

## ML Assignment – Regression

### Problem statement:

We are predicting the Insurance Charges based on the person factors - Age, Sex, BMI, No. of Children, Smoking Habit.

### Data Analyzation:

Based on the .csv data provided we have

Total number of rows = 1338

Total number of columns=6

Here, there are two Columns we are converting string to number using Nominal Data Method,

Column Sex – String Converted to number

Column Smoker - String Converted to number

### Model Outputs with R2 Score

- Generated Models and their R2 Score based on Different Algorithms below, also highlighted the best Scores from different algorithms.
- Based on R2 Score results from different algorithms – I am suggesting ‘Decision Tree’ Algorithm for this dataset with a r2 Score of 85.49%.
- Provided the results with Deployment for **Decision Tree Algorithm**.

#### 1. Decision Tree

Decision Tree				
criterion	splitter	max_depth	r2 Score	Comments
squared_error	best	5	0.8247592	
squared_error	random	5	0.8372053	
friedman_mse	best	5	0.82475929	

friedman_mse	random	5	0.83541738	
absolute_error	best	5	0.854933483	Best Model in Decision Tree
absolute_error	random	5	0.84608289	
poisson	best	5	0.8370067	
poisson	random	5	0.835954311	

## 2. Random Forest

Random Forest				
n_estimators	criterion	max_features	r2 Score	Comments
default=100	squared_error	sqrt	0.869852	
default=100	squared_error	log2	0.8698522	Best Model in Random Forest
default=100	absolute_error	sqrt	0.8657831	
default=100	absolute_error	log2	0.8657831	
default=100	friedman_mse	sqrt	0.868162	
default=100	friedman_mse	log2	0.868162	
default=100	poisson	sqrt	0.8687362	
default=100	poisson	log2	0.8687362	
default=100			0.8552594	

## 3. SVMR

Support Vector Machine			
Kernel	degree	r2 Score	Comments
poly	1	-0.088757	None of the model have best results
poly	0	-0.089709	
linear		-0.1001392	
rbf		-	
		0.08851297	
sigmoid	1	-	
		0.08992538	

precomputed			Cannot be Calculated as SVR requires a square kernel matrix as input, but your X_train is a 35x5 matrix (35 samples with 5 features).
-------------	--	--	---------------------------------------------------------------------------------------------------------------------------------------

#### 4. Multi Linear Regression

R2 Score Determined in Multi Linear Regression is 0.78947903