International Journal of Electrical Engineering, Mathematics and Computer Science
E-ISSN: 3048-1945
P-ISSN: 3048-1910

Research Article

Classification of Sales of Best-Selling Products in Ira Store Using Naive Bayes Algorithm and K-Nearest Neighbor Algorithm

Yuma Akbar [1], Kiki Setiawan [2], Muhammad Joko Umbaran Kharis Bahrudin [3], Intan Purwasih [4,*]

1 Sekolah Tinggi Ilmukomputer Cipta Karya Informatika (Stikomcki) Jakarta
2 Sekolah Tinggi Ilmukomputer Cipta Karya Informatika (Stikomcki) Jakarta
3 Sekolah Tinggi Ilmukomputer Cipta Karya Informatika (Stikomcki) Jakarta
4 Sekolah Tinggi Ilmukomputer Cipta Karya Informatika (Stikomcki) Jakarta
* Corresponding Author: e-mail: intanpurwasih780@gmail.com

Abstract: In today's world of retail and technology, competition is fiercely competitive. With the development of retail businesses increasing in number and mushrooming in a region, consumer needs are increasing, and retail business players are competing to develop their businesses by utilizing existing technology. Daily sales transaction data continues to increase, causing a lot of storage. Toko Ira has more than 228 sales transaction data records from 2023 to 2024 that have not been used. Data requires a lot of storage space. Additionally, the data has not been used in an effective way. Based on this problem, this research aims to use data mining to classify sales transaction data to determine which items are selling best. This research is a case study with a qualitative approach. This research was conducted with the Naive Bayes method and Rapidminer was used. The results of the sales transaction data classification research are the division of products into best-selling and non-selling categories. The results of this research show that the K-Nearest Neighbors (KNN) algorithm with a 50:50 data division is more effective in predicting and classifying sales of best-selling and non-selling products in IRA stores. The results show that the Naive Bayes algorithm has an accuracy of 89.91%, while the K-Nearest Neighbors (KNN) algorithm has an accuracy of 60.09%.

Keywords: Best-Selling Products, Data Mining, K-Nearest Neighbors, Naive Bayes, Sales Classification.

Received: October 10, 2024
Revised: October 30, 2024
Accepted: November 14, 2024
Published: December 16, 2024
Curr. Ver.: December 28, 2024

Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

1. Introduction

In an increasingly competitive retail environment, the ability of small culinary businesses to understand consumer preferences has become an important social and managerial issue. Food businesses such as Kedai Ira do not only compete through taste and price, but also through speed of service, product availability, and the ability to identify menus that are most demanded by customers. In practice, many micro and small enterprises still rely on manual ordering and sales recording, which may cause recording errors, slow service, overlapping orders, and limited use of sales data for decision-making. These conditions show that sales data are not merely administrative records, but social facts that reflect purchasing behavior, customer preferences, and market trends in daily transactions. Therefore, the classification of best-selling and less-selling products becomes relevant to support better stock planning, marketing strategies, and customer satisfaction. Data mining provides a systematic way to transform historical transaction data into useful knowledge for business development and operational improvement [1], [2].

Previous studies have shown that data mining methods, especially Naive Bayes, are widely used to classify best-selling products in various business contexts. [3] applied Naive Bayes to classify best-selling retail products and emphasized its usefulness for stock planning and retail management. Similarly, [4] used Naive Bayes to predict best-selling electronic products, helping companies distinguish between high-demand and low-demand items. Other studies also applied similar classification approaches to brownies tape, food menus, phone credit products, electrical equipment, and provider sales, indicating that sales classification has become a relevant topic in business data analysis. However, most previous studies focused on general retail, electronics, or non-culinary product categories, while studies on small culinary businesses with manual ordering systems remain limited. Therefore, the novelty of this study lies in its focus on Kedai Ira, a culinary business, by comparing Naive Bayes and K-Nearest Neighbor to identify best-selling and less-selling products from historical sales data and practical ordering problems.

Based on these problems, this study aims to analyze the use of data mining in classifying product sales at Kedai Ira into best-selling and less-selling categories. The study specifically examines whether Naive Bayes and K-Nearest Neighbor can be applied to support the identification of products with higher sales potential. The research also seeks to determine which algorithm produces better accuracy in classifying sales data, so the results can be used as a basis for business decision-making, stock control, and service improvement. In this context, the main research questions are: how does manual ordering affect the efficiency and accuracy of order management at Kedai Ira, and can Naive Bayes and K-Nearest Neighbor help solve the problem of identifying best-selling products? The use of these algorithms is relevant because previous studies have demonstrated their ability to classify sales data and support prediction tasks in different business and commercial domains, including retail and product demand analysis [5], [6].

This study argues that sales classification can help small culinary businesses move from intuition-based decisions toward data-driven management. By processing sales transaction data, Kedai Ira can identify product demand patterns, reduce the risk of inaccurate stock decisions, and design more effective sales strategies. The comparison between Naive Bayes and K-Nearest Neighbor is also important because each algorithm has different characteristics in handling classification problems. Naive Bayes is often considered simple and efficient for probabilistic classification, while K-Nearest Neighbor classifies data based on similarity patterns among observations. Therefore, comparing both methods provides a clearer understanding of which algorithm is more suitable for the sales data context of Kedai Ira. The contribution of this research is practical and academic: practically, it supports business owners in optimizing stock and sales strategies; academically, it enriches studies on sales classification in small culinary enterprises using comparative data mining methods, measurable accuracy evaluation, and real transaction-based evidence for decision-making [7], [8].

