

K-Nearest Neighbor (KNN) Algorithm to Determine the Stock of Building Material Store Materials

Delilla Safitri^{*}, Muhammad Fakhriza

Science and Technology, Computer Science, State Islamic University of North Sumatera, Medan, Indonesia

Email: ^{1,*}delillasafitri0411@gmail.com, ²fakhriza@uinsu.ac.id

Correspondence Author Email: delillasafitri0411@gmail.com

Submitted: 04/08/2024; Accepted: 11/08/2024; Published: 12/08/2024

Abstract—In recent months, a lot of infrastructure has been built, resulting in a shortage of goods in the warehouse due to increased demand for consumer goods and some goods not being sold. Such was the case in January and February 2024 when Riko Jaya panglong experienced a shortage of sand and cement supplies, causing losses. This makes it difficult to predict the inventory of an item in the warehouse. Inventory of goods has great strategic importance for the company. This prediction is very useful in determining the amount of goods to be shipped in the following month. Therefore, companies must implement proactive inventory management. The K-Nearest Neighbor algorithm which looks at the euclidean distance between old cases and is compared with new cases in an effort to recognize supervised data or data that already exists and has been recorded to help make decisions on the latest cases, this algorithm is very widely applied in other studies because this algorithm has very simple steps and logical reasoning processes by producing the right data and decisions. This data is processed to determine the classification of goods whether increasing or decreasing. And the K-NN algorithm with a value of $k = 3$ is used to predict stock items. The test results show that K-NN can provide accurate predictions by calculating the Euclidean distance between testing data and training data. The prediction accuracy obtained from the Confusion Matrix reached 100%, indicating the high reliability of this model. Implementation of the K-NN algorithm in RapidMiner with cross-validation technique resulted in a performance of 71.43% for decreasing classification and 67.57% for increasing classification, showing the efficiency of the algorithm in classifying stock data.

Keywords: K-Nearest Neighbor; Algorithm; RapidMiner, UML

1. INTRODUCTION

As time goes by, technology plays an increasingly important role in the business world, and technology continues to develop from the past to the present. Technology also pays attention to developments that are necessary in everyday life, and technology also plays a role in providing various conveniences for its users. Things that were once impossible are now possibilities that are slowly becoming reality one after another. As technology advances, more and more information is collected in databases and this leads to accumulation. As technology develops, so do data processing techniques [1].

In recent months, a lot of infrastructure has been built, resulting in a shortage of goods in the warehouse due to increased demand for consumer goods and some goods not being sold. Such was the case in January and February 2024 when Riko Jaya panglong experienced a shortage of sand and cement supplies, causing losses. This makes it difficult to predict the inventory of an item in the warehouse. Inventory of goods has great strategic importance for the company [2]. One of the most common problems in business is the lack of inventory when there is demand. This prediction is very useful in determining the amount of goods to be shipped in the following month. Therefore, companies must implement proactive inventory management. In this regard, companies must have the ability to anticipate existing situations and challenges in order to reduce overall business costs. Based on this, it is necessary to make systematic calculations to determine the stock of goods in the future [3].

Based on the explanation above, the building shop has never determined the next month's inventory. Even though predicting inventory has a huge impact on the company. An appropriate calculation algorithm is needed in a decision support system. The K-Nearest Neighbor algorithm is one of the simple data classification algorithms where the shortest distance calculation is used as a measure in classifying objects based on the data closest to the object [4]. The K-Nearest Neighbor algorithm is classified as a supervised algorithm, which is a process where an algorithm obtained through the learning process is formed on old records that are classified, and new records of unknown origin are classified based on the learning results [5]. In the K-nearest neighbor algorithm, new data is classified based on the distance between the new data and the degree of closest similarity of the new data to the pattern data. Basically, the K-nearest neighbor algorithm makes predictions based on how similar the data is to the data in the dataset [6].

