Assignment-Regression Algorithm

Problem Statement or Requirement:

Wants to predict the insurance charges based on the several parameters

- Stage 1 : Supervised Learning, (Problem statement is defined and input, output are clear)

Observations from Dataset:

- We have five independent variables
- has two categorical columns
- Has 1339 rows and 6 columns

Mention the pre-processing method

- We have two categorical data column (Nominal data), we need to convert that using one hot encoding

R2 Values:

Multi Linear Regression

R2= 0.789479

Support Vector Machine: (With Standard Scaler we get better r value)

S.No	C value	Kernel				
		linear	***poly***	***rbf***	***sigmoid***	***precomputed***
1	10	0.46593	0.03906	-0.03316	0.04019	This SVR instance is not fitted yet, it needs square matrics
2	100	0.63124	0.61565	0.31379	0.53142	
3	1000	0.76714	0.85205	0.81149	0.28739	

Decision Tree

S.No	**criterion**	**splitter**	**max_features**	R2 score
	"squared_error", "friedman_mse", "absolute_error", "poisson"	***"best", "random"***	***"auto", "sqrt", "log2"***	
1	friedman_mse	best	auto	0.69266
2	friedman_mse	best	sqrt	0.69236
3	friedman_mse	best	log2	0.67169
4	friedman_mse	random	log2	0.68559
5	friedman_mse	random	auto	0.70271
6	friedman_mse	random	sqrt	0.6824
7	absolute_error	random	sqrt	error
8	squared_error	random	sqrt	error
9	poisson	best	auto	error
10	mae	best	auto	0.71516
11	mae	random	auto	0.73774
12	mae	random	log2	0.68223
13	mae	random	sqrt	0.70567
14	mae	best	sqrt	0.71306
15	mae	best	log2	0.73264

Assignment-Regression Algorithm

Random Forest:

S.No	n_estimators	random_state	criterion	max_features	R2 score
1	10	0	mae	Log2	0.84834
2	10	10	mae	Log2	0.84706
3	100	10	mae	Log2	0.86262
4	100	0	mae	Log2	0.86256
5	100	0	mae	auto	0.84006
6	100	10	mae	auto	0.84428
7	10	10	mae	auto	0.83442
8	10	0	mae	auto	0.82201
9	100	0	mae	sqrt	0.86256
10	100	0	friedman_mse	sqrt	0.86688
11	100	0	mse	sqrt	0.86730
12	100	0	mse	auto	0.83857
13	100	0	mse	Log2	0.86730

Final Model:

We have chosen Random forest with following hyper parameters (n_estimators=100,random_state=0, criterion='mse', max_features="log2" or “sqrt”) it has given maximum r2 score which is **0.86730**