**Project Report**

**Details of Classification Method:**

**kNN:**

kNN for iris data set: **Training Data**

|  |  |
| --- | --- |
| **Confusion Matrix:** | **Overall Statistics** |
| Prediction setosa versicolor virginica  setosa 43 0 0  versicolor 0 39 2  virginica 0 1 35 | Accuracy: 0.975  95% CI: (0.9287, 0.9948)  No Information Rate: 0.3583  P-Value [Acc > NIR]: < 2.2e-16  Kappa: 0.9624 |

kNN for Life Expectancy Data set: **Training Data**

|  |  |
| --- | --- |
| Confusion Matrix | Overall Statistics |
| Prediction Africa Asia Europe North America South America  Africa 32 5 1 1 1  Asia 5 31 8 5 6  Europe 1 11 24 1 3  North America 0 4 2 12 2  South America 0 0 0 0 0 | Accuracy: 0.6387  95% CI: (0.5578, 0.7142)  No Information Rate: 0.329  P-Value [Acc > NIR]: 3.148e-15  Kappa: 0.5156  Mcnemar's Test P-Value : 0.1768 |

**Naïve Bayes:**

Naïve Bayes for iris data set: **Training Data**

|  |  |
| --- | --- |
| **Confusion matrix** | **Overall Statistics** |
| Prediction setosa versicolor virginica  setosa 43 0 0  versicolor 0 37 2  virginica 0 3 35 | Accuracy: 0.9583  95% CI: (0.9054, 0.9863)  No Information Rate: 0.3583  P-Value [Acc > NIR]: < 2.2e-16  Kappa: 0.9374 |

Naïve Bayes for Life Expectancy Data set: **Training Data**

|  |  |
| --- | --- |
| Confusion Matrix | Overall Statistics |
| Prediction Africa Asia Europe North America South America  Africa 32 5 0 1 2  Asia 3 33 3 0 0  Europe 1 13 30 2 0  North America 2 0 2 16 0  South America 0 0 0 0 10 | Accuracy: 0.7806  95% CI: (0.7072, 0.8431)  No Information Rate: 0.329  P-Value [Acc > NIR]: < 2.2e-16  Kappa: 0.7137  Mcnemar's Test P-Value : NA |

**C4.5:**

C4.5 for iris data set: **Training Data**

|  |  |
| --- | --- |
| **Confusion Matrix** | **Overall Statistics** |
| Prediction setosa versicolor virginica  setosa 43 0 0  versicolor 0 39 0  virginica 0 1 37 | Accuracy: 0.9917  95% CI: (0.9544, 0.9998)  No Information Rate: 0.3583  P-Value [Acc > NIR]: < 2.2e-16  Kappa : 0.987 |

C4.5 for life expectancy Data set: **Training Data**

|  |  |
| --- | --- |
| Confusion Matrix | Overall Statistics |
| Prediction Africa Asia Europe North America South America  Africa 35 5 0 1 0  Asia 3 40 9 4 4  Europe 0 3 24 1 3  North America 0 2 1 12 1  South America 0 1 1 1 4 | Accuracy: 0.7419  95% CI: (0.6656, 0.8088)  No Information Rate: 0.329  P-Value [Acc > NIR]: < 2.2e-16  Kappa: 0.6543  Mcnemar's Test P-Value: NA |

**RIPPER:**

RIPPER for iris data set: **Training data**

|  |  |
| --- | --- |
| **Confusion Matrix** | **Overall Statistics** |
| Prediction setosa versicolor virginica  setosa 43 0 0  versicolor 0 39 1  virginica 0 1 36 | Accuracy: 0.9833  95% CI: (0.9411, 0.998)  No Information Rate: 0.3583  P-Value [Acc > NIR]: < 2.2e-16  Kappa : 0.975 |

RIPPER for Life Expectancy Data set: **Training Data**

|  |  |
| --- | --- |
| Confusion Matrix | Overall Statistics |
| Prediction Africa Asia Europe North America South America  Africa 30 4 0 1 0  Asia 8 37 10 7 6  Europe 0 9 24 3 5  North America 0 1 1 8 1  South America 0 0 0 0 0 | Accuracy: 0.6387  95% CI: (0.5578, 0.7142)  No Information Rate: 0.329  P-Value [Acc > NIR]: 3.148e-15  Kappa: 0.5063  Mcnemar's Test P-Value: NA |

**Oblique:**

Oblique for iris data set: **Training data:**

#Text Plot for Oblique Classifier is in R script

|  |  |  |
| --- | --- | --- |
| setosa | versicolor | virginica |
| Min. :0.0000  1st Qu.:0.0000  Median :0.0000  Mean :0.3583  3rd Qu.:1.0000  Max. :1.0000 | Min. :0.0000  1st Qu.:0.0000  Median :0.0000  Mean :0.3333  3rd Qu.:1.0000  Max. :1.0000 | Min. :0.0000  1st Qu.:0.0000  Median :0.0000  Mean :0.3083  3rd Qu.:1.0000  Max. :1.0000 |

**Results on Test Dataset :**

**Iris Dataset**

**kNN:**

Confusion Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| Prediction | Setosa | Versicolor | Virginica |
| Setosa | 7 | 0 | 0 |
| Versicolor | 0 | 9 | 1 |
| Virginica | 0 | 1 | 12 |

Precision, Recall and F measure

|  |  |  |  |
| --- | --- | --- | --- |
|  | Setosa | Versicolor | Virginica |
| Recall/Sensitivity | 1.0000 | 0.9000 | 0.9231 |
| Positive Predictive Value/Precision | 1.0000 | 0.9000 | 0.9231 |
| F Measure | 1.0000 | 0.9000 | 0.9231 |

