

# MACHINE LEARNING REGRESSION ASSIGNMENT

## STATEMENT

### 1. PROBLEM IDENTIFICATION:

- First, we collect the data from the client, then from the collected data, we have to predict the insurance charge.
- Machine learning  Supervised learning  Regression

### 2. Info about the dataset:

- 1338 rows x 6 columns

### 3. Pre-processing method:

- sex\_male (male & female) ; smoker\_yes (yes & no)

### 4. Model - Support Vector Machine

### 5. Research values:

- Support vector machine:

S.NO	HYPER PARAMETER	LINEAR (R_score)	RBF (R_score)	POLY (R_score)	SIGMOID (R_score)
1	C10	-0.0396	-0.0568	-0.0536	-0.0547
2	C100	0.1064	-0.0507	-0.0198	-0.0304
3	C500	0.5928	-0.0243	0.1146	-0.0705
4	C1000	0.7802	0.0067	0.2661	0.1850
5	C2000	0.8467	0.0675	0.4810	0.3970
6	C100000	0.8672	0.8724	0.7326	0.6282

- Decision tree:

<u>S.NO</u>	CRITERION	MAX FEATURES	SPLITTER	R_SCORE
1	friedman_mse	sqrt	random	0.2327
2	friedman_mse	log2	random	0.6617
3	friedman_mse	log2	best	0.6590
4	friedman_mse	sqrt	best	0.6263
5	friedman_mse	Auto	Best	0.6106
6	friedman_mse	auto	random	0.6295
7	absolute_error	sqrt	best	0.6226
8	ae	sqrt	random	0.4040
9	ae	Log2	best	0.3612
10	ae	Log2	random	0.4862
11	ae	Auto	Best	0.4382
12	ae	auto	random	0.5444
13	poisson	Log2	random	0.3829
14	Poisson	Log2	best	0.6105
15	Poisson	Sqrt	Best	0.6007
16	poisson	sqrt	random	0.6963
17	poisson	Auto	random	0.5867
18	poisson	auto	best	0.6181

#### 6. FINAL MODEL NAME:

- Support Vector Machine – This is a good model that's r\_score value is near by 1 with comparing other models.