International Journal Software Engineering and Computer Science (IJSECS)
4 (2), 2024, 523-534
Published Online August 2024 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v4i2.2865.

RESEARCH ARTICLE **Open Access**

Predicting Consumer Demand Based on Retail Stock Using the K-Nearest Neighbors Algorithm

Anindya Putri N.A *
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra, Indonesia.
Corresponding Email: aputrina17@gmail.com.

Wahyu Hadikristanto
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra, Indonesia.
Email: wahyu.hadikristanto@pelitabangsa.ac.id.

Edora
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, Deli Serdang Regency, North Sumatra, Indonesia.
Email: edora@pelitabangsa.ac.id.

Received: June 28, 2024; Accepted: July 10, 2024; Published: August 1, 2024.

Abstract: Inefficient stock management, such as improper stock management, will result in excess or shortage of goods. Excess stock can cause high storage costs and the risk of unsold goods. Predict consumer needs based on stock. Analyze inefficient stock to improve shortages. One effective method for making this prediction is using the K-Nearest Neighbors (K-NN) algorithm. The K-NN algorithm is a simple but powerful machine-learning technique that can be used for classification and regression. The model scenario results show 24 objects in the Low-needs group and 14 in the High-needs group. Evaluation and performance testing using the Rapid Miner tool can also produce a relevant picture of the modelled scenario. The model implemented using the K-NN algorithm has an Accuracy value of 97.50% with a Standard Deviation of +/- 750%, then a Precision value of 100%, and a Recall value of 950%. By measuring model performance with cross-validation, the resulting accuracy has a standard deviation value, which aims to see the distance between the average accuracy and the accuracy of each experiment (iteration).

Keywords: K-NN; Data Mining; Retail Stock; Classification.

1. Introduction

Retail inventory is one of the fundamental components in supply chain management that directly impacts a store's financial and operational performance. Efficient inventory management ensures sufficient product availability to meet customer demand without experiencing excess or inventory shortage. According to Darmi and Setiawan (2016), improper inventory management can cause two main problems, namely excess inventory, which results in high storage costs and the risk of expired goods, and shortage of inventory, which causes lost sales opportunities and decreased customer satisfaction [1]. Retail inventory management involves several vital aspects, such as procurement, storage, and inventory control. An effective inventory management system allows stores to track the movement of incoming and outgoing goods and determine the right time to reorder. Research by Hasanah *et al.* (2019) shows that stores that implement an information technology-based inventory management system can significantly optimize profits and reduce operational costs [2]. However, excessive stock increases storage costs and can lead to the risk of unsold goods. On the other hand, understocking will disappoint customers and reduce loyalty, ultimately hurting the store's revenue and reputation [3].

Understanding consumer demand patterns is critical to maintaining optimal stock balance. This is important to ensure that the right products are available at the right time, increasing customer satisfaction and maximizing store profits. Predicting demand is becoming increasingly important given the high competition in the retail sector, which demands efficiency in stock management. The K-Nearest Neighbors (K-NN) algorithm is one approach that can be used to predict consumer needs based on retail stock data [4]. K-NN is a simple yet effective machine-learning technique often used for classification and regression tasks. In predicting consumer needs, this algorithm can help identify purchasing patterns and consumer behaviour by analyzing historical sales and stock data. This algorithm calculates the distance between new data and the closest historical data and then uses information from the nearest neighbour data to make predictions [5]. The selection of the K value in the K-NN algorithm is essential because it can affect the accuracy of the prediction. Research by Fahlevi (2020) revealed that a K value that is too small will make the model too sensitive to noise, while a K value that is too large will make the model too generalized [3]. Therefore, it is essential to determine the correct K value based on the characteristics of the data used. In addition, applying the K-NN algorithm in retail stock management also requires calculating the proximity between new and old cases. This helps group retail stock data into relevant categories, such as high or low stock needs. Research by Iku *et al.* (2019) shows that this approach can be used to solve problems in the field of Data Mining with high accuracy [6]. This study will use the K-NN algorithm to predict consumer needs based on retail stock. Using historical sales data, this study aims to identify patterns that can be used to optimize stock management, thereby increasing store operational efficiency and customer satisfaction. This algorithm was chosen because of its simplicity of implementation and ability to provide accurate results in classification and regression.

2. Research Method

This study begins with an observation stage to identify problems related to stock management and sales of goods at Toko Anugrah. This observation aims to understand sales patterns and identify factors that influence stock requirements. Sales data used in this study were collected during the period January-February 2024. After the observation stage, the data obtained from Toko Anugrah was transferred into Excel format (.xlsx) to facilitate data processing. The data consists of information about items sold during the study period. The next stage involves data processing using the K-Nearest Neighbor (K-NN) algorithm. This algorithm was chosen because of its ability to classify based on the closest distance between data, which is relevant for predicting stock requirements based on historical sales data. The data processing process begins with setting the K parameter in the K-NN algorithm. The selection of the K value is very important because it will affect the classification results. The data is then processed using Microsoft Excel to perform initial calculations. This process includes steps such as data normalization, Euclidean distance calculation, and classification based on the closest distance. The results of this manual calculation are then used as a basis for further testing. After the manual calculation is complete, the processed data is applied to the Rapid Miner software. Rapid Miner is used to test the accuracy of the model generated with the K-NN algorithm. This process involves testing the model by comparing the results of manual calculations with the results obtained from the software. This stage is important to ensure that the model used has an adequate level of accuracy.

