

PROBLEM STATEMENT OR REQUIREMENT:

A client's requirement is, he wants to predict the insurance charges based on the several parameters. The Client has provided the dataset of the same.

1) Identify your problem statement

- a) We have to collect the dataset from client.
- b) Insurance charge prediction
- c) identify three stages

Machine Learning → Supervised_Learning → Regression

2.) Tell basic info about the dataset (Total number of column and rows)

It has (1339-Rows) and (7-Columns)

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

Female, Male → (catogarical data convertes into numerical data (0/1))

Smoker_yes/no → (catogarical data convertes into numerical data (0/1))

4) All the research values (r2_score of the models) should be documented.

(You can make tabulation or screenshot of the results.)

1.MULTIPLE LINEAR REGRESSION

r2_score = 0.78

2.SUPPORT VECTOR MACHINE

s.no	Hyper parameter	Linear (r_score)	Rbf (r_score)	Poly (r_score)	Sigmoid (r_score)
1	10	0.46	-0.03	0.03	0.03
2	100	0.62	0.32	0.61	0.52
3	500	0.76	0.66	0.82	0.44
4	1000	0.76	0.81	0.85	0.28
5	2000	0.74	0.85	0.86	-0.59

SVM (C=2000, kernel=poly) **r2_score =0.86**

3. DECISION VECTOR MACHINE

s.no	criterion	Max_features	splitter	r_score value
1	Mse	Auto	Best	0.68
2	Mse	Auto	Random	0.70
3	Mse	Log2	Best	0.66
4	Mse	Log2	Random	0.59
5	Mse	Sqrt	Best	0.61
6	Mse	Sqrt	Random	0.62
7	Mae	Auto	Best	0.68
8	Mae	Auto	Random	0.76
9	Mae	Log2	Best	0.69
10	Mae	Log2	Random	0.72
11	Mae	Sqrt	Best	0.70
12	Mae	Sqrt	Random	0.68
13	Friedman_mse	Auto	Best	0.68
14	Friedman_mse	Auto	Random	0.69
15	Friedman_mse	Log2	Best	0.68
16	Friedman_mse	Log2	Random	0.72
17	Friedman_mse	Sqrt	Best	0.75
18	Friedman_mse	sqrt	random	0.73

The Decision Tree Regression use R2_value

(Criterion-friedman_mse,

max_features- auto,

splitter- best)

r2_value =0.76

4.DECISION TREE

s.no	criterion	Max_features	N_estimators	r_score
1	mse	Auto	100	0.85
2	Mse	Auto	50	0.85
3	Mse	Sqrt	100	0.87
4	Mse	Sqrt	50	0.86
5	Mse	Log2	100	0.87
6	Mse	Log2	50	0.86
7	Mas	Auto	100	0.85
8	Mas	Auto	50	0.85
9	Mas	Sqrt	100	0.87
10	Mas	Sqrt	50	0.86
11	Mas	Log2	100	0.87
12	mas	Log2	50	0.87
13	Friedman_mse	Auto	100	0.85
14	Friedman_mse	Auto	50	0.85
15	Friedman_mse	Log2	100	0.87
16	Friedman_mse	Log2	50	0.87
17	Friedman_mse	sqrt	100	0.87
18	Friedman_mse	sqrt	50	0.87

Many models are give the same r_score value 0.87

The random forest regression (mse, log2, 100) r2_value=0.87

5) Develop a good model with r^2 _score. You can use any machine learning algorithm; you can create many models. Finally, you have to come up with final model

Final machine learning best method of regression:

Random forest r _score value(mse, log2, 100) r^2 _value=0.87

In the random forest compare to other algorithms it give the better accuracy then others. So I have to choice Random Forest