	Random	0.6912
20	Absolute_error	auto	best	0.6961
21	Absolute_error	sqrt	Best	0.7055
22	Absolute_error	sqrt	Random	0.710
23	Absolute_error	Log2	Best	0.7396
24	Absolute_error	Log2	random	0.6168

The Decision tree Regression use R^2 -value (friedman_mse, log2, best) =0.7474

3. RANDOM FOREST REGRESSOR

S.NO	CRITERION	MAX FEATURES	N_ESTIMATORS	R VALUE
1	Squared_error	auto	50	0.8496
2	Squared_error	auto	100	0.8535
3	Squared_error	sqrt	50	0.869
4	Squared_error	sqrt	100	0.8709
5	Squared_error	Log2	50	0.8694
6	Squared_error	Log2	100	0.8709
7	Mae	auto	50	0.8536

8	Mae	auto	100	0.8526
9	Mae	sqrt	50	0.8715
10	Mae	sqrt	100	0.8713
11	Mae	Log2	50	0.8715
12	Mae	Log2	100	0.8713
13	Friedman mse	Auto	50	0.8497
14	Friedman mse	Auto	100	0.8537
15	Friedman mse	Sqrt	50	0.8704
16	Friedman mse	Sqrt	100	0.8712
17	Friedman mse	Log2	50	0.8704
18	Friedman mse	Log2	100	0.8712
19	poisson	Auto	50	0.8493
20	poisson	Auto	100	0.8527
21	poisson	Sqrt	50	0.8632
22	poisson	Sqrt	100	0.8680
23	poisson	Log2	50	0.8632
24	poisson	Log2	100	0.8680

The Random Forest Regression R^2 value (Mae, sqrt,50) =0.8715

4. The final machine learning best method of Regression:
Random forest R^2 value (mae, sqrt,50) =0.8715