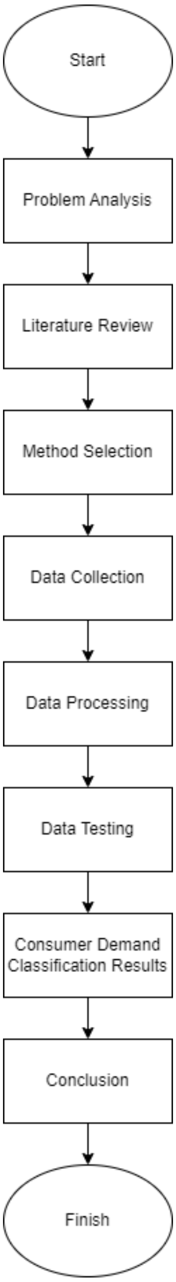

Figure 1. Research Design

Testing on the Rapid Miner software is carried out using cross-validation techniques to measure model performance in terms of accuracy, precision, and recall. Cross-validation is carried out by dividing the data into several subsets, where each subset will be used alternately as training data and test data. This technique aims to test the consistency of the model and ensure that the classification results are not influenced by random data division. The result of this testing process is a classification of consumer needs which are divided into two groups: high needs and low needs. The results of the classification are expected to provide guidance for Toko Anugrah in managing their stock more efficiently, so that they can increase sales and reduce the risk of shortages or excess stock. The conclusion of this study is drawn based on the results of the classification that has been carried out. The test results show adequate accuracy, precision, and recall values, which can be used as a basis for decision making in stock management at Toko Anugrah. Thus, the results of this study provide practical contributions for stores in optimizing their stock management based on historical sales data analysis.

3. Result and Discussion

3.1 Results

3.1.1 Implementation of the K-Nearest Neighbor Algorithm

The implementation of the K-Nearest Neighbor (K-NN) algorithm in this study aims to predict retail stock needs based on historical sales data. The study utilizes a dataset comprising 54 records, which are divided into two parts: 15 records are used as training data, while the remaining 39 records are used as testing data. This division is intended to train the K-NN model to classify stock needs based on historical patterns and to test the accuracy of the predictions generated by the model. The initial step in implementing the K-NN algorithm is to determine the value of the parameter kkk, which represents the number of nearest neighbors to be considered in the classification process. The selection of the kkk value is crucial as it can significantly impact the overall performance of the model. In this study, the kkk value is set to 2, meaning that the model will consider the two nearest neighbors in the classification process. Once the kkk value is determined, the next step is to calculate the distance between the testing data and each training data point. The distance is calculated using the Euclidean Distance formula, which is mathematically expressed as:

$$D = \sqrt{(x1 - x2)^{\,2} + (y1 - y2)^2}$$

In this formula, x_1 and y_1 are the coordinates of the testing data, while x_2 and y_2 are the coordinates of the training data. After the distance between the testing data and the training data is calculated, the results are sorted from the smallest to the largest. The testing data is then classified into a particular group based on the majority value of the nearest neighbors obtained. If most of the nearest neighbors fall into the high stock needs category, the testing data will be classified as high stock needs, and vice versa. As an example of the implementation, the testing data used in this study includes various retail products with variables such as the number of items sold and the available stock. By applying the Euclidean Distance calculation, the classification results are obtained, grouping these products into either low or high stock needs categories. Table 1 presents the Euclidean Distance calculation results for each testing data. The results obtained from the calculation of the Euclidean Distance between the retail stock requirement object with high retail stock requirement and each neighbor can be seen in Table 1 below:

Table 1. Calculation of Euclidean Distance

Product Name	Sold Units	Unsold Units	Demand Level	Distance to Demand Level	k = 3
ABC Sweet Soy Sauce Pet 135ml	11	4	Low Demand	10	3
Vidoran Kids Chocolate Tp 115ml	45	60	High Demand	61	
Sunlight Lime Ref 210/220ml	17	30	High Demand	20	
Softies Mask 3ply 5s	9	77	High Demand	64	
Sari Roti Sobek Ckt Srkaya 214g	17	51	High Demand	40	
GMP Sugar 1kg	3	1	Low Demand	12	
Pronas Corned Beef Can 198g	7	5	Low Demand	8	2
Sari Roti Special White Bread	87	70	Low Demand	99	
Sari Roti Chocolate SW 49g	82	164	High Demand	169	
Sari Roti Sobek Chocolate 214g	79	18	Low Demand	73	
My Roti Tawar Funwari 8s	35	1	Low Demand	31	
Tropical Cooking Oil Bottle 2L	32	11	Low Demand	26	1
FS Setra Wangi Premium Rice 5kg	12	16	High Demand	7	
Sosro Teh Celup 30s	7	4	Low Demand	9	
Sania Cooking Oil Pouch 2L	69	58	Low Demand	77	

