**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**

Answer: STAGE-1 --🡪 Machine Learning

STAGE-2 --🡪 Supervised Learning

STAGE-3 --🡪 Regression

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

Answer: Total number of rows: 1338

Total number of columns: 6

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

Answer: From the given dataset total columns is 6 in that sex and smoker both are nominal data, so this converted into numerical data

**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.**

Answer: The final model is from Random forest algorithm with the combination parameter Random forest regressor is (criterion=absolute\_error, max\_features=sqrt, n\_estimators=100) . The R2 value is 0.8736

**5.) All the research values (r2\_score of the models) should be documented. (You can make tabulation or screenshot of the results.)**

**\*Multiple linear regressor**: R2 value is 0.7894 for give dataset

**\*Support Vector Machine :**

The R2 value for SVR is : 0.8663 with hyper parameter combination is (kernel=rbf , C=3000) this a best model from support vector machine for the given dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| S.NO | Hyper Parameter | Linear  (R Value) | RBF  (R Value) | Poly  (R Value) | Sigmoid (R Value) |
| 1 | C 10 | 0.4624 | -0.0322 | 0.0387 | 0.0393 |
| 2 | C 100 | 0.6288 | 0.3200 | 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 2000 | 0.7440 | 0.8547 | 0.8605 | -0.5939 |
| 6 | C 3000 | 0.7414 | 0.8663 | 0.8598 | -2.1244 |

**\*Decision Tree:**

Decision tree r value is 0.69

The R2 value for Decision Tree Regressor is : 0.7589 with combination is (criterion=absolute\_error, max\_features=log2, splitter=best) this a best model from Decision Tree Regressor for the given dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Criterion | Max Features | Splitter | R Value |
| 1 | Squared Error | Auto | Best | 0.7113 |
| 2 | Squared Error | Auto | Random | 0.7179 |
| 3 | Squared Error | SQRT | Best | 0.7212 |
| 4 | Squared Error | SQRT | Random | 0.6163 |
| 5 | Squared Error | Log2 | Best | 0.7398 |
| 6 | Squared Error | Log2 | Random | 0.6712 |
| 7 | friedman\_mse | Auto | Best | 0.7012 |
| 8 | friedman\_mse | Auto | Random | 0.7293 |
| 9 | friedman\_mse | SQRT | Best | 0.7523 |
| 10 | friedman\_mse | SQRT | Random | 0.6483 |
| 11 | friedman\_mse | Log2 | Best | 0.7416 |
| 12 | friedman\_mse | Log2 | Random | 0.6923 |
| 13 | absolute\_error | Auto | Best | 0.6821 |
| 14 | absolute\_error | Auto | Random | 0.7093 |
| 15 | absolute\_error | SQRT | Best | 0.6528 |
| 16 | absolute\_error | SQRT | Random | 0.6547 |
| 17 | absolute\_error | Log2 | Best | 0.7589 |
| 18 | absolute\_error | Log2 | Random | 0.7228 |
| 19 | Poisson | Auto | Best | 0.6774 |
| 20 | Poisson | Auto | Random | 0.7262 |
| 21 | Poisson | SQRT | Best | 0.6291 |
| 22 | Poisson | SQRT | Random | 0.6400 |
| 23 | Poisson | Log2 | Best | 0.7150 |
| 24 | Poisson | Log2 | Random | 0.6888 |

**\*Random Forest:**

The R2 value for Random Forest Regressor is: 0.8736 with combination is (criterion=absolute\_error, max\_features=sqrt, n\_estimators=100) this a best model from Random Forest Regressor for the given dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No | Criterion | Max Features | N\_ Estimators | R Value |
| 1 | squared\_error | Auto | 10 | 0.8430 |
| 2 | squared\_error | Auto | 100 | 0.8523 |
| 3 | squared\_error | Sqrt | 10 | 0.8413 |
| 4 | squared\_error | Sqrt | 100 | 0.8720 |
| 5 | squared\_error | Log2 | 10 | 0.8712 |
| 6 | squared\_error | Log2 | 100 | 0.8732 |
| 7 | absolute\_error | Auto | 10 | 0.8403 |
| 8 | absolute\_error | Auto | 100 | 0.8541 |
| 9 | absolute\_error | Sqrt | 10 | 0.8517 |
| 10 | absolute\_error | Sqrt | 100 | 0.8736 |
| 11 | absolute\_error | Log2 | 10 | 0.8418 |
| 12 | absolute\_error | Log2 | 100 | 0.8721 |
| 13 | friedman\_mse | Auto | 10 | 0.8141 |
| 14 | friedman\_mse | Auto | 100 | 0.8542 |
| 15 | friedman\_mse | Sqrt | 10 | 0.8613 |
| 16 | friedman\_mse | Sqrt | 100 | 0.8715 |
| 17 | friedman\_mse | Log2 | 10 | 0.8516 |
| 18 | friedman\_mse | Log2 | 100 | 0.8725 |
| 19 | Poisson | Auto | 10 | 0.8244 |
| 20 | Poisson | Auto | 100 | 0.8355 |
| 21 | Poisson | Sqrt | 10 | 0.8008 |
| 22 | Poisson | Sqrt | 100 | 0.8276 |
| 23 | Poisson | Log2 | 10 | 0.8100 |
| 24 | Poisson | Log2 | 100 | 0.8290 |

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

Answer: The final model is from RANDOM FOREST REGRESSOR ALGORITHM with r2 value of 0.8736 . The reason to choose this model is other the algorithm is not performed with this model