Assignment-Regression Algorithm

1. Problem statement: To predict the insurance charges based on the various parameters and the dataset provided by the client. To develop a model that predicts the insurance charges.
2. The dataset consists of 6 columns and 1339 rows with the data available for various parameters such as age, sex, BMI, children, smoker and charges.
3. The pre-processing is done for converting the strings to number.

To find the correct machine learning regression method using R^2 value.

1. Multiple Linear Regression R2_score= 0.7894

2. Support Vector Machine

S. No.	kernel	degree	gamma	Coef0	tol	c	Epsilon	Shrinking	Cache_size	Verbose	Max_iter	R_Score
1	**rbf**	3	auto	0.0	0.001	1.0	0.1	True	200	False	-1	-0.0897

3. Decision Tree

S. No.	criterion	max_features	splitter	r_score
1.	friedman_mse	sqrt	best	0.7070
2.	squared_error	sqrt	best	0.6969
3.	absolute_error	sqrt	best	0.7025
4.	poisson	sqrt	best	0.5999
5.	friedman_mse	sqrt	random	0.6632
6.	squared_error	sqrt	random	0.6178
7.	absolute_error	sqrt	random	0.6781
8.	poisson	sqrt	random	0.6038
9.	friedman_mse	log2	best	0.7392
10.	squared_error	log2	best	0.7436
11.	absolute_error	log2	best	0.7515
12.	poisson	log2	best	0.6471
13.	friedman_mse	log2	random	0.6677
14.	squared_error	log2	random	0.6704
15.	absolute_error	log2	random	0.6944
16.	poisson	log2	random	0.6588
17.	friedman_mse	auto	best	0.7021

18.	squared_error	auto	best	0.7023
19.	absolute_error	auto	best	0.6817
20.	poisson	auto	best	0.6885
21.	friedman_mse	auto	random	0.6864
22.	squared_error	auto	random	0.7289
23.	absolute_error	auto	random	0.7031
24.	poisson	auto	random	0.6570

4. Random Forest

S. No.	n_estimators	criterion	max_depth	random_state	min_samples_split	min_samples_leaf	min_weight_fraction_leaf	max_features	max_leaf_nodes	min_impurity_decrease	bootstrap	oob_score	n_jobs	verbose	Warm_start	Ccp_alpha	Max_samples	R_score
1	100	squared_error	None	None	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8561
2	100	absolute_error	None	int	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8549
3	100	friedman_mse	None	int	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8551
4	100	poisson	None	int	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8348
5	100	squared_error	None	None	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8550
6	100	absolute_error	None	None	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8540
7	100	friedman_mse	None	None	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8523
8	100	poisson	None	None	2	1	0.0	1.0	None	0.0	True	False	None	0	False	0.0	None	0.8329

The final best model is identified as the **Random Forest regression** based on the parameters mentioned below.

(n_estimators=100, criterion=squared_error, max_depth=none, random_state=none, min_samples_split=2, min_samples_leaf=1.0, min_weight_fraction_leaf=0.0, max_features=1.0, max_leaf_nodes=none, min_impurity_decrease=0.0, bootstrap=True, oob_score=False, n_jobs=none, verbose=0, warm_start=False, Ccp_alpha=0.0, Max_samples=none).

Based on the various parameters this model provides the highest R^2 value of 0.8561 compared to other models. Hence, this model could be considered as the best model to predict the best insurance charges.