1.) Identify your problem statement -we have clear information for predict the value

Stage1: domain selection-machine learning

Stage2: supervised learning

Stage3: Regression

2.) Tell basic info about the dataset (Total number of rows, columns)

Total number of rows and column :1338 rows × 5 columns

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

Yes we have the categorical column so we are converting string to number using nominal data-one hot encoding method

To find following the machine learning regression method using in R2 value

1.Multiple linear regression (R2 value) = 0.7100923694262504

2.Support vector machine:

S.NO	Hyper parameter	Linear (R value)	RBF (Non linear R value)	Poly R value	Sigmoid R value
1	C=1000	0.040	-11.6	-10.9	-1.98
2	C=2000	0.340	-11.27	-9.78	-0.66
3	C=5000	0.736	-10.8	-5.68	-0.19

The SVM Regression using R2 Value (Linear and hyper parameter C5000)=0.736

3.Decision Tree

S.NO	criterion	splitter	R value
1	squared_error	best	0.73
2	friedman_mse	best	0.74
3	absolute_error	best	0.68
4	poisson	best	0.74
5	squared_error	random	0.67
6	friedman_mse	random	0.71
7	absolute_error	random	0.75
8	poisson	random	0.70

The Decision tree Regression using R2 Value (*criterion =absolute_error* and *splitter = random*)=0.75

4.Random forest regression (R2 value) = 0.8449642099190535

The final machine learning best method of regression:

4.Random forest regression (R2 value) = 0.8449642099190535