Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Application of K-Nearest Neighbor Method for Prediction of Best-Selling Fruit Sales at Ziel Kiosk

Mutiara Revilianti[1]*, Ade irma Purnamasari[2], Agus Bahtiar[3], Edi Wahyudin[4]

[1,2]*Informatics Engineering, STMIK IKMI Cirebon*
[3]*information System, STMIK IKMI Cirebon*
[4]*Computerized Accounting, STMIK IKMI Cirebon*
mutiararevilianti@gmail.com[1], irma2974@yahoo.com[2], agusbahtiar038@gmail.com[3], ediwahyudin@gmail.com[4]*

Abstract

Ziel kiosk sells various types of high-quality fresh fruits. Unfortunately, there is currently no system that manages fruit sales prediction, so there is often a buildup of goods, damaged and rotten goods, or even a shortage of goods, resulting in losses for the kiosk. The data collected is less accurate and effective because the current system is operated manually. This research conducts a data mining process on fruit sales data from Ziel Kiosk from January - December 2023. In sales prediction, Fruit Kiosks can use data mining techniques to be more proactive in managing stock items. This not only avoids the accumulation of fruit stock that can cause spoilage and damage, but also reduces the risk of stock shortages that can affect customer satisfaction. The purpose of this research is to ensure that Ziel Kiosk can see the sales rate for each product sold, so that they can avoid the accumulation of goods and concentrate on the most sold products. With an 80:20 data split, the K-Nearest Neighbor model has high accuracy. This algorithm can also predict fruit sales with an accuracy rate of 97.22% by determining the categories of fruit sales that are in demand or not in demand.

Keywords: *Data Mining, Prediction, K-Nearest Neighbor, Sales*

1. Introduction

The rapid development of information technology has encouraged various sectors to utilize technology in data management and more effective decision making. In the trade sector, especially in fruit sales, the use of technology such as data mining is very important. Data mining enables the analysis of large amounts of data to find hidden patterns. These patterns can help in making important decisions, such as finding best-selling products and predicting customer buying patterns [1].

One of the major problems in stock management and sales planning is uncertain consumer behavior, which is influenced by variables such as seasonality, price, and product availability. The complex patterns of consumer behavior often make traditional statistical-based approaches insufficient. Therefore, more advanced methods, such as machine learning algorithms, are required to make more accurate predictions. K-Nearest Neighbor (KNN) is one of the most widely used machine learning algorithms. It is known for being easy to use and effective in handling very large data. Previous research on the use of the KNN algorithm for sales prediction shows promising results.

According to research [2] they can accurately predict sales in the energy industry. Meanwhile [3] showed that KNN works well to find the best-selling products in the retail market. Another study [4] utilized this method to predict fruit and vegetable sales in a supermarket, where the results showed that the method was effective in identifying sales trends, although there were constraints in terms of prediction timeliness. However, specific research on the use of KNN to predict fruit sales is still scant. This paves the way for new research in this area.

This research aims to develop a fruit sales prediction model using the K-Nearest Neighbor algorithm by utilizing historical consumer transaction data [5]. The main focus of the research is to explore the effectiveness of the KNN algorithm in identifying purchase trends, taking into account external factors such as seasonality and price. The results of this research are expected to make a real contribution to data-driven business management, especially in the fresh food sector.

2. Literature Review

2.1. Data Mining

Data mining is a combination of a number of computer science disciplines that define the process of discovering new patterns from very large data sets, including methods that are slices of artificial intelligence, machine learning, statistics, and database systems. Based on the definition of Data Mining according to the experts mentioned, it can be concluded that Data Mining is a data search process that can automatically get a model from a large database. Data Mining is a method for finding hidden information in databases and part of the Knowledge Discovery in Databases (KDD) process to find useful information and patterns in data [6].

