

**Problem Statement:**

1. Identify Problem statement: 3 stages
  1. Machine Learning
  2. Supervised Learning (Requirement clear & I/p and O/P are also clear)
  3. Regression – O/P labels are in numerical format

2. Tell basics about dataset:

Dataset have 5 input column (age,sex,children,bmi and smoker) and 1 output (charges). The columns (sex and smoker) are in categorical value, so we have to do some preprocessing.

- 3.PreProcessing method:

As mentioned above dataset have categorical value as input and it is nominal type , so we converting to 1 and 0. For that we using “One Hot Encoding” in the code.

4. Develop Model:

Developed MultipleLinearRegression, SVM, DecisionTree,RandomFactor to find the best model, which shows below.

5. Best Model

As per report of r\_score value, Random Forest has chosen as best model and created deployment phase for the respective model. Why, because the r\_score value is given as high accuracy for the given dataset, when compared to the other model.

Please find the report below,

**Multi Linear Regression:**

R\_score value is 0.78

## 1.SVM – Support Vector Machine

<i>Kernel</i>	<i>C</i>	<i>R_score</i>
linear	0	-0.01
	10	0.46
	100	0.62
	1000	0.76
	10000	0.74
rbf	0	-0.08
	10	-0.03
	100	0.32
	1000	0.81
	10000	0.870
poly	0	-0.07
	10	0.03
	100	0.61
	1000	0.85
	10000	0.85
sigmoid	0	-0.07
	10	0.03
	100	0.52
	1000	0.28
	10000	-34.15

Hyper tuning parameter in SVM is kernel="rbf",c=10000 for given dataset

## 2. Decision Tree

<i>crit</i> <b>erion</b>	<i>spl</i> <b>itter</b>	<i>R</i> <b>_score</b>
squared_error (default)	best(default)	0.69
	Random	0.74
friedman_mse	best	0.69
	Random	0.68
absolute_error	best	0.67
	Random	0.72
poisson	best	0.72
	Random	0.71

Hyper tuning parameter in Decision tree is criterion =" squared\_error", splitter=random for given dataset

## 3. Random Forest

<i>crit</i> <b>erion</b>	<i>max</i> <b>_features</b>	<i>R</i> <b>_score</b>
squared_error	sqrt	0.872
	log2	0.866
friedman_mse	sqrt	0.862
	Log2	0.871
absolute_error	Sqrt	0.870
	Log2	0.871
poisson	Sqrt	0.871
	Log2	0.870

Hyper tuning parameter in Decision tree is criterion ="squared\_error", max\_features=sqrt for given dataset

By analysing the above hyper tuning report **RandomForest** given **high** accuracy when compared to the other model for the given data set.

So we **saving RandomForest model** for the deployment