Regression Assignment

Problem Statement or Requirement:

A 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.

1.)Identify your problem statement

Multi input one output

I need to predict Insurance charge.

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

Rows: 1338

Columns: 6

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

nominal date convert into One Hot Encoding using (pd.get_dummies)

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.

5.)All the research values (r2_score of the models) should be documented.

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

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

The Random tree Regression R2_score

(n_estimators-100, criterion-mse, max_features-sqrt)-0.87

To find the following Machine Learning regression method using R2_value

The Multi Linear Regression R2-score -0.789

The Support Vector Machine Regression R2-score - (Kernel-sigmoid)-0.013

The Decision tree Regreesion R2-score- (criterion-mse, max_features-sqrt, splitter-best) - 0.77

The Random tree Regression R2_score (n_estimators-100, criterion-mse, max_features-sqrt)-0.87

Support vector machine

SI.No:	Hyper parameter	Linear	poly	rbf	sigmoid
1	c=1	-1.436	-12266.2	-0.088	-0.073
2	c=10	-113.04	-1163348.2	-0.080	0.013
3	c=100	-146.14	-32979013.9	-0.021	-0.543
4	c=500	-152.22	-29631241.41	-6.239	-5.484
5	c=1000	-154.87	-10543590.2	-0.0158	-4.549

Decision Tree Regressor

SI:NO:	Criterion	Max Features	Splitter	R_Score
1	mse	auto	best	0.69
2	mse	auto	random	0.66
3	mse	sqrt	best	0.77
4	mse	sqrt	random	0.64
5	mse	log2	best	0.69
6	mse	log2	random	0.67
7	friedman_mse	auto	best	0.69
8	friedman_mse	auto	random	0.67
9	friedman_mse	sqrt	best	0.75
10	friedman_mse	sqrt	random	0.63
11	friedman_mse	log2	best	0.67
12	friedman_mse	log2	random	0.63
13	mae	auto	best	0.66
14	mae	auto	random	0.73
15	mae	sqrt	best	0.69
16	mae	sqrt	random	0.67
17	mae	log2	best	0.70
18	mae	log2	random	0.73

Random Forest Regressor

SI:NO:	n_estimator	criterion	max_features	r_score
1.	10	mse	auto	0.83
2	50	mse	auto	0.84
3	100	mse	auto	0.85
4	10	mse	sqrt	0.85
5	50	mse	sqrt	0.86
6	100	mse	sqrt	0.87
7	10	mse	log2	0.85
8	50	mse	log2	0.86
9	100	mse	log2	0.87