2.2. Sales

Sales are a normal economic activity in which the company receives the planned results/benefits from sales or returns the costs incurred. Sales of company products are shown after deducting discounts and returns [7]. Sales are the purchase of something (goods or services) from one party to another in exchange for money from one party. Based on the definition of sales, it can be concluded that sales is an economic activity carried out between buyers and sellers in transactions in the form of exchanging goods or services for cash.

2.3. Prediction

Prediction/forecasting is one of the methods used to predict or forecast something that has never happened. Prediction is a structured forecasting activity about something that might happen in the future that is based on information from the past and present that is owned, so that the error can be minimized. Based on the definition of prediction, it can be concluded that prediction is a forecasting activity or an estimate of a future situation to conduct a test taken from past data with the aim of minimizing errors that occur[8].

2.4. K-Nearest Neighbor

One method that uses a guided algorithm is K-Nearest Neighbor (K-NN). Where the results of new instance queries are categorized according to most K-NN categories. The K-NN approach classifies information based on how close or far an information is from other information[9].The K-NN method is quite simple yet highly accurate. The purpose of this algorithm is to classify new objects using attributes and training data. The short path between test samples and training samples serves as the basis for the direct K-NN method that calculates K-NN. The majority of the K-NN is then used as a prediction of the test sample after it has been collected.

2.5. Rapid Miner

Rapid Miner is software for data processing. By using data mining principles and algorithms, Rapid Miner extracts patterns from large data sets by combining statistical, artificial intelligence and database methods. RapidMiner makes it easy for users to calculate a lot of data by using operators. These operators are used to modify data. The data is connected to the nodes in the operator then we only need to connect it to the result node to see the results. The results shown by RapidMiner can also be displayed visually with graphs. Making RapidMiner is one of the software choices for extracting data with data mining methods.

3. Research Methods

The stages in the problemsolving process aim to make the research produce correct and precise data. The research stages can be seen in the figure below.

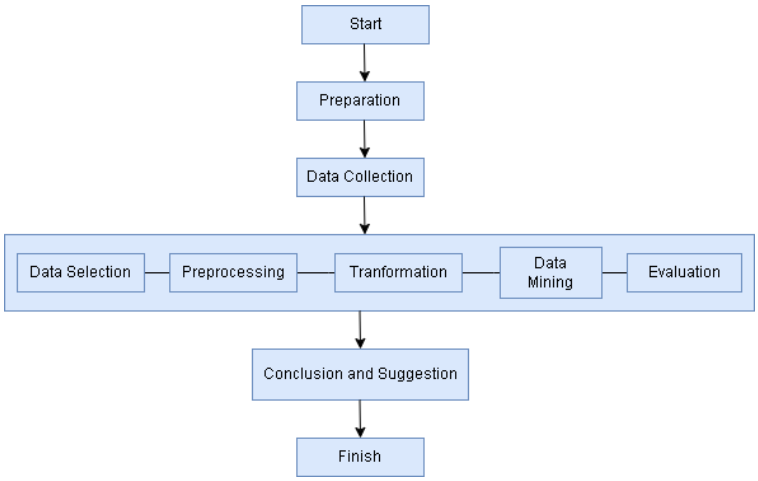

Fig. 1: Research Stages

The stages of this research are outlined in this flow chart. Includes the stages that will be taken in the research as a whole. The stages that will be carried out are:

3.1. Preparation

This stage is the stage of the Ziel Fruit Kiosk subject. The object taken is fruit sales. Limitations and compile a research plan.
.

3.2. Data Collection

Is the process of obtaining information or facts that are relevant to a specific purpose. The data obtained is primary data, namely data collected and processed by Ziel Kiosk. The data used is sales data from January to December 2023. In addition, observations and interviews are also used to obtain data.

3.3. Data Mining Processing

The data that has been collected is then processed further in accordance with the stages of Knowledge Discovery in Database (KDD).