2. Literature Review

Systematic Literature Review

Table 1. *Review PICOC.*

Classification of Best-Selling Products on Ira Stores Using the Naive Bayes Algorithm	
Population	Classification of Data Mining of the Best-Selling Products at Kedai Ira Using the Naive Bayes Method
Intervention	Recording errors, delays in order processing, and lack of structured data are some of the obstacles that are often faced.
Comparation	n/a
Outcomes	Classification of Bestsellers and Non-Sellers and Accuracy Rates of Naive Bayes Modeling
Context	Dataset Private

A Systematic Literature Review (SLR) is an organized method used to identify, evaluate, and interpret relevant studies related to a specific research topic. In this study, the SLR approach is used to strengthen the theoretical foundation of product sales classification using data mining methods. The review process is guided by the PICOC framework, which consists of Population, Intervention, Comparison, Outcomes, and Context. The population in this study refers to data mining classification of best-selling products at Kedai Ira. The intervention focuses on problems such as recording errors, delayed order processing, and the lack of structured sales data. The outcome is the classification of products into best-selling and less-selling categories, along with the accuracy level produced by the model. This approach is relevant because previous studies have shown that data mining can support product classification and assist business owners in making more accurate sales decisions [3], [4].

The literature selection process in this study was carried out by reviewing scientific journals and books related to data mining, classification, best-selling product analysis, Naive Bayes, and K-Nearest Neighbor. The review protocol included publications from journals and books published within the relevant time range, with keywords such as “data mining,” “classification,” “best-selling products,” and “Naive Bayes.” This selection strategy helps ensure that the sources used are directly related to the research problem and support the development of a suitable classification model. Previous studies have applied Naive Bayes to classify sales data in various contexts, such as brownies tape products and favorite food menus, showing that the method is suitable for analyzing consumer demand patterns. Therefore, the SLR process in this research is not only used to collect references but also to identify research gaps, compare methods, and position the Kedai Ira case within previous sales classification studies [9], [10].

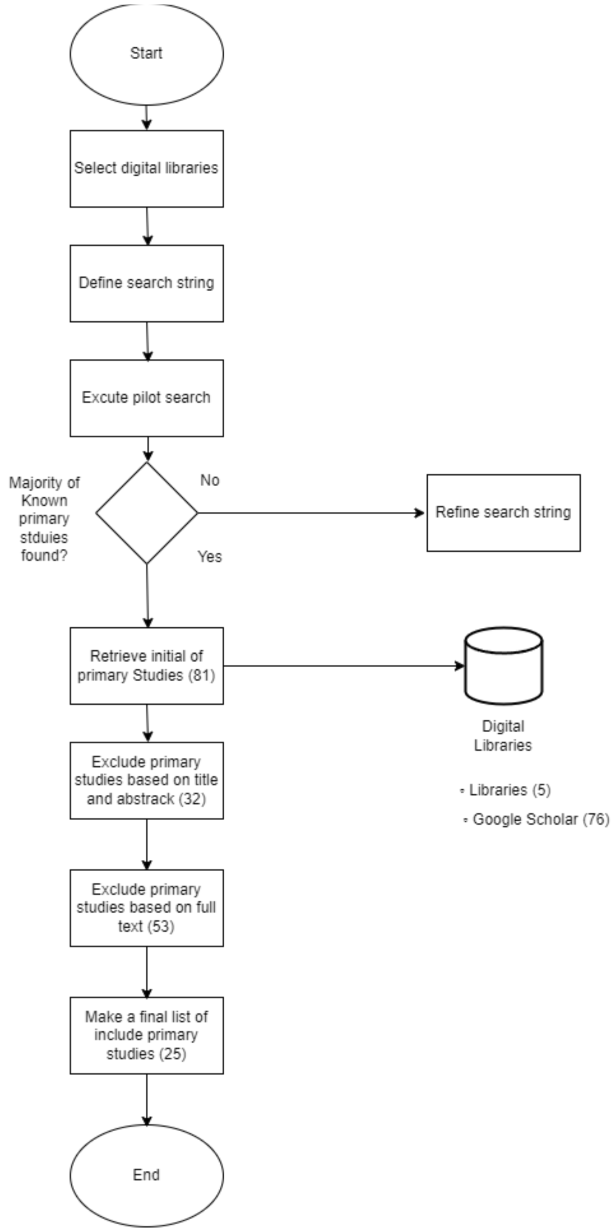


Figure 1. Study Selection Strategy.

Business, Sales, and Productivity

Business refers to activities carried out by individuals or groups to produce, sell, or exchange goods and services with the aim of gaining profit. In the culinary sector, business success depends not only on product quality but also on service efficiency, customer satisfaction, and the ability to understand market demand. Kedai Ira, as a small culinary business, needs to identify which products are most preferred by customers in order to maintain competitiveness. Best-selling products are important because they reflect consumer preferences and can be used as a basis for stock planning and marketing strategies. In data-driven business management, sales transaction records can be transformed into valuable information through data mining. This supports business owners in making decisions based on actual sales patterns rather than intuition alone. Studies on retail and small business sales classification show that data mining helps identify product demand and supports better business planning [8], [11].

