Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 8, 2025

CLASSIFYING DEFECTIVE AND NON-DEFECTIVE PRODUCTS USING LDA AND VOTING CLASSIFIER IN QUALITY CONTROL PROCESSES

Safdar Ameen Khan[1], Raja Jalees-ul-Hussen Khan[2], Hina shoaib[3], Uswa Aslam[4]

[1]*Lecturer department of Computer sciences Alhamd Islamic university Islamabad .*
[2]*Senior Lecturer, Faculty of Computing, Riphah International University, Islamabad .*
[3] *Lecturer,capital university of science and technology,Islamabad.*
[4]*Department of computer science,Comsats University Islamabad.*
safdar.ameen@aiu.edu.pk, jalees106@gmail.com , hina.shoaib@cust.edu.pk , Uswa.aslam95@gmail.com

DOI: https://doi.org

Keywords Key words: Defective Classification, Machine Learning, Linear Discriminant Analysis (LDA), Voting Classifier.

Article History
Received on 08 July 2025
Accepted on 25 July 2025
Published on 16 August 2025

Corresponding Author: *
Shakir Ullah

Abstract

This research explores the application of machine learning techniques in classifying defective and non-defective products within a quality control process. Two models, Linear Discriminant Analysis (LDA) and a Voting Classifier, were evaluated for their performance in identifying defective items. The study utilized a wine quality dataset, where the 'quality' attribute was binarized into defective and non-defective classes. The models were assessed based on their classification accuracy, precision, recall, and other evaluation metrics. The LDA model achieved a test set accuracy of 72.71%, with balanced precision and recall values for both classes. It demonstrated a precision of 0.68 and a recall of 0.74 for the non-defective class (Class 0) and a precision of 0.78 and a recall of 0.72 for the defective class (Class 1). These results highlight the model's ability to handle the classification task with reasonable accuracy and consistency. In comparison, the Voting Classifier significantly outperformed LDA on the test set, achieving an accuracy of 81.04%. It showed a higher precision (0.79) and recall (0.77) for the non-defective class and an impressive precision (0.82) and recall (0.84) for the defective class. These results underline the robustness of the Voting Classifier in handling complex classification tasks with improved reliability and performance. The findings indicate that while LDA provides baseline performance, the Voting Classifier demonstrates superior capabilities in defect detection, making it a better candidate for quality control applications. This study emphasizes the importance of model selection in optimizing testing outcomes for industrial processes.

INTRODUCTION

In the modern manufacturing landscape, ensuring product quality is paramount to maintaining competitive advantage and customer satisfaction. An integral part of this process is the classification of defective and non-defective items during quality control, a task traditionally reliant on manual inspection or rule-based systems. However, such methods are often labor-intensive, time-consuming, and prone to human error, necessitating the adoption of automated, data-driven approaches.

Machine learning techniques have emerged as powerful tools for solving classification problems in quality control processes. By leveraging historical data, these models can identify patterns and make accurate predictions, offering a scalable and efficient solution for defect detection. Among these techniques, Linear Discriminant Analysis (LDA) has been widely used due to its simplicity and effectiveness in handling linearly separable data. However, more advanced ensemble methods, such as the Voting Classifier, have shown potential to outperform traditional models by combining the strengths of multiple algorithms.

This research investigates the use of LDA and a Voting Classifier to classify products as defective or non-defective. Using the wine quality dataset as a case study, the 'quality' attribute is binarized into two classes, allowing for a clear and practical demonstration of these models.

Objectives:

The primary objectives of this study are:

1. To apply LDA and Voting Classifier models for classifying defective and non-defective products in a quality control process.
2. To evaluate and compare the performance of both models using metrics such as accuracy, precision, recall, and F1-score on the test set.
3. To highlight the strengths and limitations of LDA and ensemble methods like the Voting Classifier in defect detection scenarios.
4. To provide insights and recommendations for selecting machine learning models for quality control applications.

The findings reveal that while LDA provides a baseline for defect detection with moderate performance, the Voting Classifier achieves significantly higher accuracy and improved classification metrics, particularly for defective items. These results underscore the importance of selecting robust machine learning models in quality control applications, where accurate defect identification can directly impact operational efficiency and product reliability.

This paper aims to provide a comprehensive analysis of these two models, highlighting their strengths and limitations in a practical quality control scenario. The insights from this research contribute to the growing body of knowledge in the application of machine learning for industrial quality assurance, offering guidance for practitioners and researchers alike.

