Project Name :Insurance Charges Prediction

1 Identify your problem statement
Machine Learning ->Supervised->Regression

2 Total number of rows and columns
1338 rows × 6 columns

3 Mention the preprocessing method.

Sex and Smoker is categorical column , Machine learning models cannot work on categorical variable ,this norminal data need to be convert to numeric variable by one-hot encoding. To overcome the dummy variable trap, we drop one of the column created during this process.

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.All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)
Machine learning regression method using in r2 value

a **Multiple Linear Regression(R2 value) =** **0.79**

No	Hyper Parameter	Linear	RBF(Non Linear)	Poly	sigmoid
1	C10	-0.0016	-0.0820	-0.0931	-0.0908
2	C100	0.5433	-0.1248	-0.0998	-0.1181
3	C500	0.5433	-0.1246	-0.0820	-0.4563
4	C1000	0.6340	-0.1175	-0.0555	-1.6659
5	C3000	**0.7591**	-0.0962	0.0489	-12.0190

b **The SVM Regression use R2 value using Linear with C3000 =** **0.7591**

No	Criterion	Splitter	Max Feat	R value
1	mse	best	auto	0.69
2	mse	best	log2	0.73
3	mse	random	sqrt	0.68
4	absolute_error	best	auto	0.68
5	absolute_error	best	log2	0.57
6	absolute_error	random	sqrt	0.71
7	poisson	best	auto	0.69
8	poisson	best	log2	0.64
9	poisson	random	log2	0.58
10	mae	best	auto	0.70
11	mae	best	log2	**0.72**
12	mae	random	log2	0.63
13	friedman_mse	best	auto	0.69
14	friedman_mse	best	log2	0.66

15	friedman_mse	random	sqrt	0.71

c **The Decision Tree Regression use R2 value using nyper parameter (mae ,s** **0.72**

No	Criterion	n-Estimators	R value
1	mse	10	0.83
2	mse	100	0.85
3	mse	90	0.85
4	absolute_error	10	0.84
5	absolute_error	100	**0.85**
6	absolute_error	90	0.85
7	absolute_error	95	0.69
8	poisson	10	0.82
9	poisson	100	0.83
10	mae	90	0.85
11	mae	10	0.84
12	mae	100	0.85
13	friedman_mse	90	0.85
14	friedman_mse	10	0.83
15	friedman_mse	100	0.85

d **Random Forest Regression use R2 value** **0.85**

I would choose Random Forest algorithm model to be consider more reliable.