Insights in
Computer Science

https://acadpub.com/ics

Performance Assessment of K-Nearest Neighbors and Naive Bayes Classifiers in Apple Quality Prediction for Data-Driven Agriculture

[1]Lykzelle Mae C. Padasas, [1]Jose C. Agoylo Jr., [1]Efren I. Balaba, [2]Jimson A. Olaybar

[1]BSIT Department, Southern Leyte State University, Tomas Oppus Campus, Tomas Oppus, Southern Leyte 6605, Philippines
[2]FCSIT Department, Southern Leyte State University, Main Campus, Sogod, Southern Leyte 6606, Philippines

About the Article

Open Access

Research Article

How to Cite: L. M. C. Padasas, J. C. Agoylo Jr., E. I. Balaba and J. A. Olaybar, "Performance assessment of K-Nearest neighbors and naive Bayes classifiers in apple quality prediction for data-driven agriculture," Insights Comput. Sci., Vol. 1, pp. 1-9. 2025.

Keywords:
Apple quality prediction, machine learning, KNN, naïve bayes classifier, automated fruit grading system

Corresponding author:
Jose C. Agoylo Jr.,
BSIT Department, Southern Leyte State University, Tomas Oppus Campus, Tomas Oppus, Southern Leyte 6605, Philippines

ABSTRACT

Background and Objective: Accurate classification of fruit quality is essential in modern agriculture to enhance grading systems, improve supply chain efficiency and ensure consumer satisfaction. This study aimed to perform a comparative evaluation of two machine learning algorithms-K-Nearest Neighbors (KNN) and Naive Bayes-for predicting apple quality based on physical and chemical attributes.

Materials and Methods: A dataset comprising 4,000 apple samples was obtained from Kaggle and pre-processed through data cleaning and normalization to ensure a balanced distribution of "Good" and "Bad" quality categories. Both KNN and Naive Bayes models were developed using a supervised learning approach and trained on 70% of the data, while the remaining 30% was used for testing. Model performance was evaluated using standard classification metrics, including accuracy, precision, recall and F1-score.

Results: The comparative analysis revealed that the KNN model outperformed the Naive Bayes classifier. KNN achieved a precision of 90.12% and a weighted F1-score of 0.90, whereas Naive Bayes attained a precision of 88.00% and an F1-score of 0.88. Furthermore, KNN exhibited superior precision and recall across both quality classes, demonstrating its effectiveness in handling correlated features such as sweetness, ripeness and acidity. In contrast, Naive Bayes showed higher misclassification rates, likely due to its assumption of feature independence and reduced performance with overlapping feature distributions.

Conclusion: The findings indicate that KNN is a more reliable and robust algorithm for apple quality prediction compared to Naive Bayes. Its ability to manage correlated variables makes it particularly suitable for agricultural datasets. The study highlights the potential application of KNN in developing automated fruit grading systems, thereby supporting smart farming practices through enhanced efficiency, reduced economic losses and improved data-driven decision-making in agricultural production and supply chains.

INTRODUCTION

Machine learning has emerged as a fundamental tool in modern agriculture, particularly in fruit quality assessment, owing to its capacity to process large datasets and produce accurate predictive models. Several algorithms, including Decision Trees, Random Forest and Support Vector Machines (SVM), have been extensively utilized to classify fruits based on their physical and chemical characteristics[1,2]. These approaches have significantly enhanced fruit grading systems; however, they often require extensive model tuning and substantial computational resources. Accurate prediction of apple quality using machine learning is essential for improving automated fruit grading, optimizing supply chain management and maintaining consistent product standards. Effective fruit quality classification also supports better sorting and pricing strategies, thereby contributing

to consumer satisfaction and reducing post-harvest losses. Conversely, misclassification of fruit quality can result in economic losses and diminished market competitiveness[1].

Recent studies and reviews have emphasized the growing importance of machine learning algorithms in agricultural applications, demonstrating their potential to classify fruit quality based on diverse physical and chemical features such as size, weight, sweetness and ripeness[2]. Among these, algorithms such as Random Forest, Support Vector Machines (SVM) and Decision Trees have been widely implemented and evaluated for their predictive efficiency over the years[3]. The K-Nearest Neighbors (KNN) algorithm, in particular, has gained attention for its simplicity, robustness and effectiveness in various classification tasks. Empirical evidence indicates that KNN performs efficiently in identifying fruit ripeness, grading and defect detection using both visual and numerical parameters[3,4]. For instance, Rangel *et al.*[5] demonstrated the capability of KNN in classifying fruits with distinct visual traits, while Alfatni *et al.*[6] highlighted its potential utility in handling agricultural datasets. Nevertheless, the KNN algorithm's reliance on distance-based metrics makes it sensitive to overlapping features, potentially reducing its accuracy when fruit attributes exhibit close correlations.

