

# Hope Artificial Intelligence



## Assignment-Regression Algorithm

Download Set: •

Google Drive Link [Click here](#) .

Git Hub Link: [https://raw.githubusercontent.com/RamishaRaniK/dataset/main/insurance\\_pre.csv](https://raw.githubusercontent.com/RamishaRaniK/dataset/main/insurance_pre.csv)

### 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.

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

#### 1.) Identify your problem statement

- Domain Selection - Machine Learning
- Learning Selection- Supervised Learning
- Classification/Regression - Regression

#### 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 – So I used Label Encoding to convert categorical into numeric form for 2 columns.

#### 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.

5.) All the research values (r2\_score of the models) should be documented.

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

1. MACHINE LEARNING = **R<sup>2</sup> value**= 0.7894

2. SUPPORT VECTOR MACHINE

R value					
S.No	Hyper parameter	Linear	RBF	Poly	Sigmoid
1	C=10	0.4624	-0.0322	0.0387	0.0393
2	C=100	0.6288	0.32003	0.6179	0.5276
3	C=500	0.7631	0.6642	0.8263	0.4446
4	C=1000	0.7649	0.8102	0.8566	0.2874
5	C=5000	0.7414	0.8747	0.8595	-7.53

Support Vector Machine Regression = **R<sup>2</sup> Value** [C=5000, kernel=RBF]=0.8747

3. DECISION TREE

S.No	Criterion	Splitter	R value
1	Squared_error	Best	0.6972
2	Squared_error	Random	0.6443
3	Friedman_mse	Best	0.688
4	Friedman_mse	Random	0.6698
5	Absolute_error	Best	0.6822
6	Absolute_error	Random	0.7025
7	Poisson	Best	0.718
8	Poisson	Random	0.7111

Decision Tree Regression = **R<sup>2</sup> Value** [criterion=Poisson, Splitter=Best] = 0.7180

#### 4. RANDOM FOREST

S.No	Criterion	N_Estimators	R_value
1	Absolute_error	50	0.8523
2	Absolute_error	100	0.8513
3	Squared_error	50	0.8497
4	Squared_error	100	0.8537
5	Friedman_mse	50	0.8498
6	Friedman_mse	100	0.8539
7	Poisson	50	0.849
8	Poisson	100	0.8526

Random Forest Regression = **R<sup>2</sup> Value** [criterion=Friedman\_mse, N\_estimators=100]=0.8539

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

- Algorithm = Support Vector Machine Regression  
Parameters = [C=5000, kernel=RBF]  
**R<sup>2</sup> Value** = 0.8747
- Algorithm = Random Forest Regression  
Parameters = [criterion=Friedman\_mse, N\_estimators=100]  
**R<sup>2</sup> Value** =0.8539

This is the highest score. I tested all the algorithms, but none of them had a good model, including this one. However, compared to the others, this achieved the highest score.