Assignment-Regression Algorithm

1.) **Identify your problem statement**

Client requests to predict the insurance charges based on the several parameter using provided dataset.

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

Total number of rows: 1338
Total number of columns: 6

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

Columns "Sex "and "Smoker" are categorical columns and will be converted to numerical data. Then first column will be dropped for converted numerical data of Sex and Smoker.

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.**

Multilinear Regression, Support Vector Machine, Decision Tree and Random Forest algorithm used to create the models.

5.) **All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)**

Multiple Linear Regression:

R^2 value = 0.789479034986701

Support Vector Machine:

SL.No	Hyper Parameter	Linear (r value)	RBF (Non Linear) (r value)	Poly (r value)	SIGMOID (r value)
1		0.465930219	-0.033169074	0.03906576	0.040195212
2	C100	0.631242902	0.313798207	0.61565518	0.531425558
3	C500	0.764481266	0.661926404	0.8234642	0.444274245
4	C1000	0.767145485	0.811494921	0.85205496	0.287397825
5	C2000	0.747843338	0.858330622	0.85623714	-0.587880094
6	C3000	0.745373651	0.869845312	0.85531902	-2.091526784

R^2 value = 0.869845312

Decision Tree:

SL.No	CRITERION	SPLITTER	R VALUE
1	squared_error	best	0.679259405
2	squared_error	random	0.715221069
3	friedman_mse	best	0.68719868
4	friedman_mse	random	0.733678367
5	absolute_error	best	0.680253952
6	absolute_error	random	0.731183473
7	poisson	best	0.712080005
8	poisson	random	0.71210777

R^2 value = 0.731183473

Random Forest:

SL.No	n_estimators	Random_State	R Value
1	1	0	0.744103403
2	10	0	0.833030413
3	50	0	0.849832932
4	100	0	0.853830791

R^2 value = 0.853830791

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

Final model R^2 value = 0.869845312030898 from Support Vector Machine Regression algorithms which is closer to 1 compared to other algorithm's R^2 values for the given data from insurance_pre.csv file