



IMPLEMENTATION OF THE NAIVE BAYES METHOD FOR CATERING SALES PREDICTION AT PT NEGARA RASA INDONESIA

IMPLEMENTASI METODE NAIVE BAYES UNTUK PREDIKSI PENJUALAN KATERING DI PT NEGARA RASA INDONESIA

Benifati Gulo^{1*}, Syaeful Machfud²

^{1*}Faculty of Computer Science, Pamulang University, Email: gulobeni74@gmail.com

²Faculty of Computer Science, Pamulang University, Email: dosen02836@unpam.ac.id

*email koresponden: gulobeni74@gmail.com

DOI: <https://doi.org/10.62567/micjo.v3i2.2374>

Abstract

This study discusses the implementation of the Naïve Bayes method to predict catering sales at PT.Negara Rasa Indonesia. The background of this study is based on the problem of suboptimal sales due to the absence of a structured sales prediction system. The Naïve Bayes method was chosen because of its simplicity, speed, and ability to classify data with a high degree of accuracy. The data used in this study is historical sales data from the last two years, which has undergone cleaning, labeling, and transformation into four sales categories, namely very popular, popular, fairly popular, and less popular. The testing process was carried out using RapidMiner software by dividing the dataset into training data and test data at various ratios of 80:20. The test results showed a very high level of accuracy, with the highest value reaching 91.41%. These findings prove that the Naïve Bayes method is reliable for predicting catering sales, thereby assisting decision-making in more efficient sales management and planning at PT. Negara Rasa Indonesia.

Keywords : Naïve Bayes, Sales Prediction, Historical Data, Rapidminer, Classification.

Abstrak

Penelitian ini membahas implementasi metode Naïve Bayes untuk memprediksi penjualan catering pada PT.Negara Rasa Indonesia. Latar belakang penelitian didasari oleh permasalahan penjualan yang belum optimal akibat tidak adanya sistem prediksi penjualan yang terstruktur. Metode Naïve Bayes dipilih karena kesederhanaan, kecepatan, serta kemampuannya dalam mengklasifikasikan data dengan tingkat akurasi yang tinggi. Data yang digunakan dalam penelitian ini adalah data historis penjualan selama dua tahun terakhir, yang telah melalui proses cleaning, labeling, dan transformasi menjadi empat kategori penjualan, yaitu sangat laris, laris, cukup laris, kurang laris. Proses pengujian dilakukan menggunakan perangkat lunak RapidMiner dengan membagi dataset menjadi data latih dan data uji pada berbagai rasio 80:20, Hasil pengujian menunjukkan tingkat akurasi yang sangat tinggi, dengan nilai tertinggi mencapai 91,41% Temuan ini membuktikan bahwa metode Naïve Bayes dapat diandalkan untuk memprediksi penjualan catering, sehingga dapat membantu pengambilan keputusan dalam pengelolaan dan perencanaan penjualan yang lebih efisien di PT. Negara Rasa Indonesia.

Kata Kunci : Naïve Bayes, Prediksi Penjualan, Data Historis, Rapidminer, Klasifikasi.



1. INTRODUCTION

In the business world, sales prediction serves as a benchmark for the progress of a company itself. In the business world, the ability to predict sales is one of the important indicators used to assess the progress and health of a company. Usually, this prediction is carried out by analyzing sales data from previous months or years, which is then used as a basis to estimate performance in the future. However, PT. Negara Rasa Indonesia faces major challenges in this regard. The company experiences difficulties in predicting the percentage of sales for the coming months or years. This problem has not yet found an effective solution until now.

One of the main causes is uncertainty and lack of trust in the data analysis process. The internal analysis team often experiences difficulties because there are concerns about the potential for nepotism practices within the company. They worry that the analysis results may not be truly objective, but instead influenced by the interests of certain parties. As a result, the company does not have a strong foundation or basis for formulating strategies to deal with market fluctuations. Without accurate and reliable predictions, PT. Negara Rasa Indonesia finds it difficult to prepare anticipatory steps when there is a decline in sales or sudden changes in market conditions.

