***Accuracy***

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | Random Forest | Logistic Regression | XGBoost |
| 1 | 0.9523 | 0.9154 | 0.9458 |
| 2 | 0.9587 | 0.9217 | 0.9587 |
| 3 | 0.9630 | 0.9283 | 0.9587 |
| 4 | 0.9609 | 0.9326 | 0.9609 |
| 5 | 0.9478 | 0.9087 | 0.9543 |
| 6 | 0.9630 | 0.9370 | 0.9565 |
| 7 | 0.9543 | 0.9348 | 0.9500 |
| 8 | 0.9522 | 0.9261 | 0.9587 |
| 9 | 0.9500 | 0.9391 | 0.9587 |
| 10 | 0. 9500 | 0.9109 | 0.9457 |
| avg | 0. 9552 | 0.9255 | 0.9548 |
| stdev | 0.0054 | 0.0104 | 0.0054 |

***F-measure***

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | Random Forest | Logistic Regression | XGBoost |
| 1 | 0. 9478 | 0. 9152 | 0. 9458 |
| 2 | 0. 9630 | 0. 9214 | 0. 9586 |
| 3 | 0. 9586 | 0. 9280 | 0. 9588 |
| 4 | 0. 9674 | 0. 9325 | 0. 9608 |
| 5 | 0. 9435 | 0. 9088 | 0. 9545 |
| 6 | 0. 9630 | 0. 9369 | 0. 9565 |
| 7 | 0. 9520 | 0. 9345 | 0. 9500 |
| 8 | 0. 9563 | 0. 9256 | 0. 9588 |
| 9 | 0. 9475 | 0. 9385 | 0. 9586 |
| 10 | 0. 9497 | 0. 9104 | 0. 9454 |
| avg | 0. 9549 | 0. 9252 | 0. 9548 |
| stdev | 0. 0076 | 0. 0103 | 0. 0054 |

Like Table 12.4

***F-measure rank***

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | Random Forest | Logistic Regression | XGBoost |
| 1 | 0. 9478(1) | 0. 9152(3) | 0. 9458(2) |
| 2 | 0. 9630(1) | 0. 9214(3) | 0. 9586(2) |
| 3 | 0. 9586(2) | 0. 9280(3) | 0. 9588(1) |
| 4 | 0. 9674(1) | 0. 9325(3) | 0. 9608(2) |
| 5 | 0. 9435(2) | 0. 9088(3) | 0. 9545(1) |
| 6 | 0. 9630(1) | 0. 9369(3) | 0. 9565(2) |
| 7 | 0. 9520(1) | 0. 9345(3) | 0. 9500(2) |
| 8 | 0. 9563(2) | 0. 9256(3) | 0. 9588(1) |
| 9 | 0. 9475(2) | 0. 9385(3) | 0. 9586(1) |
| 10 | 0. 9497(1) | 0. 9104(3) | 0. 9454(2) |
| Avg rank | 1.4 | 3 | 1.6 |

Like table 12.8

R = 2

nSum (Rj−R)^2 = 15.2

1/n(k-1) Sum (Rij−R)^2 = 1

The critical value for k=3 and n=10 at the α=0.05 level is 6.2

Therefore 15.2 > 6.2 then we can reject the null hypothesis which states that there is at least one model that is has a significant difference from the others so now we preform the Nemenyi test

Critical difference = 1.0478

Compute the absolute differences between average ranks for each pair and compare that with the critical difference

* Random Forest vs Logistic Regression: ∣1.4−3.0∣=1.6
* Random Forest vs XGBoost: ∣1.4−1.6∣=0.2
* Logistic Regression vs XGBoost: ∣3.0−1.6∣=1.4

Logistic Regression performs significantly differently from both Random Forest and XGBoost, while Random Forest and XGBoost have comparable performance.

## Introduction

This report compares three supervised classification algorithms—Random Forest, XGBoost, and Logistic Regression—based on (1) training time, (2) accuracy, and (3) F-measure. These algorithms were selected for their strong baseline performance on UCI datasets. Stratified ten-fold cross-validation was used for evaluation, with the Friedman test applied to F-measure results, yielding a statistic of 15. Since this exceeded the critical value (6.2 at k=3, n=10 and α=0.05), the null hypothesis was rejected, indicating significant differences in performance.