3.1.2 Evaluation of Testing Data Using Rapid Miner Tools

After the K-NN model is applied to the dataset, the next step is to evaluate the performance of the model. The evaluation is conducted using Rapid Miner, a software tool specifically designed for data analysis and machine learning algorithm implementation. Rapid Miner is chosen for its ability to handle complex data processing and provide various functions for model evaluation. The evaluation process begins with importing the previously processed data into Rapid Miner. The data is then separated into two sets: training and testing datasets. At this stage, relevant attributes and labels are also determined to ensure that the classification process runs smoothly. After that, the K-NN model is applied to the testing dataset using the K-NN operator available in Rapid Miner. This operator allows the model to predict stock needs categories based on the data trained earlier. The next step in the evaluation process is to define the role of each attribute in the dataset. This is done by adding the "Set Role" function operator to the process view in Rapid Miner. This operator is used to assign the role of which attribute will be used as the label in the prediction process, so the prediction results can be compared with the actual data to measure the model's accuracy.

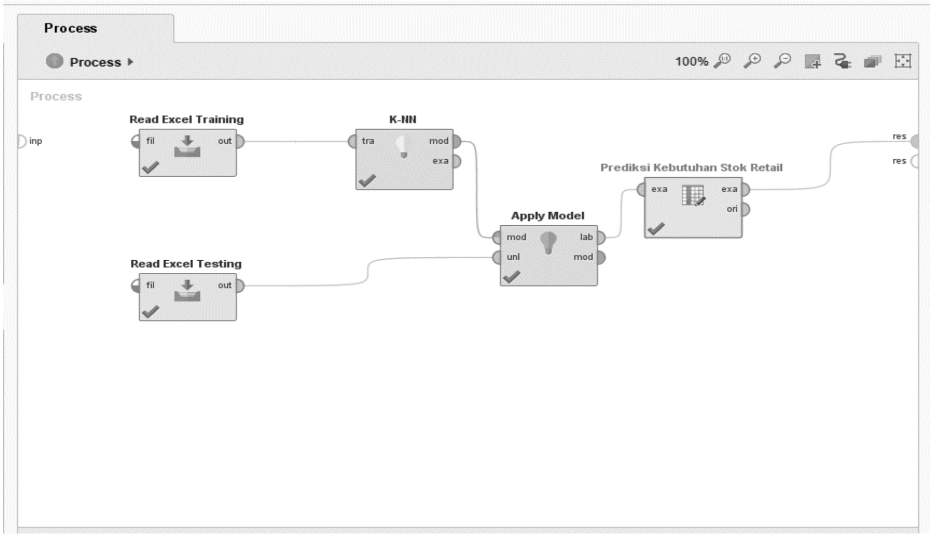

Figure 2. Implementation of the K-Nearest Neighbor Algorithm on Rapidminer

After all the settings are completed, the process is executed using the "Apply Model" and "Scoring" functions in Rapid Miner. These functions allow the model to be applied to the testing dataset and generate predictions that can then be evaluated. The results of this process are displayed in a table showing the classification of stock needs for each product. Figure 3 shows the prediction results obtained from 39 testing data, where each cluster has a confidence value that indicates the model's level of certainty regarding the generated predictions.

To measure the model's performance more thoroughly, testing is conducted using the K-Fold Cross Validation method. This method is chosen for its ability to reduce bias and overfitting in machine learning models. K-Fold Cross Validation works by dividing the dataset into several subsets, where each subset is used alternately as training data and testing data. In this study, the kkk value for Cross Validation is set to 5, meaning that the dataset is divided into five equally sized parts. The steps in this testing involve setting parameters in the Cross Validation operator in Rapid Miner. The parameters include the number of iterations (number of folds) and the type of sampling used (sampling type). In this testing, the number of folds parameter is set to 5, while the sampling type parameter is set to automatic, allowing Rapid Miner to automatically divide the dataset based on data distribution. After the parameters are set, the Cross Validation operator is executed by selecting the K-NN operator in the Training section and the Performance operator in the Testing section. The Performance operator is used to measure the model's performance based on metrics such as accuracy, precision, and recall. The testing results indicate that the K-NN model applied has a relatively high accuracy, with the values for accuracy, precision, and recall obtained presented in a Confusion Matrix table. The Confusion Matrix is a useful tool in evaluating model performance, as it allows the observation of the number of correct and incorrect predictions for each category. The accuracy calculation is performed by summing the True Positive (TP) and True Negative (TN) values, then dividing by the total number of data tested. Precision is calculated by dividing the number of correct positive predictions (True Positive) by the sum of correct positive (True Positive) and incorrect positive (False Positive) predictions. Meanwhile, recall is calculated by dividing the number of correct positive predictions (True Positive) by the sum of correct positive predictions (True Positive) and incorrect negative predictions (False Negative).

