Assignment-Regression Algorithm

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?
Machine Learning
Supervised Learning
Regression

2.) Tell basic info about the dataset (Total number of rows, columns)
1338 rows × 6 columns

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

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.
Why I am telling this is the good model, because of the the r_score value its better than other algorithms.
THE **RANDOM FOREST REGRESSION** USE **r2 VALUE**log2,absolute_error,100=0.87166

1. MULTIPLE LINEAR REGRESSION (r2 value) = 0.78947

2. SUPPORT VECTOR MACHINE

SNO	PARAMETER	R _VALUE
1	RBF&C=10	-0.081969
2	Degree&=3	-0.088427
3	verbose=False	-0.088427
4	Tol =1e-3,epsilon=0.1,C=1000000	0.808305

THE **SVM REGRESSION** USE **r2 VALUE** tol =1e-3,epsilon=0.1,(C=1000000) = 0.808305

3. DECISION TREE

S.NO	CRITERION	MAX FEATURIES	SPLITTER	R VALUE
1.	**squared_error**	**None**	**best**	0.69079
2.	**friedman_mse**	**sqrt**	**random**	0.63760
3.	**poisson**	**log2**	**random**	0.73907
4.	**absolute_error**	**Log2**	**random**	0.72579
5.	**absolute_error**	**None**	**best**	0.65506
6.	**absolute error**	**None**	**random**	0.77300
7.	**Absolute error**	**Log2**	**Best**	0.63556

THE **DECISION TREE REGRESSION** USE **r2 VALUE** absolute_error,None,random=0.77300

4. RANDOM FOREST

S.NO	n_estimators	max_features	criterion	R VALUE
1.	**100**	**1.0**	**squared_error**	0.85079
2.	**100**	**Log2**	**absolute_error**	0.87166
3.	**100**	**None**	**friedman_mse**	0.85463
4.	**100**	**Log2**	**friedman_mse**	0.87069
5.	**100**	**1.0**	**absolute_error**	0.850402

THE **RANDOM FOREST REGRESSION** USE **r2 VALUE**log2,absolute_error,100=0.87166