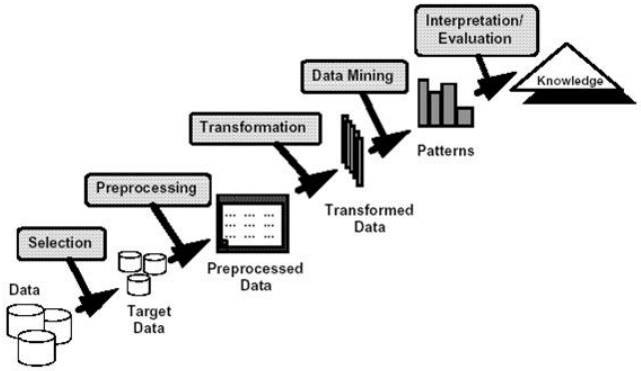

Fig. 2: Knowledge Discovery Process in Databases (KDD).

3.4. Data Selection

The operational data to be used needs to be selected before the information mining stage in KDD begins, this stage is at the data selection stage. The data used is monthly sales data from January to December 2023 at Kiosk Ziel.

Table 1: Monthly Sales Data

No	Mont	Name Fruit	Stock(kg)	Sold (kg)
1.	January	Duku	55	55
2.	January	Rambutan	35	30
3.	January	Siem Oranges	100	95
4.	January	Dragon Fruit	40	35

3.5. Data Pre-processing/Cleaning

Stages in data processing that aim to clean, normalize, and change the format of data so that it is easy to understand and use for analysis.

3.6. Transformation

The transformation stage where the selected data is changed by summarizing (aggregation). So that the data can be used for data mining.

3.7. Data Mining

Is a stage in the KDD process that aims to find patterns or useful information from the data that has been selected and processed. The method used in this research is the K-Nearest Neighbor method. To measure the performance of the K-NN algorithm using four evaluation methods, namely:

a. Accuracy, is the overall prediction ratio. Expressed as the ratio of true positives and true negatives to the total evaluated, it can be calculated using the equation
   Accuracy = $\frac{TP + TN \times 100\%}{(TP+FP+FN+TN)}$ (1)
b. Precision, is a prediction ratio that measures how many positive predictions are correct out of all the results predicted to be positive by the model. In addition, it is defined as the ratio between true positives and the total number of positive predicted results by the model. The equation is as follows
   Precision = $\frac{TP + TN \times 100\%}{}$ (2)

(TP+FP+FN+TN)

c. Recall, is a prediction ratio that measures how much positive data can be detected by the model. It can also be defined as the ratio between true positives and the total amount of data that is actually positive. The equation is

Recall = TP + TN x100% (3)
(TP+FP+FN+TN)

d. Specifity, is the correctness of predicting a comparison with all negative data using the equation

Specificity = TP + TN x100% (4)
(TP+FP+FN+TN)

3.8. Evaluation

Is a process to evaluate the quality of the model that has been created based on data that has been processed by data mining algorithms. At this stage, the fruit sales pattern is obtained from the data mining process with the K-Nearest Neighbor method, the pattern or information generated from the data mining process is in the form of rules obtained from the K-Nearest Neighbor calculation.

4. Results and Discussion

4.1. Data Selection

The data that will be used in the data mining process is data for each month. Sales data is also grouped into two categories, namely in demand and not in demand.

Table 2: Data Selection Table

Name Fruit	Month	Stock	Sold (kg)	Category
Duku	January	100	93	In Demand
Rambutan	January	125	120	In Demand
SiemOranges	January	55	55	In Demand
Salak Pondoh	January	40	20	Not Demand
……..	……..	…	…….	……..
Melon	December	80	74	In Demand

4.2. Data Pre-processing/Cleaning

In this stage there is no data merging or cleaning because the data has no missing value or value 0 (zero).