**Naïve Bayes:**

Confusion Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| Prediction | Setosa | Versicolor | Virginica |
| Setosa | 7 | 0 | 0 |
| Versicolor | 0 | 8 | 1 |
| Virginica | 0 | 2 | 12 |

Precision, Recall and F measure

|  |  |  |  |
| --- | --- | --- | --- |
|  | Setosa | Versicolor | Virginica |
| Recall/Sensitivity | 1.0000 | 0.8000 | 0.9231 |
| Positive Predictive Value/Precision | 1.0000 | 0.8889 | 0.8571 |
| F Measure | 1.0000 | 0.8421 | 0.8888 |

**C4.5:**

Confusion Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| Prediction | Setosa | Versicolor | Virginica |
| Setosa | 7 | 0 | 0 |
| Versicolor | 0 | 8 | 0 |
| Virginica | 0 | 2 | 13 |

Precision, Recall and F measure

|  |  |  |  |
| --- | --- | --- | --- |
|  | Setosa | Versicolor | Virginica |
| Recall/Sensitivity | 1.0000 | 0.8000 | 1.0000 |
| Positive Predictive Value/Precision | 1.0000 | 1.0000 | 0.8667 |
| F Measure | 1.0000 | 0.8889 | 0.9285 |

**RIPPER:**

Confusion Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
| Prediction | Setosa | Versicolor | Virginica |
| Setosa | 7 | 0 | 0 |
| Versicolor | 0 | 8 | 0 |
| Virginica | 0 | 2 | 13 |

Precision, Recall and F measure

|  |  |  |  |
| --- | --- | --- | --- |
|  | Setosa | Versicolor | Virginica |
| Recall/Sensitivity | 1.0000 | 0.8000 | 1.0000 |
| Positive Predictive Value/Precision | 1.0000 | 1.0000 | 0.8667 |
| F Measure | 1.0000 | 0.8889 | 0.9285 |

**Life Expectancy Dataset:**

**kNN:**

Confusion Matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Africa | 13 | 3 | 0 | 0 | 0 |
| Asia | 0 | 4 | 1 | 2 | 0 |
| Europe | 1 | 0 | 7 | 2 | 0 |
| North America | 1 | 3 | 1 | 1 | 0 |
| South America | 0 | 0 | 0 | 0 | 0 |

Precision, Recall and F measure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Recall | 0.8667 | 0.4000 | 0.7778 | 0.0000 | NA |
| Precision | 0.8125 | 0.5714 | 0.7000 | 0.1667 | NA |
| F Measure | 0.8387 | 0.4705 | 0.7368 | 0.1816 | NA |

**Naïve Bayes:**

Confusion Matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Africa | 13 | 2 | 0 | 0 | 0 |
| Asia | 0 | 3 | 0 | 0 | 0 |
| Europe | 0 | 1 | 7 | 3 | 0 |
| North America | 2 | 4 | 2 | 2 | 0 |
| South America | 0 | 0 | 0 | 0 | 0 |

Precision, Recall and F measure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Recall | 0.8667 | 0.3000 | 0.7778 | 0.4000 | NA |
| Precision | 0.8667 | 1.0000 | 0.6364 | 0.2000 | NA |
| F Measure | 0.8667 | 0.4615 | 0.7000 | 0.2667 | NA |

**C4.5:**

Confusion Matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Africa | 13 | 4 | 0 | 0 | 0 |
| Asia | 1 | 1 | 5 | 2 | 0 |
| Europe | 0 | 2 | 3 | 2 | 0 |
| North America | 1 | 3 | 1 | 1 | 0 |
| South America | 0 | 0 | 0 | 0 | 0 |

Precision, Recall and F measure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Recall | 0.8667 | 0.1000 | 0.3333 | 0.2000 | NA |
| Precision | 0.7647 | 0.1111 | 0.4285 | 0.16667 | NA |
| F Measure | 0.8125 | 0.1025 | 0.3749 | 0.1819 | NA |

**RIPPER:**

Confusion Matrix:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Africa | 13 | 1 | 0 | 0 | 0 |
| Asia | 1 | 4 | 1 | 2 | 0 |
| Europe | 0 | 1 | 7 | 3 | 0 |
| North America | 1 | 4 | 1 | 0 | 0 |
| South America | 0 | 0 | 0 | 0 | 0 |

Precision, Recall and F measure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Africa | Asia | Europe | North America | South America |
| Recall | 0.8667 | 0.4000 | 0.7778 | 0.0000 | NA |
| Precision | 0.9286 | 0.5000 | 0.6364 | 0.0000 | NA |
| F Measure | 0.8965 | 0.4444 | 0.7000 | 0.0000 | NA |

**Final Table:**

**Accuracy**

|  |  |  |
| --- | --- | --- |
|  | Iris Dataset-Test Accuracy | Life Expectancy Dataset-Test Accuracy |
| kNN | 0.933 | 0.641 |
| Naïve Bayes | 0.9 | 0.641 |
| C4.5 | 0.933 | 0.4615 |
| RIPPER | 0.933 | 0.6154 |

**Conclusion**

From the final table. It can be concluded that **kNN works best** for both the above given data set. For Iris dataset kNN, C4.5, RIPPER all give very high accuracy than naïve Bayes. But for life expectancy dataset, kNN and Naïve Bayes give better accuracy followed by RIPPER. So it can be concluded that, for small data sets c4.5 and ripper have shown good accuracy, but for large data sets naïve Bayes and kNN are better.