ExampleSet (39 examples, 4 special attributes, 3 regular attributes) Filter (39 / 39 examples):

Row No.	Kebutuhan ...	prediction(Kebutuh...	confidence(...	confidence(...	Nama Produk	Kebutuhan ...	Kebutuhan ...
1	29	Kebutuhan Rendah	0.804	0.196	PUCUK HAR...	58	?
2	20	Kebutuhan Rendah	0.814	0.186	SOSRO TEH ...	56	?
3	164	Kebutuhan Tinggi	0.382	0.618	AQUA AIR PE...	82	?
4	205	Kebutuhan Tinggi	0.192	0.808	FORVITA MA...	12	?
5	8	Kebutuhan Rendah	0.824	0.176	SEDAAP MIE ...	70	?
6	3	Kebutuhan Rendah	0.872	0.128	KOREA GLO...	4	?
7	1	Kebutuhan Rendah	0.875	0.125	POND'S FF P...	6	?
8	2	Kebutuhan Rendah	0.873	0.127	HELLO PAN...	4	?
9	4	Kebutuhan Rendah	0.867	0.133	DOWNY DAR...	4	?
10	146	Kebutuhan Tinggi	0.189	0.811	TEH GELAS ...	5	?
11	31	Kebutuhan Tinggi	0.393	0.607	ALE-ALE JER...	33	?
12	34	Kebutuhan Tinggi	0.370	0.630	MOLTO SB AI...	4	?
13	28	Kebutuhan Tinggi	0.370	0.630	YAKULT 5S	10	?
14	14	Kebutuhan Rendah	0.630	0.370	WALL'S P.PO...	36	?
15	53	Kebutuhan Rendah	0.651	0.349	WALL'S POP...	82	?
16	20	Kebutuhan Rendah	0.816	0.184	TELUR AYAM...	60	?
17	4	Kebutuhan Rendah	0.819	0.181	ECO BAG 38...	29	?

Figure 3. Results of Predicting Test Data Objects on Rapid Miner

The evaluation results demonstrate that the K-NN model applied to this dataset performs well, with a high accuracy rate and adequate precision and recall. This evaluation provides confidence that the K-NN model can be relied upon to predict retail stock needs based on historical sales data. With these results, the K-NN model is expected to be an effective tool in helping retail stores manage their stock more efficiently, optimizing sales and improving customer satisfaction. Through the implementation and evaluation of this model, this study successfully shows that the K-NN algorithm can be used to classify retail stock needs with a high degree of accuracy. The results of this implementation and evaluation can be used as a basis for decision-making in stock management, as well as a reference for further research in the fields of data mining and machine learning.

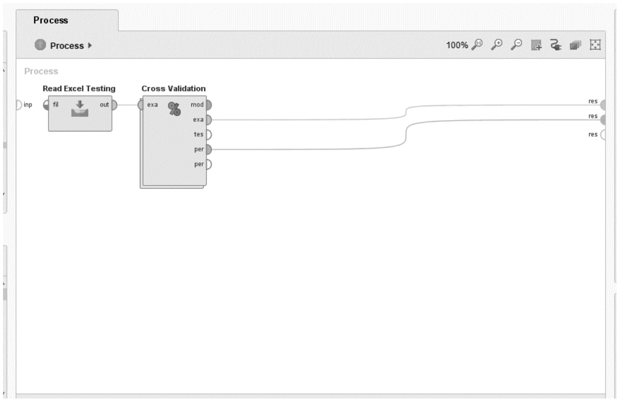

Figure 4. Evaluation of KNN Model with Cross Validation Process 1

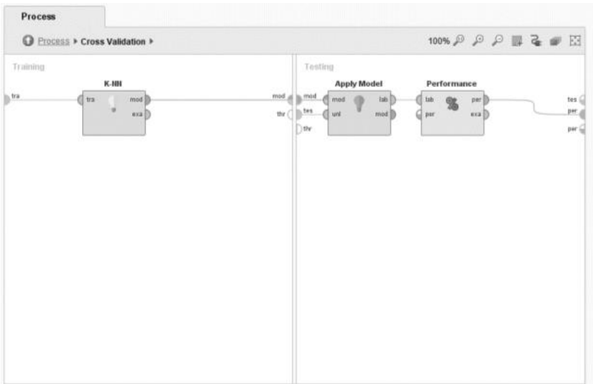

Figure 5. Evaluation of KNN Model with Cross Validation Process 2

This figure illustrates the first phase of the K-Fold Cross Validation process applied to evaluate the performance of the K-Nearest Neighbor (K-NN) model. During this phase, the dataset is divided into several subsets, where each subset is used alternately as training data and testing data. The process aims to assess the consistency of the model's performance across different segments of the data. The visualization highlights the workflow within the software, demonstrating how the data is processed and evaluated to generate the necessary accuracy metrics. This figure represents the continuation of the Cross Validation process, focusing on the subsequent iterations in the K-Fold Cross Validation. It provides a detailed view of how the model's predictions are validated against the testing subsets. Each fold in the Cross Validation process contributes to the overall evaluation, ensuring that the model is robust and not overfitted to any particular subset of the data. The figure shows the step-by-step progression of the process, which culminates in a comprehensive evaluation of the model's performance.