The Nemenyi test was then conducted to pinpoint differences. Using a critical difference of 1.047, we found that Logistic Regression performed significantly worse than Random Forest and XGBoost.

## Experimental setup

**Dataset**: <https://archive.ics.uci.edu/dataset/94/spambase>

Algorithms:

1. Random Forest: A powerful ensemble learning algorithm that builds multiple decision trees during training and combines their outputs (via majority voting or averaging) to improve accuracy and reduce overfitting.
2. XGBoost: An optimized gradient boosting algorithm that builds sequential decision trees, where each tree corrects errors from the previous one. It is known for its speed and high performance on structured data.
3. Logistic Regression: A simple and widely used linear model for binary classification that predicts probabilities using a logistic function and makes decisions based on a threshold. It works well when the data is linearly separable.

Procedure:

* Stratified 10-fold cross-validation.
* Evaluation metrics: training time, accuracy, F-measure.
* Tools: Mention Python and relevant libraries used.

## **Results and Analysis**

F-measure like table 12.4 from stratified ten-fold cross-validation tests.

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | Random Forest | Logistic Regression | XGBoost |
| 1 | 0. 9478 | 0. 9152 | 0. 9458 |
| 2 | 0. 9630 | 0. 9214 | 0. 9586 |
| 3 | 0. 9586 | 0. 9280 | 0. 9588 |
| 4 | 0. 9674 | 0. 9325 | 0. 9608 |
| 5 | 0. 9435 | 0. 9088 | 0. 9545 |
| 6 | 0. 9630 | 0. 9369 | 0. 9565 |
| 7 | 0. 9520 | 0. 9345 | 0. 9500 |
| 8 | 0. 9563 | 0. 9256 | 0. 9588 |
| 9 | 0. 9475 | 0. 9385 | 0. 9586 |
| 10 | 0. 9497 | 0. 9104 | 0. 9454 |
| avg | 0. 9549 | 0. 9252 | 0. 9548 |
| stdev | 0. 0076 | 0. 0103 | 0. 0054 |

F-measure like table 12.8

|  |  |  |  |
| --- | --- | --- | --- |
| Fold | Random Forest | Logistic Regression | XGBoost |
| 1 | 0. 9478(1) | 0. 9152(3) | 0. 9458(2) |
| 2 | 0. 9630(1) | 0. 9214(3) | 0. 9586(2) |
| 3 | 0. 9586(2) | 0. 9280(3) | 0. 9588(1) |
| 4 | 0. 9674(1) | 0. 9325(3) | 0. 9608(2) |
| 5 | 0. 9435(2) | 0. 9088(3) | 0. 9545(1) |
| 6 | 0. 9630(1) | 0. 9369(3) | 0. 9565(2) |
| 7 | 0. 9520(1) | 0. 9345(3) | 0. 9500(2) |
| 8 | 0. 9563(2) | 0. 9256(3) | 0. 9588(1) |
| 9 | 0. 9475(2) | 0. 9385(3) | 0. 9586(1) |
| 10 | 0. 9497(1) | 0. 9104(3) | 0. 9454(2) |
| Avg rank | 1.4 | 3 | 1.6 |

R = 2

nSum (Rj−R)^2 = 15

1/n(k-1) Sum (Rij−R)^2 = 1

Friedman statistic = 15

critical value at k=3, n=10 and α=0.05 = 6.2

Therefore 15.2 > 6.2 then we can reject the null hypothesis which states that there is at least one model that is has a significant difference from the others so now we preform the Nemenyi test.

To we preform the Nemenyi test we need to calculate the critical difference `formula` qα = 2.343.

We obtained the CD of 1.047

Compute the absolute differences between average ranks for each pair and compare that with the critical difference

* Random Forest vs Logistic Regression: ∣1.4−3.0∣=1.6
* Random Forest vs XGBoost: ∣1.4−1.6∣=0.2
* Logistic Regression vs XGBoost: ∣3.0−1.6∣=1.4

Logistic Regression performs significantly differently from both Random Forest and XGBoost, while Random Forest and XGBoost have comparable performance.

A screenshot of a computer

Description automatically generated

## Discussion

## Conclusion