Sales are a central activity in business because they directly influence revenue, profit, and business sustainability. Sales activities include creating demand, offering products, negotiating transactions, and delivering products to customers. In small culinary businesses, sales performance is closely related to product availability, order accuracy, and customer satisfaction. If sales records are still managed manually, errors may occur in recording orders, identifying popular products, and preparing stock. This condition can reduce productivity and affect business performance. Productivity in sales refers to the ability of a business to generate optimal sales results by using available resources efficiently. A productive sales system allows business owners to understand which products contribute most to revenue and which products need evaluation. Previous studies have demonstrated that classification models can be used to analyze sales levels, optimize inventory, and improve marketing strategies in various business sectors [12], [13].

Definition of Data Mining

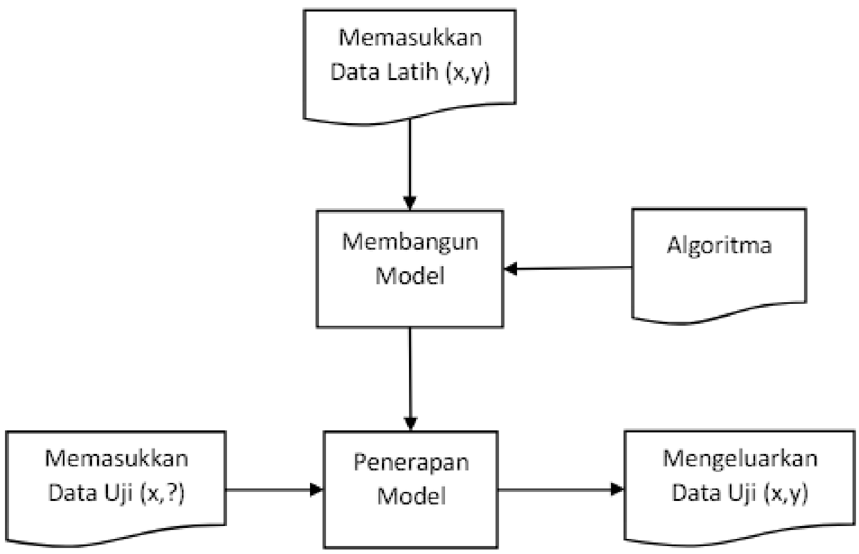


Figure 2. Classification Process.

Data mining is the process of discovering meaningful patterns, relationships, and knowledge from large or structured datasets. In the context of this study, data mining is used to process Kedai Ira's sales transaction data so that products can be classified into best-selling and less-selling categories. Classification is one of the important tasks in data mining, where a model is built from training data and then used to predict the class of new data. In sales analysis, classification helps business owners understand product performance based on historical transaction patterns. Several algorithms can be used for classification, including Naive Bayes, Decision Tree, K-Nearest Neighbor, and Artificial Neural Network. Each algorithm has its own approach to learning from data and producing predictions. Data mining is relevant for Kedai Ira because it can convert manual sales records into useful information for stock planning, product evaluation, and sales strategy development [14], [15].

Definition of CRISP-DM

The diagram below shows the six stages of CRISP DM: which are listed in table 3:

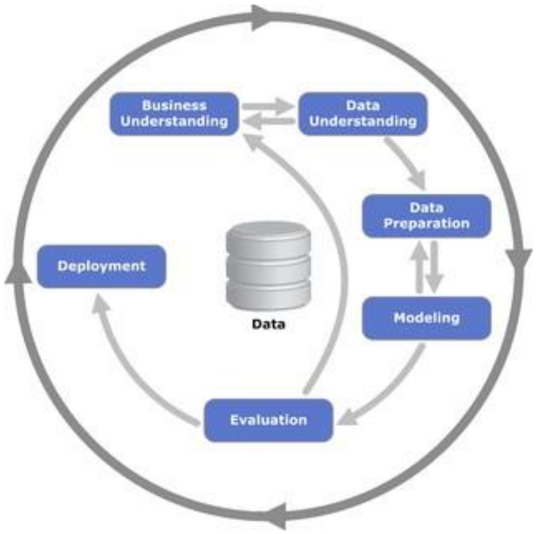
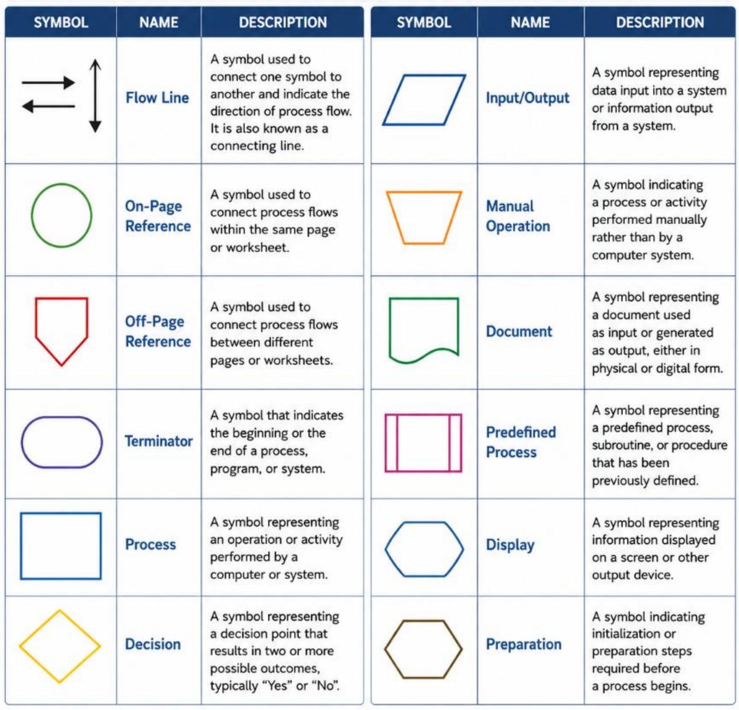


Figure 3. CRISP-DM Schema.

