1) Problem Statement:

Insurance Charge Prediction based on several parameters

2) Data set Total number of rows and columns

1338 rows × 5 columns

There are two columns are classifier, we convert string to numerical value using ordinal data (Label encoding algorithms) because compare possible.

Stage 1- Domain (Machine Learning)

Stage 2 - Supervised Learning

Stage 3 - Regression

To find following the machine learning regression method using in r2 value

1.Multiple Linear Regression:

R2 Value = 0.7891

2. Support Vector Machine:

S.NO	Hyper Parameter	Linear (r value)	RBF(Non-Linear) (r value)	Poly (r value)	Sigmoid (r value)
1	C10	0.4624	-0.0323	0.0386	0.0394
2	C100	0.6289	0.3196	0.6164	0.5268
3	C500	0.7630	0.6616	0.8285	0.4429
4	C1000	0.7648	0.8107	0.8546	0.2120
5	C2000	0.7439	0.8540	0.8583	-0.6216
6	C3000	0.7413	0.8646	0.8580	-2.1431

The **SVM Regression** use R2 value (RBF(non linear) and hyper parameter(C3000)) = 0.8646

3. Decision Tree:

S.NO	CRITERION	MAX FEATURES	SPLITTER	R VALUE
1	**squared_error**	Sqrt	Best	0.7780
2	**squared_error**	log2	random	0.6125
3	**squared_error**	Sqrt	random	0.6256
4	**squared_error**	log2	best	0.7129
5	**absolute_error**	Sqrt	Best	0.7274
6	**absolute_error**	log2	Random	0.6460
7	**absolute_error**	Log2	Best	0.7158
8	**absolute_error**	Sqrt	Random	0.6330
9	**friedman_mse**	Log2	Best	0.6980
10	**friedman_mse**	Sqrt	random	0.7009
11	**friedman_mse**	Sqrt	Best	0.6253
12	**friedman_mse**	log2	random	0.6364
13	**poisson**	Log2	Best	0.6395
14	**poisson**	Sqrt	random	0.6779
15	**poisson**	Sqrt	Best	0.7054
16	**poisson**	log2	random	0.6550
17	**poisson**	None	best	0.6955
18	**poisson**	None	random	0.6812
19	**squared_error**	None	random	0.6951
20	**squared_error**	None	best	0.7019
21	**absolute_error**	None	Best	0.6843
22	**absolute_error**	None	Random	0.7261
23	**friedman_mse**	None	Best	0.6855
24	**friedman_mse**	None	random	0.6817

The **Decision tree Regression** use **R2 value (squared_error, sqrt, best) = 0.7780**

4. Random Forest

S.NO	CRITERION	MAX_FEATURES	N_ESTIMATORS	R VALUE
1	squared_error	Auto	10	0.8453
2	squared_error	Auto	100	0.8535
3	squared_error	Sqrt	10	0.8515
4	squared_error	Sqrt	100	0.8614

5	squared_error	Log2	10	0.8469
6	squared_error	Log2	100	0.8640
7	absolute_error	Auto	10	0.8410
8	absolute_error	Auto	100	0.8585
9	absolute_error	Sqrt	10	0.8602
10	absolute_error	Sqrt	100	0.8722
11	absolute_error	Log2	10	0.8565
12	absolute_error	Log2	100	0.8676
13	friedman_mse	Auto	10	0.8482
14	friedman_mse	Auto	100	0.8514
15	friedman_mse	Sqrt	10	0.8586
16	friedman_mse	Sqrt	100	0.8645
17	friedman_mse	Log2	10	0.8501
18	friedman_mse	Log2	100	0.8654
19	poisson	Auto	10	0.8390
20	poisson	Auto	100	0.8557
21	poisson	Sqrt	10	0.8417
22	poisson	Sqrt	100	0.8642
23	poisson	Log2	10	0.8496
24	poisson	Log2	100	0.8667

The **Random Forest** Regression **R2 Value** (absolute_error,sqrt,100) = **0.8722**

The final machine learning regression best method:

The **Random Forest** Regression **R2 Value** (absolute_error,sqrt,100) = **0.8722**

(OR)

The **SVM Regression** use R2 value (RBF(non linear) and hyper parameter(C3000)) = 0.8646