THE PREDICTION OF PRODUCT SALES LEVEL USING K-NEAREST NEIGHBOR AND NAÏVE BAYES ALGORITHMS (CASE STUDY: PT KOTAMAS BALI)

Aris Setiawan Miha Djami[1*)], Nengah Widya Utami[2], A A Istri Ita Paramitha[3]

[1]Teknik Informatika, [2]Sistem Informasi Akuntansi, [3]Sistem Informasi
Primakara University
www.primakara.ac.id
[1*)]arismdj@gmail.com; [2]widya@primakara.ac.id; [3]ita@primakara.ac.id

(*) Corresponding Author

Abstract— PT Kotamas Bali is a company that operates in tableware and kitchenware, where every sale is sold at various counters. Product sales are permanently printed and entered in sales reports, and there are problems such as the fact that the product is sold very much and it is difficult to see the rate of sale of most products, lots, and not lots. Then, to see the level of product sales, it is necessary to use data mining techniques with the method of Knowledge Discovery in Databases to predict the purchase rate of products using the two algorithms K-Nearest Neighbor and Naïve Bayes. The purpose of this research is so that PT Kotamas Bali can see the sales rate of each product sold so that there is no accumulation of goods and more focus on the most marketed products. These two algorithms result in different accuracy on the 90:10 data split, where the K-Nearest Neighbor algorithm successfully predicted the sales rate of the product with a 99% accuracy rate and was categorized as an excellent classification. The Naïve Bayes algorithm failed to make predictions with an accuracy of only 54% and was classified as a failure classification. ROC performance results on the K-Nearest Neighbor algorithm with an AUC value of 99% and the Naïve Bayes algorithm with an AUC of 74%. K-Nearest Neighbor managed to obtain the highest accuracy, while the Naïve Bayes algorithm failed to conduct classification.

Keywords: Data Mining; Product Sales Level; Knowledge Discovery in Databases; K-Nearest Neighbor; Naïve Bayes.

Abstrak— *PT Kotamas Bali merupakan perusahaan yang bergerak dibidang tableware dan kitchenware dimana setiap penjualan dijual diberbagai counter. Penjualan produk selalu diinput dan dimasukkan kedalam laporan penjualan dan terdapat permasalahan seperti produk yang dijual sangatlah banyak dan susah untuk melihat tingkat penjualan produk paling laris, laris, dan tidak laris. Maka untuk melihat tingkat penjualan produk perlu dilakukan teknik data mining dengan metode Knowledge Discovery in Databases untuk memprediksi tingkat penjualan produk dengan menggunakan dua algoritma K-Nearest Neighbor dan Naïve Bayes. Tujuan penelitian ini agar pihak PT Kotamas Bali dapat melihat tingkat penjualan produk dari setiap produk yang dijual agar tidak terjadi penumpukan barang dan lebih fokus pada produk yang paling laris dan laris. Hasil klasifikasi tingkat penjualan produk dengan dua algoritma ini menghasilkan akurasi yang berbeda pada split data 90:10 dimana algoritma K-Nearest Neighbor berhasil memprediksi tingkat penjualan produk dengan tingkat akurasi 99% dan dikategorikan sebagai excellent classification dan algoritma Naïve Bayes gagal melakukan prediksi dengan akurasi hanya 54% dan dikategorikan sebagai failure classification. Hasil kinerja ROC pada algoritma K-Nearest Neighbor dengan nilai AUC sebesar 99% dan algoritma Naïve Bayes dengan nilai AUC sebesar 74%. Berdasarkan evaluasi kedua algoritma tersebut, K-Nearest Neighbor berhasil mendapatkan akurasi tertinggi dan algoritma Naïve Bayes gagal melakukan klasifikasi.*

Kata Kunci*: data mining, tingkat penjualan produk, Knowledge Discovery in Databases, K-Nearest Neighbor, Naïve Bayes.*

INTRODUCTION

Sales are the most crucial factor for a company, as they are the source of its income. If a company generates high revenue, it is expected to gain profits to sustain its operations further. Therefore, to increase sales, a company requires sales strategies.

Sales strategies are highly crucial and influential in a business or enterprise. Therefore, it is unsurprising that every company implements

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4420

various marketing strategies to enhance sales. With increased sales volume, the company can achieve the desired profits. The more effective the marketing strategies the company implements, the greater the chances it has to expand its market share (Tarmizi, 2021).

