

1) Problem Identification

Domain: Machine Learning
It comes under Supervised learning beause I/P & O/P are present
It is Regression as the O/P is continuous numerical value
Simple linear may not be not used since we have more than 2 inputs
We need to predict Insurance Charges

2) Basic info of dataset

Dataset has 1338 rows and 6 columns

3) Pre-Processing method

Dataset contains COLUMNS "Sex" & "Smoker". This data is **Nominal** data.
It has been converted as numerical values (binary values (i.e) 1, 0)
I have used **one hot encoding** method to convert it as numerical data as this is Nominal data

4) Model development

Models developed using regression algorithms and uploaded the ipython files along with this document

5) R Square Values of Models

Multiple linear Regression	R Square Value
	0.78

Support Vector Machine					
(C) Penalty or Amount of regularization default C = 1.0	R Square (linear)	R Square (defalut = rbf)	R Square (poly)	R Square (sigmoid)	R Square (precomputed) N/A for this data set
0.01	-0.0798	-0.0897	-0.0893	-0.0897	My dataset is not a square matrix
0.001	-0.0879	-0.0897	-0.0897	-0.0897	
0.0001	-0.8895	-0.0897	-0.0897	-0.0897	
0.1	-0.1221	-0.0896	-0.0862	-0.0897	
1.00	-0.1115	-0.0884	-0.0645	-0.0899	
2.00	-0.1003	-0.0873	-0.0671	-0.09	
1000.00	0.6339	-0.1176	-0.0547	-1.7112	

Decision Tree		
criterion default = squared_error	Splitter default = best	R Square
squared_error	best	0.6999
friedman_mse	best	0.6833
absolute_error	best	0.7158
poisson	best	0.6675
squared_error	random	0.7188
friedman_mse	random	0.7514
absolute_error	random	0.7308
poisson	random	0.7577

Random Forest					
criterion default = squared_error	max_features (default=1.0)	n_estimators (default =100)	min_samples_split (default=2)	random_state (default=0)	R Square
squared_error	sqrt	100	2	0	0.8661
friedman_mse	sqrt	80	10	0	0.8789
absolute_error	sqrt	50	5	0	0.8773
poisson	sqrt	50	10	0	0.8723
squared_error	log2	100	2	0	0.8661
friedman_mse	log2	80	20	0	0.8763
absolute_error	log2	100	10	0	0.8835
poisson	log2	50	5	0	0.8741
squared_error	1	100	2	0	0.8537
friedman_mse	1	80	5	0	0.8574
absolute_error	1	100	5	0	0.8637
poisson	1	50	2	0	0.8384
squared_error	2	100	2	0	0.8661
friedman_mse	2	50	5	0	0.8756
absolute_error	2	50	10	0	0.8811
poisson	2	50	5	0	0.8741

6) Final Model

I have chosen the best finalized model as the model which was created by **Random forest regression** algorithm
Because the R2_score value is 0.88 that is closer to 1