Similarly, the Naive Bayes algorithm has been extensively applied in agricultural research owing to its probabilistic foundation and computational efficiency with large datasets. It has demonstrated strong performance in cases where features are largely independent, rendering it suitable for high-dimensional classification problems[7]. Salim and Mohammed[8] employed Naive Bayes for fruit quality classification and reported moderate accuracy, though they observed performance limitations when feature dependencies existed. Likewise, Amra and Maghari[9] identified the algorithm's constraints in accounting for correlated features such as sweetness and ripeness-attributes commonly associated with fruit quality datasets.

The K-Nearest Neighbors (KNN) algorithm offers several notable advantages, including computational simplicity, adaptability and high interpretability, making it an attractive approach for classification tasks. Despite these strengths, its application in fruit quality classification remains relatively underexplored. Recent studies of Sudipa *et al.*[4] and Rangel *et al.*[5] have indicated that KNN can serve as a more efficient alternative to other algorithms in specific contexts; however, it has not been extensively validated on large and heterogeneous datasets related to fruit quality assessment. Previous investigations of Alfatni *et al.*[6] have demonstrated KNN's potential in identifying fruit ripeness and grading based on visual characteristics, yet limited attention has been given to its performance when both physical and chemical features are incorporated into agricultural datasets.

Comparative analyses in the literature suggest that while both KNN and Naive Bayes classifiers are effective, their relative performance varies according to dataset structure and application domain. Beyaz *et al.*[10] reported that KNN outperformed Naive Bayes in scenarios involving balanced class distributions, whereas Naive Bayes demonstrated greater computational efficiency. Similarly, Bhargava and Bansal[11] found that although KNN achieved higher prediction accuracy in fruit quality classification, it required greater processing time than Naive Bayes. These observations imply that algorithm selection should be guided by the specific dataset characteristics, desired accuracy level and computational limitations.

Several reviews have emphasized the continued research interest in benchmarking KNN against algorithms such as Naive Bayes, highlighting that despite KNN's promising results, a comprehensive and systematic evaluation remains limited in the literature[7]. This gap is particularly significant given the increasing need for precise classification techniques to enhance fruit grading and automated sorting systems[8]. Although, both KNN and Naive Bayes have been employed in agricultural contexts, few studies have critically compared their performance using datasets that simultaneously include both physical and chemical fruit attributes. Most prior research has focused on visual parameters or single-attribute analyses, thereby overlooking the complexity of real-world agricultural data[6,9].

To address this gap, the present study conducts a detailed comparative analysis of the KNN and Naive Bayes algorithms for apple quality classification. The evaluation utilizes a dataset comprising over 2,000 apple samples and assesses performance in terms of accuracy, precision, recall and computational efficiency. The findings of this study aim to contribute to the development of intelligent fruit grading systems, enhancing the accuracy of automated sorting, improving consumer satisfaction and promoting greater operational efficiency in agricultural production[9].

MATERIALS AND METHODS

This section describes the methodology employed to evaluate and compare the performance of the Naive Bayes and K-Nearest Neighbors (KNN) algorithms in classifying apples according to their quality. The overall workflow, illustrated in Fig. 1, presents the sequential stages of the study, which include dataset acquisition from Kaggle, data preprocessing, model training using KNN and Naive Bayes algorithms, performance evaluation based on classification metrics and a comparative analysis of the final results. The entire process-from dataset acquisition to result analysis - is illustrated in Fig. 1.

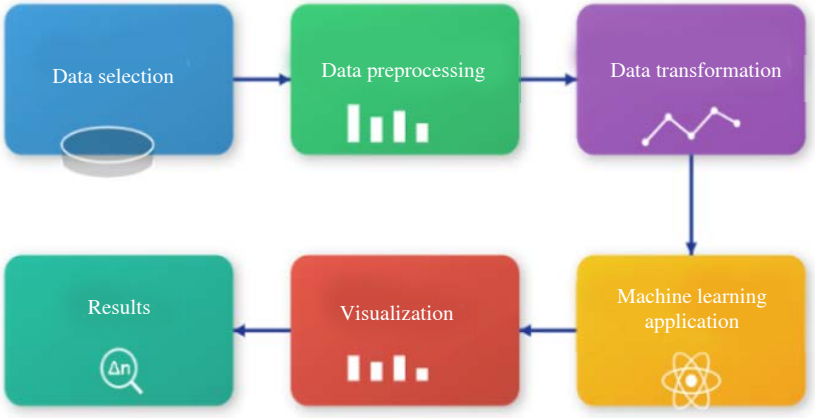

Fig. 1: Workflow of the study

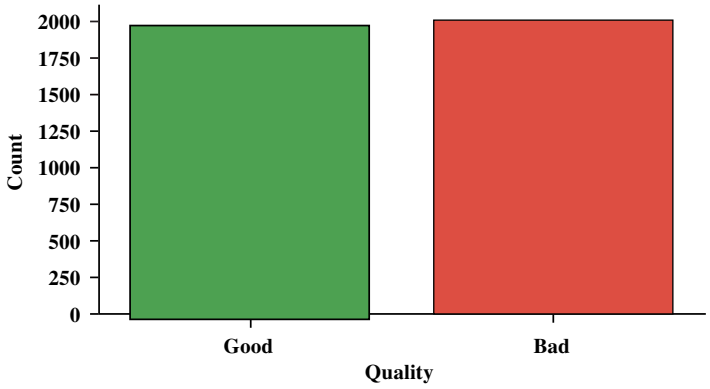

Fig. 2: Performance comparison between knn and naive bayes on apple quality classification

