

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

To predict the insurance charges based on age ,bmi ,children ,sex and smoker with the given dataset

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

There are 1338 rows and 6 columns. The 6 columns are age , bmi , children , sex and smoker.

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

Using dtype I converted the objects to int.The dataset now has 8 columns with age , bmi , children , charges , sex_female , sex_male , smoker_no , smoker_yes .

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.

- In SVM , the kernel 'poly' has better result.
- In Decision tree, the criterion 'squared_error' and splitter 'random' has better result.
- In random forest,the criterion 'squared_error' has better result.

5.) All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

SVM

| Kernel | r_score |
|---------------|---------------------|
| <i>linear</i> | 0.6357732776408936 |
| rbf | 0.38611924616280535 |
| poly | 0.7482218876571232 |

| | |
|-------------|--------------------|
| sigmoid | 0.5299539606103874 |
| precomputed | - |

Decision Tree

| Criterion | Splitter | r_score |
|-----------------------|----------|--------------------|
| <i>squared_error</i> | best | 0.7027781432656248 |
| <i>squared_error</i> | random | 0.7409706061888768 |
| <i>friedman_mse</i> | best | 0.7195404091493405 |
| <i>friedman_mse</i> | random | 0.7144790260350603 |
| <i>absolute_error</i> | best | 0.6873812016237725 |
| <i>absolute_error</i> | random | 0.7289429693573095 |
| <i>poisson</i> | best | 0.7135606977025122 |
| <i>poisson</i> | random | 0.7200810611845658 |

Random forest

| Criterion | r_score |
|-----------------------|--------------------|
| <i>squared_error</i> | 0.8610425386278729 |
| <i>absolute_error</i> | 0.8535991721828784 |
| <i>friedman_mse</i> | 0.8605677607190321 |

| | |
|----------------|--------------------|
| <i>poisson</i> | 0.8604602294238679 |
|----------------|--------------------|

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

The final model is created using 'Random Forest' with the criterion 'squared_score'. It has the highest r_score of 0.8610425386278729