

## Assignment-Regression Algorithm

### 1. Problem Statement Identification

<b>PREDICTION</b>	Insurance Charges
<b>DOMAIN SELECTION</b>	Machine Learning / Deep Learning
<b>LEARNING SELECTION</b>	Supervised-Learning
<b>REGRESSION / CLASSIFICATION</b>	Supervised Learning - Regression

### 2. Tell basic info about the dataset (total num of Rows & columns )

Total num of Rows : 1338

Total num of Columns : 6

Inputs / Features : Age, sex, bmi, children, smoker

Output / Target : Charges

Categorical Data columns : Sex, smoker

### 3. Pre-processing Method

One-Hot Encoding performed on : Columns (Sex, smoker) [ Nominal → Numerical Data ]

#### 4. Model Creation

I. Simple Linear Regression (r\_score value) = **0.7894**

II. Multiple Linear Regression (r\_score value) = **0.7894**

III. Support Vector Machine

Default Parameters: ( *kernel='rbf', C=1.0*)

S.No	Rbf (r_score)	Linear (r_score)	Poly (r_score)	Sigmoid (r_score)	C
1.	0.8102	0.7649	0.8566	0.2874	1000
2.	0.8547	0.7440	0.8605	-0.5939	2000
3.	0.8663	0.7414	0.8598	-2.1244	3000
4.	<b>0.8717</b>	0.7414	0.8600	-5.5103	4000

R\_score Value = **0.8717**

(Kernel ="rbf" , c=4000)

(Dataset FileName : Insurance\_pre (1))

#### IV. Decision Tree

Default Parameters: (criterion="squared\_error", splitter="best")

S.NO	CRITERION	SPLITTER	R_SCORE VALUE
1.	Squared_error	Best	0.6850
2.	Squared_error	random	0.7124
3.	friedman_mse	Best	0.7028
4.	friedman_mse	random	0.6553
5.	absolute_error	Best	0.6711
6.	absolute_error	random	0.7465
7.	poisson	Best	0.7332
8.	poisson	random	0.6941

R\_score Value = 0.7465

(criterion='absolute\_error', splitter='random')

## V. Random Forest

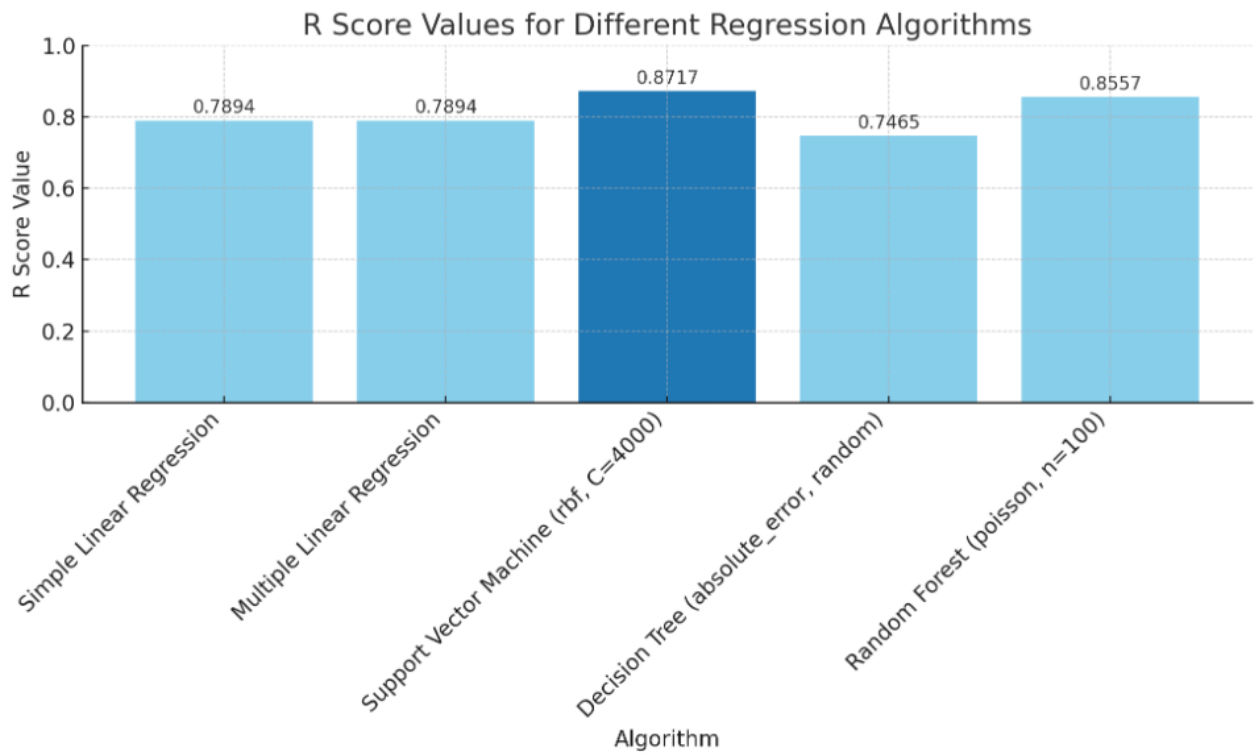
Default Parameters: ( criterion="squared\_error", n\_estimators=100 )

S.NO	CRITERION	n_estimators	R_SCORE VALUE
1.	squared_error	100	0.8494
2.	squared_error	200	0.8542
3.	absolute_error	100	0.8529
4.	absolute_error	200	0.8519
5.	friedman_mse	100	0.8554
6.	friedman_mse	200	0.8526
7.	poisson	100	0.8557
8.	poisson	200	0.8536

R\_score Value = 0.8557

( criterion='poisson', n\_estimators=100 )

(Dataset FileName : Insurance\_pre (1))



**Final Model:** Support vector Machine Algorithm (r\_score = 0.8717)

**Justification:** Based on the above analysis **SVM** performs comparatively better when compared to other algorithms