International Journal of Applied Mathematics and Computing
E-ISSN: 3047-146X
P-ISSN: 3048-1988
IJAMC

Research Article

Prediction of Credit Sales Value with the Naive Bayes Algorithm on Sujase Cell Jakarta

Veri Arinal [1*], Untung Surapati [2], Sugiyono Sugiyono [3], Dita Safira [4]

1 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikomcki) Jakarta
2 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikomcki) Jakarta
3 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikomcki) Jakarta
4 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (Stikomcki) Jakarta
* Corresponding Author: e-mail: veriarinal@stikomcki.ac.id

Abstract: Background: The rapid growth of mobile phone usage has significantly increased the demand for prepaid credit services (mobile airtime), creating large volumes of transaction data that require effective analysis for business decision-making. Sujase Cell, a mobile credit retailer in Jakarta, faces challenges in predicting future sales performance and customer purchasing interest due to the accumulation of transaction records over time and the limitations of manual analysis. **Objective:** This study aims to identify customer purchasing interest and predict mobile credit sales values by implementing the Naive Bayes algorithm as a data mining approach to support sales forecasting and business development strategies. **Methods:** The research employed a quantitative predictive approach using a private dataset obtained from Sujase Cell. Data collection was conducted through observation and literature review. The dataset consisted of historical mobile credit sales transactions and sales balance records collected during the study period. The data underwent preprocessing stages, including normalization using the Min-Max Scaler technique, followed by data partitioning into training and testing datasets. The Naive Bayes classification method was then applied to analyze sales patterns and generate predictions. Model performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and confusion matrix-based assessment metrics. Several experimental scenarios involving different training-testing ratios and parameter configurations were conducted to determine the most effective predictive model. **Results:** The findings indicate that the Naive Bayes method successfully identified sales trends and customer purchasing behavior patterns. The best-performing model was obtained using a 90% training dataset and 10% testing dataset, resulting in the lowest prediction error. Experimental results demonstrated that the generated prediction model was capable of following actual sales patterns and producing reliable forecasting outcomes. The implementation of Naive Bayes provides valuable support for sales planning, inventory management, and marketing decision-making at Sujase Cell, enabling the business to improve operational efficiency and anticipate future market demand more effectively.

Keywords: Customer Interest; Mobile Credit; Naive Bayes; Sales Forecasting; Transaction Analysis.

Received: April 15, 2024
Revised: Mey 12, 2024
Accepted: Mey 25, 2024
Published: July 15, 2024
Curr. Ver.: July 25, 2024

1. Introduction

The rapid expansion of mobile communication technology has transformed mobile phone services into an essential component of daily life, significantly increasing the demand for prepaid credit and digital communication products. Mobile credit remains a critical resource for accessing voice calls, text messaging, and internet services, particularly in developing economies where prepaid subscriptions dominate the telecommunications market. Consequently, mobile credit retailers play an important role in facilitating digital connectivity and supporting consumer communication needs. However, many small and medium-sized mobile credit outlets still rely on manual transaction recording and sales

monitoring processes, which often lead to inefficiencies, inaccurate reporting, and difficulties in evaluating sales performance. These challenges become more pronounced as transaction volumes increase over time, making it difficult for business owners to identify purchasing trends and forecast future demand. According to [1], digital sales management systems can substantially improve transaction monitoring and reporting accuracy. Furthermore, effective sales forecasting enables businesses to optimize inventory planning and improve decision-making processes by utilizing historical transaction data as a strategic business asset [2].

Recent studies have demonstrated the effectiveness of machine learning and data mining techniques in solving various prediction and classification problems across different domains. Naive Bayes, in particular, has gained considerable attention due to its computational efficiency, simplicity, and ability to produce reliable classification results from relatively small datasets. Previous research has successfully applied Naive Bayes to classify customer purchasing interest, determine best-selling products, and support marketing decisions in retail environments [3], [4]. In the telecommunications sector, Naive Bayes has also been utilized to identify popular voucher products and predict consumer demand patterns. Despite these advancements, limited research has focused specifically on forecasting mobile credit sales values at small-scale prepaid credit retailers using actual transactional data. Most prior studies concentrated on product classification or customer segmentation rather than integrating sales prediction with customer purchase interest analysis. Therefore, this study addresses an important research gap by applying the Naive Bayes algorithm to predict mobile credit sales performance while simultaneously examining customer purchasing behavior within the operational context of Sujase Cell Jakarta.