CRISP-DM, or Cross-Industry Standard Process for Data Mining, is a structured framework commonly used to guide data mining projects. It consists of several stages, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. In this study, CRISP-DM is relevant because it provides a systematic process for analyzing sales data at Kedai Ira. The first stage begins with understanding the business problem, namely the need to identify best-selling and less-selling products. The next stage involves understanding and preparing sales data so that it can be processed by classification algorithms. After that, the Naive Bayes and K-Nearest Neighbor models can be built and evaluated based on accuracy results. The use of CRISP-DM helps ensure that data analysis is not conducted randomly, but follows clear and measurable steps. This framework is also suitable for practical data mining applications in industry, services, and small business contexts [16], [17].

Definition of Flowchart

Shown in the table Figure 4:



SYMBOL	NAME	DESCRIPTION	SYMBOL	NAME	DESCRIPTION
	Flow Line	A symbol used to connect one symbol to another and indicate the direction of process flow. It is also known as a connecting line.		Input/Output	A symbol representing data input into a system or information output from a system.
	On-Page Reference	A symbol used to connect process flows within the same page or worksheet.		Manual Operation	A symbol indicating a process or activity performed manually rather than by a computer system.
	Off-Page Reference	A symbol used to connect process flows between different pages or worksheets.		Document	A symbol representing a document used as input or generated as output, either in physical or digital form.
	Terminator	A symbol that indicates the beginning or the end of a process, program, or system.		Predefined Process	A symbol representing a predefined process, subroutine, or procedure that has been previously defined.
	Process	A symbol representing an operation or activity performed by a computer or system.		Display	A symbol representing information displayed on a screen or other output device.
	Decision	A symbol representing a decision point that results in two or more possible outcomes, typically "Yes" or "No".		Preparation	A symbol indicating initialization or preparation steps required before a process begins.

Figure 4. CRISP-DM Scheme.

A flowchart is a visual representation used to describe the sequence of steps in a process or system. It uses standard symbols to represent activities, decisions, inputs, outputs, and process flows. In data mining research, a flowchart is useful for explaining the stages of data processing, starting from data collection, data cleaning, data preparation, model testing, and evaluation of classification results. For this study, the flowchart can help readers understand how sales data from Kedai Ira are processed using Naive Bayes and K-Nearest Neighbor. Compared with textual explanations, flowcharts make the research procedure easier to understand because they show the logical order of each activity. Flowcharts are also useful in system design because they clarify how data move from one stage to another. In sales classification studies, process visualization supports methodological transparency and helps ensure that the classification procedure is systematic, repeatable, and aligned with the research objectives [15], [18].

3. Materials and Method

Research Data

The data used in this study are qualitative in nature and consist of information obtained from observations, respondents, and supporting documents collected for research purposes. The research was conducted through several stages. The first stage was observation, during which the researcher directly observed business activities and sales processes at Kedai Ira to gain a comprehensive understanding of the operational conditions and challenges faced by the store. Following the observation stage, data collection was carried out in May 2024 with the assistance of the store owner and sales staff. The collected dataset contains 228 sales transaction records and is categorized as a private dataset. The variables included in the dataset are transaction date, transaction time, menu item, quantity sold, price, and product category. All data were recorded and stored in a notebook before being organized for analysis. The qualitative data in this research consist of descriptive information related to sales activities and product characteristics that cannot be directly measured using numerical values but can be observed, interpreted, and analyzed to support the classification of best-selling and less-selling products. The structure of the dataset is presented in Table 2.

Table 2. Initial Data Attributes.

Attribute	Remarks
Date	On what date of purchase.
Jam	Time of purchase.
Menu	Pilihan Makanan.
Qty	The amount purchased.
Pricing	Price paid.
Category	Selling and Not Selling Out.

Methodology Implementation

This study classifies the sales status of products at Kedai Ira by applying the Naive Bayes classification method. The research process began with data collection from sales transaction records, followed by data cleaning and data transformation stages. Data cleaning was performed to remove incomplete, inconsistent, or duplicate records that could affect the quality of the analysis. Afterward, data transformation was conducted to convert and organize the data into a suitable format for the classification process and to ensure that the dataset met the requirements of the analytical model. To process and analyze the data systematically, this study adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. CRISP-DM consists of six main phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. These phases provide a structured approach for transforming raw sales data into meaningful information and knowledge. Through the implementation of the CRISP-DM framework and the Naive Bayes algorithm, the study aims to classify products into best-selling and less-selling categories while evaluating the performance of the classification model based on its accuracy and predictive capability. The results are expected to support more effective sales management and decision-making at Kedai Ira.

The following are the stages of the research methodology used can be seen in figure 5.

