**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. problem statement

Stage 1: Machine Learning

Stage 2: Supervised Learning

Stage 3: Regression

1. basic info about the dataset

6 rows 1339 columns

1. pre-processing method

converts categorical data into dummy variables.

1.Multiple Linear Regression (R value) = 0.7894

2.Support Vector Machine

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SN** | **HYPER PARAMETER** | **LINEAR** | **RBF (NON LIEAR)** | **SIGMOLD** | **POLY** |
| 1 | C100 | 0.6288 | 0.3200 | 0.5276 | 0.6179 |
| 2 | C500 | 0.7631 | 0.6641 | 0.4446 | 0.8263 |
| 3 | C1000 | 0.7649 | 0.8102 | 0.2874 | 0.8566 |
| 4 | C2000 | 0.7440 | 0.8547 | -0.5939 | 0.8605 |
| 5 | C3000 | 0.7414 | 0.8663 | -2.1244 | 0.8598 |
| 6 | C4000 | 0.7414 | 0.8717 | -5.5103 | 0.8600 |

Support Vector Machine Regression use R value RBF((nonlinear) and parameter ((C4000)) = 0.8717

3.Decision Tree

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SN** | **CRITERION** | **SPLITTER** | **MAX FEATURES** | **R VALUE** |
| 1 | friedman\_mse | best | sqrt | 0.7010 |
| 2 | friedman\_mse | random | log2 | 0.6488 |
| 3 | friedman\_mse | best | log2 | 0.7067 |
| 4 | friedman\_mse | random | sqrt | 0.6311 |
| 5 | absolute\_error | best | sqrt | 0.7309 |
| 6 | absolute\_error | random | log2 | 0.7575 |
| 7 | absolute\_error | random | sqrt | 0.6263 |
| 8 | absolute\_error | best | log2 | 0.7343 |
| 9 | poisson | random | sqrt | 0.7435 |
| 10 | poisson | best | log2 | 0.7135 |
| 11 | poisson | random | log2 | 0.6718 |
| 10 | poisson | best | sqrt | 0.7049 |

The Decision Tree Regression use R value (absolute\_error , random , log2 ) = 0.7575

4.Random Forest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SN** | **CRITERION** | **MAX FEATURES** | **N\_ESTIMATORS** | **R VALUE** |
| 1 | squared\_error | sqrt | 50 | 0.8695 |
| 2 | squared\_error | log2 | 100 | 0.8680 |
| 3 | absolute\_error | sqrt | 50 | 0.8708 |
| 4 | absolute\_error | log2 | 100 | 0.8710 |
| 5 | friedman\_mse | sqrt | 50 | 0.8702 |
| 6 | friedman\_mse | log2 | 100 | 0.8710 |
| 7 | poisson | sqrt | 50 | 0.8632 |
| 8 | poisson | log2 | 100 | 0.8680 |

The Random Forest Regression R value (absolute\_error, log2, 100) =0.8710

(or)

The Random Forest Regression R value (absolute\_error, sqrt, 50) =0.8708