Companies can analyze and predict sales using various methods, one of which is data mining. Data mining is extracting information from large datasets stored in databases, warehouses, or other storage media. (Budiyanto & Dwiasnati, 2018) Data mining has been widely utilized in various industries, including retail, finance, healthcare, transportation, manufacturing, and aerospace. Data mining tools and techniques have proven beneficial, especially in strategic decision-making. Using pattern recognition technologies, statistical methods, and mathematical algorithms to sift through stored information, data mining has enabled analysts to identify facts, relationships, trends, patterns, exceptions, and anomalies that may have otherwise gone unnoticed (Arhami & Muhammad, 2020).

The classification technique is a crucial component of data mining. It involves learning from a dataset to generate rules that can classify or recognize new, unseen data. Classification is the process of assigning data objects to pre-defined categories or classes. The classification application is wide-ranging, including customer profiling, fraud detection, medical diagnosis, sales prediction, and more (Suyanto, 2017). Data mining uses Various methods for classification, such as K-Nearest Neighbor, Naïve Bayes, Decision Tree, Neural Network, and others. The author employed two classification algorithms, KNN and Naïve Bayes, for predicting product sales levels.

K-Nearest Neighbors (KNN) is a method used to classify objects based on the training data points that are closest in distance to the thing. The training data is projected into a high-dimensional space, where each dimension represents a feature of the data (Soepriyanto, 2021). The advantages of the KNN algorithm include its ease of understanding and implementation, effectiveness with large training datasets, and ability to produce accurate results (Saputra & Ajeng Kristiyanti, 2022).

In addition to KNN, the Naïve Bayes algorithm is a simple probabilistic classification method that calculates a set of probabilities by summing the frequencies and combinations of values from the given dataset as the probabilities of having specific characteristics or interactions with other category characteristics (Budiyanto & Dwiasnati, 2018). The advantages of the Naïve Bayes algorithm include its ease of understanding, suitability for both qualitative and quantitative data, not requiring a large amount of data, not needing extensive data training, fast and efficient calculations, the ability to ignore missing values in the math, and applicability to both binary and multiclass classification problems (Saputra & Ajeng Kristiyanti, 2022).

Many studies have been conducted regarding these methods. Research on data mining applications for predicting Unilever product sales using the KNN method by Aisha Alfani et al. (2021) found high and low accuracy prediction levels. The highest accuracy in predicting product sales was 86.66% (Alfani WPR et al., 2021). Another study by Yulianto (2019) examined product sales prediction using the Naïve Bayes algorithm, with a correctly classified accuracy rate of 63.414% (Yulianto, 2019). To obtain a good prediction of product sales levels, it is advisable to compare two algorithms to determine which algorithm is more effective. It can serve as a benchmark for the best algorithm's performance.

PT Kotamas Bali is a subsidiary of PT Kotamas Makmur, a company established in 1972 and operating in the tableware and kitchenware category. It is a major distributor and retailer of international brands such as Corelle, Tefal, Neoflam, Lumerpa, and Ocean Glassware. PT Kotamas Bali is one of the company's branches in Denpasar, Bali (*Kotamas Makmur - Houseware & Lifestyle Products*, 2022). Sales are made through various counters, including Hypermart, Grandlucky Sunset, Grandlucky Sanur, Transmart, and others. Sales are also conducted in several star-rated hotels. Each deal is permanently recorded and entered into reports using Excel format.

After conducting interviews and observations with the branch manager of PT Kotamas Bali, Ni Made Rai Indrawati, there are issues regarding the large number of products being sold, making it difficult to determine the sales levels of frequently purchased and less popular products. To avoid accumulating excessive stock, it is necessary to make predictions that can assist the company in identifying which products will be bought next, allowing for the reduction of stock for less popular items and increasing supply for frequently purchased items. PT Kotamas Bali inputs sales data into MS Excel, involving over 1,000 types of products sold, ensuring monthly product sales are permanently recorded. Data mining techniques can help address existing issues at PT Kotamas Bali.

Based on the issues presented, the author researched "Prediction Of Product Sales Level Using K-Nearest Neighbors and Naïve Bayes Algorithms (Case Study: PT Kotamas Bali)." This research aims to assist PT Kotamas Bali in

predicting product sales levels, which can serve as a basis for stock management decisions.

MATERIALS AND METHODS

The data mining research method used is the Knowledge Discovery in Databases (KDD) method, where this method is used as a stage in the research flow. The secondary data source in this research is the branch manager of PT Kotamas Bali, who provided a dataset consisting of product sales data from 2020-2022.