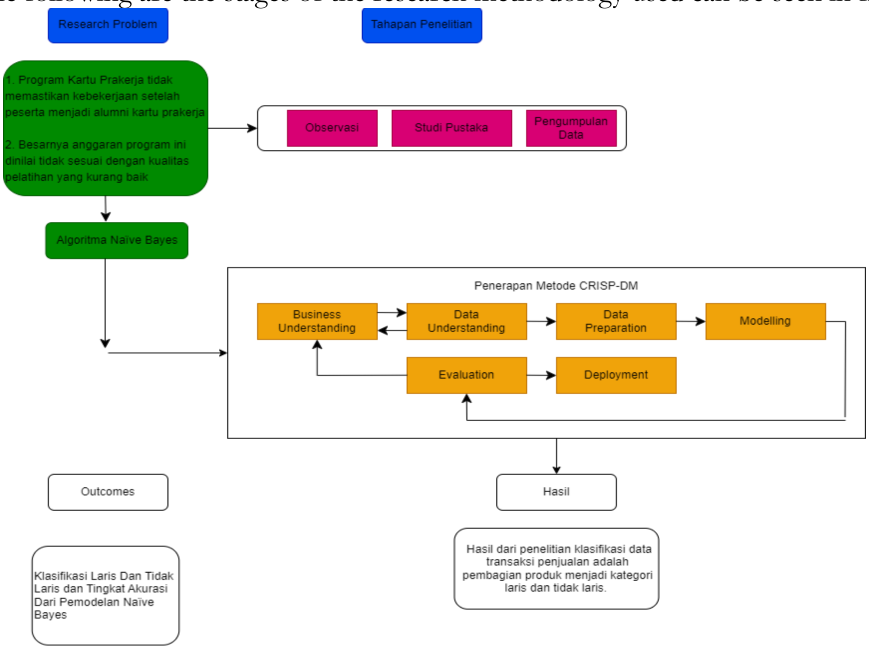


Figure 5. Stages of Methodology Application.

Research Design

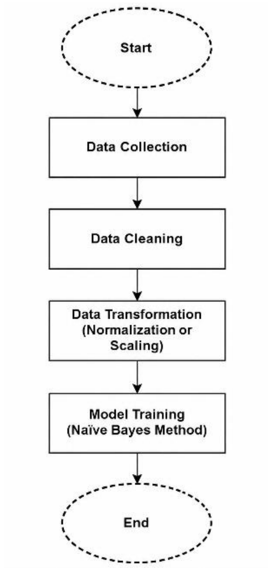


Figure 6. Stages of Classification Research.

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework as the primary research methodology because it provides a structured and systematic approach for conducting data mining projects. The research process follows six main phases of CRISP-DM: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. During the Business Understanding phase, the research objectives and business problems are clearly identified. The main objective of this study is to classify product sales at Kedai Ira using the Naive Bayes and K-Nearest Neighbor algorithms and to evaluate their classification performance. The expected outcome is to determine the most suitable algorithm for identifying best-selling and less-selling products. In the Data Understanding phase, the researcher examines the characteristics of the dataset obtained from Kedai Ira. The dataset consists of 228 sales transaction records with six attributes: Date, Time, Menu, Quantity, Price, and Category. This phase provides an analytical foundation by exploring the dataset, identifying potential issues, and preparing the information required for further processing and model development.

The next stage is Data Preparation, which involves transforming raw transaction data into a format suitable for classification analysis. This phase includes data collection, data cleaning, and data transformation. Data collection was conducted using sales transaction records obtained directly from Kedai Ira. Subsequently, data cleaning was performed to identify and remove inaccurate, incomplete, duplicated, or inconsistent records that could negatively affect the classification results. Data transformation was then carried out to organize the dataset into a structured format that could be processed efficiently by the selected algorithms. This stage is particularly important because the quality of the classification model depends heavily on the quality of the prepared data. Through these preparation activities, the dataset becomes more reliable and suitable for analytical processing. The final output of this phase is a clean and structured dataset that can be used for training and testing classification models. Proper data preparation also reduces the possibility of bias and improves the overall performance of the predictive model.

The Modeling phase applies two classification algorithms, namely Naive Bayes and K-Nearest Neighbor (KNN), to classify products into best-selling and less-selling categories. The dataset is divided into training and testing data to build and evaluate the classification models. After the models are developed, the Evaluation phase is conducted using a confusion matrix to measure model performance. The evaluation process generates values such as True Positive, True Negative, False Positive, and False Negative, which are used to calculate Accuracy, Precision, and Recall. These performance metrics help determine how effectively each algorithm predicts product sales categories based on historical transaction data. Finally, in the Deployment phase, the knowledge and findings generated from the analysis are documented and prepared for practical use. The research results are presented in the form of reports and scientific publications, enabling business owners to utilize the classification outcomes for decision-making, inventory planning, and sales strategy improvement. The deployment stage ensures that the knowledge extracted from the data can provide practical benefits and support future business development.

4. Results and Discussion

Research Tools

To support the research, tools are needed. Software and hardware are the research tools used.

Software

Table 3. Indicates the software version, tools, and functions. Table 3.

Table 3. Software Specifications.

No	*Sofware*	*Versi*	Fungsi
1	Microsoft Excel	2286	To identify the feelings resulting from the results of Twitter data crawling to be put into the training data and data storage.
2	Rapidminer Studio	10.0	To implement the text mining method.
3	Windows	10 Pro	For business use, with features that can improve security and productivity.

Hardware

Table 4. indicates the hardware tool, its type, and its specifications.

Table 4. Hardware Specifications.