PerformanceVector

```
PerformanceVector:
accuracy: 97.50% +/- 7.50% (micro average: 97.44%)
ConfusionMatrix:
True:   Kebutuhan Rendah        Kebutuhan Tinggi
Kebutuhan Rendah:       25      1
Kebutuhan Tinggi:       0       13
precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Kebutuhan Tinggi)
ConfusionMatrix:
True:   Kebutuhan Rendah        Kebutuhan Tinggi
Kebutuhan Rendah:       25      1
Kebutuhan Tinggi:       0       13
recall: 95.00% +/- 15.00% (micro average: 92.86%) (positive class: Kebutuhan Tinggi)
ConfusionMatrix:
True:   Kebutuhan Rendah        Kebutuhan Tinggi
Kebutuhan Rendah:       25      1
Kebutuhan Tinggi:       0       13
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Kebutuhan Tinggi)
AUC: 0.950 +/- 0.150 (micro average: 0.950) (positive class: Kebutuhan Tinggi)
AUC (pessimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Kebutuhan Tinggi)
```

Figure 6. Results of KNN Model Evaluation with Cross Validation

This figure presents the final results obtained from the K-Fold Cross Validation applied to the K-NN model. The results include key performance metrics such as accuracy, precision, and recall, calculated over all the folds in the validation process. The figure encapsulates the overall effectiveness of the model, demonstrating how well it can generalize to unseen data. The results provide insight into the reliability of the model in predicting retail stock needs based on historical data.

Table View   Plot View

accuracy: 97.50% +/- 7.50% (micro average: 97.44%)

	true Kebutuhan Rendah	true Kebutuhan Tinggi	class precision
pred. Kebutuhan Rendah	25	1	96.15%
pred. Kebutuhan Tinggi	0	13	100.00%
class recall	100.00%	92.86%	

Figure 7. Accuracy K-NN

Figure 7 shows the accuracy metric of the K-NN model after the evaluation process. Accuracy is calculated as the ratio of correct predictions (both True Positives and True Negatives) to the total number of predictions. This figure highlights the model's ability to correctly classify the retail stock needs into either high or low demand categories. The higher the accuracy, the better the model is at making correct predictions across the dataset. This figure depicts the precision of the K-NN model, which is a measure of the accuracy of the positive predictions. Precision is calculated as the ratio of True Positives to the sum of True Positives and False Positives. A high precision indicates that the model is effective at minimizing false positives, ensuring that when it predicts high demand, it is usually correct. This figure underscores the reliability of the model in making specific and accurate predictions.

Table View   Plot View

precision: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Kebutuhan Tinggi)

	true Kebutuhan Rendah	true Kebutuhan Tinggi	class precision
pred. Kebutuhan Rendah	25	1	96.15%
pred. Kebutuhan Tinggi	0	13	100.00%
class recall	100.00%	92.86%	

Figure 8. Percision K-NN

Table View Plot View

recall: 95.00% +/- 15.00% (micro average: 92.86%) (positive class: Kebutuhan Tinggi)

	true Kebutuhan Rendah	true Kebutuhan Tinggi	class precision
pred. Kebutuhan Rendah	25	1	96.15%
pred. Kebutuhan Tinggi	0	13	100.00%
class recall	100.00%	92.86%	

Figure 9. Recall K-NN

Figure 9 illustrates the recall metric for the K-NN model, representing the model's ability to capture all relevant instances. Recall is calculated as the ratio of True Positives to the sum of True Positives and False Negatives. A high recall means that the model is successful in identifying most of the high demand instances, even if some of the predictions might include false positives. This figure emphasizes the model's capacity to cover all potential high demand cases, contributing to a comprehensive stock management strategy.

3.2 Discussion

The testing results from the dataset of 39 records indicated that the K-Nearest Neighbor (K-NN) algorithm effectively classified the retail stock needs into high and low-demand categories. The model was applied to the testing dataset, which contained 39 records, and predictions were made based on the criteria established by the K-NN algorithm in this study. The results showed that 25 objects were predicted to fall into the low-demand category, while 14 objects were classified as high-demand. This classification was achieved using the algorithm's defined parameters and was further validated through testing using the Rapid Miner tool. The evaluation conducted with Rapid Miner yielded results that aligned with the initial predictions, confirming the model's effectiveness. Specifically, the model correctly classified 25 objects as low demand and 14 as high demand, achieving an accuracy rate of 97.50%. This high level of accuracy suggests that the K-NN algorithm is highly reliable in predicting retail stock needs based on historical data. Further analysis of the model's performance was carried out using the cross-validation function in Rapid Miner. Cross-validation is a robust method that allows repeated testing of data samples and divides the dataset into equally sized parts, ensuring that each part is used for training and validation. By applying cross-validation to both the training and testing datasets, a Confusion Matrix was generated, which provided detailed metrics such as Accuracy, Precision, and Recall. The results from the 10-fold cross-validation process indicated that the model achieved an Accuracy of 97.50%, with a Standard Deviation of +/- 7.50%. Additionally, the Precision was measured at 100%, and the Recall was at 95.0%. These metrics demonstrate the model's strong performance across various tests, with the low Standard Deviation indicating consistent accuracy across different iterations. The Precision of 100% highlights the model's effectiveness in minimizing false positives, while the Recall of 95.0% suggests that the model successfully identified most of the high-demand cases. By evaluating the model's performance with cross-validation, the resulting accuracy metrics were reliable and showed minimal variance, ensuring that the average accuracy closely matched the accuracy of each iteration. This rigorous evaluation process confirms the robustness of the K-NN model in predicting retail stock needs, making it a valuable tool for efficient retail stock management.