Catering is a service industry that provides food and beverages for various events or activities, both indoors and outdoors. Catering includes planning, preparation, processing, transportation, presentation, and service of food and beverages. Simply put, catering consists of a chef and a team who prepare, cook, and serve food. Catering is a food service in which the cooking process is carried out at a location different from where it is served. The prepared food is then delivered to the event location such as parties, meetings, or industrial canteens. This service can include the provision of snacks to complete meals, according to customer requests [8].

PT Negara Rasa Indonesia is a company engaged in the catering business with various market segments, ranging from wedding events, corporate meetings, to the provision of daily healthy meals. In facing intense competition, the company needs to implement data analysis methods that can assist in business decision-making, such as sales prediction. One method that can be used is Naïve Bayes, a probability-based classification method that is effective in predicting events based on historical data.

2. RESEARCH METHOD

Research Design

This study uses a quantitative approach with an experimental method to implement the Naïve Bayes method in predicting catering sales based on historical sales data. The research design was developed to evaluate the effectiveness of the Naïve Bayes classification method in categorizing sales into four classes, namely Very Popular, Popular, Fairly Popular, and Less Popular. This approach was chosen because of the ability of Naïve Bayes to handle categorical and numerical data simultaneously, as well as its computational efficiency for medium-sized datasets.

The research methodology framework was built based on the stages of Knowledge Discovery in Database (KDD), which include data collection, data cleaning, attribute selection, data transformation, and model evaluation. Each stage was designed to ensure the quality of the input data and the validity of the prediction results generated by the Naïve Bayes model. The validation process was carried out using the hold-out validation technique with an 80:20 data split for training data and testing data to measure the generalization performance of the model.

Data Sources and Types

The research data were obtained from historical sales records of products at Toko Sukses Kunci Pamulang over a two-year period, covering 713 sales transaction records whose completeness and consistency had been verified. The dataset consists of temporal variables (transaction date), product identity variables (item name and product code), quantitative variables (selling price, sales quantity, total sales), and categorical variables (unit code). The selection of this data source was based on the availability of comprehensive data that are representative of retail product sales patterns.

The type of data used is secondary data in the form of time series sales transactions that have



been structured in digital format. The data characteristics include diverse product variations with heterogeneous ranges of prices and sales volumes, providing sufficient complexity for testing the robustness of the Naïve Bayes method. Data validity was ensured through a verification process with the store management to confirm the accuracy and relevance of the data to actual business conditions.

Data Preprocessing

The data preprocessing stage was carried out systematically to ensure optimal dataset quality for the implementation of the Naïve Bayes method. This process includes three main stages, namely data cleaning, data selection, and data transformation, which were conducted sequentially to produce a structured dataset that meets the requirements of the classification model. Data cleaning was performed to identify and handle missing values, duplicate records, and inconsistencies in data format across all critical attributes. The examination was conducted on the columns Date, Menu, Vendor, Menu Price, Event, Order Quantity, and Amount Sold (Amount) using statistical profiling techniques, which indicated that there were no missing values or duplicates in the dataset.

Data selection was carried out based on relevance to the objective of predicting catering sales by selecting seven main attributes as input features for the model, namely the columns Date, Menu, Vendor, Menu Price, Event, Order Quantity, and Amount Sold (Amount). Data transformation was performed by converting the numeric attribute Sales Quantity into ordinal categories using four intervals: “Very Popular” for \leq Rp. 7,000,000, “Popular” for Rp 5,000,000 – Rp 6,999,999, “Fairly Popular” for Rp 3,000,000 – Rp 4,999,999, and “Less Popular” for $>$ Rp 3,000,000. The transformation results produced a balanced class distribution with 189 data in the “Less Popular” category (19.12%), 343 data in the “Fairly Popular” category (34.71%), 234 data in the “Popular” category (23.68%), and 222 data in the “Very Popular” category (22.46%), supporting the optimal performance of the method in learning the pattern of each class.