Based on the identified research gap, this study seeks to investigate how customer purchasing interest can be analyzed through mobile credit sales prediction and how the Naive Bayes algorithm can be implemented to support sales forecasting activities. The primary research problem arises from the difficulty experienced by Sujase Cell in predicting monthly and annual sales performance using conventional approaches. As transaction records accumulate over time, business owners require a systematic method to transform historical sales data into meaningful information that can support future planning. The research is guided by two main questions: how customer purchasing interest can be identified through sales prediction, and how effectively the Naive Bayes algorithm can be applied to forecast mobile credit sales. To address these questions, historical sales data are analyzed using a probabilistic classification approach that leverages patterns observed in previous transactions. According to [5], Naive Bayes offers strong predictive capabilities in classification tasks, while [6] emphasize that data-driven forecasting models can significantly enhance sales prediction accuracy and business intelligence performance.

This research argues that predictive analytics can provide substantial benefits for small-scale mobile credit retailers by improving the accuracy of sales forecasting and supporting more informed business decisions. Through the implementation of the Naive Bayes algorithm, historical transaction data can be transformed into actionable insights that assist retailers in understanding customer demand patterns, anticipating future sales fluctuations, and developing more effective marketing strategies. The expected contribution of this study extends beyond the technical implementation of a classification algorithm. Practically, the proposed approach can help Sujase Cell and its reseller network monitor sales performance more efficiently, reduce reliance on manual reporting, and enhance operational effectiveness. The study also contributes to the growing body of literature on data mining applications in retail sales prediction by providing empirical evidence from the mobile credit industry, a sector that remains underrepresented in predictive analytics research. As highlighted by [7], predictive models can strengthen marketing and sales strategies, while [8] argue that intelligent prediction systems improve organizational responsiveness and decision-making capabilities in dynamic business environments.

2. Literature Review

Sales Prediction and Forecasting

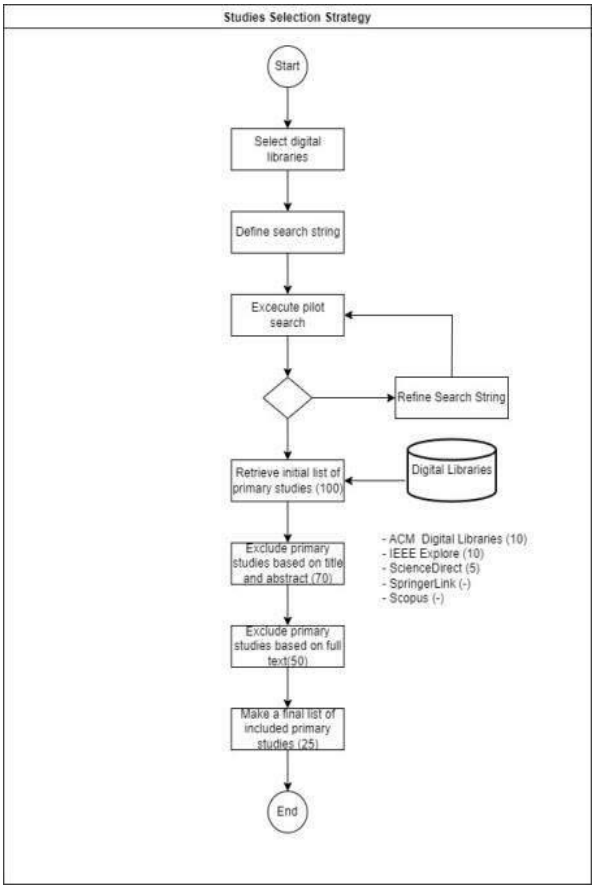

Figure 1. Studies Selection Strategy.

Sales prediction is a critical component of business intelligence because it enables organizations to anticipate future demand, optimize inventory levels, and formulate effective marketing strategies. Forecasting relies on historical data and analytical techniques to estimate future business outcomes under varying market conditions. In retail and telecommunication industries, accurate sales prediction helps organizations reduce uncertainty and improve operational efficiency. The concept of forecasting emphasizes the use of past transaction records to estimate future events with a reasonable degree of confidence rather than absolute certainty. Modern predictive systems integrate data mining and machine learning approaches to identify hidden patterns within large datasets and generate more reliable forecasts. The increasing availability of transactional data has encouraged researchers to develop advanced predictive models capable of supporting managerial decision-making processes. According to [2], machine learning-based forecasting models significantly improve demand prediction accuracy by leveraging historical business data. Similarly, [6] argue that predictive analytics provides valuable insights into consumer behavior and sales performance, allowing organizations to respond more effectively to changing market demands.

