

# Hope Artificial Intelligence

## Assignment-Regression Algorithm

### 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
- 2.) Tell basic info about the dataset (Total number of rows, columns)
- 3.) Mention the pre-processing method if you're doing any (like converting string to number –nominal 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.
- 5.) All the research values ( $r^2$  score of the models) should be documented. (You can make tabulation or screenshot of the results.)
- 6.) Mention your final model, justify why you have chosen the same.

.

1. Need to predict the insurance charges based on given dataset.

Problem Identification stages,

- Stage 1 - Machine Learning
- Stage 2 – Supervised
- Stage 3 – Regression

2. The dataset contains 1338 rows and 6 columns

3. The dataset contains string data so converted into number – nominal (one hot encoding)

## Models with r2\_score

Algorithm	Model	Hyper tuning	R score
Multiple Linear Regression	Linear Regression	-	0.7894790349867009
SVM	kernel='rbf'	C=10	-0.03227329390671052
SVM	kernel='rbf'	C=100	0.3200317832050831
SVM	kernel='linear'	C=10	0.4624684142339678
SVM	kernel='linear'	C=100	0.6288792857320358
SVM	kernel='poly'	C=10	0.038716222760231456
SVM	kernel='poly'	C=100	0.6179569624059797
SVM	kernel='sigmoid'	C=10	0.03930714378274347
SVM	kernel='sigmoid'	C=100	0.5276103546510411
Decision Tree	Criterion=mse	Splitter=best	0.6884080648214863
Decision Tree	Criterion=mse	Splitter=random	0.71415026473638
Decision Tree	Criterion=mae	Splitter=best	0.6675886309869457
Decision Tree	Criterion=mae	Splitter=random	0.7004229734128404
Decision Tree	Criterion=friedman_mse	Splitter=best	0.7152972147900241
Decision Tree	Criterion=friedman_mse	Splitter=random	0.6907398130703057
Random Forest	Criterion=mse	estimators=100	0.8537966429381433
Random Forest	Criterion=mse	estimators=10	0.8467876446012808
Random Forest	Criterion=mae	estimators=100	<b>0.8551921332671869</b>
Random Forest	Criterion=mae	estimators=10	0.8473477881812199
Random Forest	Criterion=friedman_mse	estimators=100	0.8514659685951451
Random Forest	Criterion=friedman_mse	estimators=10	0.8252814250782897

I have created many models using machine learning algorithm and tested with r2\_score. Finally, I have selected "Random Forest" model which gives high score by comparing other models.

So, I saved "Random Forest" model with parameter(criterion=mae) and Hyper tuning(estimators=100) in a filename = "finalized\_model\_random\_forest.sav"

Also, deployment for end user.