The K-Nearest Neighbor algorithm works by keeping all the training data in memory. An algorithm is used to complete the task of calculating input data. Riko Jaya Building Shop is a construction company located at Jalan Marelan 1 Pasar 4 Barat No.3, Rengas Pulau, Medan Marelan District, where the company sells various kinds of building materials such as cement, iron, nails, wall paint, and other building materials. The company is open from 08:00 WIB to 18:00 WIB. Panglong Riko Jaya has been established since 2006 which directly serves the needs of its buyers. problems in the management of materials that cannot be predicted properly. Therefore, the company must be able to manage its inventory availability in such a way that there is no overstock or shortage. If the inventory runs out or there is often excess inventory, the company suffers losses. As a building

materials company, it is not uncommon for us to have difficulty predicting which inventory will be in the next few months.

A decision support system is a system that aims to support meaningful decision making in semi-structured decision situations. A decision support system is a set of model-based procedures that work in data processing and evaluation to assist decision makers in making decisions [7]. The requirements for the success of this system must be fast, simple, and easy to use for communication. From this report it can be concluded that a decision support system is not a decision-making tool, but rather a support system that helps decision makers by completing the information that is processed appropriately and necessary, so that decisions can be made related to problems. be faster and more accurate [8].

To improve the quality of the decision-making process and results, a decision support system (DSS) was created, where DSS can combine data and knowledge to improve the efficiency and effectiveness of the decision-making process [9]. Information is used to support decision making in an organization or business. It can also be said that it is a computer system that processes data into information to make decisions about some semi-structured problems [10].

The K-Nearest Neighbor (KNN) algorithm is a method for classifying new objects based on the value (K) of their nearest neighbors [11]. The K-Nearest Neighbor (KNN) algorithm is a supervised learning algorithm, where the results of a new query instance are classified based on the majority of the categories in KNN. The class that appears the most, will be the class of the classification result [12].

The K-Nearest Neighbor (KNN) algorithm is a method for classification, this method classifies an object based on the training data that has the closest distance from the object. The K-Nearest Neighbor algorithm or commonly called KNN is a data classification method that works relatively in a simpler way compared to other data classification methods [5]. This algorithm tries to classify new data that has no known class by selecting the K number of data that is closest to the new data. The largest class of the K closest data is chosen as the predicted class for the new data. K is generally determined by the number of equal distances in the classification process. The K-Nearest Neighbor (KNN) algorithm is a classification method that is robust to noisy training data and is effective when the data is large [13].

Previous research conducted by Mohammad kafil (2019) entitled "application of the K-Nearest neighbor method for web-based sales prediction at boutiq dealove bondowoso" this research is a support system to facilitate sales forecasting Erratic consumer demand along with higher needs, resulting in a shortage of stock in the warehouse [5].

Previous research conducted by Abdul Azis, Ahmad Turmudi, Aswan S Sunge (2024) entitled "Prediction of Sales of Best-Selling Drugs and Medical Devices Using the K-Nearest Neighbor Algorithm" this study shows that the K-Nearest Neighbors (KNN) algorithm, provides accurate predictions with value in the classification of drugs and medical devices, so that it can be relied upon in supporting hospitals in more efficient drug stock management [14].

In this study, the K-nearest neighbor algorithm helps select new customers who want to make transactions in the form of payments so that there are no losses to the business that provides and continues to help the process of optimizing operational costs. The K-Nearest Neighbor algorithm which looks at the euclidean distance between old cases and is compared with new cases in an effort to recognize supervised data or data that already exists and has been recorded to help make decisions on the latest cases, this algorithm is very widely applied in other studies because this algorithm has very simple steps and logical reasoning processes by producing the right data and decisions [1]. The K-Nearest Neighbor (K-NN) method has the advantage of being effectively applied to large amounts of data and is robust to a noise training data which is a data that has the most distant value compared to other data. K-Nearest Neighbor (K-NN) has a concept that is intuitive and easy to understand [2] In addition, K-Nearest Neighbor (K-NN) is also relatively simple in various programming languages. K-Nearest Neighbor (K-NN) has few hyperparameters that need to be set, mainly only the parameter k (number of nearest neighbors). This is different from other algorithms that may have many hyperparameters that need to be adjusted [8].

