

**1.) Identify your problem statement:**

Domain – Machine Learning

Dataset have a clear requirement –Supervised Learning

More than one input – MLR/SVM/Decision Tree/Random forest

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

1338 rows x 6 columns are available

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

Doing some pre-processing method

Convert categorical to nominal using one-hot-encoding

Drop/reduce some unnecessary columns

**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.**

Random Forest given good r2\_score compare to other models

**R2\_score=0.8508** (criterion="mae", max\_features="log2", n\_estimators=100)

**Multiple Linear Regression:**

R2\_score=0.71

## Support Vector Machine:

S.No	C	gamma	kernel	R_score
1.	10	auto	rbf	-0.062665
2.	10	auto	poly	-0.0146317
3.	10	auto	sigmoid	-0.0181639
4.	10	auto	linear	0.3045647
5.	10	scale	rbf	-0.0626749
6.	10	scale	poly	-0.0145354
7.	10	scale	sigmoid	-0.0181457
8.	10	scale	linear	0.3045647
9.	100	auto	rbf	0.1984809
10.	100	auto	poly	0.4621191
11.	100	auto	sigmoid	0.4478397
12.	100	auto	linear	0.5780588
13.	100	scale	rbf	0.1984411
14.	100	scale	poly	0.462532
15.	100	scale	sigmoid	0.4480129
16.	100	scale	linear	0.5780588
17.	1000	auto	rbf	0.6751839
18.	1000	auto	poly	0.7859069
19.	1000	auto	sigmoid	0.3636938
20.	1000	auto	linear	0.6584714
21.	1000	scale	rbf	0.6751124
22.	1000	scale	poly	0.7859516
23.	1000	scale	sigmoid	0.350668
24.	1000	scale	linear	0.6584714
25.	2000	auto	rbf	0.7737705
26.	2000	auto	poly	0.8053768
27.	2000	auto	sigmoid	-0.1322524
28.	2000	auto	linear	0.6554113
29.	2000	scale	rbf	0.7737209
30.	2000	scale	poly	0.8051435
31.	2000	scale	sigmoid	-0.0429397
32.	2000	scale	linear	0.6554113
33.	3000	auto	rbf	0.7915783
34.	3000	auto	poly	0.8074628
35.	3000	auto	sigmoid	-0.8331917
36.	3000	auto	linear	0.6548732
37.	3000	scale	rbf	0.8439
38.	3000	scale	poly	0.82829
39.	3000	scale	sigmoid	-0.1016
40.	3000	scale	linear	0.76337

### Decision Tree Regression:

	criterion	max_features	splitter	R2_score
0	mse	auto	best	0.71
1	mse	auto	random	0.7320
2	mse	sqrt	best	0.7359
3	mse	sqrt	random	0.653
4	mse	log2	best	0.773
5	mse	log2	random	0.7223
6	mae	auto	best	0.714
7	mae	auto	random	0.7704
8	mae	sqrt	best	0.7288
9	mae	sqrt	random	0.6727
10	mae	log2	best	0.6776
11	mae	log2	random	0.712
12	friedman_mse	auto	best	0.7467
13	friedman_mse	auto	random	0.703
14	friedman_mse	sqrt	best	0.7432
15	friedman_mse	sqrt	random	0.708
16	friedman_mse	log2	best	0.6784
17	friedman_mse	log2	random	0.6779

### Random Forest:

criterion	max_features	n_estimators	R2_score
mse	auto	10	0.829
	auto	100	0.8459
mse	sqrt	10	0.8318
mse	sqrt	100	0.8473
mse	log2	10	0.8318
mse	log2	100	0.8473
mae	auto	10	0.8378
mae	auto	100	0.8508
mae	sqrt	10	0.8378
mae	sqrt	100	0.8508
mae	log2	10	0.8378
mae	log2	100	0.8508