Assignment-Regression Algorithm

1.) Identify your problem statement

Machine Learning → Supervised learning → Regression

2.) Tell basic info about the dataset (Total number of rows, columns)
Client requirement is insurance charges prediction
6 Rows and 1338 Colum

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

This dataset is not for ordering .this dataset is 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

Find the machine learning regression using r2 value

Multiple linear regression (r2 value= 0.1497)

1. Support vector machine :

S.NO	***C Parameter***	***Linear (r value)***	***Poly (r value)***	***Rbf (r value)***	***Sigmoid (r value)***
1		-0.1244	-0.0870	-0.0884	-0.08700
2	C=0.01	-0.0847	-0.0896	-0.0896	-0.0896
3	C=10	-0.1243	-0.0756	-0.0817	-0.0819
4	C=100	-0.1241	-0.0819	-0.1111	-0.1100
5	C=1000	-0.1241	-0.0774	-0.1221	-1.0147

The svm Regression use r2 value. (-0.1244)

2. Decision Tree Regression

S.NO	***Criterion***	***Splitter***	***Max_features***	***r2 Value***
1	Squared _error	best	sqrt	-0.5897
2	Squared _error	best	log2	-0.5856
3	Squared _error	random	sqrt	-0.5398
4	Squared _error	random	log2	-0.6694
5	Friedman _mse	best	sqrt	-0.7258
6	Friedman _mse	best	log2	-0.6669

7	Friedman _mse	random	sqrt	-0.5639
8	Friedman _mse	random	log2	-0.5278
9	Absolute _error	best	sqrt	-0.4658
10	Absolute _error	best	log2	-0.6272
11	Absolute _error	random	sqrt	-0.2580
12	Absolute _error	random	log2	-0.3001
13	Poisson	best	sqrt	-0.6589
14	Poisson	best	log2	-0.7345
15	Poisson	random	sqrt	-0.8232
16	Poisson	random	log2	-0.6216
17	Squared _error	best		-0.554
18	Squared _error	random		-0.554
19	Friedman _mse	best		-0.533
20	Friedman _mse	random		-0.625
21	Absolute _error	Best		-0.6203
22	Absolute _error	random		-0.6203
23	Poisson	Best		-0.601
24	Poisson	Random		-0.509

The Decision Tree Regression r2 (Poisson _error ,random ,sqrt=-0.8232)

3. RandomForest Regression

S.NO	***Criterion***	***N_ estimators***	***Max _features***	***r2 Value***
1	Squared _error	10		*-0.0569*
2	Squared _error	50		*-0.0043*
3	Squared _error	100		*0.02067*
4	Friedman _mse	10		*-0.0777*
5	Friedman _mse	50		*0.0116*
6	Friedman _mse	100		*0.03035*
7	Absolute _error	10		*0.0267*
8	Absolute _error	50		*0.0244*
9	Absolute _error	100		*0.0092*
10	Poisson	10		*-0.050*
11	Poisson	50		*0.0465*
12	Poisson	100		*0.0275*
13	Squared _error	10	*sqrt*	*0.0778*
14	Squared _error	50	*sqrt*	*0.0426*
15	Squared _error	100	*sqrt*	-0.0634
16	Squared _error	10	Log2	0.0227
17	Squared _error	50	Log2	0.0469
18	Squared _error	100	Log2	0.0514
19	Absolute _error	10	*sqrt*	0.0250
20	Absolute _error	50	*sqrt*	0.0712
21	Absolute _error	100	*sqrt*	0.0641
22	Absolute _error	10	Log2	0.0603

23	Absolute _error	50	Log2	0.0576
24	Absolute _error	100	Log2	0.0821
25	Friedman _mse	10	*sqrt*	0.0722
26	Friedman _mse	50	*sqrt*	0.0390
27	Friedman _mse	100	*sqrt*	0.0589
28	Friedman _mse	10	Log2	0.0480
29	Friedman _mse	50	Log2	0.0500
30	Friedman _mse	100	Log2	0.0577
31	Poisson	10	*sqrt*	0.0324
32	Poisson	50	*sqrt*	0.0250
33	Poisson	100	*sqrt*	0.0528
34	Poisson	10	Log2	0.0465
35	Poisson	50	Log2	0.0618
36	Poisson	100	Log2	0.0627

The finally machine learning best model of Regression

The Random Forest Regression r2 value Multiple linear regression (r2 value= 0.1497)

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

This dataset is being compared to another model for predictions. A better choice would be Multiple linear. As it works for predicting values is better.