

1) Problem Statement

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.

Machine Learning - Supervised Learning - Regression

Name of the project : Insurance Charge Prediction

2) Basic info about dataset:

Total no. of rows : 6

Total no. of columns : 1339

3) Nominal Data

converting String to number

5) To find r^2 score using Machine learning algorithm_regression

1. Multiple Linear Regression :

r^2 score : 0.7894

3. Support Vector Machine :

S.NO	HYPER PARAMETER	LINEAR	RBF	SIGMOID
1	C10	0.4624	-0.0322	0.0393
2	C100	0.628	0.3200	0.5276
3	C500	0.7631	0.6642	0.4446
4	C1000	0.7649	0.81020	0.2874
5	C2000	0.7440	0.854	-0.5939
6	C3000	0.7414	0.8663	-2.1244

r^2 score(C3000,rbf) : 0.8663

4. Decision Tree :

S.NO	CRITERION	MAX_FEATURES	SPLITTERS	R2_SCORE
1	friedman_mse	auto	best	0.6942
2	friedman_mse	auto	random	0.732

3	friedman_mse	log2	best	0.7672
4	friedman_mse	log2	random	0.6439
5	friedman_mse	sqrt	best	0.6117
6	friedman_mse	sqrt	random	0.61380
7	mse	auto	best	0.7031
8	mse	auto	random	0.7372
9	mse	log2	best	0.6777
10	mse	log2	random	0.6854
11	mse	sqrt	best	0.7019
12	mse	sqrt	random	0.6536
13	mae	auto	best	0.6617
14	mae	auto	random	0.6592
15	mae	log2	best	0.6919
16	mae	log2	random	0.7100
17	mae	sqrt	best	0.7010
18	mae	sqrt	random	0.7144

r2_score (mse,auto,random):0.7372

5. Random Forest :

S.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R2_VALUE
1	mse	auto	10	0.999
2	mse	auto	100	0.9997
3	mse	log2	10	0.993
4	mse	log2	100	0.996
5	mse	sqrt	10	0.9934
6	mse	sqrt	100	0.996
7	mae	auto	10	0.999
8	mae	auto	100	0.999
9	mae	log2	10	0.995
10	mae	log2	100	0.996
11	mae	sqrt	10	0.995
12	mae	sqrt	100	0.996

Random Forest Algorithm gives best

r2_score(mse,auto,10) :0.999

r2_score(mae,auto,10):0.999