Hope Artificial Intelligence

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 charges.

1.) Identify your problem statement:

Ans:

Problem Statement :

Selecting the Domain Selection -> Machine Learning -> Supervised Learning->Regression

2.) Tell basic info about the dataset (Total number of rows, columns)
Ans:
Basic Info about the dataset:
Total number of rows: 1338
Total number of columns: 6
Basic info about the we need to predict the insurance charges

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

Ans:

Yes we need to convert string to number using one hot Encoding Method

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.

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
R2_Score Value: **0.7270899256752571**

Support Vector Machine

S.No	Hyper Parameter (C Value)	Linear R value	Non Linear RBF R Value	Non Linear Poly R Value	Non Linear Sigmoid R Value

1	C10	-0.012461582644439195	-0.11873899370466034	-0.10803377096131506	-0.10779233871909377
2	C100	0.5181800407397958	-0.04234296926646963	0.048188515693572076	0.04889805767086575
3	C500	0.7138107410620067	0.2267930527759766	0.4784875063175211	0.4966481152104939

Decision Tree

S.No	Criterian	Max Features	Splitter	R Value
1	*squared_error*	None	best	0.5512147657194636
2	friedman_mse	None	best	0.5379092989055643
3	*absolute_error*	None	best	0.5527943851542783
4	*squared_error*	None	random	0.5140450245389471
5	friedman_mse	None	random	0.42005668407075847
6	*absolute_error*	None	random	0.4611776889956879

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

Final Model : Multiple Linear Model is the Final Model because the R Value is **0.7270899256752571**