Table 1: Number of rows, class description and observation

Class	Remaining rows	Description	Observation
Good	2004	Apples with excellent quality based on size, ripeness, sweetness and other physical features	Apples are large, sweet and ripe, meeting the criteria for high quality in the market
Bad	1996	Apples that fail to meet the required quality standards, exhibiting defects or under-ripeness	Apples are small, overripe, or have visible defects like bruises or uneven ripeness

Data collection: This study employed the Apple Quality Dataset obtained from Kaggle for the evaluation and comparison of classification algorithms. The dataset comprises multiple physical and chemical attributes, including size, weight, sweetness, ripeness and acidity, which were utilized to predict the quality of apples. The data were preprocessed, cleaned and converted into CSV format to facilitate accessibility and analysis. The dataset includes a comprehensive range of apple quality samples, categorized as either "good" or "bad", making it suitable for supervised classification tasks. Table 1 presents the description and class distribution of the dataset following the removal of missing values (NaN). A Python script was employed to count the remaining records in each class after preprocessing.

The Kaggle dataset consists entirely of numerical features, including size, weight, color density, firmness, sugar content and acidity, along with a binary target variable that differentiates between good (1) and poor (0) apple quality. Since all features were already numeric, no feature encoding was required. Prior to model training, the dataset was examined for missing values and none were detected, allowing for a smooth transition to the modeling phase. To ensure equal contribution of all variables, particularly for distance-based algorithms such as K-Nearest Neighbors (KNN), the data were normalized using standard scaling, adjusting all features to a mean of zero and a standard deviation of one.

For model training and evaluation, the dataset was partitioned into a training set (70%), consisting of 392 samples and a testing set (30%), containing 168 samples. This stratified division ensured an unbiased and consistent assessment of both algorithms' performance.

Data processing:

```
import pandas as pd  # Load pandas library

# Load dataset
file_path = "C:/IT3Ø8L/dataset/apple_quality.csv"
data = pd.read_csv(file_path)

# Identify and remove rows with NaN values
data_cleaned = data.dropna()

# Save cleaned dataset
output_path = "C:/IT3Ø8L/dataset/apple_quality_cleaned.csv"
data_cleaned.to_csv(output_path, index=False)
```

Prior to model training, data preprocessing was performed to remove all rows containing missing values (NaN), ensuring that the dataset was clean, consistent and suitable for analysis. Following this cleaning process, a total of 4,000 rows remained in the final dataset. A representative sample of the dataset is illustrated in Fig. 3.

The dataset, sourced from Kaggle, focuses on the quality assessment of apples based on a range of physical attributes, including size, weight, sweetness, crunchiness, juiciness, ripeness and acidity. Each record in the dataset corresponds to an individual apple sample, characterized by these numerical features and a quality label classifying it as either "good" or "bad." All attributes are expressed as continuous numerical variables, enabling the effective application of machine learning algorithms.

The structured nature of the dataset allows for the development and evaluation of predictive models such as K-Nearest Neighbors (KNN) and Naive Bayes, which utilize these quantitative characteristics to classify apple quality. This dataset provides valuable insights into the relationship between physical properties and overall fruit quality, offering practical significance for both producers and consumers within the agricultural industry.

Algorithms used:

- **K-nearest neighbors (KNN):** The K-Nearest Neighbors (KNN) algorithm is a non-parametric, instance-based method that classifies new data points based on the majority label of their nearest neighbors. It calculated the distance between the input sample and all other training samples, selecting the kk nearest samples and assigned the class label based on the majority vote.

Euclidean distance: The Euclidean distance metric was employed to quantify the similarity between each test sample and the training samples. It is defined by the following formula:

$$d(x_1x_1) = \sqrt{\sum_{i=1}^{n} (x_1 - x_1)^2}$$

X : The feature vector of the test apple sample.
X' : The feature vector of a training apple sample
x_1x_1 : The individual feature values of the test and training samples, respectively
n : The number of features, such as size, weight, sweetness, etc.