Method of Collecting Data

By conducting observations, interviews, and unstructured documentation with the head of the branch of PT Kotamas Bali, where the form of the questions given is open, depending on the understanding of the researchers and the data obtained at the time. The primary source of data in this research is the head of the branch of PT Kotamas Bali, which provides a set of product sales data from each counter.

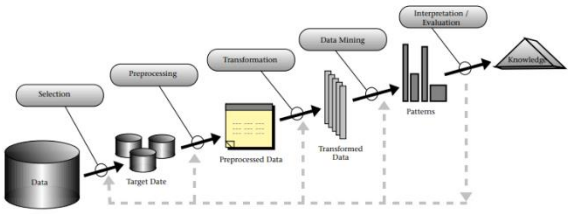
Datasets

The data used is the product sales data from 2020-2022. The author selected the data using Microsoft Excel with the file format DataPenjualanProduk20-22.csv. An example of the chosen product sales data from 2020-2022 can be seen in Table 1.

Table 1. Product Sales Dataset 2020-2022

No	Items	Product Quantity	Sold Quantities	Category
1	6 Cup Muffin Pan Grey	432	44	Tidak Laris
2	12 Cup Muffin Pan Grey	1540	770	Tidak Laris
3	12 Cup Mini Muffin Pan Grey	1320	1050	Laris
4	6 Cup Donut Pan Grey	516	96	Tidak Laris
5	Large Loaf Pan Grey	144	21	Tidak Laris
6	Small Loaf Pan Grey	400	374	Paling Laris
7	Square Pan Grey	72	24	Tidak Laris
8	Rectangular Pan Grey	214	118	Tidak Laris
9	Medium Cookie Sheet Grey	321	121	Tidak Laris
10	Round Pan Grey	36	0	Tidak Laris
....	Round Tart Pan Grey	48	38	Laris
1.375	EC-DC-W28GI	545	345	Tidak Laris
1.376	EC-DC-W30I	760	560	Laris
1.377	EC-DC-C20I	450	250	Tidak Laris

Knowledge Discovery in Databases



Sumber : (Fayyad et al., 1996)
Figure 1. KDD Stage

Data Selection

At the data selection stage, the data obtained from the sales data at PT Kotamas Bali is analyzed to extract the necessary parameters for performing classification prediction. The attributes taken from the data include Items, Product Quantity, Sold Quantities, and Category.

Data Preprocessing

The next step is data preprocessing, where the data obtained will be cleared so that the data classification process becomes more accurate. Data removed is missing or null, duplicate data, and so on. After preprocessing, the data that was initially 1,377 now becomes 1,090 net data, as shown in Table 2. Split data will be divided into 70:30, 80:20, and 90:10.

Table 2. Data Cleaning

Items	1090
Product Quantity	1090
Sold Quantities	1090
Category	1090

Data Transformation

Data that has been cleared will be transformed. Processing the data with Python for classification will convert the data into a format that can be read and processed by Python. Transformation by converting data into numerical data by marking the Category column into three classes: 0 as the Laris class, 1 as the Paling Laris, and 2 as the Tidak Laris class.

Data Mining

In this section, data mining will be carried out to find patterns or information from product

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4420

sales data. The data mining algorithms are K-Nearest Neighbor and Naïve Bayes.

1. K-Nearest Neighbor Algorithm

The formula used for calculations using the K-Nearest Neighbor algorithm is the Euclidean Distance. (Muqorobin et al., 2020)

$$D(x,y) = \sqrt{\Sigma_{k-1}^{n}}(x_k - y_k)^2 \quad \text{..... (1)}$$

a. Choose the value of k
b. Calculate the distance of data testing and training using the Euclidean distance formula.

X_k = training data
Y_k = testing data
D (x, y) = distance
k = variable data
n = dimension data

c. Calculate all the distances.
d. Select the nearest distance with the parameter k
e. Select the highest number of classes and then classify

2. Naïve Bayes Algorithm

The formula used for calculations using the Naïve Bayes algorithm is the Bayes Theorem. (Romadhon & Kurniawan, 2021)

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad \text{..... (1)}$$

P(A|B) : The Posterior value of AwhenB is occurred
P(B|A) : The Likelihood value of B when A is occurred
P(A) : The Prior value of class A
P(B) : The Evidence value of a class

Interpretation / Evaluation

After the data mining process is completed, patterns or information can be detected in the performance of both the K-Nearest Neighbor and Naïve Bayes algorithms. By examining the level of accuracy of both algorithms using the confusion matrix, it can be used as an informational tool in predicting the rate of product sales and comparing the performance of the two algorithms.