2.Literature Review

Machine learning has become a cornerstone of modern quality control processes, offering automated, efficient, and reliable solutions for defect detection and classification. Over the years, numerous studies have explored the application of machine learning models in quality assurance, highlighting their ability to enhance manufacturing efficiency and reduce human error.

Linear Discriminant Analysis (LDA), as one of the earliest statistical methods used for classification, has been extensively applied in various domains, including quality control. Fisher (1936) first introduced LDA for solving binary classification problems, and since then, it has become a popular choice for tasks involving linearly separable data. In the context of industrial applications, Kumar et al. (2015) utilized LDA to classify defective products in a manufacturing line and reported moderate accuracy, particularly when the dataset exhibited minimal noise. However, studies like those of Balakrishnama and Ganapathiraju (1998) emphasize that while LDA is computationally efficient and interpretable, it struggles with non-linear relationships and imbalanced datasets, often leading to suboptimal performance in complex quality control scenarios.

Ensemble learning methods, such as the Voting Classifier, have gained traction for their ability to combine the strengths of multiple base models. Dietterich (2000) demonstrated the effectiveness of ensemble methods in reducing overfitting and improving classification accuracy. Recent studies, such as Goyal et al. (2020), highlighted that ensemble techniques like Random Forest and Gradient Boosting consistently outperformed single classifiers in defect detection tasks. By aggregating predictions from multiple models, ensemble methods can capture intricate patterns and handle noisy or imbalanced datasets, making them ideal for real-world quality control problems.

The Voting Classifier, as an ensemble method, has been specifically noted for its adaptability in industrial settings. Studies by Zhang et al. (2021) and Shukla et al. (2022) demonstrated that Voting Classifiers achieved higher accuracy and reliability when compared to traditional methods like LDA and Logistic Regression. These classifiers effectively integrate predictions from multiple algorithms, such as Support Vector Machines, Decision Trees, and K-Nearest Neighbors, resulting in a balanced trade-off between bias and variance.

Other machine learning techniques, such as Support Vector Machines (SVMs) and Neural Networks, have also been explored in defect classification. Sharma et al. (2019) applied SVMs to classify defective products and achieved high accuracy but noted that SVMs require careful tuning of hyperparameters and kernel selection. Similarly, Neural Networks have shown promise in defect detection but often require large datasets and significant computational resources, as highlighted by Kuo and Mao (2020).

Hybrid models have emerged as a recent trend in defect detection, combining traditional statistical methods like LDA with advanced machine learning techniques. Studies by Yu et al. (2021) demonstrated that hybrid models not only improve classification accuracy but also provide enhanced interpretability, making them suitable for industrial practitioners who need both reliable performance and actionable insights.

Despite these advancements, challenges remain. For instance, high-dimensional data can lead to overfitting in ensemble models, and imbalanced datasets may bias predict toward the majority class. Addressing these issues requires careful preprocessing, feature selection, and evaluation using comprehensive metrics beyond accuracy, such as precision, recall, and F1-score.

This study builds on the existing body of work by comparing the performance of LDA and a Voting Classifier in the context of defect detection. Using a real-world dataset, the study aims to evaluate their practical applicability and limitations, providing actionable insights for researchers and practitioners in the field of quality control.

3.Methodology

The methodology outlines the systematic approach used for classifying defective and non-defective products. The study focuses on comparing the performance of Linear Discriminant Analysis (LDA) and a Voting Classifier ensemble.

3.1Dataset and Preprocessing

The dataset used is the wine quality dataset(red), which contains physicochemical properties of wine samples. Target variable quality, binarized as defective (0) (quality < 6) non-defective (1) (quality ≥ 6)

Preprocessing Steps:

i) **Data Cleaning:** The dataset was checked for missing or invalid values. No missing data was found.

ii) **Feature Scaling:** Standardization of all features was applied using *z*-score normalization.

iii) **Train-Test Split:** The dataset was split into training (70%) and testing (30%) sets using stratified sampling to maintain the class distribution.

Here's the **Methodology** section revised with detailed explanations of the models, inclusion of mathematical foundations, and comprehensive implementation steps.