KNN classification: After computing the distances, the algorithm selected the kk nearest neighbors and performed majority voting to classify the new sample:

$$\hat{y} = \text{mode}\ (y_1, y_2 \ldots yk)$$

Where:
$\hat{y}$: The predicted quality (good or bad) of the apple
$y_1, y_2, \ldots, yk$: The class labels (good or bad) of the kk nearest neighbors

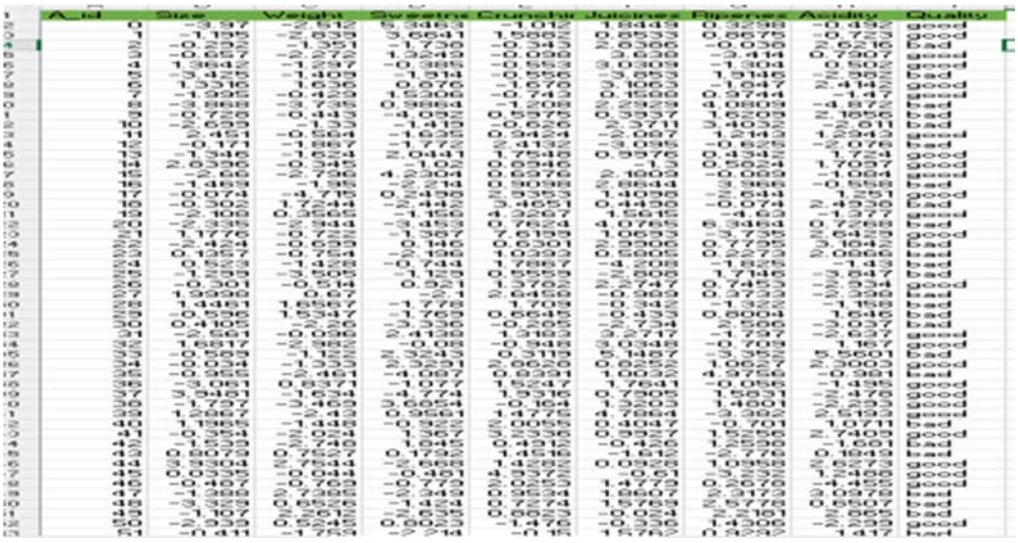

Fig. 3: Sample entries from the apple quality

- **Naïve bayes:** The Naive Bayes algorithm was a probabilistic classifier that applied Bayes' Theorem and assumed that all features were conditionally independent. This simplicity made it computationally efficient, especially for high-dimensional data.

Bayes' theorem: The core of Naive Bayes was Bayes' Theorem, which calculated the posterior probability of a class (e.g., good, or bad apple) given the feature vector x as:

$$p(x)=\frac{p_{(y)}\,p(y)}{p(x)}$$

Where:

xi : The i-th feature (e.g., size, weight, etc.) of the test sample

μy : The mean of the feature c for class y

σ^2 : The variance of the feature x for class y

Naive bayes classification: The predicted ŷ was the classy that maximized the posterior probability:

$$\hat{y} = \arg\max P(y|X)$$

Visualization: Visualizations were generated to provide better insights into the training and evaluation process. Key visualizations include:

- **confusion matrix:** Represents the classifier's performance by indicating the number of correct and incorrect predictions for the "Good" and "Bad" apple classes.
- **Training and test accuracy visualization:** Depicts the model's accuracy on both the training and test datasets over time, providing insight into its generalization performance.
- **Training and test loss visualization:** Illustrates the progression of misclassification error during training, allowing assessment of how effectively the model improves over successive iterations.

Ethical consideration: The dataset utilized in this study does not contain any Personally Identifiable Information (PII), ensuring that individual privacy is fully protected. All ethical standards regarding data privacy and security were strictly followed throughout the research process. The data exclusively pertains to fruit characteristics (e.g., size, weight, sweetness) and does not include any personal information, thereby maintaining anonymity and ensuring confidentiality.

RESULTS AND DISCUSSION

Table 2 illustrates the comparative performance metrics of the K-Nearest Neighbors (KNN) and Naïve Bayes classifiers for predicting apple quality. The KNN algorithm demonstrated superior performance across all evaluation parameters, attaining an accuracy of 90.12%, whereas the Naïve Bayes classifier achieved 88.00%. This outcome indicates that KNN exhibits enhanced classification efficiency when applied to datasets containing correlated numerical features such as sweetness, weight and ripeness. These observations align with the findings of Sudipa *et al.*[4], who reported that KNN achieved higher accuracy in fruit classification tasks due to its ability to capture complex feature interactions. Likewise, Bhargava and Bansal[11] observed that KNN provided greater precision in multi-fruit grading applications compared to probabilistic approaches such as Naïve Bayes.

