

# Assignment Regression Algorithm

Step1: 3 stages of problem identification

Stage 1 – Machine Learning

Stage 2 – Supervised Learning

Stage 3 – Regression

Step 2 : Details about dataset

There are 6 columns,1339 rows available.

Inputs – age,sex,bmi,children,smoker

Output- charges

Step 3 : Preprocessing

Input data sex and smoker is a nominal data , so we have to use one hot ending. After preprocessing there are 8 columns,1339 .

Step 4 :

Developing a model.

Step 5 : Find r\_score value for all models

1) Multiple Linear Regression **r\_score value = 0.78**

## 2) Decision Tree :

<b>criterion</b>	<b>splitter</b>	<b>max_features</b>	<b>r_score</b>
mse	best	sqrt	0.68
mse	random	sqrt	0.54
mse	best	Log2	0.58
mse	random	Log2	0.69
Mse	best	auto	0.68
mse	random	auto	0.72
friedman_mse	best	sqrt	0.77
friedman_mse	random	sqrt	0.67
friedman_mse	best	Log2	0.71
friedman_mse	random	Log2	0.74
mse	best	none	0.70

Decision tree regression **r2\_score values** (friedman\_mse,best,sqrt) = 0.77

## 3) Random Forest :

<b>Bootstrap</b>	<b>criterion</b>	<b>max_features</b>	<b>R2_score</b>
<i>true</i>	<i>mse</i>	auto	0.83
bool	<i>mse</i>	auto	0.82
<i>Bool</i>	<i>mse</i>	sqrt	0.85
True	mse	Sqrt	0.85
true	mse	Log2	0.851
<b>n_estimators</b>			
100	mse	Log2	0.86
100	mse	auto	0.855
100	mse	sqrt	0.86
10	mse	sqrt	0.85

Random Forest **r2\_score values** (n\_estimators=100,mse,log2) = 0.86

Step 6 :

The final best machine Learning method of regressor

Random Forest **r2\_score values** (n\_estimators=100,mse,log2) = **0.86**