Comparing Classifiers

SFWR TECH 4DA3

|  |  |  |
| --- | --- | --- |
| Group Member | Student ID | Student # |
| Arifur Rahman | Rahmaa48 | 400300356 |
| Satinder Singh | Sings227 | 400484436 |
| Tushar Joshi | Joshit8 | 400447818 |

In this project, we explored the banknote authentication dataset using two distinct machine learning models: Linear Discriminant Analysis and Linear Support Vector Machine. The primary objectives were to determine the authenticity of banknotes based on selected features and compare the results of the selected classifiers.

**Dataset Details:**

The Banknote Authentication dataset is publicly available and contains 1,372 samples, each with four numerical features derived from image data. These features are variance, skewness, kurtosis, and entropy. Each sample is labeled as either "0" (counterfeit) or "1" (genuine), indicating the authenticity of the banknote.

Lohweg,Volker. (2013). banknote authentication. UCI Machine Learning Repository. <https://doi.org/10.24432/C55P57>.

**Data Organization:**

We followed a common practice of splitting the dataset into training and test sets using Sci-kit train\_test\_split() function, with a standard 75-25 ratio. This resulted in approximately 1,029 samples for training and 343 samples for testing. The training set was used to train the models, while the test set was reserved for evaluating their performance on unseen data.

**Model training and Evaluation:**

Both the LDA and SVM models were trained using the training dataset. We used the scikit-learn library in Python for implementing these models. After training, the models were evaluated on the test dataset using confusion matrix, accuracy, precision, recall, and F1-score as performance metrics.

**Results and Analysis:**

|  |  |  |
| --- | --- | --- |
| **Model** | **Training Time (sec)** | **Testing Time (sec)** |
| LDA | 0.00164 | 0.00022 |
| SVM | 0.01252 | 0.00111 |

A screenshot of a computer

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **TP** | **FP (Type I Error)** | **FN (Type II Error)** | **TN** |
| LDA | 184 | 0 | 7 | 152 |
| SVM | 189 | 2 | 2 | 150 |

* The LDA model achieved an accuracy of approximately 98 % on the test set. The linear SVM model, on the other hand, achieved an accuracy of around 99%.
* SVM model took longer than LDA model for prediction and training.
* Based on the provided confusion matrix data, both models exhibit strong performance with relatively low false positives and false negatives. The Linear Discriminant Analysis (LDA) model achieved 184 true positives and only 7 false negatives, indicating good performance in classifying genuine banknotes. The Linear Support Vector Machine (SVM) model also performed well, achieving 189 true positives and 2 false negatives.

**Conclusion:**

In this analysis, we utilized the Banknote Authentication dataset to build and evaluate Linear Discriminant Analysis and Linear Support Vector Machine models. Both models demonstrated strong performance in distinguishing between genuine and counterfeit banknotes. The SVM model exhibited slightly higher accuracy as compared to LDA model. Further optimization and exploration of more complex models could potentially enhance the overall classification accuracy.