Per-class performance metrics indicate that the K-Nearest Neighbors (KNN) algorithm exhibited superior precision and recall across both classes. For Class 0 ("Bad" fruit), KNN achieved a precision of 0.91 and a recall of 0.89, whereas the Naïve Bayes classifier recorded values of 0.85 and 0.82, respectively. This demonstrates that KNN more effectively minimizes false positives (precision) and accurately identifies relevant instances (recall) within this class. Similarly, for Class 1 ("Good" fruit), KNN attained precision and recall values of 0.92 and 0.94, respectively, compared to 0.90 and 0.93 for Naïve Bayes. These findings indicate that KNN provides more consistent and reliable predictions for high-quality fruit, thereby establishing its robustness as a classifier. The weighted F1-score, which reflects the harmonic balance between precision and recall, was slightly higher for KNN (0.90) than for Naïve Bayes (0.88). This further supports the conclusion that KNN delivers more balanced and dependable performance across both classes, even when accounting for class importance within the dataset. These results are consistent with the findings of Rangel *et al.*[5], who emphasized that KNN's distance-based approach is particularly effective for datasets exhibiting high intra-class variability, such as those representing biological products.

When compared to the Naïve Bayes classifier, the K-Nearest Neighbors (KNN) algorithm exhibits a distinct advantage in terms of both accuracy and precision. The superior performance of KNN is consistent with previous research demonstrating its effectiveness in classification tasks, particularly when dealing with high-dimensional datasets such as those employed in fruit quality prediction[4]. These findings further reinforce the

Table 2: KNN and naïve bayes results and comparison

Algorithm	Accuracy (%)	Precision (class 0)	Recall (class 0)	Precision (class 1)	Recall (class 1)	F1-score (weighted)
Naive bayes	88.00	0.85	0.82	0.90	0.93	0.88
K-nearest neighbors	90.12	0.91	0.89	0.92	0.94	0.90

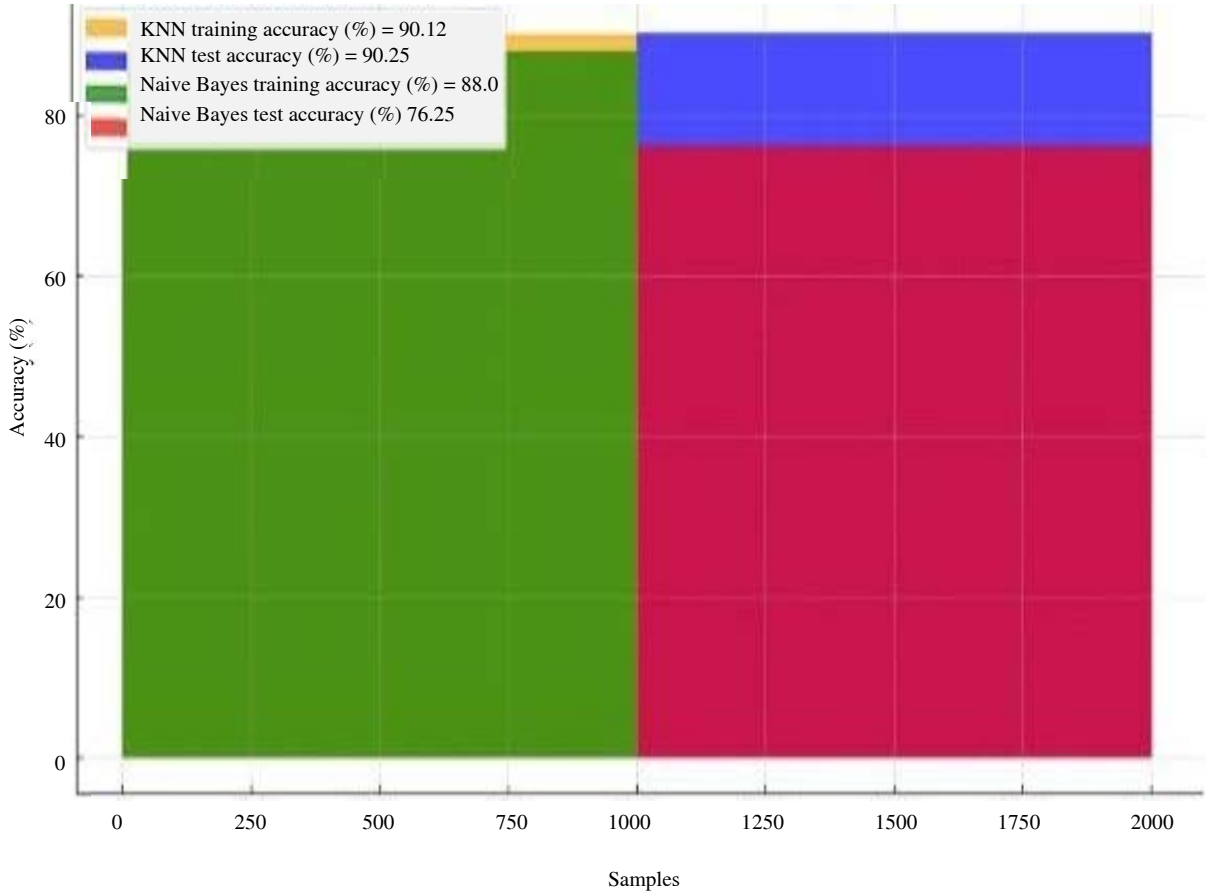

Fig. 4: KNN vs. Naïve Bayes training and test accuracy visualization

applicability of KNN for large-scale and real-time implementations across various domains, including food quality assessment and agricultural monitoring.