Rapid miner is a data mining tool developed java as a programming language. Java language is considered as one of the languages that has various advantages, such as simplicity, security, power, impact, high-level object-oriented capabilities, and many other advantages. Rapid miner is developed for data mining tasks adopted by several license options (starter, personal, professional, enterprise) [4].

Rapid miner is an open source software, one of the solutions to analyze data mining. Rapid miner is a data science software platform developed by the same name company that provides an integrated environment for data preparation, machine learning, deep learning, text mining, predictive analysis. It is used for business and commercial, as well as for research, education, training, rapid prototyping, and application development and supports all steps in the machine learning process including data preparation, visualization results, model validation and optimization [15].

2. RESEARCH METHODOLOGY

2.1 Research Method and Materials

In the research methodology there is a sequence of work that must be followed, this framework sequence is a description of the steps that must be taken so that this research can run well. To complete this research, of course, this research must be carried out systematically. Which is where the data collection technique is divided into two, namely supporting data and main data. Supporting data consists of library research data or literature studies. The main data is data that will be used with interviews and observations.

The needs analysis stage is a process to get information on what needs will be used and aims to build a system so that the system runs. This needs analysis stage that will be carried out is an activity to identify the needs and specifications of the system to be built. Aims to clearly understand how the needs of the system to be built in accordance with the objectives of the problem to be determined. This stage is the stage after collecting data and case information in this study. After collecting data, the next stage is the stage of analyzing the system to be created based on the K-Nearest Neighbor algorithm [16].

The K-Nearest Neighbor (KNN) algorithm is a method for classifying new objects based on the value (K) of their nearest neighbors. The K-Nearest Neighbor (KNN) algorithm is a supervised learning algorithm, where the results of a new query instance are classified based on the majority of the categories in KNN. The class that appears the most, will be the class of classification results. In this study using Euclidean distance to find the nearest neighbor, so the formula for calculating the distance with Euclidean is as follows [17].

The K-Nearest Neighbor (KNN) algorithm is a method for classification, this method classifies an object based on the training data that has the closest distance from the object. The K-Nearest Neighbor algorithm or commonly called KNN is a data classification method that works relatively in a simpler way compared to other data classification methods. This algorithm tries to classify new data that has no known class by selecting the K number of data that is closest to the new data. The largest class of the K closest data is chosen as the predicted class for the new data. K is generally determined by the number of equal distances in the classification process. The K-Nearest Neighbor (KNN) algorithm is a classification method that is robust to noisy training data and effective when the data is large [18].

At this stage of analyzing input requirements, activities are carried out to describe information in the form of datasets used by the author to design the system to be built [19]. The input requirements that are included in the system determine the stock of sales items in the form of datasets obtained from the Riko Jaya building shop :

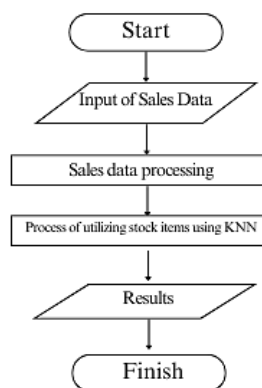


Figure 1. Flowchart System

Based on the system flowchart image above, this chart describes the process flow that will be built using the K-Nearest Neighbor algorithm. In the picture above, it is explained that when the program is run, the User will input sales material data, then the data is processed using the K-Nearest Neighbor algorithm, after that the fifth chart results in determining the number of stock items and the last is a chart stating the process has been completed.