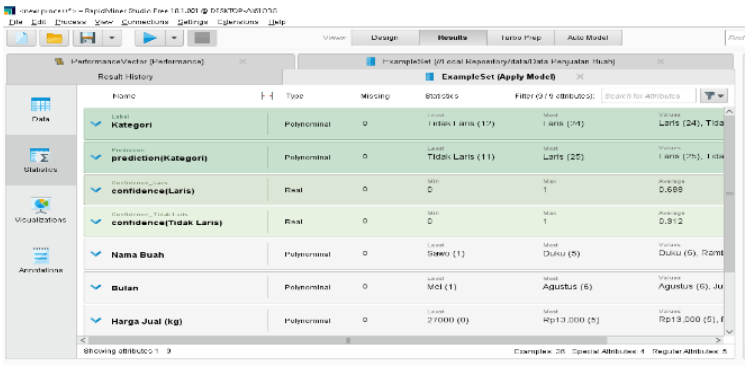

Fig. 3: Preprocessing and Cleaning Data

4.3. K-Nearest Neighbor

Testing methods using Confusion Matrix after calculating Accuracy, Precision, Recall, and Specificity. It can be concluded that the results obtained can be seen in Table 3.

Table 3: Method Testing using Confuision Matrix

Accuracy	*Precision*	*Recall*	*Specifity*
97,22%	96,00%	100%	91,67%

Performance Measurement:

a. *Accuracy*

$$Accuracy = \frac{TP + TN}{(TP+FP+FN+TN)} x100\%$$

$$= \frac{24 + 11}{24+1+11+0} x100\% \quad (1)$$

= 97,22%

b. *Precision*

$$Precision = \frac{TP}{TP+FP} x100\%$$

$$= \frac{24}{24+1} x\ 100\% \quad (2)$$

= 96,00%

c. *Recall*

$$Recall = \frac{TP}{TP+FN} x100\%$$

$$= \frac{24}{24+0} x100\% \quad (3)$$

= 100%

d. *Specifity*

$$Specifity = \frac{TN}{TN+FP} x100\%$$

$$= \frac{11}{11+1} x100\% \quad (4)$$

= 91,67%

4.4. Application of KNN Algorithm with Rapidminer

At this stage begins with data selection and then determining the class label, can be seen in Figure 4.

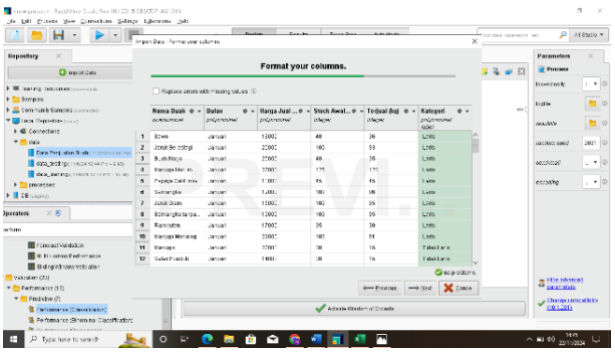

Fig.4: Class Label Determination

The model used in accordance with the K-NN algorithm consists of the retrive dataset operator, the split data operator as a transformation stage to set the training data at 80% and the testing data at 20% of the total data set, the K-NN operator as a form of the K-NN algorithm process itself, to combine the KNN and split data operators into a single model and the performance operator to measure the performance accuracy of the K-NN model that has been formed using the apply model operator.