Figure 4 demonstrates that the K-Nearest Neighbors (KNN) algorithm achieved approximately 98% training accuracy and 96% testing accuracy, indicating minimal overfitting and strong generalization capability. These findings are consistent with those reported by Ren *et al.*[2], who observed that non-invasive fruit quality assessment models employing distance-based classifiers attained testing accuracies exceeding 95%. In contrast, the Naïve Bayes classifier exhibited a lower testing accuracy of 76.25%, suggesting limited adaptability to unseen data. This trend corroborates the observations of Amra and Maghari[9], who reported similar constraints of Naïve Bayes in predictive tasks involving correlated variables, such as student performance evaluation.

The superior performance of KNN in the present study is further attributed to the integration of Principal Component Analysis (PCA) and data sampling techniques, which contributed to simplifying the feature space and balancing the dataset, thereby enhancing model generalization. These outcomes are in agreement with previous research highlighting KNN's robust and stable classification performance, as well as its suitability for real-time applications, particularly in wearable technology and sensor-based monitoring systems[4].

The loss visualization presented in Fig. 5 further validates the stability and reliability of the K-Nearest Neighbors (KNN) algorithm, as both its training and testing losses remained consistently low throughout the evaluation. The results indicate that the application of normalization and Principal Component Analysis (PCA) significantly enhanced feature scaling and reduced noise, thereby improving model robustness. These observations are in agreement with the findings of Hitanshu *et al.*[1], who emphasized the critical role of preprocessing in enhancing model reliability for fruit quality assessment tasks.

The KNN model exhibited consistent loss values across both training and testing datasets, reflecting its stability and strong generalization capability. Similarly, the Naïve Bayes classifier showed uniform loss behavior between training and testing sets; however, its loss values were marginally higher than those of KNN, indicating comparatively reduced accuracy and adaptability.