4. Related Work

The literature on applying K-Nearest Neighbor (K-NN) and other data mining techniques for predictive modelling in various fields provides valuable insights into the adaptability and efficacy of these methods in different contexts. The work by Fahlevi (2020) on implementing a genetically modified K-Nearest Neighbor for classifying recipients of government assistance illustrates the flexibility of the K-NN algorithm when adapted to specific needs. Fahlevi research demonstrates that changing the K-NN algorithm can enhance its performance in specialized applications, particularly in socio-economic settings where classification based on demographic and economic criteria is crucial [3]. This adaptation shows that K-NN is helpful in its traditional form and can be tailored to meet the demands of more complex datasets that require nuanced classification criteria. Similarly, Iku, Mustofa, and Kumala (2019) employed the K-NN method to predict retail rice prices in traditional markets in Gorontalo, demonstrating its applicability in an economic forecasting [6]. Their study highlights how K-NN can effectively predict market prices based on historical sales data. The accuracy of their predictions reinforces the value of K-NN in economic and retail applications, where price predictions can significantly impact decision-making processes for vendors and buyers alike. The ability of K-NN to provide reliable price forecasts suggests that this method can be a powerful tool in managing supply chains and inventory, particularly in environments where price volatility is a concern. The research conducted by Hasanah *et al*. (2019) delves into using data mining techniques, including K-NN, for clustering and classification tasks. Their study emphasizes the importance of data mining in extracting valuable patterns from large datasets, which can then be used to inform strategic decisions in various domains. Hasanah's work highlights the role of clustering in grouping similar data points, which is essential for adequate classification. By clustering data before applying K-NN, researchers and practitioners can enhance the accuracy of their predictions, making this combination a potent tool in data analysis [2].

Darmi and Setiawan (2016) explored the application of the K-Means clustering algorithm in categorizing product sales. Their research is particularly relevant in retail, where understanding customer purchasing behaviour and product demand is vital. By using K-Means clustering, they were able to segment products into different categories based on sales data, which can help retailers optimize their inventory and marketing strategies [1]. The study underscores the synergy between clustering techniques like K-Means and classification algorithms like K-NN, where clustering can preprocess data before classification, leading to more refined and accurate results. The study by Hamdani (2020) on customer loyalty segmentation using a combination of RFM (Recency, Frequency, Monetary) analysis and K-Means clustering further supports the integration of clustering techniques with predictive algorithms [7]. The research demonstrated how targeted marketing strategies could be developed to enhance customer retention by segmenting customers based on their purchase history. This approach is highly applicable in retail and service industries, where understanding customer behaviour is critical to sustaining competitive advantage. Pramana *et al*. (2023) applied the Naive Bayes algorithm to predict best-selling products at CV Akusara Jaya Abadi in the context of predictive modelling for product sales. Their findings highlight the importance of selecting the suitable predictive algorithm for specific datasets, as different algorithms may yield varying degrees of accuracy depending on the nature of the data. Naive Bayes, known for its simplicity and efficiency, proved effective in this context, suggesting that while K-NN and clustering techniques are powerful, other algorithms like Naive Bayes also have significant roles in predictive analytics [8].

The application of K-Means clustering in inventory management was further explored by Prastiwi, Pricilia, and Rasywir (2022), who focused on determining stock levels in a mini-market setting. Their study demonstrated how clustering could segment inventory based on turnover rates, which helps maintain optimal stock levels. This research is precious for small to medium-sized enterprises where inventory management can significantly impact profitability [9]. The research by Febriyanti, Bancin, and Amanda (2022) on clustering sales data at PT Swasti Tunggal Mandiri reinforces the utility of K-Means in managing large datasets typical in retail environments. By grouping products based on sales performance, their study provided insights into which products should be prioritized in restocking efforts, thus helping to optimize inventory management [10]. Sallaby *et al*. (2022) extended the application of K-Means clustering to analyze product sales in Bengkulu, Indonesia, further validating the algorithm's effectiveness in diverse geographical and market contexts. Their study indicates the universal applicability of K-Means across different retail environments, reinforcing that clustering is a fundamental step in data preprocessing before applying classification techniques like K-NN [11]. Fakhriza and Umam (2021) applied K-means clustering to analyze top-selling products at PT Sukanda Djaya, emphasizing how clustering can reveal underlying patterns in sales data that are not immediately apparent. Their study demonstrates the value of clustering in uncovering hidden relationships within data, which can be leveraged for more effective decision-making [12]. Muttaqin *et al*. (2020) utilized the K-NN algorithm to predict sales based on web traffic data in online retail [4].

