Assignment-Regression Algorithm

Problem Statement or Requirement:-
  A Client's requirement is, he wants to predict the insurance charges based on several parameters. The Client has provided the dataset of the same
.

1.Identify Your Problem Statement:-
  Client need to predict the insurance charges based on various types of people
  1.Domain Selection:
   Requirement is fully clear from the client. So it comes under Machine Learning.
  2.Learning Selection:
   Input and respective output data is present. So it comes under Supervised learning.
  3. Regression & Classification:-
   The output data values comes under numbers. So regression is the perfect method to sort this problem.

2.Tell basic info about the dataset:-
  There are 6 Columns and 1339 Rows are present in the dataset

3.Pre-Processing Method:-
  Yes, Here converted categorical data into nominal data which is called **One Hot Coding**. Because AI could not understand words, It only can understand **"0 and 1"**

4.Develop a good model with r2_score:- MULTIPLE LINEAR

  Multiple linear r2_score = 0.7894

 Develop a good model with r2_score:- SVM-REGRESSION

KERNEL	C=0.01 r2_Score	C=10 r2_Score	C=100 r2_Score
rbf	-0.0897	-0.0897	-0.0897
poly	-0.0897	-0.0897	-0.0897
sigmoid	-0.0897	-0.0897	-0.0897
linear	-0.0897	-0.0897	-0.0897

Develop a good model with r2_score:- DECISION TREE

CRITERION	SPLITTER	MAX_FEATURES	r2_Score
friedman_mse	best	auto	0.7178
friedman_mse	best	sqrt	0.7138
friedman_mse	best	log2	0.7521
friedman_mse	random	auto	0.7538
friedman_mse	random	sqrt	0.7162
friedman_mse	random	log2	0.7213

Develop a good model with r2_score:- RANDOM FOREST

n_estimators	criterion	max_features	random_state	r2_Score
20	friedman_mse	sqrt	0	0.8590
20	friedman_mse	log2	0	0.8590
50	friedman_mse	sqrt	0	0.8700
50	friedman_mse	log2	0	0.8700
100	friedman_mse	sqrt	0	0.8709
100	friedman_mse	log2	0	0.8709

5.Reason for choosing this model:-

   I did four types of algorithms, which are **Multiple Linear,SVM,Decision tree,Random Forest.** All r2_score values are noted. First three algorithms are given poor value while random forest gives the best **"r2=0.8709"** value. The accuracy is perfect compared to other algorithms,So I chose the **"RANDOM FOREST"** algorithm.