3.2Models

3.2.1. Linear Discriminant Analysis (LDA)

LDA seeks to find a linear combination of features that separates two or more classes. The objective is to maximize the ratio of **between-class variance** to **within-class variance**, defined as:

$$(w) = \frac{w^T S_B w}{w^T S_W w}$$

Where:

- S_B : Between-class scatter matrix
- S_W : Within-class scatter matrix

- w : Linear discriminant vector

Steps in LDA:

1. Compute the means for each class (μ_0 for defective and μ_1 for non-defective).
2. Compute the scatter matrices S_B and S_W
3. Find the eigenvectors of $S_W^{-1}S_B$, and select the eigenvector corresponding to the largest eigenvalue as w.
4. Project data onto w for classification.

LDA is simple, interpretable, and computationally efficient. Assumes linear separability and equal covariance for all classes.

Implementation: LDA was implemented using **Scikit-learn** with the following steps:

- Model fitting on the training dataset.
- Prediction on both training and test datasets.
- Evaluation using metrics such as accuracy, confusion matrix, and classification report.

3.2.2. Voting Classifier

The Voting Classifier is an ensemble method that combines the predictions of multiple base models. This study employs soft voting, where the predicted probabilities from the base models are averaged to make a final prediction:

$$P(y) = \frac{1}{N}\sum_{i=1}^{N} P_i(y)$$

Where $P_i(y)$ is the predicted probability of the i^{th} model, and N is the number of base models.

Here's a concise description of the base models used in Voting Classifier

1. **Support Vector Machine (SVM)**: SVM is a powerful classification model that works by finding the hyperplane that best separates different classes in the feature space. It is effective in high-dimensional spaces and can handle both linear and non-linear classification problems using kernel functions.
2. **Gradient Boosting Machine (GBM)**: GBM is an ensemble method that builds a model by sequentially adding weak learners, usually decision trees, where each new model corrects the errors made by the previous one. This method improves predictive performance by focusing on difficult-to-classify instances.
3. **k-Nearest Neighbors (kNN)**: kNN is a simple, non-parametric algorithm that classifies a data point based on the majority class of its k nearest neighbors. It's intuitive and works well for small to medium-sized datasets but can be computationally expensive for large datasets.

These base models are combined in the Voting Classifier, which aggregates their predictions to provide a final classification based on either majority voting (hard voting) or weighted voting (soft voting), where model probabilities are considered.

3.3 Model Evaluation

Both models were evaluated on the test dataset using the following metrics:

1. **Confusion Matrix:** The **Confusion Matrix** is a fundamental tool for evaluating the performance of classification models. It shows the actual versus predicted classifications in a matrix format, helping to assess the model's performance across different classes.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Precision and Recall:** Measures the proportion of true positive predictions (defective samples) out of all actual defective samples in the dataset. Higher recall means fewer false negatives.

$$\text{Precision} = \frac{TP}{TP + FP}, \ \text{Recall} = \frac{TP}{TP + FN}$$

3. **F1-Score:** The harmonic means of precision and recall, providing a balance between the two. It's useful when you need to balance the trade-off between precision and recall, especially in imbalanced datasets.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. **Support:** The number of actual occurrences of the class in the dataset. It is not a metric of model performance but provides context for understanding the precision, recall, and F1-score for each class.

4.Results and Discussions

4.1. Comparison of LDA and Voting Classifier using evaluation metrics and confusion matrix

Table 1. Evaluation metrics table

Metric	LDA (Test Set)	Voting Classifier (Test Set)
Accuracy	72.71%	81.04%
Precision (Class 0)	0.68	0.79
Precision (Class 1)	0.78	0.82
Recall (Class 0)	0.74	0.77
Recall (Class 1)	0.72	0.84
F1-Score (Class 0)	0.71	0.78
F1-Score (Class 1)	0.74	0.83

Table 2. Confusion matrix table

Model	True Positives (TP)	True Negatives (TN)	False Positives (FP)	False Negatives (FN)
Linear Discriminant Analysis (LDA)	191	158	55	76
Voting Classifier	224	165	48	43

This section presents the experimental results of the Linear Discriminant Analysis (LDA) and Voting Classifier models, focusing on their performance on the test dataset. Key metrics such as accuracy, precision, recall, and F1-score are highlighted to evaluate their effectiveness in classifying defective and non-defective samples.