Jenis *Hardware*	Spesifikasi
Model	HP Notebook
Processor	Intel Celeron N3060
Memory	DDR3 4Gb
GPU	Graphics HD
Storage	500Gb HDD
Display	14 inch High Definition (HD)

Implementation and Testing

Below is a series of implementations and tests of this research that have been carried out:

Research Implementation

The data used in this study consist of sales transaction records obtained from Kedai Ira. These data were processed through several stages, including data correction, integration, selection, and preparation to ensure their suitability for classification analysis. The research implementation was conducted using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which provides a systematic approach for data mining projects. The first stage, Business Understanding, focused on identifying the research objectives and understanding the available sales data. The original sales records were collected manually and later converted into Microsoft Excel spreadsheets. The dataset contains six attributes: Date, Time, Menu, Quantity (Qty), Price, and Category, with a total of 228 transaction records. The second stage, Data Understanding, involved examining the characteristics and quality of the collected data. Sales information was initially recorded in the store's transaction book and then transferred into Excel format for analysis. During this stage, the data were identified, selected, and cleaned to ensure that only relevant attributes were retained for the classification process. Figures 7 and 8 present examples of the original handwritten sales records and the corresponding Excel-based dataset used in this research.

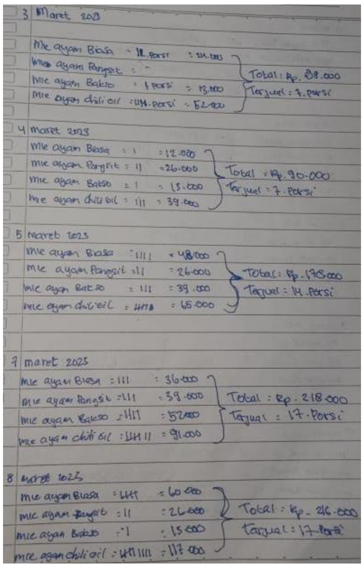


Figure 7. Raw Data on Sales at Kedai Ira.

1	Tanggal	Jam	Menu	Qty	Harga	Kategori
2	03/02/2023	11:00	Mie ayam pangsit	24	306.000	Laris
3	04/02/2023	11:00	Mie ayam bakso	39	497.000	Laris
4	05/02/2023	11:00	Mie ayam chili oil	67	859.000	Laris
5	14/02/2023	11:00	Mie ayam chili oil	38	491.000	Laris
6	15/02/2023	11:00	Mie ayam bakso	20	258.000	Laris
7	16/02/2023	11:00	Mie ayam pangsit	32	411.00	Laris
8	17/02/2023	11:00	mie ayam spesial	20	400.000	Laris
9	18/02/2023	11:30	Sate usus	25	125.000	Laris
10	01/03/2023	11:30	Chesee Roll	5	50.000	Tidak Laris
11	02/03/2023	11:30	Mie ayam chili oil	39	506.000	Laris
12	03/03/2023	11:30	Es tee	20	100.000	Laris
13	04/03/2023	11:30	Leci tea	7	84.000	Tidak Laris
14	05/03/2023	11:30	Mie ayam pangsit	14	178.000	Kadang Tidak Laris
15	07/03/2023	11:30	Mie ayam bakso	17	218.000	Tidak Laris
16	08/03/2023	11:30	mie ayam spesial	17	340.000	Tidak Laris
17	09/03/2023	11:00	Es jeruk	10	50.000	Tidak Laris
18	10/03/2023	11:00	Soda gembira	14	139.000	Kadang Tidak Laris
19	11/03/2023	11:00	Mie ayam chili oil	20	260.000	Laris
20	14/03/2023	11:00	Mie ayam bakso	9	117.00	Tidak Laris
21	15/03/2023	11:00	mie ayam polos	10	120.000	Kadang Tidak Laris
22	16/03/2023	11:00	Es campur	8	96.000	Tidak Laris

Figure 8. Excel Input Data.

The research process was conducted following the CRISP-DM framework, beginning with the Data Understanding phase, where sales transaction records were collected directly from Kedai Ira. A total of 228 transaction records were gathered and analyzed, consisting of six main attributes: Date, Time, Menu, Quantity (Qty), Price, and Category. After understanding the characteristics of the dataset, the Data Preparation phase was carried out through data selection, preprocessing, and transformation to ensure that the data were accurate, consistent, and suitable for classification analysis. The Modeling phase involved applying two classification algorithms, namely K-Nearest Neighbor (KNN) and Naive Bayes, using RapidMiner software. The models were configured, trained, and tested using the prepared dataset to classify products into best-selling and less-selling categories. Subsequently, the Evaluation phase was performed using a confusion matrix to measure model performance based on Accuracy, Precision, and Recall metrics. Cross-validation was also applied to ensure model reliability. Finally, the Deployment phase involved exporting and preparing the best-performing model for practical implementation, enabling Kedai Ira to support product classification, inventory management, and sales decision-making based on historical transaction data.

Research Testing

To ensure that the system can meet standards and reduce errors or errors, system testing is performed. To test this research, researchers have tested the data in Excel format. The process of classifying the K-Nearest Neighbors and Naive Bayes algorithm on the Rapidminer application is as follows:

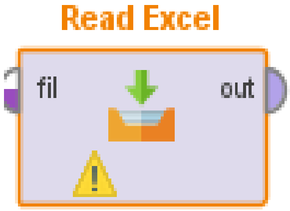


Figure 9. Test Data Import.