RESULTS AND DISCUSSION

Knowledge Discovery from Data (KDD) will be processed using Python. Python is a powerful programming language with more straightforward and common scripting, and it is becoming popular in various fields of computer science. (Swamynathan, 2019)

Metrix Evaluation

In the concept of data mining, a method is used to measure the accuracy of data and apply it to a decision support system called a confusion matrix. The confusion matrix has four terms that describe the results of measurement performance classifications: True Negative (TN), False Positive (FP), True positive (TP), and False negative (FN). (Rahmad, F. Suryanto, Y. Ramli, 2020). The results of the metric evaluation below use a 90:10 data split.

1. Confusion Matrix K-Nearest Neighbor

The multiclass matrix confusion visualization in Figure 3 shows the successful prediction result of the KNN model with a value of k = 2. The KNN model can predict multiclass well in each class on a 90:10 data split with 109 test data. In class 0, it was possible to predict 40 data and only wrongly predict 1 data; in class 1, the model could predict all 22 data; and in class 2, the model could also correctly forecast all 46 data.

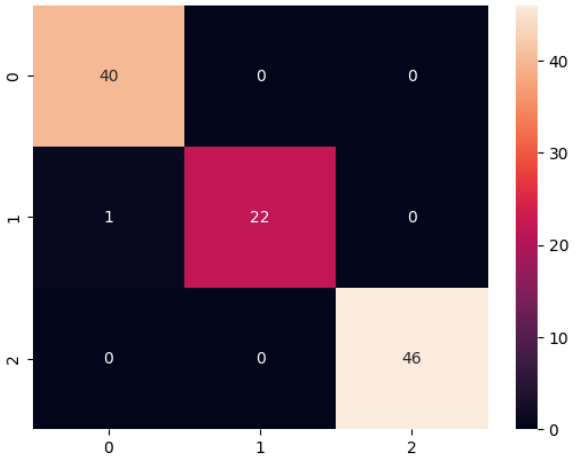


Figure 2. Confusion Matrix KNN

The following are examples of manual computations of evaluation metrics like accuracy, precision, recall, and fi-score on the K-Nearest Neighbor confusion matrix.

$$\text{Accuracy} = \frac{40+22+46}{109} = \frac{108}{109} = 99\ \%$$

$$\text{Precision} = 0 = \frac{40}{40+1+0} = \frac{40}{41} = 98\ \%$$

$= 1 = \frac{22}{22+0+0} = \frac{22}{22} = 100\ \%$
$= 2 = \frac{46}{46+0+0} = \frac{46}{46} = 100\ \%$

Recall $= 0 = \frac{40}{40+0+0} = \frac{40}{40} = 100\ \%$
$=1 = \frac{22}{1+22+0} = \frac{22}{23} = 96\ \%$
$= 2 = \frac{46}{46+0+0} = \frac{46}{46} = 100\ \%$

F1-Score $= 0 = \frac{2(0,98\times1)}{0,98+1} = 99\ \%$
$= 1 = \frac{2(1\times0,96)}{1+0,96} = 98\ \%$
$= 2 = \frac{2(1\times1)}{1+1} = 100\ \%$

2. Confusion Matrix Naïve Bayes

The results of the multiclass matrix confusion visualization in Figure 4 show the results of successful predictions in the Naïve Bayes model. Naïve Bayes could only predict two classes on a 90:10 data split with 109 test data in each model class. In class 0, models predicted 16 data points and mispredicted 16. In class 1, models could not predict data at all, and in class 2, models could predict 43 data but mispredicted 34 data.

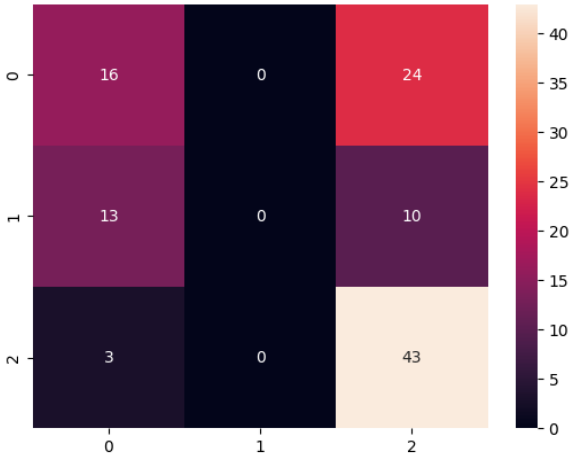


Figure 3. Confusion Matrix Naïve Bayes

The following are examples of manual computations of evaluation metrics such as accuracy, precision, recall, and fi-score on the Naïve Bayes confusion matrix.