Implementation of the Naïve Bayes Algorithm

The implementation of the Naïve Bayes algorithm begins with calculating the prior probability for each target class using the formula $P(C_i) = S_i/s$, where the calculation results show $P(\text{Very Popular}) = 0.224$, $P(\text{Popular}) = 0.236$, $P(\text{Fairly Popular}) = 0.347$, and $P(\text{Less Popular}) = 0.191$. The likelihood probability calculation is performed for each combination of feature values and target classes using the formula $P(X|C) = n(X|C)/\Sigma(C)$, producing a comprehensive probability table for all input features. The classification process for the training data involves building a probabilistic model based on 790 training data, while 197 testing data are used to evaluate the model performance through prediction and comparison with the actual categories.

Validation of the algorithm implementation was conducted using the RapidMiner Studio application to ensure the consistency of the manual calculation results with the software implementation. The classification process produces predicted demand categories for each testing data based on the combination of prior and likelihood probabilities from all input features. The constructed model is then used to predict product demand by calculating the probability of each class and selecting the class with the highest probability as the final prediction result.

Model Evaluation

The evaluation of the model performance was carried out using accuracy and error rate metrics calculated based on the prediction results on the testing data. Accuracy was calculated using the formula $(\text{number of correct predictions})/(\text{number of testing data}) \times 100\%$, providing the percentage of the model’s success in correctly classifying product demand categories. The error rate was calculated as the complement of accuracy using the formula $(\text{number of incorrect predictions})/(\text{number of testing data}) \times 100\%$, indicating the model’s prediction error level. The evaluation process was conducted by comparing the model’s prediction results with the actual labels in the testing data to calculate the number of correct and incorrect classifications.

The analysis of the evaluation results includes measuring the overall model performance as well as identifying patterns of prediction errors in each demand class. The evaluation implementation used a hold-out validation strategy to ensure the objectivity of the results and the generalizability of the model



to new data. The evaluation results serve as the basis for assessing the effectiveness of the Naïve Bayes method in the context of retail product demand prediction and provide recommendations for practical implementation in business decision-making systems. **Research Design**

This study uses a quantitative approach with an experimental method to implement the Naïve Bayes method in predicting catering sales based on historical sales data. The research design was developed to evaluate the effectiveness of the Naïve Bayes classification method in categorizing sales into four classes, namely Very Popular, Popular, Fairly Popular, and Less Popular. This approach was chosen because of the ability of Naïve Bayes to handle categorical and numerical data simultaneously, as well as its computational efficiency for medium-sized datasets.

The research methodology framework was built based on the stages of Knowledge Discovery in Database (KDD), which include data collection, data cleaning, attribute selection, data transformation, and model evaluation. Each stage was designed to ensure the quality of the input data and the validity of the prediction results generated by the Naïve Bayes model. The validation process was carried out using the hold-out validation technique with an 80:20 data split for training data and testing data to measure the generalization performance of the model.

Data Sources and Types

The research data were obtained from historical sales records of products at Toko Sukses Kunci Pamulang over a two-year period, covering 713 sales transaction records whose completeness and consistency had been verified. The dataset consists of temporal variables (transaction date), product identity variables (item name and product code), quantitative variables (selling price, sales quantity, total sales), and categorical variables (unit code). The selection of this data source was based on the availability of comprehensive data that are representative of retail product sales patterns.

The type of data used is secondary data in the form of time series sales transactions that have been structured in digital format. The data characteristics include diverse product variations with heterogeneous ranges of prices and sales volumes, providing sufficient complexity for testing the robustness of the Naïve Bayes method. Data validity was ensured through a verification process with the store management to confirm the accuracy and relevance of the data to actual business conditions.

Data Preprocessing

The data preprocessing stage was carried out systematically to ensure optimal dataset quality for the implementation of the Naïve Bayes method. This process includes three main stages, namely data cleaning, data selection, and data transformation, which were conducted sequentially to produce a structured dataset that meets the requirements of the classification model. Data cleaning was performed to identify and handle missing values, duplicate records, and inconsistencies in data format across all critical attributes. The examination was conducted on the columns Date, Menu, Vendor, Menu Price, Event, Order Quantity, and Amount Sold (Amount) using statistical profiling techniques, which indicated that there were no missing values or duplicates in the dataset.