The testing stage is needed to be a measure that the system can be run in accordance with the objectives. Method testing is a test that aims to show the results of the system using the K-Nearest Neighbor algorithm and manually. In this study using the rapid miner data processing platform from the previous explanation of the steps of use and the resulting output. The next step is to explain the data validation used using the rapid miner system. Rapid miner is an open-source data science platform that can be used to perform analysis. By using rapidminer, users can use descriptive and predictive techniques to get a deep understanding of the data they have. Rapid miner serves as an important tool in navigating and reshaping data, facilitating researchers in carrying out tasks such as predictive modeling, data clustering, and fluid data visualization. With its user-friendly and adaptable features, rapid miner helps in this research to extract valuable insights from complex data sets, accelerate the analysis process, and improve the quality of decision making through a strong information base [14].

The application or use of this system is that building stores can predict stock items based on sales results using the K-Nearest Neighbor algorithm and is also useful for helping building stores predict stock items [20]. Material sales data at building stores are grouped and processed using the K-NN algorithm to determine sales. Furthermore, the process continues with the division of data into two segments called the training dataset and the testing dataset. The attributes used in this research are item name, month, item price and category.

3. RESULT AND DISCUSSION

At the data analysis stage, the data needed in this study is in the form of material sales data at the Riko Jaya building shop. The first stage is to collect data, the data used in this study is purchase data based on the last 1 year, then the data is selected and will be used to determine the number of stock items in the following year. At the data analysis stage, the determination of the stock of sales items is carried out, then implemented using the K-neighbor nearest algorithm to determine the stock of goods using rapid miner [12].

In this section it will be explained in general how to determine the best-selling sales in building materials which include in demand and not in demand using the k-neighbour nearest algorithm.

The building material sales data contained in this study can be obtained through several stages of data collection, namely direct interviews and using building material sales data at Riko Jaya, namely increased sales and decreased sales.

3.1 Stage of K-Nearest Neighbor Algorithm

In this study, the data used in testing accuracy manually using Microsoft Excel is 16 data. Where 12 of the data becomes training data and 4 becomes testing data. The following are the calculation steps using K- nearest neighbor.

Table 1. Stock Selection Data of Cement In 2023

No	Material	Unit	Month	Incoming Items	Outgoing Items	Last Stock	Description
1.	Cement	sak	January	500	900	100	Decreased
2.	Cement	sak	February	900	700	200	Decreased
3.	Cement	sak	March	800	350	650	Increased
4.	Cement	Sak	April	200	700	150	Decreased
5.	Cement	Sak	May	900	500	550	Increased
6.	Cement	Sak	June	900	600	850	Increased
7.	Cement	Sak	July	200	850	250	Decreased
8.	Cement	Sak	August	800	750	300	Increased
9.	Cement	Sak	September	500	645	155	Decreased
10.	Cement	Sak	October	700	700	155	Decreased
11.	Cement	Sak	November	800	800	155	Decreased
12.	Sand	kubik	January	50	20	20	Increased
13.	Sand	kubik	February	13	26	7	Increased
14.	Sand	kubik	March	33	21	38	Increased
15.	Sand	kubik	April	10	40	8	Increased
16.	Sand	kubik	May	20	18	8	Increased
17.	Sand	kubik	June	20	15	13	Increased
18.	Sand	kubik	July	30	23	10	Increased
19.	Sand	kubik	August	25	20	15	Increased
20.	Sand	kubik	September	25	17	23	Increased
21.	Sand	kubik	October	22	23	22	Increased
22.	Sand	kubik	November	18	20	20	Increased

The training data used in this study is taken from January 2023 to December 2023. As for the testing data used for testing data (test) used for predictions in January 2024 by entering data in December 2023.

