1. **PROBLEM STATEMENT:**

A client's requirement is, he wants to predict the insurance charges based on the parameters like age, sex, BMI, childrens and smoking.

STAGE-1:Domain Selection

Machine Learning

STAGE-2:Learning Selection

The requriment is clear also both input and output is presented in the dataset. So learning selection is **Supervised** .

STAGE-3:Regression

2. **BASIC INFO ABOUT DATASET:**

Total Rows: 1338

Total Columns: 6

3. **Data Preprocessing:**

Converted categorical data like sex, smoker into numerical using **One Hot Encoding** and also string into integer using code **dtype=int.**

4. **R values of the models:**

LINEAR REGRESSION: 0.7894

SUPPORT VECTOR MACHINE:

KERNEL	GAMMA	C	R values
Rbf	Scale	100	-0.1248
Rbf	Auto	100	-0.0745
Linear	Scale	100	0.5432
linear	Auto	100	0.5432

DECISION TREE:

S.NO	**CRITERION**	**MAX FEATURES**	**SPLITTER**	**R values**
1	Squared error	Auto	Best	0.68067
2	Squared error	Auto	Random	0.74004
3	Squared error	Sqrt	Best	0.5855
4	Squared error	Sqrt	Random	0.75528
5	Squared error	Log2	Best	0.6525
6	Squared error	Log2	Random	0.7127

7	Absolute error	Auto	Best	0.6933
8	Absolute error	Auto	Random	0.6880
9	Absolute error	Sqrt	Best	0.6916
10	Absolute error	Sqrt	Random	0.7485
11	Absolute error	Log2	Best	0.7032
12	Absolute error	Log2	Random	0.7048
13	Friedman mse	Auto	Best	0.6847
14	Friedman mse	Auto	Random	0.6625
15	Friedman mse	Sqrt	Best	0.70060
16	Friedman mse	Sqrt	Random	0.6023
17	Friedman mse	Log2	Best	0.7331
18	Friedman mse	Log2	random	0.6948

RANDOM FOREST:

S.NO	CRITERION	MAX FEATURES	N_estimators	R values
1	Squared error	None	50	0.8498
2	Squared error	None	100	0.8538
3	Squared error	Sqrt	50	0.8695
4	Squared error	Sqrt	100	0.8710
5	Squared error	Log2	50	0.8695
6	Squared error	Log2	100	0.8710
7	Absolute error	None	50	0.8526
8	Absolute error	None	100	0.8520
9	Absolute error	Sqrt	50	0.8708
10	Absolute error	Sqrt	100	0.8710
11	Absolute error	Log2	50	0.8708
12	Absolute error	Log2	100	0.8710

13	Friedman mse	None	50	0.8500
14	Friedman mse	None	100	0.8540
15	Friedman mse	Sqrt	50	0.8702
16	Friedman mse	Sqrt	100	0.8710
17	Friedman mse	Log2	50	0.8702
18	Friedman mse	Log2	100	0.8710
19	Poisson	None	50	0.8491
20	Poisson	None	100	0.8526
21	Poisson	Sqrt	50	0.8632
22	Poisson	Sqrt	100	0.8680
23	Poisson	Log2	50	0.8632
24	Poisson	Log2	100	0.8680

5. FINAL MODEL SELECTION:

Random forest R^2 value (Fierdman, sqrt&log2, 100)=0.8701