The application of the K-Nearest Neighbor (K-NN) algorithm has been explored in various domains, each study contributing unique insights into its effectiveness and adaptability. Alfani, Rozi, and Sukmana (2021) examined the prediction of sales for Unilever products using the K-NN algorithm. Their study demonstrated the algorithm's capability to accurately forecast sales trends, highlighting its utility in the consumer goods sector where sales volume, product type, and consumer behavior are critical variables. This research aligns with broader efforts to use K-NN for predicting market dynamics, especially in environments characterized by fluctuating demand. What distinguishes their work is the focus on a well-established consumer brand, where the ability to predict demand can directly impact inventory management and marketing strategies.

Rismala, Ali, and Rinaldi (2023) applied K-NN to predict the best-selling motorcycle models, considering its effectiveness in niche markets such as the automotive industry [14]. Their study, similar to Alfani et al. in its focus on sales prediction but differs in its application to a more specialized product category. The automotive sector presents unique challenges, such as brand loyalty and seasonality, which require appropriate prediction models [13]. The study's focus on a specific and competitive market segment adds value to the existing body of knowledge, particularly in understanding how K-NN can be adapted to different types of products with different demand drivers. In a related study, Irfayanti and Satria (2020) used K-NN to predict instant concrete sales at PT Decon Multi Industri. This study further confirms the versatility of K-NN across industries, demonstrating its applicability in the construction sector, which is subject to cyclical demand and economic conditions [15]. Compared to the consumer goods and automotive sectors, building materials such as concrete pose different forecasting challenges, such as the impact of large-scale projects on demand. Irfayanti and Satria’s research thus contributes to the literature by demonstrating the application of K-NN in sectors where demand forecasting is influenced by macroeconomic factors. Saepudin et al. (2020) explores the optimization of K-NN through Particle Swarm Optimization (PSO) for sentiment analysis in social media. This study stands out by integrating K-NN with another optimization technique, PSO, to improve the accuracy of the algorithm in non-traditional applications such as sentiment analysis [16]. Unlike the sales prediction models discussed earlier, this study focuses on text data rather than numeric sales data. The integration of K-NN with PSO is an innovative approach to improve the performance of the algorithm, especially in processing and analyzing large amounts of unstructured data, such as social media posts. Similarly, Assistance (2021) focuses on optimizing K-NN parameters to classify development assistance priorities in rural areas [17]. This study, unlike the commercial applications discussed in other studies, addresses socio-economic challenges, where accurate classification can lead to more effective resource allocation. By optimizing parameters such as the number of neighbors and distance metrics, this study improves the accuracy of the model in making decisions that directly impact the welfare of the community. The emphasis on social impact distinguishes this study from other, more commercially oriented studies. Roni, Crysdian, and History (2022) review the potential of marketplace platforms to increase sales, discussing the role of K-NN in optimizing marketplace operations [18]. This review is broader in scope, covering a variety of data mining techniques, including K-NN, and their applications in e-commerce. The focus on digital marketplaces provides a different perspective from the more traditional retail environment examined in other studies. The study’s comprehensive analysis of K-NN applications in e-commerce offers insights into how digital platforms can leverage predictive algorithms to improve customer segmentation and enhance sales strategies. Imron (2020) improves the accuracy of K-NN through Z-Score normalization and PSO in predicting customer churn [19]. This study differs from other studies in that it focuses on customer retention, rather than sales prediction or product demand forecasting. The use of normalization techniques and optimization algorithms highlights the importance of preprocessing and algorithmic improvement in improving prediction accuracy. Imron’s work contributes to the understanding of how K-NN can be optimized for specific tasks, particularly in competitive industries where customer retention is critical. Desyanti and Wulandari (2022) applied K-NN to predict motorcycle stock levels, contributing to the literature on inventory management. Their study is particularly relevant in industries where inventory costs are significant, and accurate demand forecasting can yield substantial cost savings [20]. Unlike studies that focus on sales prediction, this study highlights the utility of K-NN in operational aspects of a business, such as inventory control. The focus on optimizing stock levels shows how K-NN can be applied beyond sales forecasting to ensure that operational efficiency is maintained. Their research showcases the algorithm's versatility in adapting to different data types, such as web-based metrics, and its ability to provide actionable insights for online retailers looking to optimize their digital marketing strategies. These studies underscore the adaptability and effectiveness of K-NN and clustering algorithms in various domains, from retail and marketing to socio-economic classification and inventory management. The consistent findings across different studies demonstrate that these algorithms are versatile and crucial in extracting actionable insights from large datasets, thereby supporting more informed decision-making processes across industries.

5. Conclusion

Based on the author's research results, the data on the need for goods for the classification method approach can be applied in analyzing data to predict the need for retail stock goods with low and high levels of product needs. The results of the K-Nearest Neighbor algorithm can be implemented where the results also show a new insight, namely predicting the grouping of the level of need. There are 54 samples tested to indicate the level of need in 2 groups, namely low or high. The model scenario results show 24 objects in the Low-needs group and 14 in the High-needs group. Evaluation and performance testing using the Rapid Miner tool can also produce a relevant picture of the modelled scenario. The model implemented using the K-NN algorithm has an Accuracy value of 97.50% with a Standard Deviation of +/- 750%, then a Precision value of

100%, and a Recall value of 950%. By measuring model performance with cross-validation, the resulting accuracy has a standard deviation value, which aims to see the distance between the average accuracy and the accuracy of each experiment (iteration).

