

## Regression Assignment

### 1. Identify the Problem statement

There are three stages of AI Prediction

Stage 1:

Domain Selection: Machine Learning

Stage 2:

Learning: Supervised

Stage 3:

Classification: Regression

### 2. Information about the dataset

Insurance\_pre.csv contains 1338 rows × 6 columns

### 3. Preprocessing Method

The Insurance\_pre.csv has Nominal Data

### 4. r2 values for each algorithm and their respective max\_features, C parameter

1. Multiple Linear Regression **r2 value ( )** is 0.7894790349867009

### 2. Support Vector Machine

serial.No	Hyper parameter	Linear R value	RBF(Non Linear R value)	Poly (r value)	Sigmoid (r value)
1	C10	0.462468414233968	-0.03227329390671052	0.038716222760231345	0.03930714378274347
2	C100	0.6288792857320359	0.3200317832050831	0.6179569624059795	0.5276103546510407
3	C500	0.763105797597537	0.6642984645143137	0.826368354126896	0.44460610338694795
4	C1000	0.7649311738608445	0.8102064851758545	0.8566487675946551	0.28747069486976695
5	C2000	0.7440418308107487	0.854776642539298	0.860557927577388	-0.5939509731283508

6	C3000	0.74142365 99249815	0.86633939 53081686	0.8598930 084494408	-2.124419 478668986 3

Support Vector Machine best **r2 value** ( **RBF(Non Linear)** and HyperParameter (**C3000**)) is **0.8663393953081686**

### 3. Decision Tree

Serial.No	<i>criterion</i>	Max_features	Splitter	R_score Value
1	Mse	auto	best	0.6871945632 68898
2	Mse	auto	random	0.7058331807 027536
3	Mse	sqrt	best	0.5780449056 811394
4	Mse	Sqrt	random	0.6717749850 77185
5	Mse	Log2	best	0.6799833114 025472
6	Mse	Log2	random	0.6801072400 316461
7	Mae	auto	best	0.6648432062 010856
8	Mae	auto	random	0.7500551403 583591
9	Mae	sqrt	best	0.7339690095 918607
10	Mae	sqrt	random	0.6941169283 379627
11	Mae	Log2	best	0.5658811496 163172
12	Mae	Log2	random	0.7028754036 814628
13	<b><i>friedman_mse</i></b>	auto	best	0.7188171695 197177
14	<b><i>friedman_mse</i></b>	auto	random	0.6783110980 709615

15	<b><i>friedman_mse</i></b>	sqrt	best	0.7055531104 846458
16	<b><i>friedman_mse</i></b>	sqrt	random	0.5952600931 249437
17	<b><i>friedman_mse</i></b>	Log2	best	0.7963033250 752416
18	<b><i>friedman_mse</i></b>	Log2	random	0.6226609825 934093
19	<b><i>poisson</i></b>	auto	best	0.6792193646 949658
20	<b><i>poisson</i></b>	auto	random	0.6982855340 536165
21	<b><i>poisson</i></b>	sqrt	best	0.6106388856 607337
22	<b><i>poisson</i></b>	sqrt	random	0.6542909381 256115
23	<b><i>poisson</i></b>	Log2	best	0.5227100972 953048
24	<b><i>poisson</i></b>	Log2	random	0.6467486767 283614

Descision Tree best **r2 value** (criterion (***friedman\_mse***), Max\_features (Log2), splitter (Best)) is **0.7963033250752416**

#### 4.Random Forest

Serial.No	<i>criterion</i>	Max_features	<b>n_estimators</b>	R_score Value
1	Mse	auto	10	0.8446897393 177992
2	Mse	auto	100	0.8582064953 815349
3	Mse	sqrt	10	0.8535877981 749798
4	<b>Mse</b>	<b>sqrt</b>	<b>100</b>	<b>0.8722476146</b> <b>127074</b>

5	Mse	Log2	10	0.8584178716708083
6	Mse	Log2	100	0.8706412205393324
7	Mae	auto	10	0.8205612752559868
8	Mae	auto	100	0.850151429569389
9	Mae	sqrt	10	0.8500750883670041
10	Mae	sqrt	100	0.8717456643495153
11	Mae	Log2	10	0.8543698071729846
12	Mae	Log2	100	0.8736539550478307

Random Forest best **r2 value** (*criterion (mae), Max\_features (Log2), N\_estimator is 100*) is **0.8736539550478307**

The final Machine Learning Best Model of Regression:

From implementing various algorithm in order to find the best model .I found that both **Random Forest** and **Support Vector Machine** has best **R\_score values**

1.Random Forest best **r2 value** (*criterion (mae), Max\_features (Log2), N\_estimator is 100*) is **0.8736539550478307**  
or

2.Support Vector Machine best **r2 value**(*RBF(Non Linear) and HyperParameter (C3000)*) is **0.8663393953081686**