The growing adoption of predictive analytics has encouraged researchers to explore various machine learning techniques for forecasting sales performance across different industries. Studies have demonstrated that predictive models can successfully identify future purchasing trends, classify product demand, and support strategic planning. In the context of retail sales, forecasting methods are increasingly used to estimate customer demand, minimize stock shortages, and reduce inventory-related costs. The effectiveness of these models depends largely on the quality of historical data and the ability of algorithms to capture relationships among influencing variables. Recent studies have shown that prediction systems are capable of supporting business operations by providing timely and actionable information for decision-makers. Advanced predictive approaches have also been applied in financial risk

assessment and energy consumption forecasting, illustrating their versatility across domains. [9] demonstrated that machine learning techniques significantly improve predictive performance when supported by appropriate data transformation methods. Likewise, [10] reported that intelligent prediction models can achieve high forecasting accuracy and assist organizations in optimizing resource allocation and operational planning.

Within the telecommunications retail sector, forecasting sales performance is particularly important because consumer demand for prepaid credit products fluctuates based on user behavior, market competition, and technological developments. Mobile credit retailers accumulate large volumes of transactional data that can be utilized to predict future sales and identify customer purchasing patterns. Previous studies indicate that data-driven forecasting can improve business sustainability by enabling retailers to anticipate product demand and adjust marketing strategies accordingly. Sales prediction is not only useful for inventory management but also serves as a foundation for customer relationship management and business expansion initiatives. In small-scale retail environments, predictive systems can help overcome the limitations of manual reporting and subjective decision-making processes. [11] found that predictive models integrating multiple data sources can improve sales forecasting performance and support more accurate business planning. Furthermore, [12] emphasized that intelligent prediction systems contribute to better decision-making by identifying optimal sales patterns and supporting demand-oriented resource management strategies.

Naive Bayes Algorithm

The Naive Bayes algorithm is one of the most widely used probabilistic classification methods in data mining and machine learning applications. Based on Bayes' theorem, the algorithm assumes conditional independence among predictor variables, allowing efficient classification even when working with relatively small datasets. Its simplicity, computational efficiency, and ability to generate reliable predictions have made Naive Bayes a popular choice for various classification and forecasting tasks. The algorithm estimates the probability of a particular outcome by analyzing the occurrence of observed attributes within historical data. Despite its independence assumption, which may not always reflect real-world conditions, Naive Bayes often performs competitively compared to more complex machine learning methods. Its effectiveness has been demonstrated in multiple domains, including image classification, sentiment analysis, educational assessment, and customer behavior prediction. According to [5], Naive Bayes achieved excellent predictive performance in accreditation classification tasks, demonstrating its practical value in decision-support systems. Similarly, [13] reported satisfactory classification accuracy when applying Naive Bayes to agricultural image analysis problems.

The versatility of Naive Bayes has encouraged its application in commercial and marketing environments where rapid and accurate classification is required. In sales-related studies, the algorithm has been employed to identify customer purchasing interests, classify best-selling products, and support marketing strategy development. One of its primary advantages is the ability to process large amounts of historical transaction data while maintaining relatively low computational complexity. This characteristic makes Naive Bayes particularly suitable for small and medium-sized enterprises that may lack extensive computational resources. Furthermore, the algorithm provides interpretable probabilistic outputs that facilitate managerial understanding of prediction results. Previous studies involving telecommunication products have shown that Naive Bayes can effectively classify customer preferences and forecast sales outcomes based on transaction histories. [4], [14] demonstrated that the algorithm successfully identified customer purchasing interest in internet package products. Likewise, [7] found that Naive Bayes can support marketing decision-making by classifying customer behavior and identifying product demand trends.

Several studies have specifically investigated the application of Naive Bayes in telecommunications and digital product sales, highlighting its effectiveness in identifying high-demand products and supporting business decisions. Research conducted in mobile credit and voucher retail environments suggests that the algorithm can accurately classify products into categories such as best-selling and non-best-selling, enabling retailers to improve stock management and marketing effectiveness. Compared with manual analysis methods, Naive Bayes provides a systematic and data-driven approach for transforming historical transaction records into actionable business insights. Moreover, its implementation can help retailers better understand customer purchasing behavior and anticipate future demand patterns. Although more advanced machine learning algorithms are available, Naive

Bayes remains attractive due to its simplicity, transparency, and strong predictive capability in practical business scenarios. [3] demonstrated that Naive Bayes successfully classified product sales performance within a mobile phone retail environment. Similarly, [15] reported high classification accuracy in identifying best-selling voucher products, confirming the suitability of Naive Bayes for telecommunications sales prediction applications.

