Assignment-Regression Algorithm

Answer:

- Identify your problem statement
  Ans: predict the insurance charges

- Tell basic info about the dataset (Total number of rows, columns)
  Ans: 6 columns and 1338 rows
- Mention the pre-processing method if you're doing any (like converting string to number – nominal data)
  Ans: nominal data
- 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.

  Ans: Decision Tree Regression

- All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)
  Ans:
  1. Multiple Linear Regression (R_value = 0.7894790349867009)
  2.Support Vector Machine:

S.NO	HYPER PARAMETER	LINEAR	RBF	POLY	SIGMOID
1	C10	0.462468414	0.032273294	0.038716223	0.039307144
2	C100	0.628879286	0.320031783	0.617956962	0.527610355
3	C500	0.763105798	0.664298465	0.826368354	0.444606103
4	C1000	0.764931174	0.810206485	0.856648768	0.287470695
5	C2000	0.744041831	0.854776643	0.860557928	0.593950973
6	C3000	0.74142366	0.866339395	0.859893008	0.124419479

The svm Regression uses R value is 0.866339395

3. Decision Tree:

S.NO	criterion	Splitter	max_features	R-value
1	squared error	best	sqrt	0.768921
2	squared error	random	log2	0.688899
3	squared error	best	auto	0.692868
4	squared error	random	sqrt	0.670826
5	squared error	best	log2	0.684652
6	squared error	random	auto	0.732212
7	friedman_mse	best	sqrt	0.689123
8	friedman_mse	random	log2	0.647678
9	friedman_mse	best	auto	0.689123
10	friedman_mse	random	sqrt	0.647678
11	friedman_mse	best	log2	0.685368
12	friedman_mse	random	auto	0.680438
13	absolute error	random	sqrt	0.639714
14	absolute error	best	log2	0.731998
15	absolute error	random	auto	0.715823
16	absolute error	best	sqrt	0.658806
17	absolute error	random	log2	0.708018
18	absolute error	best	auto	0.682231
19	poisson	best	sqrt	0.715514
20	poisson	random	log2	0.713314
21	poisson	best	auto	0.715514
22	poisson	random	sqrt	0.712932
23	poisson	best	log2	0.701665
24	poisson	random	auto	0.725705

Decision Tree Regression uses R value is 0.731998

4. Random Forest:

s.no	criterion	max features	n_estimators	r value
1	squared_error	sqrt	10	0.852000635
2	squared_error	log2	100	0.87102719
3	squared_error	auto	10	0.833030413
4	absolute_error	sqrt	100	0.871068586
5	absolute_error	log2	10	0.857429008
6	absolute_error	auto	100	0.852009362
7	friedman_mse	sqrt	10	0.850277799
8	friedman_mse	log2	100	0.870054402
9	friedman_mse	auto	10	0.833166268
10	poisson	sqrt	100	0.868015698
11	poisson	log2	10	0.854495529
12	poisson	auto	100	0.852633426

Random Forest Regression uses R value is 0.871068585634151

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

Ans: Decision Tree Regression value is the highest R value. so, I choose this model