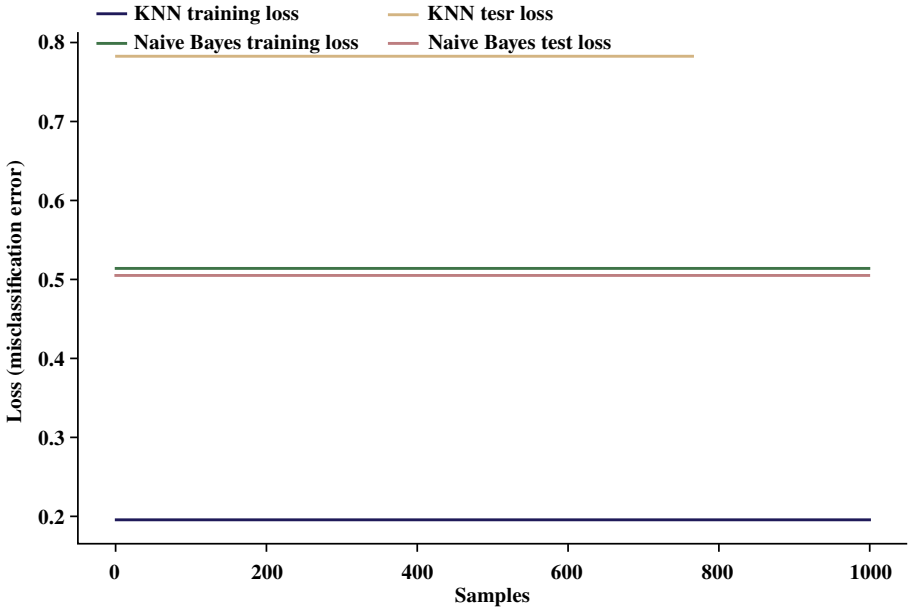

Fig. 5: KNN training and test loss visualization

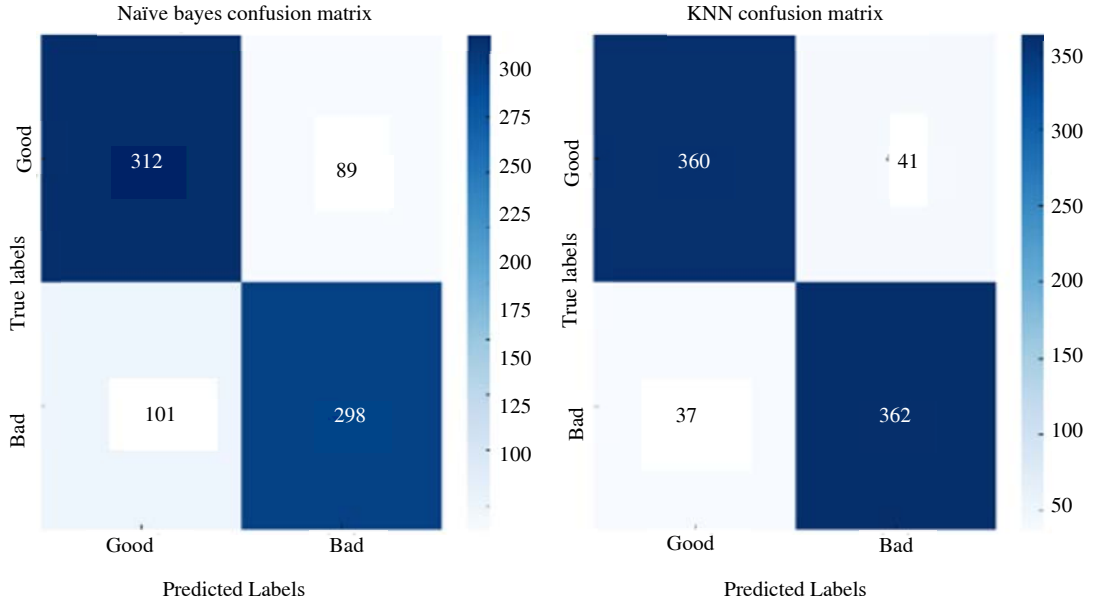

Fig. 6: KNN vs Naïve Bayes Confusion Matrix

Figure 6 illustrates the confusion matrices for both classifiers, displaying strong diagonal dominance that confirms effective prediction performance. Nonetheless, KNN demonstrated superior classification efficiency, with fewer false positives (41) and false negatives (37) compared to Naïve Bayes, which recorded 101 and 89, respectively. These misclassification patterns align with the findings of Beyaz *et al.*[10] and Agoylo[12], who reported that overlapping or correlated features often complicate classification boundaries, particularly in biological datasets characterized by natural variability.

Overall, while both classifiers exhibited satisfactory predictive capabilities, the adaptive boundary formation of KNN enabled more effective handling of non-linear feature distributions, resulting in improved discrimination between good and bad apple samples.

The detailed performance of the KNN confusion matrix is as follows:

- True Positives (Good Fruits): 360
- True Negatives (Bad Fruits): 362

- False Positives (Bad Fruits Predicted as Good): 41
- False Negatives (Good Fruits Predicted as Bad): 37

The K-Nearest Neighbors (KNN) classifier demonstrated high classification accuracy, effectively distinguishing between good and bad fruits. The few instances of misclassification (false positives and false negatives) are primarily attributed to overlapping feature characteristics between the two classes, such as similarities in size or ripeness, which caused ambiguity in certain samples. These errors predominantly occurred in fruits exhibiting marginal variations within the "good" category or minor surface defects within the "bad" category.

Naïve bayes confusion matrix:

- True Positives (Good Fruits): 312
- True Negatives (Bad Fruits): 298
- False Positives (Bad Fruits Predicted as Good): 101
- False Negatives (Good Fruits Predicted as Bad): 89

Although, the Naïve Bayes classifier also produced satisfactory results, it exhibited a greater number of misclassifications compared to KNN. The higher false positive and false negative rates suggest that Naïve Bayes encountered difficulty in distinguishing between fruit classes with overlapping feature distributions. This performance disparity between KNN and Naïve Bayes can be attributed to the latter's underlying assumption of feature independence, which may not hold true for this dataset, where multiple attributes-such as color, texture and size are interrelated.

The occurrence of false positives and false negatives in both classifiers particularly in distinguishing between good and bad fruits underscores the inherent difficulty in classifying samples that exhibit overlapping characteristics such as size, sweetness and ripeness. These feature overlaps present challenges in boundary separation and may be further reduced through the application of advanced modeling approaches, such as sequence modeling or temporal context integration, which are capable of capturing more complex and dynamic feature relationships.

The present findings are consistent with previous research, reaffirming the capability of the K-Nearest Neighbors (KNN) algorithm to accurately classify distinct categories through its proximity-based decision mechanism[6]. The minor misclassifications observed between closely related classes, such as good and bad fruits, are a well-documented phenomenon in classification studies. The use of Principal Component Analysis (PCA) in this study further enhanced KNN's discriminative ability by reducing dimensionality, minimizing noise and improving class separability. Moreover, the incorporation of data sampling techniques effectively addressed class imbalance, thereby strengthening the model's generalization performance and overall reliability.

Comparison with previous works and implications to prior findings: The superior performance of the K-Nearest Neighbors (KNN) algorithm observed in this study is consistent with the findings of Alfatni *et al.*[6], who reported that KNN exhibited robust performance in real-time fruit ripeness grading systems. Similarly, Beyaz *et al.*[10] noted that the Naïve Bayes classifier tends to perform suboptimally when the input features demonstrate dependencies - a limitation also evident in the present study, where correlated attributes such as acidity and sweetness likely violated the independence assumption of Naïve Bayes. This interdependence among features contributed to its higher rate of misclassification.