Data selection was carried out based on relevance to the objective of predicting catering sales by selecting seven main attributes as input features for the model, namely the columns Date, Menu, Vendor, Menu Price, Event, Order Quantity, and Amount Sold (Amount). Data transformation was performed by converting the numeric attribute Sales Quantity into ordinal categories using four intervals: "Very Popular" for \leq Rp. 7,000,000, "Popular" for Rp 5,000,000 – Rp 6,999,999, "Fairly Popular" for Rp 3,000,000 – Rp 4,999,999, and "> Rp 3,000,000" "Less Popular". The transformation results produced a balanced class distribution with 189 data in the "Less Popular" category (19.12%), 343 data in the "Fairly Popular" category (34.71%), 234 data in the "Popular" category (23.68%), and 222 data in the "Very Popular" category (22.46%), supporting the optimal performance of the method in learning the pattern of each class.

Implementation of the Naïve Bayes Algorithm

The implementation of the Naïve Bayes algorithm begins with calculating the prior probability for each target class using the formula $P(C_i) = S_i/s$, where the calculation results show $P(\text{Very Popular}) = 0.224$, $P(\text{Popular}) = 0.236$, $P(\text{Fairly Popular}) = 0.347$, and $P(\text{Less Popular}) = 0.191$. The likelihood



probability calculation is performed for each combination of feature values and target classes using the formula $P(X|C) = n(X|C)/\Sigma(C)$, producing a comprehensive probability table for all input features. The classification process for the training data involves building a probabilistic model based on 790 training data, while 197 testing data are used to evaluate the model performance through prediction and comparison with the actual categories.

Validation of the algorithm implementation was conducted using the RapidMiner Studio application to ensure the consistency of the manual calculation results with the software implementation. The classification process produces predicted demand categories for each testing data based on the combination of prior and likelihood probabilities from all input features. The constructed model is then used to predict product demand by calculating the probability of each class and selecting the class with the highest probability as the final prediction result.

Model Evaluation

The evaluation of the model performance was carried out using accuracy and error rate metrics calculated based on the prediction results on the testing data. Accuracy was calculated using the formula $(\text{number of correct predictions})/(\text{number of testing data}) \times 100\%$, providing the percentage of the model's success in correctly classifying product demand categories. The error rate was calculated as the complement of accuracy using the formula $(\text{number of incorrect predictions})/(\text{number of testing data}) \times 100\%$, indicating the model's prediction error level. The evaluation process was conducted by comparing the model's prediction results with the actual labels in the testing data to calculate the number of correct and incorrect classifications.

The analysis of the evaluation results includes measuring the overall model performance as well as identifying patterns of prediction errors in each demand class. The evaluation implementation used a hold-out validation strategy to ensure the objectivity of the results and the generalizability of the model to new data. The evaluation results serve as the basis for assessing the effectiveness of the Naïve Bayes method in the context of retail product demand prediction and provide recommendations for practical implementation in business decision-making systems.

3. RESULT AND DISCUSSION

Implementation of Demand Forecasting Systems

The implementation of the catering sales prediction system was carried out using the Naïve Bayes algorithm by utilizing Microsoft Excel for data preprocessing and RapidMiner Studio version 9.10.001 as the main data mining platform. The system was built on hardware with specifications MSI Modern 14 B11MOU Processor Gen Intel(R) Core(TM) i3, Memory RAM 8.00 GB (7.71 GB usable), and ROM SSD 256 GB running on the Windows 10 operating system.

The dataset used in this study consists of 988 catering sales data from PT Negara Rasa Indonesia covering the last two years. The data contain the attributes Date, Menu, Vendor, Menu Price, Event, Order Quantity, and Amount Sold (Amount). All data have undergone a cleaning process to ensure that there are no missing values or duplicate data, so that all 988 data are declared valid and ready to be used in the classification process.

