**ASSIGNMENT CLASSIFICATION**

**1.Problem Statement:**

Stage1: Machine Learning (Dataset are numeric)

Stage2: Supervised Learning ( Requirements are clear)

Stage3: Classification( Dataset have categorical data)

**2.Info about dataset:**

Total number of rows:399

Total number of columns:25

**3.Preprocessing method:**

Here in dataset contains nominal data (i.e column names - rbc, pc, pcc, ba, htn, dm, cad, appet, pe, ane) so one hot encoding is proceed by using get\_dummies and converted in to numeric values

**4. Good model :**

* Logistic Regression: f1\_macro value ('penalty': 'l2', 'solver': 'newton-cg') = 0.9924946382275899
* Support Vector Machine: f1\_macro value (c=10, gamma= auto, kernel=sigmoid) = 0.99
* Decision Tree: f1\_macro value (crieterion=entropy, splitter=random, max\_features=log2 ) = 0.99
* Random Forest: f1\_macro value (crieterion=gini, max\_features=sqrt, n\_estimators=10) = 0.98
* K Nearest Neighbors: f1\_macro value (algorithm=auto, n\_neighbors=1, p=1, weights=uniform) = 0.96
* Navie bayes GaussianNB: f1\_macro value (var\_smoothing=1e-08) =0.97
* Navie bayes MultinomialNB: f1\_macro value (alpha=0.1) = 0.81
* Navie bayes BernoulliNB: f1\_macro value (alpha=0.5, binarize= 0.0) =0.97

**5. Research values of all models:**

1. **LOGISTIC REGRESSION:**
   * + f1\_macro value: 0.99
     + The confusion Matrix: [[51 0]

[1 81]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.98 | 1.00 | 0.99 | 51 |
| 1 | 1.00 | 0.99 | 0.99 | 82 |
| Accuracy |  |  | 0.99 | 133 |
| Macro average | 0.99 | 0.99 | 0.99 | 133 |
| Weighted average | 0.99 | 0.99 | 0.99 | 133 |

* + - roc\_auc\_score : 1.0

1. **SUPPORT VECTOR MACHINE:**
   * + f1\_macro value: 0.99
     + The confusion Matrix: [[51 0]

[1 81]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.98 | 1.00 | 0.99 | 51 |
| 1 | 1.00 | 0.99 | 0.99 | 82 |
| Accuracy |  |  | 0.99 | 133 |
| Macro average | 0.99 | 0.99 | 0.99 | 133 |
| Weighted average | 0.99 | 0.99 | 0.99 | 133 |

* + - roc\_auc\_score : 1.0

1. **DECISION TREE:**
   * + f1\_macro value: 0.99
     + The confusion Matrix: [[51 0]

[1 81]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.98 | 1.00 | 0.99 | 51 |
| 1 | 1.00 | 0.99 | 0.99 | 82 |
| Accuracy |  |  | 0.99 | 133 |
| Macro average | 0.99 | 0.99 | 0.99 | 133 |
| Weighted average | 0.99 | 0.99 | 0.99 | 133 |

* + - roc\_auc\_score : 0.99

1. **RANDOM FOREST:**
   * + f1\_macro value: 0.98
     + The confusion Matrix: [[50 1]

[1 81]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.98 | 0.98 | 0.98 | 51 |
| 1 | 0.99 | 0.99 | 0.99 | 82 |
| Accuracy |  |  | 0.98 | 133 |
| Macro average | 0.98 | 0.98 | 0.98 | 133 |
| Weighted average | 0.98 | 0.98 | 0.98 | 133 |

* + - roc\_auc\_score : 0.99

1. **K NEAREST NEIGHBOR:**
   * + f1\_macro value: 0.96
     + The confusion Matrix: [[51 0]

[5 77]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.91 | 1.00 | 0.95 | 51 |
| 1 | 1.00 | 0.94 | 0.97 | 82 |
| Accuracy |  |  | 0.96 | 133 |
| Macro average | 0.96 | 0.97 | 0.96 | 133 |
| Weighted average | 0.97 | 0.96 | 0.96 | 133 |

* + - roc\_auc\_score : 0.95

1. **NAVIE BAYES GAUSSIANNB:**
   * + f1\_macro value: 0.97
     + The confusion Matrix: [[51 0]

[3 79]]

* The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.94 | 1.00 | 0.97 | 51 |
| 1 | 1.00 | 0.96 | 0.98 | 82 |
| Accuracy |  |  | 0.98 | 133 |
| Macro average | 0.97 | 0.98 | 0.98 | 133 |
| Weighted average | 0.98 | 0.98 | 0.98 | 133 |

* + - roc\_auc\_score : 0.5

1. **NAVIE BAYES MULTINOMIALNB:**
   * + f1\_macro value: 0.81
     + The confusion Matrix: [[51 0]

[25 57]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.67 | 1.00 | 0.80 | 51 |
| 1 | 1.00 | 0.70 | 0.82 | 82 |
| Accuracy |  |  | 0.81 | 133 |
| Macro average | 0.84 | 0.85 | 0.81 | 133 |
| Weighted average | 0.87 | 0.81 | 0.81 | 133 |

* + - roc\_auc\_score : 0.93

1. **NAVIE BAYES BERNOULLINB:**
   * + f1\_macro value: 0.97
     + The confusion Matrix: [[51 0]

[3 79]]

* + - The report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1\_macro | Support |
| 0 | 0.94 | 1.00 | 0.97 | 51 |
| 1 | 1.00 | 0.96 | 0.98 | 82 |
| Accuracy |  |  | 0.98 | 133 |
| Macro average | 0.97 | 0.98 | 0.98 | 133 |
| Weighted average | 0.98 | 0.98 | 0.98 | 133 |

* + - roc\_auc\_score : 0.99

**6. Final model:**

The final model I have chosen by using algorithm with crietrion= , max\_features= n\_estimators= and found r2 value is 0.92. Here the model has highest performance metrics and lowest complexity to the model