1.Identify your problem statement:

Client requirement is he want to predict the insurance charge, **insurance charge** is the output, so have clear requirement it's belongs to **supervised machine learning**

Three Stages

- Machine Learning- numbers
- Supervised Machine Learning – requirement is very clear
- Regression – output is continuous values

2.Basic info about the data set:

Total 6 columns and 1338 rows.

Input: age, sex, bmi, children, smoker

Output: charge

3. Pre-processing method:

Pre-processing method is Nominal data (one hot encoding) because input contains text, we are using get_dummies method and the parameters we are passing dataset, dtype=int, drop_first=true

4. Develop a good model with r2_score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model.

Random Forest

Sno	N_ESTIMATORS	CRITERION	MAXFEATURES	R2 VALUE
1	50	***squared_error***	***Sqrt***	0.8701028385180851
2	100	***squared_error***	***Sqrt***	0.8718001587457894
3	50	***squared_error***	***log2***	0.8681996057614894
4	100	***squared_error***	***log2***	0.8695327331318028
5	50	***squared_error***	***None***	0.852554550341361
6	100	***squared_error***	***None***	0.8554279128535067
7	50	***absolute_error***	***Sqrt***	0.8658715988246394
8	100	***absolute_error***	***Sqrt***	0.8716449460920973
9	50	***absolute_error***	***log2***	0.8697756502668389
10	100	***absolute_error***	***log2***	0.8719872939568064
11	50	***absolute_error***	***None***	0.849152753559274
12	100	***absolute_error***	***None***	0.8545582387239524
13	50	***friedman_mse***	***Sqrt***	0.8688820744852159
14	100	***friedman_mse***	***Sqrt***	0.8713481734076576
15	50	***friedman_mse***	***log2***	0.8622413870992105
16	100	***friedman_mse***	***log2***	0.8700619856800212
17	50	***friedman_mse***	***None***	0.8465958994853551
18	100	***friedman_mse***	***None***	0.8592750560471338
19	50	***Poisson***	***Sqrt***	0.8705070945780351
20	100	***Poisson***	***Sqrt***	0.8711288410480799
21	50	***Poisson***	***log2***	0.8638015283716634
22	100	***Poisson***	***log2***	0.869178851414411
23	50	***Poisson***	***None***	0.8522570024085658
24	100	***Poisson***	***None***	0.848362395429203

5. All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

Multiple Linear Regression:

R2 value: 0.7894790349867009

Support Vector Machine:

Hyper Parameter	Linear (r value)	RBF (NON LINEAR) (r value)	POLY (r value)	SIGMOID (r value)
C10	0.462468414233968	-0.032273293906710 52	0.038716222760231456	0.03930714378274347
C100	0.628879285732034 6	0.320031783205083 2	0.617956962405979 5	0.52761035465104 07
C500	0.763105797597539 3	0.664298464514313 7	0.826368354126893 4	0.44460610338694795
C1000	0.764931173864967 2	0.810206485175854 5	0.856648767594652 4	0.28747069486978516
C1500	0.744048708098290 5	0.842749470197823 5	0.858088921156268 6	-0.06744112500716382
C2000	0.744041830810786 6	0.854776642539298	0.860557928865969	-0.59395097312835 08

C3000	0.7414236599248631	0.8663393953081686	0.8598930084494358	-2.1244194786689863

Decision Tree

No	CRITERION	SPLITTER	MAX FEATURES	R VALUE
1	***squared_error***	Best	Sqrt	0.7304275213183836
2	***squared_error***	Random	Sqrt	0.7045493394573541
3	***squared_error***	Best	Log2	0.6960452317124439
4	***squared_error***	Random	Log2	0.7287209056740023
5	***squared_error***	Best	None	0.7300613818699182
6	***squared_error***	random	None	0.7209408801405768
7	***friedman_mse***	Best	Sqrt	0.7002606314979025
8	***friedman_mse***	Random	Sqrt	0.6651239025650819
9	***friedman_mse***	Best	Log2	0.704450536546119
10	***friedman_mse***	Random	Log2	0.7216578813305338
11	***friedman_mse***	Best	None	0.7306632458144676
12	***friedman_mse***	Random	None	0.7226076851495364
13	***absolute_error***	Best	Sqrt	0.6996688282460393
14	***absolute_error***	Random	Sqrt	0.6726554796535373
15	***absolute_error***	Best	Log2	0.7205577192254702
16	***absolute_error***	Random	Log2	0.6920500725802841
17	***absolute_error***	Best	None	0.7139233210656045
18	***absolute_error***	Random	None	0.7374643598233536
19	***poisson***	Best	Sqrt	0.4958275379209969
20	***poisson***	Random	Sqrt	0.6737218629045889
21	***poisson***	Best	Log2	0.6794969091733786
22	***poisson***	Random	Log2	0.7144451697711103
23	***poisson***	Best	None	0.747166745095021
24	***poisson***	Random	None	0.7066731783283752

Random Forest

Sno	N_ESTIMATORS	CRITERION	MAXFEATURES	R2 VALUE
1	50	***squared_error***	***Sqrt***	0.8701028385180851
2	100	***squared_error***	***Sqrt***	0.8718001587457894
3	50	***squared_error***	***log2***	0.8681996057614894
4	100	***squared_error***	***log2***	0.8695327331318028
5	50	***squared_error***	***None***	0.852554550341361
6	100	***squared_error***	***None***	0.8554279128535067
7	50	***absolute_error***	***Sqrt***	0.8658715988246394
8	100	***absolute_error***	***Sqrt***	0.8716449460920973
9	50	***absolute_error***	***log2***	0.8697756502668389
10	100	***absolute_error***	***log2***	0.8719872939568064
11	50	***absolute_error***	***None***	0.849152753559274
12	100	***absolute_error***	***None***	0.8545582387239524
13	50	***friedman_mse***	***Sqrt***	0.8688820744852159
14	100	***friedman_mse***	***Sqrt***	0.8713481734076576
15	50	***friedman_mse***	***log2***	0.8622413870992105
16	100	***friedman_mse***	***log2***	0.8700619856800212
17	50	***friedman_mse***	***None***	0.8465958994853551
18	100	***friedman_mse***	***None***	0.8592750560471338
19	50	***Poisson***	***Sqrt***	0.8705070945780351
20	100	***Poisson***	***Sqrt***	0.8711288410480799
21	50	***Poisson***	***log2***	0.8638015283716634
22	100	***Poisson***	***log2***	0.869178851414411
23	50	***Poisson***	***None***	0.8522570024085658
24	100	***Poisson***	***None***	0.848362395429203

6. Mention your final model, justify why u have chosen the same.

Best model is Random Forest because compare to other algorithm r2_value's, Random Forest Algorithm r2_value is highest (0.8718001587457894) and parameters are given below.

100	***squared_error***	***Sqrt***	0.8718001587457894

7. Kindly create Repository in the name Regression Assignment.

https://github.com/SENTHILKUMARTH/Week-3-Machine-Learning-Regression/tree/main/Assignment