Table 2. Data Testing

No	Material	Unit	Month	Incoming Items	Outgoing Items	Last Stock	Description
1.	Cement	Sak	December	650	760	45	Decreased
2.	Cement	Sak	December	17	22	15	Decreased
3.	Cement	Sak	December	11000	15000	1000	Increased
4.	Cement	Sak	December	100	105	33	Decreased

Based on the k nearest neighbor stage, the steps are:

1. In this case, the value of k. the determination used does not have a large rule. Only the k value will have a good accuracy value. In this study, the k value used is 3.
2. Calculate the distance according to the testing data one by one training data using the eucliden distance formula as follows:

$$d(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

First Testing Data

$$\text{Distance} = \sqrt{(1000 - 650)^2 + (900 - 760)^2 + (100 - 45)^2} = 380.88$$

$$\text{Distance} = \sqrt{(900 - 650)^2 + (700 - 760)^2 + (200 - 45)^2} = 300.21$$

$$\text{Distance} = \sqrt{(800 - 650)^2 + (350 - 760)^2 + (650 - 45)^2} = 746.08$$

$$\text{Distance} = \sqrt{(200 - 650)^2 + (700 - 760)^2 + (150 - 45)^2} = 465.91$$

$$\text{Distance} = \sqrt{(900 - 650)^2 + (500 - 760)^2 + (550 - 45)^2} = 620.60$$

$$\text{Distance} = \sqrt{(900 - 650)^2 + (600 - 760)^2 + (850 - 45)^2} = 858.47$$

$$\text{Distance} = \sqrt{(200 - 650)^2 + (850 - 760)^2 + (250 - 45)^2} = 498.62$$

$$\text{Distance} = \sqrt{(800 - 650)^2 + (750 - 760)^2 + (300 - 45)^2} = 296.02$$

$$\text{Distance} = \sqrt{(500 - 650)^2 + (645 - 760)^2 + (155 - 45)^2} = 218.67$$

$$\text{Distance} = \sqrt{(700 - 650)^2 + (700 - 760)^2 + (155 - 45)^2} = 132.75$$

$$\text{Distance} = \sqrt{(800 - 650)^2 + (800 - 760)^2 + (155 - 45)^2} = 190.47$$

Second Testing Data

$$\text{Distance} = \sqrt{(50 - 17)^2 + (30 - 22)^2 + (20 - 15)^2} = 34.33$$

$$\text{Distance} = \sqrt{(13 - 17)^2 + (26 - 22)^2 + (7 - 15)^2} = 9.80$$

$$\text{Distance} = \sqrt{(33 - 17)^2 + (21 - 22)^2 + (38 - 15)^2} = 28.03$$

$$\text{Distance} = \sqrt{(10 - 17)^2 + (40 - 22)^2 + (8 - 15)^2} = 20.54$$

$$\text{Distance} = \sqrt{(20 - 17)^2 + (18 - 22)^2 + (8 - 15)^2} = 8.60$$

$$\text{Distance} = \sqrt{(20 - 17)^2 + (15 - 22)^2 + (13 - 15)^2} = 7.87$$

$$\text{Distance} = \sqrt{(30 - 17)^2 + (23 - 22)^2 + (10 - 15)^2} = 13.96$$

$$\text{Distance} = \sqrt{(25 - 17)^2 + (20 - 22)^2 + (15 - 15)^2} = 8.25$$

$$\text{Distance} = \sqrt{(25 - 17)^2 + (17 - 22)^2 + (23 - 15)^2} = 12.37$$

$$\text{Distance} = \sqrt{(22 - 17)^2 + (23 - 22)^2 + (22 - 15)^2} = 8.66$$

$$\text{Distance} = \sqrt{(18 - 17)^2 + (20 - 22)^2 + (20 - 15)^2} = 5.48$$

Table 3. Distance Sequence 1

Rangking	Ecludien Distance	Classification
1	132.75	Increased
2	190.47	Increased
3	218.67	Increased
4	296.02	Increased
5	300.21	Increased
6	380.88	Increased
7	465.91	Increased
8	498.62	Increased
9	620.60	Increased
10	746.08	Increased

Rangking	Ecludien Distance	Classification
11	858.47	Increased

Table 4. Distance Sequence 2

Rangking	Ecludien Distance	Classification
1	5.48	Increased
2	7.87	Increased
3	8.25	Increased
4	8.60	Increased
5	8.66	Increased
6	9.80	Increased
7	12.37	Increased
8	13.96	Increased
9	20.54	Increased
10	28.03	Increased
11	34.33	Increased

From the data using the K-nearest neighbor category of the most popular data, it can be predicted that the classification that appears a lot is the "increase" classification.

