

## REGRESSION ASSIGNMENT

### 1. Identify the problem Statement

Output ( dependent variable ) : to predict insurance charges

Input (independent variables): input columns are (age,sex,bmi,children,smoker)

### 2. Tell basic info about the dataset (total number of rows & columns)

Total No of rows: 1338

Total No of columns: 6

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

Stage 1: Domain Selection : Machine Learning

Stage 2: Supervision

Stage 3: Regression

### 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 the following Machine Learning Regression Method by using r2 value

#### 1. MULTIPLE LINEAR REGRESSION :

R2 Value = 0.7894790349867009

#### 2. SUPPORT VECTOR MACHINE:

S.NO	HYPER PARAMETERS	LINEAR (r value)	RBF (Non Linear ) (r value)	POLY (r value)	SIGMOID (r value)
1	Linear	-0.0101026653160	-0.0833823859361	-0.0756996557086	-0.075429242811
2	C=0.01	-0.0888313343916	-0.0896455373986	-0.0895682848767	-0.089565015934
3	C=0.1	-0.0809599684278	-0.0890745152104	-0.0883023765541	-0.088269914504
4	C=1	-0.0101026653160	-0.083382385936	-0.0756996557086	-0.075429242811
5	C=10	0.46246841423396	-0.0322732939067	0.03871622276023	0.0393071437827
6	C=100	0.62887928573203	0.32003178320508	0.61795696240597	0.5276103546510
7	C=1000	0.76493117385963	0.81020648517585	0.85664876759465	0.2874706948697

In SVM Regression R2 Score Value (Linear (r value) and hyper parametor (Poly & C=1000))= 0.856648

### 3. DECISION TREE:

S No	CRITERION	MAX FEATURES	SPLITTER	R VALUE
1	<i>squared_error</i>	Default	default	0.686745635245084
2	Mse	None	best	0.6924537181553276
3	Mse	Auto	best	0.702980391971537
4	Mse	Auto	Random	0.7575632633742704
5	Mse	Sqrt	Best	0.6794948859933209
6	Mse	Sqrt	Random	0.6808270492236784
7.	Mse	Log2	Best	0.6855143731532924
8	Mse	Log2	random	0.6972795146736963
9	<i>friedman_mse</i>	Auto	best	0.7123759726722464
10	<i>friedman_mse</i>	Auto	Random	0.7253697854630787
11	<i>friedman_mse</i>	Sqrt	Best	0.7343636735035739
12	<i>friedman_mse</i>	Sqrt	Random	0.6165920296252206
13	<i>friedman_mse</i>	Log2	Best	0.7116166223739979
14	<i>friedman_mse</i>	Log2	random	0.7090787160685144

In Decision tree Regression R2 Score Value (criterion: Mse (r value) and hyper paramotor (random))= 0.757563

### 4. RANDOM FOREST :

S No	CRITERION	N_estimator	Max Features	Random State	R VALUE
1	Mse	10	auto	0	0.833812800619115
2	Mse	100	Auto	0	0.855040880868016
3	Mse	10	Sqrt	0	0.8529790384841727
4	Mse	100	Sqrt	0	0.8707803670428841
5	Mse	10	Log2	0	0.8529790384841727
6	Mse	100	Log2	0	0.8707803670428841
7	Default	Default	Default	default	0.8374216433248816
8	Mae	10	Auto	0	0.8368344668703614
9	Mae	100	Auto	0	0.8533534639399472
10	Mae	10	Sqrt	0	0.8581144568424336
11	Mae	100	Sqrt	0	0.872574534340602
12	Mae	10	Log2	0	0.8581144568424336
13	Mae	100	Log2	0	0.872574534340602

In Random Forest Regression R2 Score Value (Mae,100,sqrt&log2)=0.872574

The final Machine Learning best method in Regression are mentioned below

1. Random Forest Regression R2 Score Value (Mae,100,sqrt&log2)=0.872574
2. SVM Regression R2 Score Value (Linear (r value) and hyper paramotor (Poly & C=1000))= 0.856648