Assignment – Regression Algorithm

1. a)Machine Learning

b) Supervised Learning

c) Regression

2. Our dataset has 1,338 Rows and 6 Columns.

3. Use One Hot Encoding method to change nominal data.

We change sex column and smoker column.

a) Multiple Linear Regression (R2value)= 0.7894

b) Support Vector Machine:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| s. no | Hyper Parameter | Linear  (r2value) | RBF(non linear)  (r2value) | Poly  (r2value) | Sigmoid  (r2value |
| 1 | C 10 | -0.0016 | -0.0874 | - | -0.0897 |
| 2 | C 100 | 0.5432 | -0.0745 | - | -0.0897 |
| 3 | C 500 | 0.6270 | -0.0495 | - | -0.0897 |
| 4 | C 1000 | 0.6340 | -0.0407 | - | -0.0897 |
| 5 | C 2000 | 0.6893 | 0.0002 | - | -0.0897 |
| 6 | C 3000 | 0.7590 | 0.0436 | - | -0.0897 |

The SVM Regression R2value (Linear) and hyper Parameter (3000) = 0.7590

c) Decision Tree:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s. no | Criterion | Max features | Splitter | R2 value |
| 1 | mse | auto | best | 0.6932 |
| 2 | mse | auto | random | 0.7434 |
| 3 | mse | sqrt | best | 0.7248 |
| 4 | mse | sqrt | random | 0.7395 |
| 5 | mse | Log2 | best | 0.7329 |
| 6 | mse | Log2 | random | 0.6476 |
| 7 | mae | auto | best | 0.6728 |
| 8 | mae | auto | random | 0.7615 |
| 9 | mae | sqrt | best | 0.6988 |
| 10 | mae | sqrt | random | 0.7184 |
| 11 | mae | Log2 | best | 0.6873 |
| 12 | mae | Log2 | random | 0.6750 |
| 13 | Friedman mse | auto | best | 0.6935 |
| 14 | Friedman mse | auto | random | 0.6605 |
| 15 | Friedman mse | sqrt | best | 0.7295 |
| 16 | Friedman mse | sqrt | random | 0.7319 |
| 17 | Friedman mse | Log2 | best | 07386 |
| 18 | Friedman mse | Log2 | random | 0.6608 |

The decision tree Regression R2value (mae, auto, random) = 0.7615

d) Random Forest

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s. no | Criterion | Max features | N\_estimators | R2 Value |
| 1 | mse | auto | 10 | 0.8435 |
| 2 | mse | auto | 100 | 0.8533 |
| 3 | mse | sqrt | 10 | 0.8465 |
| 4 | mse | sqrt | 100 | 0.8688 |
| 5 | mse | Log2 | 10 | 0.8515 |
| 6 | mse | Log2 | 100 | 0.8692 |
| 7 | mae | auto | 10 | 0.8473 |
| 8 | mae | auto | 100 | 0.8555 |
| 9 | mae | sqrt | 10 | 0.8515 |
| 10 | mae | sqrt | 100 | 0.8718 |
| 11 | mae | Log2 | 10 | 0.8424 |
| 12 | mae | Log2 | 100 | 0.8733 |
| 13 | Friedman mse | auto | 10 | 0.8242 |
| 14 | Friedman mse | auto | 100 | 0.8568 |
| 15 | Friedman mse | sqrt | 10 | 0.8551 |
| 16 | Friedman mse | sqrt | 100 | 0.8702 |
| 17 | Friedman mse | Log2 | 10 | 0.8522 |
| 18 | Friedman mse | Log2 | 100 | 08726 |

The Random Forest Regression R2value (mae, log2, 100) = 0.8733

The final Machine Learning best method of Regression:

Random Forest R2  value (mae, log2,100)

The best model is Random Forsest , because it has highest R2 value. So we choose random forest for deployment phase.