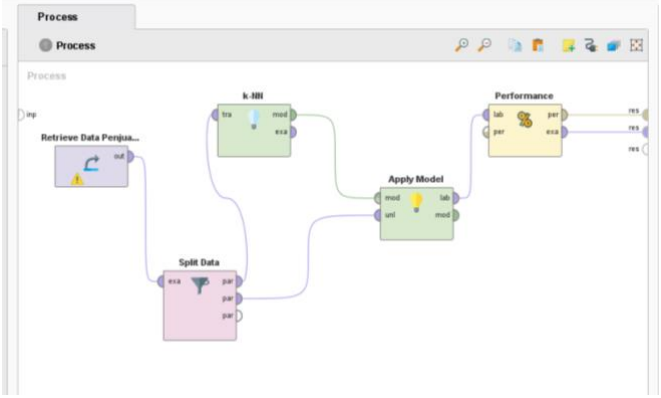

Fig. 5: K-NN Model

Fig. 6: Division of Training Data and Testing Data

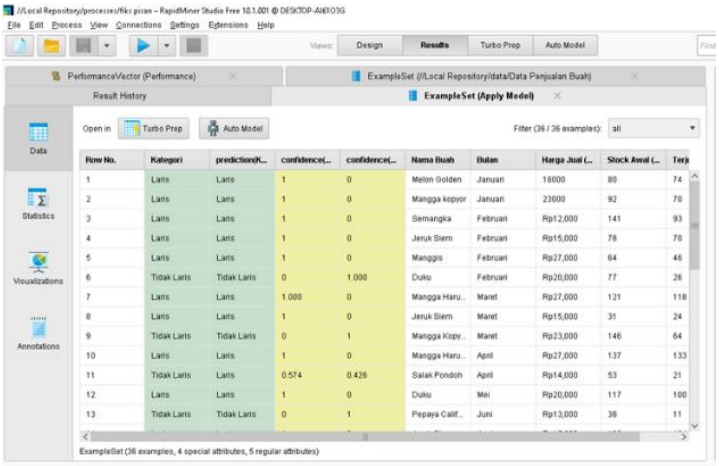

Fig. 7: Data Prediction Result

After completing the implementation in rapidminer using the K-NN algorithm to predict the best-selling fruit sales, it will provide information about fruits that are in demand and not in demand can be seen in the visualization below.

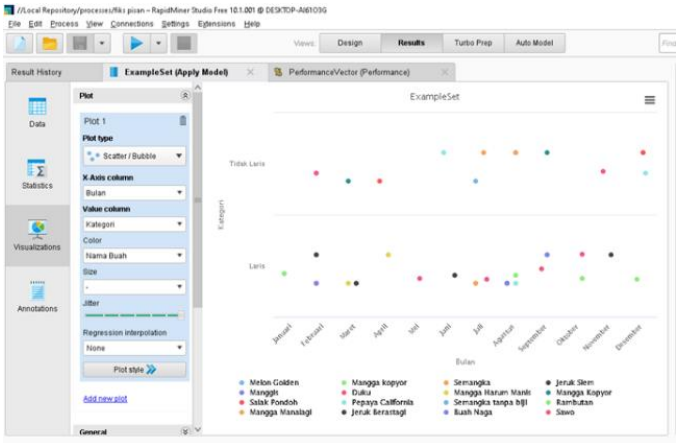

Fig. 8: Visualization Results

From the results of the sales prediction visualization above, the best-selling fruits in January are Melon Golden and Mango Kopyor, in February are Watermelon, Siem Orange and Mangosteen, in March are Mango Harumanis and Siem Orange, in April are Mango Harumanis and Salak Pondoh, in May only Duku, in June only Siem Orange, in July, Rambutan, Mango Manalagi, and Duku, in August, Berastagi Oranges, Dragon Fruit, California Papaya and Rambutan, in September, Dragon Fruit, Seedless Watermelon and Duku, in October, Rambutan and Duku, in November only Berastagi Oranges, while in December the bestseller is only Rambutan.
After the K-NN model is applied to the testing data, the last step is to measure the performance of the model using the performance operator. This operator is used to calculate the performance accuracy value of the K-NN model that has been formed by comparing the model prediction results with the actual values in the testing data.

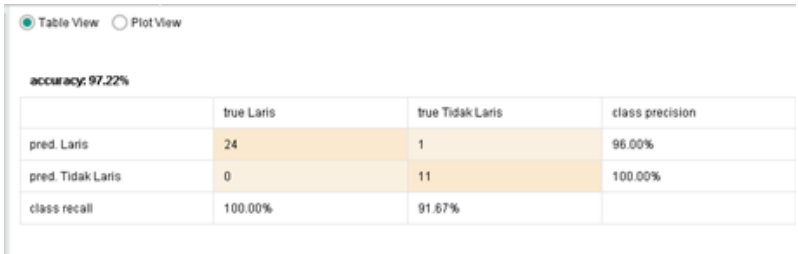

Fig. 9: Accuracy Value Results

It can be seen from the results of trials using the K-NN algorithm with prediction performance said to be in demand 24 true in demand and 1 true not in demand with class precision 96.00%, while the prediction of not in demand is true in demand 0 and 11 true not in demand with class precision 100%, while the total class recall is true in demand 100% and true not in demand 91.67%.