The dataset was divided with a composition of 80% as training data (790 data) and 20% as testing data (197 data) according to standard practice in machine learning. The labeling process was carried out by grouping products into four sales categories based on the number of sales: the category "Very Popular" for products with sales above $< \text{Rp } 7,000,000$, "Popular" for the range $\text{Rp } 5,000,000 - 6,999,999$, "Fairly Popular" for $\text{Rp } 3,000,000 - 4,999,999$, and "Less Popular" $> \text{Rp } 3,000,000$ and below. This produced a balanced class labeling with 189 data in the "Less Popular" category (19.12%), 343 data in the "Fairly Popular" category (34.71%), 234 data in the "Popular" category (23.68%), and 222 data in the "Very Popular" category (22.46%), supporting the optimal performance of the method in learning the pattern of each class.

Implementation Using RapidMiner

Preprocessing and Model Configuration



The implementation using RapidMiner was carried out through a series of systematically designed operators. The process began with importing the training and testing data, followed by the Change Type operator to convert attribute types into binomial or polynomial according to the characteristics of the data. The Change Role operator was used to assign the "kategori_penjualan" attribute as the classification target label.

The Naïve Bayes model was built using the Naive Bayes operator, which automatically calculates the prior and likelihood probabilities for each attribute. The probability calculation results produced by RapidMiner showed consistency with the manual calculations, thereby validating the accuracy of the algorithm implementation.

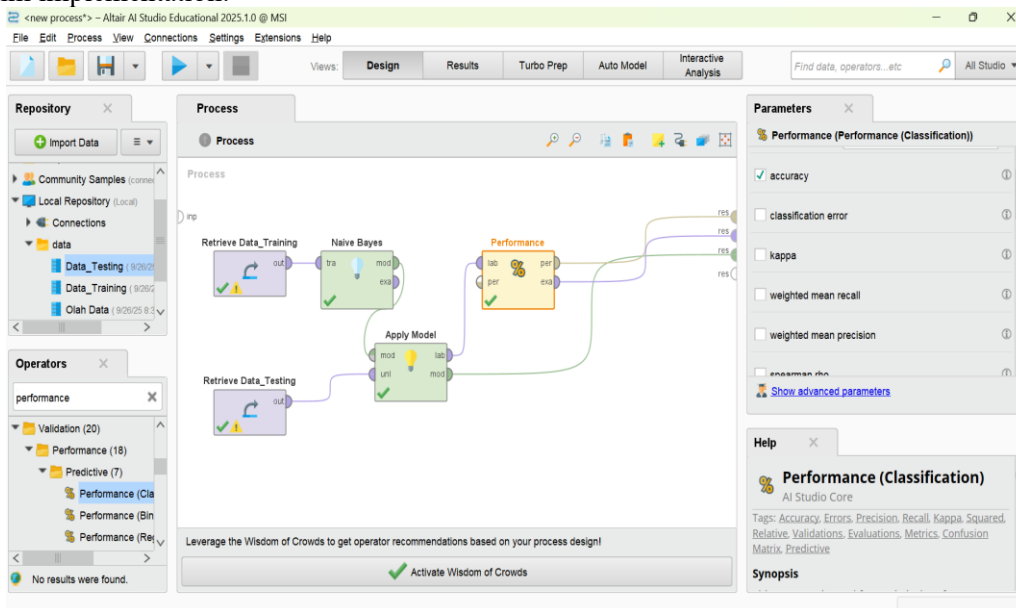


Figure 1. Rapidminer Operator Sequence

Table 1. Results from the Naive Bayes model

| Metric | Value | Interpretation |
|-----------|--------|---|
| Accuracy | 91.41% | The Naïve Bayes method has a high level of classification and prediction accuracy, showing that 91.41% of predictions match the actual data. |
| Precision | 91.42% | The Naïve Bayes method performs well in predicting positive classes; of all correctly predicted data, 91.42% are actually correct. |
| Recall | 92.72% | The Naïve Bayes method is able to recognize approximately 92.72% of the total actual positive data, demonstrating fairly good detection capabilities. |

