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

Stage 1-Machine learning.
Stage 2-Supervised learning.
Stage 3-Regression.

2.) Tell basic info about the dataset (Total number of rows, columns)

1338 rows and 6 columns.

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

Yes, inputs are converting string to number. It is a ORDINAL 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.

Final model:

In RANDOM FOREST R2(Squared_error,sqrt,100)=0.87099

5.) All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

1.Multilinear Regression (R2 value) =0.78947.

2.SUPPORT VECTOR MACHINE(SVM):

S/NO	HYPER PARAMETER	LINEAR	RBF	POLY	SIGMOID
1	C10	0.46246	-0.03227	0.03871	0.03930
2	C100	0.62887	0.32003	0.61795	0.52791
3	C500	0.76310	0.66429	0.82636	0.44460
4	C1000	0.76493	0.81020	0.85664	0.28747
5	C2000	0.74404	0.85477	0.86055	-0.5939
6	C3000	0.74142	0.86633	0.85989	-2.1244

The **SVM REGRESSION** use **R2 value (**Linear and Hyper parameter C3000=**0.86633).**

3.DECISION TREE:

S/NO	CRITERION	SPLITTER	MAX_FEATURES	R VALUE
1	Squared_error	best	none	0.68506
2	Friedman_mse	best	none	0.70471
3	Friedman_mse	random	none	0.72993
4	Friedman_mse	best	auto	0.69673
5	Friedman_mse	random	auto	0.71997
6	Friedman_mse	best	sqrt	0.72098
7	Friedman_mse	random	sqrt	0.69498
8	Friedman_mse	best	log2	0.74674
9	Friedman_mse	random	log2	0.67397

The **DECISION TREE REGRESSION** use R2 value(friedman_mse,best,log2)=**0.74674.**

4.RANDOM FOREST:

S/NO	CRITERION	N_estimators	MAX_FEATURES	R VALUE
1	Squared_error	50	Default(1.0)	0.84988
2	Squared_error	100(default)	Default(1.0)	0.85392
3	Squared_error	10	Default(1.0)	0.83315
4	Squared_error	50	Log2	0.86949
5	Squared_error	100	Log2	0.87099
6	Squared_error	50	sqrt	0.86949
7	Squared_error	100	sqrt	0.87099
8	Friedman_mse	50	log2	0.87004
9	Friedman_mse	50	sqrt	0.87004

The **RANDOM FOREST** use R2 value(squared_error,100,sqrt)=**0.87099.**

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

The final machine learning best method of regression:

The **RANDOM FOREST** use R2 value(squared_error,100,sqrt)=**0.87099.**

In random forest algorithm are gives the highest value to predict the insuranc charges(output).So, it is a best model in regression.