**HOPE ARTIFICIAL INTELLIGENCE**

***Assignment-Regression***

1. Identify your problem statement:

3 Stages of Problem Identification

• Machine Learning

• Supervised Learning

• Regression

1. Tell basic info about the dataset (Total no of rows,columns)

Total no of rows- 1338 nos

Total no of columns- 6 nos

1. Mention the pre-processing method if you are doing any (like converting string to number-nominal data)

Converting string sex ( male & female= 1 & 0 )

Converting string smoker ( yes & no = 1 & 0 )

1. 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 the final model.

The final best model is Support Vector Machine Regression

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

Please refer the tabulation below.

1. Mention your final model,justify why you have chosen the same.

The best model is Support Vector Machine Regression use **R2 value (rbf and hyper parameter (C=10000)=0.87799**

**To find out the best Machine Learning Regression method using in r2 value:**

1. **MULTIPLE LINEAR REGRESSION (R2 Value)= 0.78947**
2. **SUPPORT VECTOR MACHINE**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **S.NO** | **HYPER PARAMETER** | **LINEAR**  **(r value)** | **RBF(NON LINEAR)**  **(r value)** | **POLY**  **(r value)** | **SIGMOID**  **(r value)** |
| 1 | C=0.01 | -0.08883 | -0.08964 | -0.08956 | -0.08956 |
| 2 | C=100 | 0.62887 | 0.32003 | 0.61795 | 0.52761 |
| 3 | C=500 | 0.76310 | 0.66429 | 0.82636 | 0.44460 |
| 4 | C=1000 | 0.76493 | 0.81020 | 0.85664 | 0.28747 |
| 5 | C=10000 | 0.74142 | 0.87799 | 0.85917 | -34.15153 |

**The SVM Regression use R2 value (rbf and hyper parameter (C=10000)=0.87799**

1. **DECISION TREE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **CRITERION** | **SPLITTER** | **MAX FEATURES** | **R VALUE** |
| 1 | squared\_error | best | sqrt | 0.69481 |
| 2 | squared\_error | best | log2 | 0.74565 |
| 3 | squared\_error | best | None | 0.69547 |
| 4 | squared\_error | random | sqrt | 0.67167 |
| 5 | squared\_error | random | log2 | 0.69668 |
| 6 | squared\_error | random | None | 0.70182 |
| 7 | friedman\_mse | best | sqrt | 0.67869 |
| 8 | friedman\_mse | best | log2 | 0.71673 |
| 9 | friedman\_mse | best | None | 0.69391 |
| 10 | friedman\_mse | random | sqrt | 0.66853 |
| 11 | friedman\_mse | random | log2 | 0.63213 |
| 12 | friedman\_mse | random | None | 0.76305 |
| 13 | absolute\_error | best | sqrt | 0.69277 |
| 14 | absolute\_error | best | log2 | 0.69878 |
| 15 | absolute\_error | best | None | 0.68608 |
| 16 | absolute\_error | random | sqrt | 0.75212 |
| 17 | absolute\_error | random | log2 | 0.73840 |
| 18 | absolute\_error | random | None | 0.76057 |
| 19 | poisson | best | sqrt | 0.65280 |
| 20 | poisson | best | log2 | 0.70589 |
| 21 | poisson | best | None | 0.73321 |
| 22 | poisson | random | sqrt | 0.59301 |
| 23 | poisson | random | log2 | 0.70208 |
| 24 | poisson | random | None | 0.69662 |

**The Decision Tree Regression use R2 value(friedman\_mse,random,None)=0.76305**

1. **RANDOM FOREST**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.NO** | **N\_ESTIMATORS** | **CRITERION** | **MAX FEATURES** | **R VALUE** |
| 1 | 50 | squared\_error | sqrt | 0.86476 |
| 2 | 50 | squared\_error | log2 | 0.87149 |
| 3 | 50 | squared\_error | None | 0.85408 |
| 4 | 100 | squared\_error | sqrt | 0.87000 |
| 5 | 100 | squared\_error | log2 | 0.86806 |
| 6 | 100 | squared\_error | None | 0.85219 |
| 7 | 50 | absolute\_error | sqrt | 0.87201 |
| 8 | 50 | absolute\_error | log2 | 0.87082 |
| 9 | 50 | absolute\_error | None | 0.85357 |
| 10 | 100 | absolute\_error | sqrt | 0.87549 |
| 11 | 100 | absolute\_error | log2 | 0.87455 |
| 12 | 100 | absolute\_error | None | 0.85750 |
| 13 | 50 | friedman\_mse | sqrt | 0.86679 |
| 14 | 50 | friedman\_mse | log2 | 0.86864 |
| 15 | 50 | friedman\_mse | None | 0.85623 |
| 16 | 100 | friedman\_mse | sqrt | 0.86867 |
| 17 | 100 | friedman\_mse | log2 | 0.87169 |
| 18 | 100 | friedman\_mse | None | 0.85367 |
| 19 | 50 | poisson | sqrt | 0.87174 |
| 20 | 50 | poisson | log2 | 0.86442 |
| 21 | 50 | poisson | None | 0.85203 |
| 22 | 100 | poisson | sqrt | 0.87027 |
| 23 | 100 | poisson | log2 | 0.87301 |
| 24 | 100 | poisson | None | 0.85334 |

**The Random Forest Regression use R2 value(n\_estimators=100,absolute\_error,sqrt)=0.87549**