3. Materials and Method

Research Data

Data collection is a fundamental stage in this research because it provides the information required to develop and evaluate the sales prediction model. This study utilizes a private dataset obtained from Sujase Cell and applies two primary data collection methods: direct observation and literature review. Observation was conducted by visiting the research site to gain a comprehensive understanding of the business processes, transaction activities, and sales recording practices implemented by the mobile credit retailer. Through direct observation, the researcher was able to identify operational challenges related to sales monitoring, transaction management, and demand forecasting. This approach also enabled the collection of relevant sales data and business information that reflect actual operational conditions. In addition to observation, a literature review was carried out to support the theoretical foundation of the study and to identify appropriate analytical methods for sales prediction. Various academic sources, including journals, conference papers, books, and previous studies related to sales forecasting, data mining, and machine learning, were examined to provide a strong conceptual basis for the research and to ensure the selection of a suitable prediction approach.

Table 1. Primary data from Sujase Cell.

No	Types of Pulses	Total Sales for November 2023				Total
		Sunday 1	Sunday 2	Sunday 3	Sunday 4	
1.	Telkomsel	50	40	35	40	165
2.	Indosat	55	55	45	50	205
3.	XL	25	30	40	20	115
4.	Axis	30	35	30	45	140
5.	Tri	45	40	50	50	185
6.	Smartfren	40	45	50	60	195

A literature review was conducted as a supporting data collection method to establish the theoretical and conceptual foundation of this research. This process involved examining a wide range of academic sources, including international journals, national journals, local publications, conference proceedings, books, and other scholarly references relevant to sales prediction, data mining, machine learning, and business intelligence. The purpose of the literature review was to gather information, identify research trends, understand existing methodologies, and evaluate previous findings related to predictive analytics and classification techniques. In addition to written publications, supporting materials such as electronic documents, reports, figures, tables, and digital resources were also reviewed to enrich the understanding of the research topic. The collected literature served as a valuable source of knowledge for developing the research framework, selecting appropriate methods, and interpreting the study results. Furthermore, the literature review helped identify research gaps and opportunities for further investigation, ensuring that the study was conducted based on established scientific knowledge while contributing meaningful insights to the field of sales prediction and decision support systems.

Methodology Implementation

The methodology applied in this study is the Naive Bayes algorithm, a probabilistic classification technique designed to predict outcomes based on historical data patterns. Naive Bayes operates by calculating the probability that a particular data instance belongs to a specific class, assuming that all predictor variables are conditionally independent of one another. This characteristic makes the algorithm computationally efficient and suitable for handling large datasets while maintaining reliable predictive performance. In this research, the

algorithm is utilized to analyze historical mobile credit sales data and identify patterns that can be used to forecast future sales performance. The learning process focuses on estimating probabilities derived from training data, allowing the model to classify and predict sales trends effectively. To evaluate the performance of the prediction model, a confusion matrix is employed as a measurement tool. This evaluation method assesses classification results through several metrics, including accuracy, precision, and recall. Accuracy measures the closeness of prediction results to actual outcomes, precision evaluates the relevance of predicted results, and recall measures the model's ability to correctly identify relevant instances. Together, these metrics provide a comprehensive assessment of the effectiveness and reliability of the proposed Naive Bayes-based prediction model.

Table 2. Confusion Matrix

Actual Class	Predicted Negative	Predicted Positive	Total
Negative	True Negative (TN)	False Positive (FP)	Total Negative
Positive	False Negative (FN)	True Positive (TP)	Total Positive
Total	Total Predicted Negative	Total Predicted Positive	Grand Total

Research Design

The research design was developed based on the identified research problems, objectives, and study scope to ensure that the investigation remains focused and achieves its intended outcomes. This design serves as a structured framework that guides each stage of the research process. The study begins with data collection using a private dataset obtained from Sujase Cell, consisting of customer transaction records and mobile credit sales data collected during the observation period. After the data were gathered, a preprocessing stage was performed to organize, clean, and prepare the dataset for further analysis. This stage aimed to improve data quality and ensure that the information was suitable for predictive modeling. The prepared data were then analyzed using data mining techniques to identify patterns and relationships within historical sales records. The Naive Bayes algorithm was selected as the primary prediction method because of its ability to classify data and estimate future outcomes based on previous observations. The model was designed to address challenges related to sales forecasting and customer purchase interest analysis. The results are expected to support business planning, improve sales management, enhance marketing activities, and provide valuable insights for making more effective and data-driven business decisions.

