REGRESSION ASSIGNMENT

1. Identify your problem statement
   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.
   Stage 1: Problem Domain: Machine Learning
   Stage 2: Learning Method: Supervised Learning
   Stage 3: Regression

2. Tell basic info about the dataset (Total number of rows, columns)
   Total number of rows:1338
   Total number of columns:6

3. Mention the pre-processing method if you're doing any (like converting string to number – nominal data)
   Preprocessing method (Conversion of ordinal to nominal) is required for 2 columns such as, Sex column and Smoker column.

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.

TO FIND R2VALUE BY USING THE FOLLOWING MACHINE LEARNING ALGORITHMS

1. MULTIPLE LINEAR REGRESSION (R2 value = 0.7894)

2. SUPPORT VECTOR MACHINE:

S.No	HYPER PARAMETER	LINEAR (r value)	RBF (NON LINEAR) (r value)	POLY (r value)	SIGMOID (r value)
1	WITHOUT STANDARDIZATION	-0.1116	-0.0894	Executes for long time and couldn't find	-0.0897
WITH STANDARDIZATION					
1	C10	0.7737	-0.0322	0.0387	0.0393
2	C100	0.6288	0.3200	0.6179	0.5276
3	C500	0.7631	0.6642	0.8263	0.4446
4	C1000	0.7649	0.8102	0.8566	0.2874
5	C2000	0.7440	0.8547	0.8605	-0.5939

6	<mark>C3000</mark>	0.7414	<mark>0.8663</mark>	0.8598	-2.1244

<mark>The SVM Regression use R2 value (rbf and hyper parameter (C3000))R2 Value=0.8663</mark>
<mark>(poly and hyper parameter (C2000))R2 Value=0.8605</mark>

<mark>3. DECISION TREE:</mark>

S.No	CRITERION	MAX FEATURES	SPLITTER	R VALUE
1	mse	auto	best	0.6910
2	mse	auto	random	0.7242
3	mse	sqrt	best	0.7537
4	mse	sqrt	random	0.5862
5	mse	log2	best	0.7141
6	mse	log2	random	0.6885
7	Mae	auto	best	0.6748
8	Mae	auto	random	0.7360
9	mae	sqrt	best	0.6407
10	mae	sqrt	random	0.7209
11	mae	log2	best	0.7080
12	mae	log2	random	0.6793
13	friedman_mse	auto	best	0.6956
14	friedman_mse	auto	random	0.6729
15	<mark>friedman_mse</mark>	<mark>sqrt</mark>	<mark>best</mark>	<mark>0.7591</mark>
16	friedman_mse	sqrt	random	0.5973
17	friedman_mse	log2	best	0.6910
18	friedman_mse	log2	random	0.6706

4. RANDOM FOREST

S.No	CRITERION	MAX FEATURES	N_ESTIMATORS	R VALUE
1	Mse	auto	10	0.8331
2	Mse	auto	100	0.8539
3	Mse	sqrt	10	0.8520
4	Mse	sqrt	100	0.8709
5	Mse	Log2	10	0.8520
6	Mse	Log2	100	0.8709
7	Mse	auto	10	0.8331
8	mae	auto	10	0.8355
8	mae	auto	100	0.8521
9	mae	sqrt	10	0.8574
10	mae	sqrt	100	0.8717
11	mae	Log2	10	0.8574
12	mae	Log2	100	0.8717

The Random Forest Regression ((mae,sqrt,100),,,,,(mae,log2,100)) R2 value =0.8717

Best Model for this problem is Random Forest
Because R2-Score of this model is best among all other MLR,SVM,DT algorithms.
The Random Forest Regression ((mae,sqrt,100),,,,,(mae,log2,100)) R2 value =0.8717

Insurance Charge Prediction

For age=28, bmi=33, children=3, Male, Non-Smoker

As per MLR, charges will be = Rs. 7123.15

As per SVM, charges will be = Rs. 16589.18

As per Decision Tree, charges will be=Rs. `4449.46`

As per Random Forest, charges will be=Rs. `11177.28`

Best Model: Random Forest, charges will be=Rs. `11177.28`