References

[1] Darmi, Y., & Setiawan, A. (2016). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*, Universitas Muhammadiyah Bengkulu, 12(2), 148–157.

[2] Hasanah, H., Larasati, W., & Komputer, F. I. (2019). Pemanfaatan Data Mining Untuk Mengelompokkan. *Universitas Duta Bangsa Surakarta*, 292–300.

[3] Fahlevi, R. (2020). Penerapan Genetic Modified K-Nearest Neighbor Dalam Klasifikasi Penerima Beras Sejahtera.

[4] Muttaqin, A. G., Auliasari, K., & Wahyuni, F. S. (2020). Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada PT Wika Industri Energy. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 4(2), 1–6. https://doi.org/10.36040/Jati.V4i2.2728

[5] Mujilahwati, S., & Windasari, L. D. (2024). Implementasi Metode K-Nearest Neighbor (K-Nn) Untuk Memprediksi Penjualan Buah Di Indonesia Berbasis Website. *Seminar Nasional Teknologi Sains*, 3(1), 7–14. https://doi.org/10.29407/Stains.V3i1.4077

[6] Iku, M. H., Mustofa, Y. A., & Kumala, I. S. (2019). Metode K-Nearest Neighbor Untuk Memprediksi Harga Eceran Beras Di Pasar Tradisional Gorontalo. *Jurnal Cosphi*, 3(2), 50–53.

[7] Hamdani, R. (2020). Pengelompokkan Loyalitas Pelanggan Dengan Menggunakan Kombinasi Rfm Dan Algoritma K-Means. *Journal Of Informatic Pelita Nusantara*, *5*(1), 7-13.

[8] Pramana, I., & Sudiarsa, I. W. (2023). Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada Cv Akusara Jaya Abadi. *Jatisi (Jurnal Teknik Informatika)*, 10(4), 518–534.

[9] Prastiwi, H., Pricilia, J., & Rasywir, E. (2022). Implementasi Data Mining Untuk Menentukan Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informasi dan Rekayasa Komputer (Jakakom)*, 2(1), 141–148. https://doi.org/10.33998/Jakakom.2022.2.1.34

[10] Febriyanti, A., Bancin, P. V., & Amanda, S. (2022). Penerapan Metode K-Means Clustering Dalam Pengelompokan Data Penjualan Barang Pada PT Swasti Tunggal Mandiri. *J-Com (Journal of Computer Science)*, 2(3), 205–210. https://doi.org/10.33330/J-Com.V2i3.1936

[11] Sallaby, A. F., Alinse, R. T., Sari, V. N., & Ramadani, T. (2022). Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil Penjualan Di Toko Widya Bengkulu. *Jurnal Media Infotama*, 18(1).

[12] Fakhriza, M. H., & Umam, K. (2021). Analisis Produk Terlaris Menggunakan Metode K-Means Clustering Pada PT Sukanda Djaya. *Jika (Jurnal Informasi dan Komputer Akuntansi)*, 5(1), 8. https://doi.org/10.31000/Jika.V5i1.3236

[13] Alfani, W. P. R., Rozi, F., & Sukmana, F. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *Jipi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 6(1), 155–160. https://doi.org/10.29100/Jipi.V6i1.1910

[14] Rismala, R., Ali, I., & Rinaldi, A. R. (2023). Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 585–590. https://doi.org/10.36040/Jati.V7i1.6419

[15] Irfayanti, Y., & Satria, M. I. (2020). Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Beton Instan Pada PT Decon Multi Industri. *Sentinel*, 3(2), 288–300. https://doi.org/10.56622/Sentineljournal.V3i2.25

[16] Saepudin, A., Aryanti, R., Fitriani, E., & Dahlia, D. (2020). Optimasi Algoritma SVM Dan k-NN Berbasis Particle Swarm Optimization Pada Analisis Sentimen Fenomena Tagar# 2019GantiPresiden. *Jurnal Khatulistiwa Informatika*, *6*(1), 95-102.

[17] Assistance, V. D. (2021). Optimasi Parameter K Pada Algoritma K-Nn Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa. *Jurnal Teknik Informatika*, 20(1), 83–96.

[18] Roni, S., Crysdian, C., & History, A. (2022). Jurnal Teknologi Dan Manajemen Informatika Studi Literature Analisis Potensi Pasar Marketplace Terhadap Penjualan. *Jurnal Teknologi dan Manajemen Informatika*, 8(2), 134–142.

[19] Imron, M. A. (2020). Peningkatan Akurasi Algoritma K-Nearest Neighbor Menggunakan Normalisasi Z-Score dan Particle Swarm Optimization untuk Prediksi Customer Churn.

[20] Desyanti, D., & Wulandari, D. (2022). Implementasi Algoritma K-Nearest Neighbour Dalam Memprediksi Stok Sepeda Motor. *Building Informatics Technology Science*, 4(3), 1576–1581. https://doi.org/10.47065/Bits.V4i3.2579.