

Problem Statement or Requirement:

A client's requirement is, he wants to predict the insurance charges based on

No	Kernel	C	Epsilon	R*R
----	--------	---	---------	-----

the several parameters. The Client has provided the dataset of the same.

As a data scientist, you must develop a model which will predict the insurance charges.

1.) Identify your problem statement

ML – Regression – Supervised

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

Rows – 1338

Columns – 6

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

Ordinal data

4.) 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.

Multiple linear Regression:

$R^2 = 0.78$

SVM R^2 value Comparision:

Default:

Kernel=rbf, $c=1000$ $R^2 = 0.8102$

No	kernel	C	Epsilon	R*R
1	Linear	1	0.1	-0.0101
2	Linear	1	1	-0.0101
3	Linear	0.5	0.1	-0.0477
4	Linear	0.5	1	-0.0477
5	Poly	1	0.1	-0.0756
6	Poly	1	1	-0.0756
7	Poly	0.5	0.1	-0.0826
8	Poly	0.5	1	-0.0826
9	Rbf	1	0.1	-0.0833
10	Rbf	1	1	-0.0833
11	Rbf	0.5	0.1	-0.0865
12	Rbf	0.5	1	-0.0865
13	Sigmoid	1	0.1	-0.0754
14	Sigmoid	1	1	-0.0754
15	Sigmoid	0.5	0.1	-0.0825
16	Sigmoid	0.5	1	-0.0825
17	Precomputed	1	0.1	Error: Input is a 936x5 matrix.
18	Precomputed	1	1	Error: Input is a 936x5 matrix.
19	Precomputed	0.5	0.1	Error: Input is a 936x5 matrix.
20	Precomputed	0.5	1	Error: Input is a 936x5 matrix.

Decision Tree R*R Value Comparision:

Default:

Criterion= friedman_mse , splitter = Random , R*R= 0.7088

No	criterion	splitter	R*R
1	squared_error	Best	0.68
2	squared_error	Random	0.69
3	friedman_mse	Best	0.68
4	friedman_mse	Random	0.74
5	absolute_error	Best	0.65
6	absolute_error	Random	0.77
7	Poisson	Best	0.72
8	poisson	random	0.68

Random forest:

$R^2 = 0.84$

5.) All the research values (r^2_{score} of the models) should be documented.

All updated list above

Model	r^2_{score}
Multiple Linear Regression	0.78
SVM	0.81
Decision Tree	0.77
Random Forest	0.84

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

Random Forest has best model r^2_{score} value = 0.84