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.

S.No	Question	Answer
1	Identify your problem statement	Stage1:Machine learning
		Stage2:Supervised learning
		Stage3: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 is ordinal data - because it is comparable and no column expand
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	Good model ,final model attached in Github
5	All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)	Yes,Documented in Excel
6	Mention your final model, justify why u have chosen the same.	Choosen Random forest regression(absolute_error,log2) = 0.8705 is a best model ,because its having error is less

1)Multilinear Regression

Multilinear = 0.7891 a good model.

2)Support Vector Machine Regression

Insurance Charges Prediction Assignment - 07.05.2024		
SUPPORT VECTOR MACHINE		
Kernel : linear, poly, rbf, sigmoid, precomputed		
S.No	**Parameters**	**R2 Score**
1	Linear	-0.1115
2	Poly	-0.064
3	Rbf	-0.0884
4	Sigmoid	0.0899
5	Precomputed	Must be square matrix
Best model in SVM is poly		-0.064

3)Random Forest Regression

Insurance Charges predict Assignment - 07.05.2024				
RANDOM FOREST				
Creterion: squared_error, friedman_mse, absolute_error, poisson				
Parameter: n_estimators				
Max_features: sqrt, log2				
S.No	**Parameter**	**Creterion**	**Max_features**	**R_Score**
1	_	squared_error	*sqrt*	0.8665
2	_	squared_error	*log2*	0.8676
3	_	friedman_mse	*sqrt*	0.8645
4	_	friedman_mse	*log2*	0.8634
5	_	absolute_error	*sqrt*	0.8697
6	_	absolute_error	*log2*	0.8705
7	_	poisson	*sqrt*	0.8608
8	_	poisson	*log2*	0.8629
9	n_estimators=10	_	_	0.8254
10	n_estimators=50	_	_	0.8519
11	n_estimators=80	_	_	0.8514
12	n_estimators=100	_	_	0.8524
Best model in Random forest is absolute_error , log2				0.8705

4)Decision Tree Regression

Insurance Charges Prediction Assignment - 07.05.2024				
DECISION TREE				
Creterion: squared_error, friedman_mse, absolute_error, poisson				
Splitter: best ,random				
Max_features: sqrt, log2				
S.No	**Creterion**	**Splitter**	**Max_features**	**R2 Score**
1	squared_error	best	_	0.6866
2	friedman_mse	best	_	0.6666
3	absolute_error	best	_	0.7315
4	poisson	best	_	0.6866
5	squared_error	random	_	0.7284
6	friedman_mse	random	_	0.6464
7	absolute_error	random	_	0.6810
8	poisson	random	_	0.6624
9	squared_error	best	*sqrt*	0.6836
10	squared_error	best	*log2*	0.7077
11	squared_error	random	*sqrt*	0.7354
12	squared_error	random	*log2*	0.6529
13	friedman_mse	best	*sqrt*	0.6907
14	friedman_mse	best	*log2*	0.6826
15	friedman_mse	random	*sqrt*	0.7417
16	friedman_mse	random	*log2*	0.7300
17	absolute_error	best	*sqrt*	0.7184
18	absolute_error	best	*log2*	0.7121
19	absolute_error	random	*sqrt*	0.7153
20	absolute_error	random	*log2*	0.6733
21	poisson	best	*sqrt*	0.7136
22	poisson	best	*log2*	0.6576
23	poisson	random	*sqrt*	0.7239
24	poisson	random	*log2*	0.6832
Best model in decision tree is friedman_mse,random,sqrt				0.7417

In overall regression best model is Random Forest (absolute_error,log2) – 0.8705