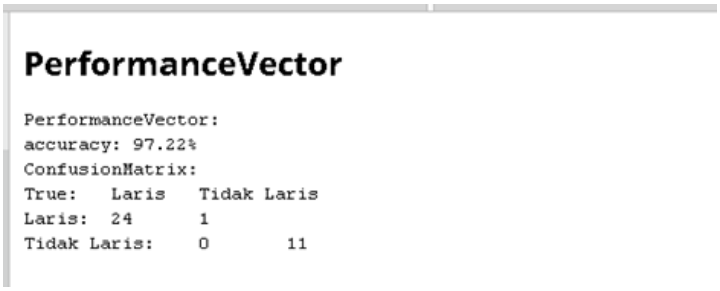

Fig. 10: Avg Results

5. Conclusion

By using the KDD stage and RapidMiner tools, the K-Nearest Neighbor (KNN) algorithm produces a prediction model of the best-selling fruit sales in January to December, the results of the best-selling fruit prediction model are Duku, Rambutan, Siem Orange, Dragon Fruit, Mango Harumanis, and Berastagi Orange. While those that are not in demand are Mango Kopyor, Sawo, and Salak Pondoh. The results of predicting the best-selling fruit sales using the K-Nearest Neighbor algorithm obtained an accuracy rate of 97, 22%, in the prediction said to be in demand 24 true in demand and 1 true not in demand with a class precision of 96.00%, while the prediction of not in demand is true in demand 0 and 11 true not in demand with a class precision of 100%, while the total class recall of true in demand is 100% and true not in demand is 91.67%. This shows that this algorithm can be used to predict the best-selling fruit sales.

Acknowledgement

The author's gratitude goes to the rector of the STMIK IKMI Cirebon campus who has lightened the financial burden for the author. Not to forget the author would like to thank the supervisors Mrs. Ade Irma Purnama Sari, M.Kom and Mr. Agus Bahtiar, M.Kom who patiently guided me until this research was completed.

References

[1] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, 2024, doi: 10.47233/jteksis.v6i1.1078.

[2] A. Ghani Muttaqin, K. Auliasari, and F. Santi Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt.Wika Industri Energy," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 1–6, 2020, doi: 10.36040/jati.v4i2.2728.

[3] M. Danny, A. Muhidin, and A. Jamal, "Application of the K-Nearest Neighbor Machine Learning Algorithm to Preduct Sales of Best-Selling Products," *Brill. Res. Artif. Intell.*, vol. 4, no. 1, pp. 255–264, 2024, doi: 10.47709/brilliance.v4i1.4063.

[4] A. A. Putri, "Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor (Studi Kasus : PT . Central Brastagi Utama)," vol. 1, no. 6, pp. 354–361, 2021.

[5] Anwar Pauji, S. Aisyah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, pp. 21–27, 2022, doi: 10.32485/kopertip.v4i1.114.

[6] R. H. Saragih *et al.*, "Penerapan Data Mining Untuk Memprediksi Penjualan Di Erigo Store Dengan K-Nearest Neighbor," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 8, no. 2, pp. 649–661, 2023, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

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

[8] L. Suryadi, N. Ngajiyanto, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel)," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 7, no. 2, pp. 174–184, 2022, doi: 10.32767/jusim.v7i2.1697.

[9] R. Rismala, I. Ali, and A. Rizki Rinaldi, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 585–590, 2023, doi: 10.36040/jati.v7i1.6419.