The next step is the confusion matrix. Accuracy states the amount of data that is classified correctly after the testing process. The formula used to calculate accuracy is as follows:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+FN} \times 100$$

Table 5. Confusion Matrix

Data	Identified Increase	Identified Decrease
Test data increased	19	0
Test data decreased	0	25

From the table above, the level of accuracy can be measured from the following classification:

$$TP = 19, FN = 25, TN = 0, FP = 0$$

$$Accuracy = ((TP+TN)/(TP+FP+TN+FN)) \times 100\%$$

$$Accuracy = ((19+25)/(19+0+25+0)) \times 100$$

$$Accuracy = 100\%$$

The next step by using the K - nearest neighbor category which is the majority, the next material stock will be predicted using rapid miner.

3.2 Implementation Data

Rapid miner is one of the data set processing software for data patterns according to the purpose of the data processing. Not all existing algorithms can be suitable or can manage existing data sets, but data patterns must be done and in accordance with the purpose of the data processing.

After classifying, the next step is to drag and drop the read excel operator then import the Microsoft excel table into the process to read the training data in excel format, because what will be predicted in month 12 or December.

Data is entered into the system based on the stock of materials every month, the original data is used as an integer, while the data to be predicted is used as a label. Next, drag and drop the data split operator to split the testing data and training data. In split data, enter 80% entry as training data, and 20% or 0.2 as testing data. Next, add the KNN operator, KNN by entering the value of k = 3 as drawn.

Then add the apply model operator, which means we want to apply the model we created to predict the stock of building materials in December. Then the output in the testing data or data that has no label is connected to the apply model, after that the data that has a label is connected to the recall. A display appears which has a decreasing and increasing class. meaning that the stock of materials in each month is either increasing or decreasing.

Then the results were materials, namely cement, sand, bricks, and 2x3 wood. It turns out that the result is that the cement material is predicted to decrease the number of stock items, the sand material is predicted to increase the number of stock items, the brick material is predicted to decrease the number of stock items, and the 2x3 wood material is predicted to decrease the number of stock items.

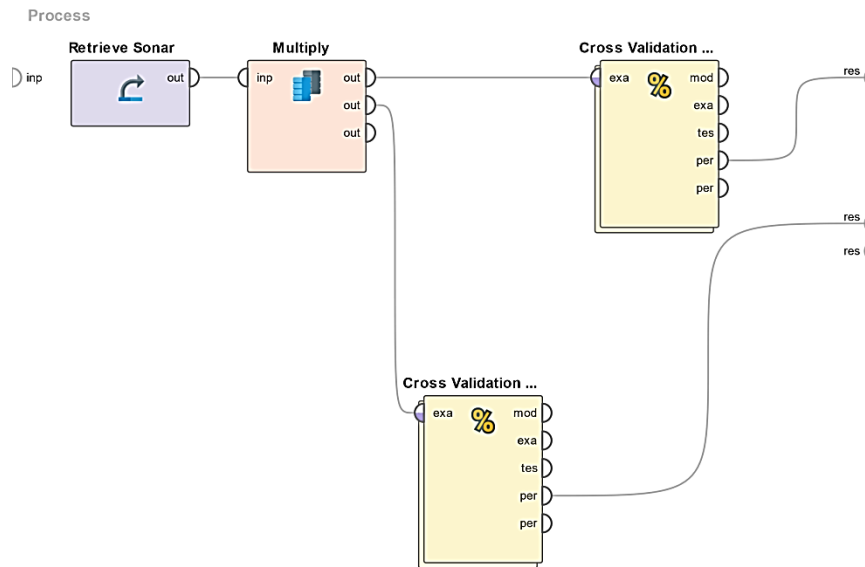


Figure 2. Cross Validation

Then to validate that this method is accurate or the method we use KNN is accurate we use the cross validation technique. We type in the operator section type cross then we connect the output information. After that we click in the cross validation section 2 times.

