

1. Identify your problem statement

Machine learning ----> supervised learning ----> regression

2. Tell basic info about the dataset (Total number of rows, columns)

Rows-1338 and columns-6

3. Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

Nominal data-male or female

Converting string to number- 0 or 1

4. Develop a good model with r2_score. You can use any machine learning algorithm.

You can create many models. Finally, you have to come up with final model

Create machine learning models are,

1. Multiple linear regression, Support vector machine, Decision tree and

Random forest

5. All the research values (r2_value of the models) should be documented.

(You can make tabulation or screenshot of the results.)

1. Multiple linear regression r2_value=0.7894

2. Support vector machine:

s.no	Hyper parameter	gamma	linear	RBF (non Linear)	poly	sigmoid
1	C=10	auto	0.289920	-0.072174	-0.21091	-0.029987
2	C=10	sqrt	0.279920	-0.072183	-0.02097	-0.030027
3	C=100	auto	0.579559	0.179784	0.450389	0.437982
4	C=100	sqrt	0.579559	0.179737	0.450940	0.439498
5	C=1000	auto	0.651037	0.661057	0.781298	0.323072
6	C=1000	sqrt	0.651037	0.660834	0.799868	0.352361
7	C=2000	auto	0.669330	0.765505	0.799868	-0.105737
8	C=2000	sqrt	0.669330	0.765279	0.799964	-0.125605
9	C=3000	auto	0.656065	0.788715	0.800758	-0.972247
10	C=3000	sqrt	0.656065	0.788610	0.800069	-0.964699

R2_value for best parameter {'c':3000,'gamma': auto,' kernel': poly} =0.800758

3.Decision Tree:

s.no	criterion	Max_features	splitter	R2_value
1	mse	auto	best	0.634346
2	mse	auto	random	0.635812
3	mse	sqrt	best	0.599496
4	mse	sqrt	random	0.612868
5	mse	log2	best	0.639382
6	mse	log2	random	0.514986
7	mae	auto	best	0.654093
8	mae	auto	random	0.639160
9	mae	sqrt	best	0.580100
10	mae	sqrt	random	0.547679
11	mae	log2	best	0.418125
12	mae	log2	random	0.634102
13	Friedman_mse	auto	best	0.638071
14	Friedman_mse	auto	random	0.640416
15	Friedman_mse	sqrt	best	0.592293
16	Friedman_mse	sqrt	random	0.604382
17	Friedman_mse	log2	best	0.649644
18	Friedman_mse	log2	random	0.560100

The R_score value for best parameter {'criterion': 'mae', 'max_features': 'auto', 'splitter': 'random'}:
0.7447774688348948

4.Random Forest:

s.no	criterion	Max_features	N_estimators	R2_value
1	mse	auto	10	0.777248
2	mse	auto	100	0.793324
3	mse	sqrt	10	0.780009
4	mse	sqrt	100	0.804203
5	mse	log2	10	0.775311
6	mse	log2	100	0.799471
7	mae	auto	10	0.757291
8	mae	auto	100	0.787581
9	mae	sqrt	10	0.765042

10	mae	sqrt	100	0.708964
11	mae	log2	10	0.780068
12	mae	log2	100	0.798548

The R_score value for best parameter {'criterion': 'mse', 'max_features': 'log2', 'n_estimators': 100}:
0.8722821294720762

6.Mention your final model, justify why you have chosen the same.

The best final model for Random Forest algorithm=0.8722821