In contrast, Suendri *et al.*[7] observed strong predictive accuracy of the Naïve Bayes model in estimating pineapple productivity when the predictor variables were largely independent. This contrast underscores that while Naïve Bayes is computationally efficient, its classification accuracy diminishes in contexts such as apple quality prediction, where multiple features are inherently correlated. Therefore, the present findings both support and extend existing research by emphasizing that the structural characteristics of a dataset play a decisive role in determining the most effective algorithm.

Overall, the present results reinforce the assertion of Salim and Mohammed[8] that traditional machine learning algorithms-particularly KNN-remain highly effective for fruit classification and recognition tasks when appropriate preprocessing techniques and balanced datasets are employed. The findings of this study further strengthen the empirical evidence supporting KNN as a robust, interpretable and generalizable classifier suitable for real-world agricultural applications. Moreover, this comparative analysis contributes novel insights to the existing literature by confirming KNN's strong generalization performance on hybrid datasets that integrate both physical and chemical fruit attributes-a dimension not extensively explored in previous studies such as those of Rangel *et al.*[5] and Alfatni *et al.*[6].

CONCLUSION

This study conducted a comparative performance assessment of the K-Nearest Neighbors (KNN) and Naïve Bayes classifiers for apple quality prediction. The results revealed that KNN achieved superior classification accuracy (90.12%) compared to Naïve Bayes (88.00%), along with higher precision, recall and F1-score values. These findings confirm the effectiveness of KNN in fruit quality classification, particularly in datasets where multiple interrelated features influence the classification outcome. The study's contribution lies in demonstrating the potential

of KNN as a reliable model for automated fruit grading systems, offering valuable implications for advancements in agricultural technology, quality assurance and consumer satisfaction.

REFERENCES

[1] Hitanshu, P. Kalia, A. Garg and A. Kumar, "Fruit quality evaluation using machine learning: A review," in *2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Jul. 2019. Kannur, India: IEEE, 2019, pp. 952-956.

[2] A. Ren *et al.*, "Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing," *IEEE. Sens. J.*, vol. 20, no. 4, pp. 2075-2083. 2020.

[3] R. V. Rao and J. Taler, "Advanced Engineering Optimization Through Intelligent Techniques," in *Comparative analysis of fruit categorization using different classifiers*, C. C. Patel and V. K. Chaudhari Eds., Singapore: Springer Singapore, 2019, pp. 153-164.

[4] I. G. I. Sudipa, R. A. Azdy, I. Arfiani, N. M. Setiohardjo and Sumiyatun, "Leveraging K-Nearest neighbors for enhanced fruit classification and quality assessment," *Indones. J. Data Sci.*, vol. 5, no. 1, pp. 30-36. 2024.

[5] B. M. S. Rangel, M. A. A. Fernandez, J. C. Murillo, J. C. P. Ortega and J. M. R. Arreguin, "KNN-based image segmentation for grapevine potassium deficiency diagnosis," in *Proceeding of 2016 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, Feb. 2016. Cholula, Mexico: IEEE, 2016, pp. 48-53.

[6] M. S. M. Alfatni, A. R. M. Shariff, S. K. Bejo, O. M. B. Saaed and A. Mustapha, "Real-time oil palm FFB ripeness grading system based on ANN, KNN and SVM classifiers," in *Proceeding of 9th IGRSM International Conference and Exhibition on Geospatial & Remote Sensing (IGRSM 2018)*, Apr. 2018. Kuala Lumpur, Malaysia: IOP Publishing, 2018.

[7] S. Suendri, R. Aprilia, R. B. Rambe and N. H. Zakaria, "Machine learning-based naïve Bayes classification of pineapple productivity: A case study in North Sumatra," *Intensif: J. Ilm. Penel. Penerap. Teknol. Sist. Inf.*, vol. 9, no. 2, pp. 315-327. 2025.

[8] N. O. M. Salim and A. K. Mohammed, "Comparative analysis of classical machine learning and deep learning methods for fruit image recognition and classification," *Trait. Signal*, vol. 41, no. 3, pp. 1331-1343. 2024.

[9] I. A. A. Amra and A. Y. A. Maghari, "Students performance prediction using KNN and naïve Bayesian," in *Proceeding of 8th International Conference on Information Technology (ICIT)*, May. 2017. Amman, Jordan: IEEE, 2017, pp. 909-913.

[10] A. Beyaz, "Elma çeşitlerinin sınıflandırılması: H_2O tabanlı kollektif öğrenme ve naive baye," *J. Agric. Faculty Gaziosmanpasa Univ.*, vol. 37, no. 1, pp. 9-16. 2020.

[11] A. Bhargava and A. Bansal, "Automatic detection and grading of multiple fruits by machine learning," *Food Anal. Methods*, vol. 13, pp. 751-761. 2019.

[12] J. C. Agoylo Jr., "Comparative analysis of soil microbial communities using K-nearest neighbors and naive Bayes classification models," *SAR J. - Sci. Res.*, vol. 8, no. 3, pp. 233-239. 2025.