Now in the training section, the k- nearest neighbor is drawn, then from this training section draw a line in the knn box, from this model information also connected to the model on the right. Now in the testing section type apply model, meaning we want to apply the model we choose, so the model we choose has a cross validation calculation and performance. Then we connect the model description here to the apply model. Then we calculate the performance type in the performance operator section, then we connect it to recall, when there is no red mark then it can be run.

Table 6. Prediction Output

Accuracy 69%			
	True Menurun	True Meningkatkan	Class Recall
Pred Menurun	5	2	71.43%
Pred Meningkatkan	12	25	67.57%
Class Recall	29.41%	92.59%	

After we run it, we can see the performance of the k- nearest neighbor algorithm for the case study of classifying stock data. For prediction, the classification decreased by 71.43%. And for prediction the classification increased by 67.57%.

3.3 Discussion

The results of the analysis using the K-Nearest Neighbor (K-NN) algorithm show that this algorithm is effective in predicting stock of goods based on sales data of building materials at Toko Riko Jaya. With the value of $k=3$, sales data in 2023 was processed to predict stock in December. The test showed that the prediction of cement and brick stock decreased, while sand stock increased. The prediction of 2x3 wood also shows a decrease in stock. The prediction accuracy obtained from the Confusion Matrix reached 100%, indicating the high reliability of this model. Implementation of the K-NN algorithm in RapidMiner with cross-validation technique resulted in a performance of 71.43% for decreasing classification and 67.57% for increasing classification, showing the efficiency of the algorithm in classifying stock data.

4. CONCLUSION

In this study, the K-nearest neighbor algorithm helps select new customers who want to make transactions in the form of payments so that there are no losses to the business that provides and continues to help the process of optimizing operational costs. The K-Nearest Neighbor algorithm which looks at the euclidean distance between old cases and is compared with new cases in an effort to recognize supervised data or data that already exists and has been recorded to help make decisions on the latest cases, this algorithm is very widely applied in other studies because this algorithm has very simple steps and logical reasoning processes by producing the right data

and decisions. Based on the research that has been done using the K-nearest neighbor (KNN) algorithm on the sales data of materials at the Riko Jaya building shop, several things can be concluded, the author concludes that use of Sales Data This research utilizes sales data for the past year to analyze and predict the stock of goods in the following year. This data is processed to determine the classification of goods whether increasing or decreasing. And the K-NN algorithm with a value of $k = 3$ is used to predict stock items. The test results show that K-NN can provide accurate predictions by calculating the Euclidean distance between testing data and training data. Implementation with Rapid Miner Using Rapid Miner to implement K-NN helps in managing and predicting building material stock data. This process involves processing the data, dividing the training and testing data, and applying the model for prediction. This section contains conclusions that answer all the problems contained in the research.