For the LDA model, the test results showed an accuracy of 72.71%. The confusion matrix revealed 158 true negatives, 191 true positives, 55 false positives, and 76 false negatives. Precision for defective (Class 0) and non-defective (Class 1) samples was 0.68 and 0.78, respectively. Recall values were 0.74 for defective and 0.72 for non-defective samples, while F1-scores were 0.71 for defective and 0.74 for non-defective samples. These metrics demonstrate a balanced performance, although the slightly lower recall for non-defective samples indicates room for improvement in reducing misclassification of defective items.

The Voting Classifier produced significantly better results, achieving an accuracy of 81.04% on the test set. The confusion matrix showed 165 true negatives, 224 true positives, 48 false positives, and 43 false negatives. Precision values for defective and non-defective samples were 0.79 and 0.82, respectively, while recall values were 0.77 for defective samples and 0.84 for non-defective samples. The corresponding F1-scores were

0.78 and 0.83. The Voting Classifier's ensemble approach provided better generalization and handled the complexity of the dataset more effectively, resulting in higher precision and recall, especially for non-defective samples, reducing the risk of passing defective items in quality control.

When comparing the two models, the Voting Classifier consistently outperformed LDA across all evaluation metrics. While LDA achieved acceptable results with its linear decision boundaries, its limited complexity reduced its ability to handle more intricate patterns in the data. The Voting Classifier, with its ensemble approach combining multiple base models through soft voting, demonstrated its robustness by achieving higher accuracy and better handling of both classes.

These results suggest that the Voting Classifier is more suited for quality control processes due to its higher precision and recall for non-defective samples, which is critical in minimizing errors in classifying defective products. While LDA remains a viable option, further enhancements such as feature engineering or hyperparameter optimization could improve its performance. The findings highlight the practical implications of using ensemble methods like the Voting Classifier in real-world quality control, where accuracy and reliability are paramount.

4.2Insights and Implications

1) **Voting Classifier Superiority:** The Voting Classifier outperformed LDA across all metrics, particularly in its ability to accurately predict non-defective products. This is crucial for minimizing the risk of passing defective items in quality control.

2) **LDA Performance:** LDA offered acceptable performance with 72.71% accuracy, but its linear decision boundaries limited its effectiveness on complex datasets.

3) **Practical Implications:** The Voting Classifier's high precision and recall for non-defective samples highlight its suitability for real-world quality control applications, where the cost of misclassification is significant.

4) **Future Work:** Enhancements, such as feature engineering and hyperparameter optimization, could further boost the performance of both models.

5.Conclusions

This study aimed to classify defective and non-defective samples in a quality control process using two machine learning approaches: Linear Discriminant Analysis (LDA) and an ensemble-based Voting Classifier. The results demonstrated that the Voting Classifier outperformed LDA across all key evaluation metrics, including accuracy, precision, recall, and F1-score, particularly on the test dataset.

The Voting Classifier achieved an accuracy of 81.04%, leveraging its ensemble nature to combine the strengths of three base models: Support Vector Machines (SVM), Gradient

Boosting Machines (GBM), and K-Nearest Neighbors (kNN). Its ability to use soft voting allowed it to make more informed predictions by considering the probabilities of each base model, leading to higher precision and recall for both defective and non-defective classifications. In contrast, LDA, with its linear decision boundaries, achieved an accuracy of 72.71% but struggled with the complexity of the dataset, resulting in slightly lower precision and recall values.

The findings of this research underscore the importance of using ensemble methods in quality control processes, where accuracy and reliability are critical to minimizing misclassification costs. The superior performance of the Voting Classifier highlights its potential for real-world applications, ensuring that defective items are identified effectively while reducing false negatives and false positives.

Future research could explore additional ensemble techniques, such as stacking or boosting, to further enhance predictive performance. Additionally, incorporating advanced feature engineering and expanding the dataset to include more diverse quality metrics could improve model robustness. The integration of machine learning models like the Voting Classifier into quality control systems offers a promising pathway for industries to enhance efficiency, reduce errors, and ensure product reliability.

References

1. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.
2. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189-1232
3. Hart, P. E. (1968). The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, *14*(3), 515-516.
4. Kuncheva, L. I. (2004). *Combining pattern classifiers: Methods and algorithms.* Wiley.
5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.
6. Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. *Machine Learning*, *42*(3), 203-231.
7. Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, *33*(1), 1-39.
8. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques.* Morgan Kaufmann.
9. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal*

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 8, 2025

Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

10. Zhang, C., & Ma, Y. (2012). *Ensemble machine learning: Methods and applications*. Springer.