Assignment - Regression

1. Problem Statement:

By using insurance_pre this data we are going to predict, When the customer need the insurance amount and we need to give them the insurance charges.

2. Basic Info:

The dataset is in Excel sheet and it is number so we use Machine Learning Domain. Input and Output is clear so we use Supervised Learning. Output is in Numerical Form so we use Regression.

Machine Learning--→Supervised Learning-→Regression.

This dataset has 1339 Rows and 6 Columns.

3. Preprocessing Method:

In this data sex column and smoker column is in string type. This name will not be comparable so we use Nominal method to change string into numerical data.

4. Good Model:

I have used all the algorithm of machine learning Like MLR , SVM , Decision Tree, and Random Forest. I got the best model creation in Random Forest for this data.

5.Research Values:

Assignment - Regression

A. Decision Tree:

S No	Criterion	Max Features	Splitter	R-value
1	mse	auto	best	0.6989439315848283
2	mse	auto	random	0.7844482749258906
3	mse	sqrt	best	0.7209933620273439
4	mse	sqrt	random	0.6958343776511821
5	mse	log2	best	0.5707284915756702
6	mse	log2	random	0.7542854327899727
7	Mae	auto	best	0.6768709786342195
8	mae	auto	random	0.7331061564452007
9	mae	sqrt	best	0.7206115179262382
10	mae	sqrt	random	0.7169409323485165
11	mae	log2	best	0.7129887794608034
12	mae	log2	random	0.64713317125108
13	friedman_mse	auto	best	0.7192459036805052
14	friedman_mse	auto	random	0.7013814962540014
15	friedman_mse	sqrt	best	0.7262562323994066
16	friedman_mse	sqrt	random	0.6644955069320679
17	friedman_mse	log2	best	0.5823669200812815
18	friedman_mse	log2	random	0.6968958892484836

Assignment - Regression

B. Multiple Linear Regression(R-value)= 0.7894790349867009

C. Support Vector Machine:

S NO	Hyper Parameter	Linear (r_value)	rbf(Non-linear)r_value	Poly	Sigmoid
1	C=0.1	-0.080959968427891	-0.08907451521042731	-0.08830237655410711	-0.08826991450485111
2	C=10	0.462468414233968	-0.03227329390671052	0.038716222760231456	0.039307143782 74347
3	C=3000	0.7414236599249314	0.8663393950916756	0.8598930084494388	-2.1244194786689854
4	C=4000	0.7414198803066961	0.8717407869762046	0.8600049580588055	-5.510333547108606
5	C=5000	0.7414179301704098	0.8747778174940977	0.8595656394397817	-7.530043237831949
6	C=7000	0.7414223482452932	0.87769236904483	0.8596647468701699	-17.954807014731475
7	C=10000	0.7414230132428099	<mark>0.8779952426221569</mark>	0.8591715079473912	-34.151535978496256

D. Random Forest(R-value): 0.8539235792996565

6.My Final Model Creation done for this dataset is by using Support Vector Machine(C=10000,'rbf').It has good model when compared to others.