REFERENCES

- [1] S. Silvillestari, "Data Mining Menggunakan Algoritma K-Nearest Neighbor Dalam Menentukan Kredit Macet Barang Elektronik," *mib*, vol. 5, no. 3, p. 1063, Jul. 2021, doi: 10.30865/mib.v5i3.3100.
- [2] D. Handoko, H. S. Tambunan, and J. T. Hardinata, "Analisis Penjualan Produk Paket Kuota Internet Dengan Metode K-Nearest Neighbor," *jurasiik*, vol. 6, no. 1, p. 111, Feb. 2021, doi: 10.30645/jurasiik.v6i1.275.
- [3] Amanda Pratiwi, Ananto Tri Sasongko, and D. K. Pramudito, "ANALISIS PREDIKSI GILINGAN PLASTIK TERLARIS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DI CV MENEMBUS BATAS," *JINTEKS*, vol. 5, no. 3, pp. 437–445, Aug. 2023, doi: 10.51401/jinteks.v5i3.3323.
- [4] A. A. Arif, M. Firdaus, and Y. Maruhawa, "Comparison of Data Mining Methods for Prediction of Rainfall with C4.5, Naïve Bayes, and KNN Algorithm," 2022.
- [5] M. Kafil, "PENERAPAN METODE K-NEAREST NEIGHBORS UNTUK PREDIKSI PENJUALAN BERBASIS WEB PADA BOUTIQ DEALOVE BONDOWOSO," *jati*, vol. 3, no. 2, pp. 59–66, Sep. 2019, doi: 10.36040/jati.v3i2.860.
- [6] D. Desyanti and D. Wulandari, "Implementasi Algoritma K-Nearest Neighbour dalam Memprediksi Stok Sepeda Motor," *bits*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2579.
- [7] Y. Findawati and M. Rosid Alfian, *Buku Ajar Text Mining*. Sidoarjo: UMSIDA Press, 2020.
- [8] I. L. Putra, "IMPLEMENTASI ALGORITMA PARTICLE SWARM OPTIMIZATION(PSO) DAN K-NEAREST NEIGHBOR(K-NN) DALAM MEMPREDIKSI KEBERHASILAN ANAK SMK MENDAPATKAN KERJA," *Technologia*, vol. 13, no. 4, p. 339, Oct. 2022, doi: 10.31602/tji.v13i4.8167.
- [9] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," 2023.
- [10] I. S. Muallif, H. Budiman, and N. Ransi, "Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor" 2023.
- [11] D. M. Meliala and P. Hasugian, "Perbandingan Algoritma K-Nearest Neighbor Dengan Decision Tree Dalam Memprediksi Penjualan Makanan Hewan Peliharaan Di Petshop Dore Vet Clinic," *jtir*, vol. 15, no. 3, p. 35, Nov. 2020, doi: 10.35842/jtir.v15i3.369.
- [12] V. P. Virza, G. Tri Pranot, and F. Eko Putra, "Klasifikasi Kebutuhan Sparepart Dengan Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Sparepart," *bit*, vol. 4, no. 3, pp. 287–293, Sep. 2023, doi: 10.47065/bit.v4i3.729.
- [13] D. A. Saputri, S. Oyama, and S. Wardani, "Implementasi Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Menentukan Stok Obat-Obatan Pada Apotek: Studi Kasus Apotek Salaam," 2021.
- [14] 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*, vol. 7, no. 2, pp. 174–184, Dec. 2022, doi: 10.32767/jusim.v7i2.1697.
- [15] A. Nurfajri and M. Subroto, "Hubungan Komunikasi Interpersonal Antar Narapidana Perempuan Dalam Merestorasi Mental di Dalam Lapas Perempuan Kelas II B Bengkulu," *Jurnal Pendidikan Kewarganegaraan UNDIKSHA*, vol. Volume 9, No 3, Sep. 2021.
- [16] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *INSYST*, vol. 1, no. 1, pp. 43–49, Aug. 2019, doi: 10.52985/insyst.v1i1.36.
- [17] Amanda Pratiwi, Ananto Tri Sasongko, and D. K. Pramudito, "ANALISIS PREDIKSI GILINGAN PLASTIK TERLARIS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR DI CV MENEMBUS BATAS," *JINTEKS*, vol. 5, no. 3, pp. 437–445, Aug. 2023, doi: 10.51401/jinteks.v5i3.3323.
- [18] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, and B. Priyatna, "Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS*, vol. 8, no. 1, p. 11, Mar. 2023, doi: 10.31328/jointecs.v8i1.4265.
- [19] E. B. Pratama and U. Saparingga, "Pemodelan UML Sistem Informasi Administrasi Kependudukan Untuk Kantor Desa," *jims*, vol. 15, no. 2, pp. 107–118, Oct. 2021, doi: 10.33998/mediasisfo.2021.15.2.1085.
- [20] M. Ikhsan and R. Kurniawan, "Penerapan Text Mining pada Sistem Rekomendasi Pembimbing Skripsi Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier," *Indonesian Journal of Computer Science*, vol. 12, no. 6, 2023.