4. Results and Discussion

Research Tools

The successful implementation of this research required appropriate hardware and software resources to support data processing, model development, analysis, and documentation activities. The hardware environment consisted of a computer system equipped with a processor, storage device, and memory resources capable of handling data analysis tasks and running the required applications. The processor was used to execute computational operations and support the overall performance of the system, while the hard drive provided storage space for datasets, project files, application components, and research documentation. Random Access Memory (RAM) played an important role in temporarily storing data and instructions needed by the processor during application execution and data analysis processes. In addition to hardware, several software tools were utilized throughout the study. MATLAB was employed for data analysis, algorithm implementation, modeling, and prediction experiments. The Windows operating system managed system resources and provided the platform for application execution. Mozilla Firefox was used to access online resources and references, while Draw.io supported the creation of diagrams and research illustrations. Furthermore, Mendeley was utilized as a reference management tool to organize academic sources and facilitate the preparation of research documentation and bibliographic materials.

Implementation and Testing

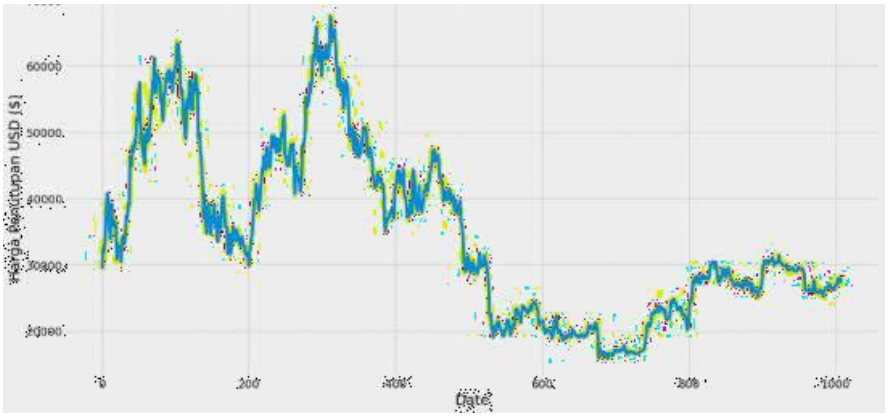

Figure 2. Credit price chart from 2021-2023.

The implementation and testing phase began with the collection and preparation of a private dataset obtained from Sujase Cell Jakarta. The dataset consisted of historical mobile credit sales records collected between 2021 and 2023 and included attributes such as transaction date and sales balance value. All records were converted into CSV format to facilitate data processing and analysis. A total of 1,010 records were compiled and subsequently examined to identify the variables required for prediction. Before model development, the dataset underwent a preprocessing stage that included data cleaning, validation, and normalization using the Min-Max scaling technique to transform numerical values into a standardized range. The processed dataset was then divided into training and testing sets using several experimental scenarios, including 70:30, 80:20, and 90:10 proportions. The Naive Bayes algorithm was applied to learn patterns from historical sales transactions and generate predictions regarding future sales performance. Model effectiveness was evaluated by comparing predicted results with actual sales data. The testing process demonstrated that the developed model was capable of identifying sales patterns accurately and providing reliable predictions that can support business planning and decision-making activities at Sujase Cell.

5. Conclusion

This study investigated the application of the Naive Bayes method for predicting mobile credit sales based on historical sales transaction data collected from Sujase Cell. The research demonstrated that historical sales records can be effectively utilized to identify sales patterns and support future sales forecasting. Through the implementation of data preprocessing, model training, and prediction procedures, the developed model was able to generate reliable forecasting results and provide valuable insights into sales trends. The experimental results showed that the prediction model achieved a high level of performance, reflected by low prediction error values and strong forecasting accuracy. Furthermore, the findings indicated that future mobile credit sales are expected to experience a positive trend, suggesting an increase in customer demand during the forecast period. The use of predictive analytics in this study contributed to a better understanding of customer purchasing behavior and sales dynamics, enabling more informed business decisions. Overall, the implementation of the Naive Bayes approach proved to be a practical and effective solution for supporting sales prediction, improving business planning, and enhancing decision-making processes within the mobile credit retail environment. The proposed model can serve as a valuable tool for assisting retailers in managing sales activities, optimizing operational strategies, and responding proactively to future market demands.

