Insurance Charge Prediction

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. **Problem Statement:**
   Input and Output are clear. – SuperVised Learning
   Input is many
   Output is Insurance Charge which is Regression

2. **DataSet:**
   a. Input is age, sex, bmi, children, smoker, charges
   b. Output is Insurance Charges
   c. Total Number of rows are :1338
   d. Total Number of columns are: 6

3. **Pre-Processing Method:**
   The input data has both Categorical and Nominal Data
   Categorical Data are : Sex, smoker
   Nominal Data are : age, bmi, children, charges
   Therefore convert Categorical Data into Nominal data

4. **Model:**
   R2_SCORE :

   Multiple Linear R_Score : 0.76568
   SVM: R_Score : -0.9063

Sno	*HyperParam*	*linear*	*poly*	*rbf*	*sigmoid*
1	C10	-0.0466	-0.0963	-0.0710	-0.07675
2	C100	0.49318	-0.1152	-0.11826	-0.09201
3	C500	0.60134	-0.1051	-0.1334	-0.26652
4	C0.5	-0.1348	-0.0683	-0.0753	-0.07611
5	C0.1	-0.1306	-0.0742	-0.07584	-0.07599
6	C1000	0.6131	-0.08850	-0.1297	-0.9063

Decision Tree

Sno	Criterion	Max_features	Splitter	R Score
1	squared_error	sqrt	best	0.5839
2	friedman_mse	sqrt	best	0.618650
3	absolute_error	sqrt	best	0.69077
4	poisson	sqrt	best	0.62056
5	squared_error	sqrt	random	0.65209
6	friedman_mse	sqrt	random	0.66199
7	absolute_error	sqrt	random	0.60450
8	poisson	sqrt	random	0.6213
9	squared_error	log2	best	0.7199
10	friedman_mse	log2	best	0.56909
11	absolute_error	log2	best	0.72402
12	poisson	log2	best	0.73210
13	squared_error	log2	random	0.57171
14	friedman_mse	log2	random	0.65473
15	absolute_error	log2	random	0.70884
16	poisson	log2	random	0.665280
17				
18				

RandomForest

Sno	Criterion	Max_features	n_estimators	R Score
1	squared_error	sqrt	50	0.8417
2	friedman_mse	sqrt	50	0.8438
3	absolute_error	sqrt	50	0.8408
4	poisson	sqrt	50	0.8392
5	squared_error	log2	50	0.8417
6	friedman_mse	log2	50	0.8438
7	absolute_error	log2	50	0.8408
8	poisson	log2	50	0.8392
9	squared_error	None	50	0.82793
10	friedman_mse	None	50	0.82811

11	*absolute_error*	None	50	0.8198
12	*poisson*	None	50	0.82726
13	*squared_error*	*sqrt*	100	0.8425
14	*friedman_mse*	*sqrt*	100	0.84371
15	*absolute_error*	*sqrt*	100	0.8443
16	*poisson*	*sqrt*	100	0.8417
17	*squared_error*	*log2*	100	0.84259
18	*friedman_mse*	*log2*	100	0.8437
19	*absolute_error*	*log2*	100	0.8443
20	*poisson*	*log2*	100	0.84173
21	*squared_error*	None	100	0.8333
22	*friedman_mse*	None	100	0.8337
23	*absolute_error*	None	100	0.8259
24	*poisson*	None	100	0.83453

5. **Best Model :**

Multiple Linear R_Score : **0.76568**

SVM: R_Score : **-0.9063**

Decision: **0.73210**

Random Forest: 0.84371

Good Model has to be nearly 1, but here largest number nearing 1 is:

Random Forest: R_Score : 0.84371