The implementation of the classification model in RapidMiner began by creating a new project using the Blank Process option. The sales transaction dataset stored in Microsoft Excel format was then imported into RapidMiner using the Read Excel operator. During the import process, the data attributes were configured, and the target attribute was assigned as the class label for classification purposes. After the dataset was successfully loaded, the Cross Validation operator was added to the process panel and connected to the imported data to enable model training and testing. The cross-validation subprocess was then opened to configure the training and testing stages. In the training section, the K-Nearest Neighbor (KNN) and Naive Bayes operators were inserted to build the classification models. In the testing section, the Apply Model and Performance operators were added to evaluate model performance. These operators were connected appropriately to ensure that the trained models could generate predictions and calculate evaluation metrics. Finally, the process was executed by clicking the Run icon on the toolbar, allowing RapidMiner to generate classification results and performance measurements such as Accuracy, Precision, and Recall for both algorithms.

Data Processing Using K-Nearest Neighbor and Naive Bayes Algorithms in RapidMiner

The data processing stage in RapidMiner began by importing the prepared Excel dataset into the application. During the import process, each attribute and data type was defined according to its role in the classification model. The target attribute was assigned as the label, which served as the classification objective, while the remaining attributes were designated as predictive variables. After the dataset configuration was completed, the data were stored in the Repository and then dragged into the Process View workspace. Next, the Cross Validation operator was added to facilitate model training and testing using the same dataset. By opening the Cross Validation operator, separate training and testing subprocesses were automatically generated. In the training section, the K-Nearest Neighbor (KNN) and Naive Bayes algorithms were inserted to build classification models based on historical sales data. In the testing section, the Apply Model and Performance operators were added to evaluate model performance. All operators were connected through their respective ports to ensure proper data flow. Finally, the process was executed using the Run button, enabling RapidMiner to calculate classification results and generate performance metrics for both algorithms. Figures 10 and 11 illustrate the configuration and implementation of the KNN and Naive Bayes classification models used in this study.

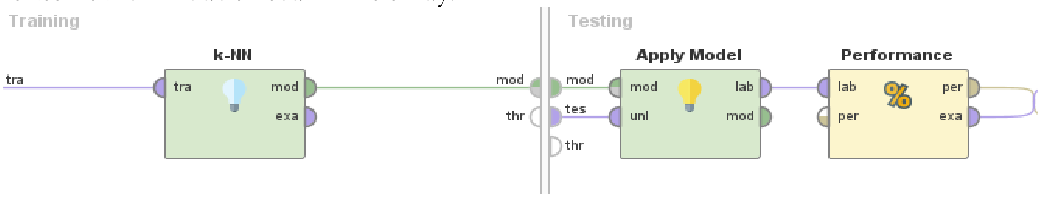


Figure 10. Model K-Nearest Neighbor.

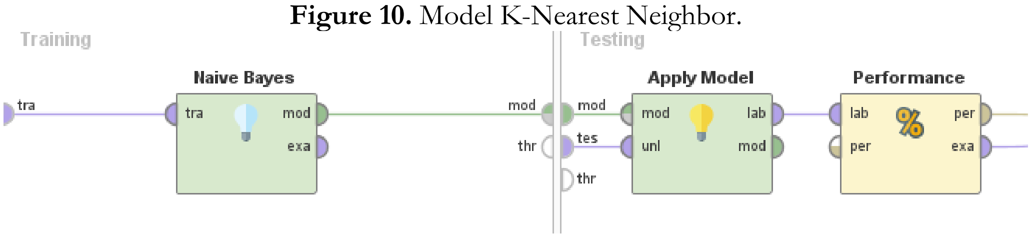


Figure 11. Model Naive Bayes.

The figure below shows the results of the accuracy of the classification process of the K-Nearest Neighbors and Naive Bayes algorithm after a while: Figure 12 and Figure 13

accuracy: 60.18% +/- 10.36% (micro average: 60.09%)

	true Laris	true Tidak Laris	true Kadang Tidak ...	true Laris	class precision
pred. Laris	33	13	3	1	66.00%
pred. Tidak Laris	28	90	29	0	61.22%
pred. Kadang Tida...	2	15	14	0	45.16%
pred. Laris	0	0	0	0	0.00%
class recall	52.38%	76.27%	30.43%	0.00%	

Figure 12. Accuracy Algoritma K-Nearest Neighbor.

accuracy: 89.94% +/- 3.53% (micro average: 89.91%)

	true Laris	true Tidak Laris	true Kadang Tidak ...	true Laris	class precision
pred. Laris	62	3	6	1	86.11%
pred. Tidak Laris	0	109	6	0	94.78%
pred. Kadang Tida...	1	6	34	0	82.93%
pred. Laris	0	0	0	0	0.00%
class recall	98.41%	92.37%	73.91%	0.00%	

Figure 13. Accuracy Algoritma Naive Bayes.

The RapidMiner testing results demonstrated the classification performance of the K-Nearest Neighbor (KNN) and Naive Bayes algorithms using 228 sales transaction records collected from Kedai Ira. The results showed that the Naive Bayes algorithm achieved an accuracy of 89.91%, significantly outperforming the K-Nearest Neighbor algorithm, which achieved an accuracy of 60.09%. The findings indicate that Naive Bayes was more effective in classifying product sales data, particularly when additional attributes were included in the dataset. Although KNN produced lower overall accuracy, it still demonstrated satisfactory classification performance under certain testing conditions. Based on the evaluation results, Naive Bayes proved to be the more reliable algorithm for predicting and classifying best-selling and less-selling products within the Kedai Ira sales dataset.