References

[1] R. Afriansyah, "Pengembangan Sistem Informasi Pelaporan Transaksi Penjualan Dengan Multilokasi dan Multi Harga Produk Pada Konter," *Manutech J. Teknol. Manufaktur*, vol. 12, no. 2, 2020.

[2] M. A. Khan and S. Saqib, "Effective Demand Forecasting Model Using Business Intelligence Empowered With Machine Learning," *IEEE Access*, vol. 8, pp. 116013–116023, 2020, doi: 10.1109/ACCESS.2020.3003790.

[3] W. A. Purnama and T. A. Putra, "Klasifikasi Penjualan Produk Menggunakan Algoritma Naive Bayes pada Konter HP Bayu Cell," *REMIK Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 8, no. 1, pp. 286–292, 2024, doi: 10.33395/remik.v8i1.13207.

[4] A. S. Mubarok, D. R. Prehanto, and M. Ali, "Deteksi Minat Beli Pelanggan Terhadap Produk Paket Internet Menggunakan Algoritma Naive Bayes," *Inovate*, vol. 3, no. 2, pp. 58–63, 2019.

[5] L. Genisa and D. I. Mulyana, "Implementasi Penerapan Metode C4.5 dan Naive Bayes Dalam Tingkat Kelulusan Akreditasi Lembaga PAUD Pada Badan Akreditasi Nasional," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1595–1602, 2021, doi: 10.30865/mib.v5i4.3267.

[6] L. Huang and Z. Dou, "Online Sales Prediction: An Analysis With Dependency SCOR-Topic Sentiment Model," *IEEE Access*, vol. 7, pp. 79791–79797, 2019, doi: 10.1109/ACCESS.2019.2919734.

[7] R. P. Pratiwi, I. Tazro, and C. Juliane, "Penerapan Algoritma Naive Bayes untuk Mengidentifikasi Strategi Marketing dalam Penjualan Deposit E-Money," *Coopetition J. Ilm. Manaj.*, vol. 13, no. 1, pp. 65–72, 2022, doi: 10.32670/coopetition.v13i1.896.

[8] W. Yang, Y. Chen, and Y. Chen, "Intelligent Agent-Based Predict System With Cloud Computing for Enterprise Service Platform in IoT Environment," *IEEE Access*, vol. 9, 2021.

[9] M. Z. Abedin, G. Chi, M. M. Uddin, S. Satu, I. Khan, and P. Hajek, "Tax Default Prediction Using Feature Transformation-Based Machine Learning," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2020.3048018.

[10] A. Abdelaziz, V. Santos, and M. S. Dias, "Convolutional Neural Network with Genetic Algorithm for Predicting Energy Consumption in Public Buildings," *IEEE Access*, vol. 11, pp. 64049–64069, 2023, doi: 10.1109/ACCESS.2023.3284470.

[11] P. Pai and C. Liu, "Predicting Vehicle Sales by Sentiment Analysis of Twitter Data and Stock Market Values," *IEEE Access*, vol. 6, pp. 57655–57662, 2018, doi: 10.1109/ACCESS.2018.2873730.

[12] T. Mingjie, "Research on Commodities Constraint Optimization Based on Graph Neural Network Prediction," *IEEE Access*, vol. 11, pp. 90131–90142, 2023, doi: 10.1109/ACCESS.2023.3302923.

[13] C. B. Roring, D. I. Mulyana, and Y. T. Lestari, "Klasifikasi Tingkat Kematangan Buah Jambu Bol Berdasarkan Warna Kulit Menggunakan Metode Naive Bayes," *J. Pendidik. Tambusai*, vol. 6, no. 1, pp. 2938–2948, 2022.

[14] M. Tavasoli, E. Lee, and Y. Mousavi, "WIPE: A Novel Web-Based Intelligent Packaging Evaluation via Machine Learning and Association Mining," *IEEE Access*, vol. 12, pp. 45936–45947, 2024, doi: 10.1109/ACCESS.2024.3376478.

[15] I. Nawangsih and A. Setyaningsih, "Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Voucher Kuota di Edi Cell," *J. SIGMA*, vol. 3, pp. 14902–14914, 2019.