1. Implementation Language: Python

Lab\_Assignment 2: (Classification) Implementation and Comparison of Decision Tree and Naïve Bayesian Classification Algorithms.

CSE\_4225 Fall 2020

**Prepared by- Pankaj Chandra Kar Roll: 49**

Pankaj

1. Tested Datasets: 8 (*Iris, Adult, Agaricus-lepiota, Ballon, BCW, Diabetes, Letter-recognition, Car*)
2. 70% data are used for training randomly and remaining 30% data are used for testing for each dataset.

**Iris Dataset**

* **Length:** 149
* **Trained Data:** 104
* **Tested Data:** 45

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 0.33 | 0.33 | 0.33 | 45 |  | 0.33 | 0.29 | 0.31 | 45 |  |
| **Micro Avg** | 1.0 | 1.0 | 1.0 | 45 | 100% | 0.87 | 0.87 | 0.87 | 45 | 100% |
| **Weighted Avg** | 1.0 | 1.0 | 1.0 | 45 | 1.0 | 0.87 | 0.93 | 45 |

**lenses Dataset**

* **Length:** 24
* **Trained Data:** 17
* **Tested Data:** 7

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 1.0 | 1.0 | 1.0 | 7 | 100% | 0.24 | 0.33 | 0.28 | 7 | 71% |
| **Weighted Avg** | 1.0 | 1.0 | 1.0 | 7 | 0.51 | 0.71 | 0.60 | 7 |

**Adult Dataset**

* **Length:** 32560
* **Trained Data:** 22792
* **Tested Data:** 9768

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 0.65 | 0.59 | 0.66 | 9768 | 69% | 0.74 | 0.77 | 0.75 | 9768 | 81% |
| **Weighted Avg** | 0.67 | 0.7 | 0.69 | 9768 | 0.82 | 0.81 | 0.81 | 9768 |

**Abalone Dataset**

* **Length:** 4177
* **Trained Data:** 22791
* **Tested Data:** 1253

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Micro Avg** | 0.77 | 0.71 | 0.73 | 1253 | 82% | 0.22 | 0.22 | 0.22 | 1253 | 21% |
| **Weighted Avg** | 0.81 | 0.82 | 0.81 | 1253 | 0.20 | 0.22 | 0.17 | 1253 |

**Ballon Dataset**

* **Length:** 19
* **Trained Data:** 13
* **Tested Data:** 6

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 1.0 | 1.0 | 1.0 | 6 | 100% | 0.33 | 0.50 | 0.40 | 6 | 67% |
| **Weighted Avg** | 1 | 1.0 | 1.0 | 6 | 0.44 | 0.67 | 0.53 | 6 |

**BCW Dataset**

* **Length:** 698
* **Trained Data:** 490
* **Tested Data:** 210

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 0.39 | 0.50 | 0.44 | 210 | 78% | 0.95 | 0.97 | 0.96 | 210 | 97% |
| **Weighted Avg** | 0.60 | 0.78 | 0.68 | 210 | 0.97 | 0.97 | 0.97 | 210 |

**Car Dataset**

* **Length:** 1727
* **Trained Data:** 1208
* **Tested Data:** 519

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 0.96 | 1.0 | 0.98 | 518 | 99% | 0.26 | 0.34 | 0.30 | 519 | 65% |
| **Weighted Avg** | 0.99 | 0.99 | 0.99 | 518 | 0.51 | 0.65 | 0.57 | 519 |

**Japanese credit data set**

* **Length:** 690
* **Trained Data:** 483
* **Tested Data:** 207

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Target Values** | **Decision Tree** | | | | | **Naïve Bayes** | | | | |
| Avg  Precision | Avg  Recall | F1-Score | Support | Accuracy | Precision | Recall | F1-Score | Support | Accuracy |
| **Macro Avg** | 0.94 | 1.0 | 0.97 | 207 | 99% | 0.84 | 0.84 | 0.84 | 207 | 84% |
| **Weighted Avg** | 0.98 | 0.99 | 0.98 | 207 | 0.84 | 0.84 | 0.84 | 207 |