Validity and Reliability Testing in Rapidminer

Validity testing is conducted to determine whether an instrument is considered valid or invalid in measuring a research variable [9]. Reliability testing is used to determine the consistency of a measurement instrument, specifically whether the measuring tool used is dependable and remains consistent when the measurement is repeated. A measuring instrument is considered reliable if it produces the same results even when measurements are conducted multiple times. Validity is a process used to measure the extent to which an instrument or variable truly measures what it is intended to measure [10]. In other words, data or questionnaires are considered valid if the questions, indicators, or variables used are relevant and aligned with the research objectives. In this study, the researcher predicts Sales Categories (Less Popular, Moderately Popular, Popular, Very Popular) based on variables such



as Menu Price, Event, and Number of Orders. Therefore, validity testing is carried out to determine whether each of these variables has a significant relationship with the sales category. If the results are significant and strong, the variables are considered valid to be used as predictors.

The evaluation measures how well and consistently the prediction model performs (accuracy, kappa, AUC, recall, and precision) in the RapidMiner application. Model validity testing is represented by the metrics of Accuracy, AUC, and Recall, while model reliability testing is represented by the Kappa value and the stability of Cross Validation results.

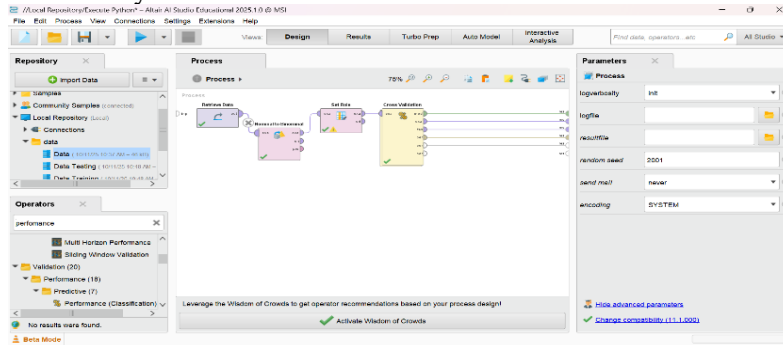


Figure 2. Order of Validity and Reliability Test Operators

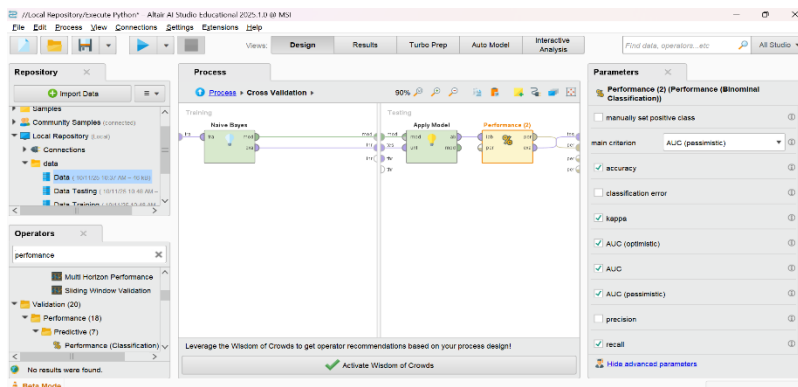


Figure 3. Order of Validity and Reliability Test Operators using the Naïve Bayes Method

Table 2. Results of the Validity and Reliability Test of the Naive Bayes Method

| Metric | Value | Interpretation |
|-----------|-----------------|--|
| Accuracy | 85.03% ± 3.27% | The Naïve Bayes method has a high level of classification and prediction accuracy, showing that 85% of predictions match the actual data. |
| Precision | 74.86% ± 9.24% | The Naïve Bayes method has good accuracy in predicting positive classes; of all correctly predicted data, 75% are actually correct. |
| Recall | 57.32% ± 10.39% | The Naïve Bayes method is able to recognize about 57% of the total actual positive data, indicating fairly good detection capabilities. |
| AUC | 0.897 ± 0.049 | The AUC value is close to 1, indicating excellent classification capabilities and that the Naïve Bayes method is able to distinguish classes well. |