5. Conclusion

This study successfully applied data mining techniques to classify product sales at Kedai Ira using the K-Nearest Neighbor (KNN) and Naive Bayes algorithms. The analysis was conducted on a private dataset consisting of 228 sales transaction records containing information on date, time, menu items, quantity, price, and category. The results revealed that products could be effectively classified into best-selling and less-selling categories, providing valuable insights into sales patterns and customer preferences. Based on the evaluation results obtained through RapidMiner, the Naive Bayes algorithm achieved an accuracy of 89.91%, while the K-Nearest Neighbor algorithm achieved an accuracy of 60.09%. These findings indicate that Naive Bayes performed significantly better in classifying sales data and was more suitable for identifying product sales categories within the Kedai Ira dataset. The study demonstrates that data mining can support decision-making by helping business owners monitor sales trends, optimize inventory management, and improve sales strategies. Overall, the implementation of classification algorithms provides an effective approach for transforming sales transaction data into useful business knowledge.

References

[1] R. Fitriana, *Data Mining dan Aplikasinya: Contoh Kasus di Industri Manufaktur dan Jasa*. Wawasan Ilmu, 2022.

[2] I. G. M. D. I. N. S. W. W. G. A. Pradnyana, *Data Mining: Menemukan Pengetahuan dalam Data*. Depok, 2020.

[3] N. W. Wardani, P. G. S. C. Nugraha, and G. S. Mahendra, "Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel," *J. Sains dan Teknol.*, vol. 12, no. 3, 2024, doi:

10.23887/jstundiksha.v12i3.38605.

[4] N. Pransiska and A. Mirza, "Penerapan Data Mining Prediksi Penjualan Barang Elektronik Terlaris Menggunakan Algoritma Naïve Bayes (Studi Kasus: Planet Cash and Credit Cabang Muara Enim)," in *Bina Darma Conference on Computer Science*, 2023.

[5] N. N. F. Adzani, W. Witanti, and F. R. Umbara, "Klasifikasi Tingkat Penjualan Video Game Dengan Menggunakan Metode K-Nearest Neighbors," *INFOTECH J.*, vol. 9, no. 2, pp. 618–625, 2023, doi: 10.31949/infotech.v9i2.7371.

[6] N. M. A. Novitadewi, P. Sugiartawan, and Y. P. Fittryani, "Klasifikasi Data Penjualan Dengan Metode K-Nearest Neighbor Pada PT. Terang Abadi Raya," *J. Sist. Inf. dan Komput. Terap. Indones.*, vol. 5, no. 1, pp. 11–20, 2022, doi: 10.33173/jsikti.173.

[7] N. Musfita, N. Fitriyani, and Z. W. Baskara, "Klasifikasi Penjualan Provider Pulsa di Kecamatan Masbagik Lombok Timur Menggunakan Metode Naïve Bayes," *ESTIMASI J. Stat. Its Appl.*, vol. 4, no. 2, pp. 261–272, 2023, doi: 10.20956/ejsa.v4i2.27890.

[8] I. R. Pratama, Maimunah, and E. R. Arumi, "Sistem Klasifikasi Penjualan Produk Alat Listrik Terlaris Untuk Optimasi Pengadaan Stok Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2135, 2022, doi: 10.30865/mib.v6i4.4418.

[9] U. Darwan and others, "Menentukan Menu Makanan Favorit di Outlet Barbar Sampit Menggunakan Algoritma Naive Bayes," 2022.

[10] K. Mukhlisin, "Penerapan Algoritma Naïve Bayes Pada Produk Brownies Tape Dalam Menentukan Tingkat Penjualan Laris Dan Kurang Laris," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 5, no. 2, pp. 51–56, 2021, doi: 10.32485/kopertip.v5i2.168.

[11] I. Nawangsih and A. Setyaningsih, "Penerapan Algoritma Naive Bayes Untuk Menentukan Klasifikasi Produk Terlaris Pada Penjualan Pulsa," *SIGMA J. Teknol. Pelita Bangsa*, pp. 195–207, 2020.

[12] E. Widodo and P. Saadah, "Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris Untuk Optimasi Strategi Pemasaran di Toko Citramedia Menggunakan Metode Naive Bayes," *SIGMA J. Teknol. Pelita Bangsa*, 2022.

[13] R. W. Abdullah, D. Hartanti, H. Permatasari, A. W. Septyanto, and Y. A. Bagaskara, "Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti)," *J. Ilm. Inform. Glob.*, vol. 13, no. 1, 2022, doi: 10.36982/jiig.v13i1.2060.

[14] M. N. Arhami, *Data Mining Algoritma dan Implementasi*. Yogyakarta, 2022.

[15] Rahmaniar, "Analisis Klasifikasi Model Pemilihan Lokasi Perumahan Dalam Meningkatkan Strategi Manajemen Penjualan Menggunakan Model K-Nearest Neighbor," *J. Sist. Inf.*, pp. 108–120, 2023.

[16] M. A. Muslim and others, *Data Mining Algoritma C4.5 Disertai Contoh Kasus dan Penerapannya dengan Program Komputer*. Semarang, 2022.

[17] F. Rizki, A. Faisol, and F. S. Wahyuni, "Penerapan Metode Naïve Bayes Untuk Memprediksi Penjualan Pada UD. Hikmah Pasuruan Berbasis Web," 2020.

[18] R. Rindiyani, A. Primadewi, Maimunah, and A. H. Purwantini, "Klasifikasi Penjualan Berdasarkan Platform Pada UMKM Omah Branded Menggunakan Random Forest," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1520, 2022, doi: 10.30865/jurikom.v9i5.4949.