4. CONCLUSION

Based on the results of the experimental implementation of the Naïve Bayes method in predicting the level of catering sales based on historical data at PT. Negara Rasa Indonesia, the conclusions that the author can present from this research are as follows: By using the Naïve Bayes method, this method has proven to be applicable for conducting prediction processes at PT. Negara Rasa Indonesia. In addition, the validity and reliability tests produced consistent and accurate values. Based on the experimental results, the classification of catering sales categories using the Naïve Bayes method achieved an average value of 91.42%, with an accuracy rate of 91.41%, a validity value of 85.03% with a deviation of $\pm 3.27\%$, and a reliability value of 0.550 ± 0.095 .

From a practical perspective, the implementation of this prediction system provides a concrete solution to address sales management problems that have been faced by PT. Negara Rasa Indonesia, particularly in reducing the risk of less popular catering products that could negatively impact business operations. This study demonstrates that the Naïve Bayes method, with its simplicity of implementation and computational efficiency, can serve as a viable alternative to conventional methods in strategic decision-making in the retail sector. The stability of historical data patterns, which has been proven consistent between the two-year period and the most recent year, indicates that the model can be relied upon for long-term predictions, providing a strong foundation for the digital transformation of inventory management systems in the MSME retail sector.

5. REFERENCES

- S. Alam and S. Sunardi, "Analisis Prediksi Penjualan Kue Menggunakan Metode Naive Bayes," *SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, vol. XII, no. 2, pp. 54–67, 2023.
- N. Ayuningtyas, N. R., and F. M. Basysyar, "Penerapan Data Mining Pada Penjualan Produk MS Glow Menggunakan Metode Naive Bayes Untuk Strategi Pemasaran," *Jurnal Accounting Information System (AIMS)*, vol. 5, no. 2, pp. 157–166, 2022, doi:10.32627/aims.v5i2.503.
- M. Damara S. Deifa, I. F. Nur, and J. Sahertian, "Sistem Prediksi Minat Penjualan Jaket di Grosir Murah Kediri Menggunakan Metode Naive Bayes," *Seminar Nasional Inovasi Teknologi*, vol. 1, no. 1, pp. 310–314, 2021.
- F. Harahap, W. Fahrozi, R. Adawiyah, E. T. Siregar, and A. Y. N. Harahap, "Implementasi Data Mining Dalam Memprediksi Produk AC Terlaris Untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes," *Jurnal Unitek*, vol. 16, no. 1, pp. 41–51, 2023, doi:10.52072/unitek.v16i1.541.
- A. P. Husaini and A. Lisdiyanto, "Sistem Prediksi Penjualan Produk APD Terlaris di PT A3 Karunia Sidoarjo Menggunakan Metode Naive Bayes," *Jurnal Teknologi dan Sistem Informasi Bisnis*, vol. 6, no. 2, pp. 431–437, 2024, doi:10.47233/jteksis.v6i2.1266.
- M. J. Saputra and M. I. Herdiansyah, "Penerapan Naive Bayes Dalam Memprediksi Penjualan dan Persediaan Kain Jumputan Pada Toko Batiq Colet Tuan Kentang Palembang," *Jurnal Mantik*, vol. 6, no. 2, pp. 2502–2507, 2022.
- E. U. Putri, S. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penjualan Pestisida Pada CV Mitra Artha Sejati Menggunakan Algoritma Naive Bayes," *Kesatria: Jurnal Penerapan Sistem Informasi dan Manajemen*, vol. 2, no. 1, pp. 39–46, 2021.
- M. B. Hajiboyo and A. Mansah, *Manajemen Usaha Catering*. Yogyakarta: Deepublish, 2020.
- R. Slamet and S. Wahyuningsih, *Metodologi Penelitian Kuantitatif*. Yogyakarta: Deepublish, 2023.
- Nursalam, *Metodologi Penelitian Ilmu Keperawatan: Pendekatan Praktis*, 6th ed. Jakarta: Salemba Medika, 2023.