Accuracy $= \frac{16+0+43}{109} = \frac{59}{109} = 0,54 = 54\ \%$

Precision $= 0 = \frac{16}{16+3+3} = \frac{16}{32} = 0,5 = 50\ \%$
$= 1 = \frac{0}{0+0+0} = \frac{0}{0} = 0 = 0\ \%$
$= 2 = \frac{43}{43+10+24} = \frac{43}{77} = 56\ \%$

Recall $= 0 = \frac{16}{16+0+24} = \frac{16}{40} = 40\ \%$
$= 1 = \frac{0}{13+0+10} = \frac{0}{23} = 0\ \%$
$= 2 = \frac{43}{43+0+3} = \frac{43}{46} = 93\ \%$

F1-Score $= 0 = \frac{2(0,5\times0,40)}{0,5+0,40} = 44\ \%$
$= 1 = \frac{2(0\times0)}{0+0} = 0\ \%$
$= 2 = \frac{2(0,56\times0,93)}{0,56+0,93} = 70\ \%$

ROC Analysis

ROC Analysis is a curve that depicts the probability using the variables of sensitivity and specificity, with values ranging from 0 to 1. The closer the curve follows the left boundary and the top boundary of the ROC space, the more accurate the classification is. ROC Analysis shows the performance of both algorithms in measuring or analyzing data (Astuti et al., 2022)

The classification range for data mining classification methods is as follows:
a. 0.90 – 1.00 = Excellent Classification
b. 0.80 – 0.90 = Good Classification
c. 0.70 – 0.80 = Fair Classification
d. 0.60 – 0.70 = Poor Classification
e. 0.50 – 0.60 = Failure Classification

1. ROC Curve K-Nearest Neighbor

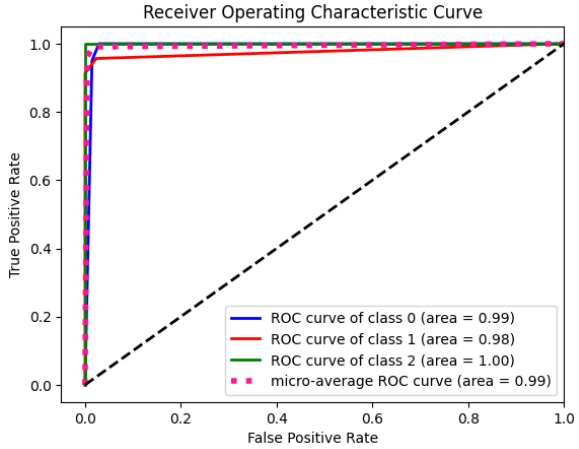


Figure 4. ROC Curve KNN

From the result, Figure 5 is a visualization of the matplotlib ROC Curve, where the AUC of the KNN algorithm obtained is 99% with a value of ROC class 0 = 99%, a matter of ROC grade 1 = 98%, and a value of ROC Class 2 = 99%.

2. ROC Curve Naïve Bayes

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4420

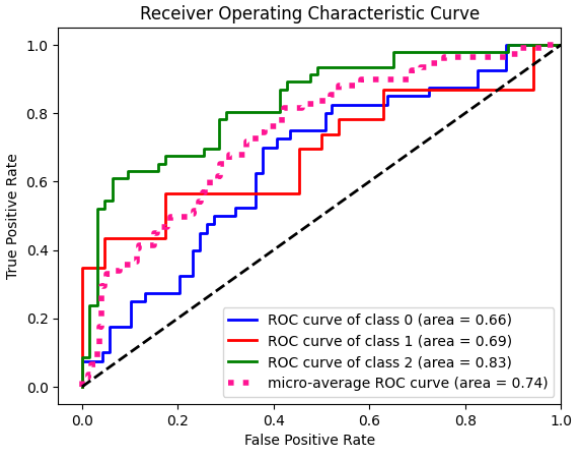


Figure 5. ROC Curve Naïve Bayes

From the result, figure 6 is a visualization of the matplotlib ROC Curve, where the AUC of the Naïve Bayes algorithm obtained is 74% with a value of ROC class 0 = 66%, a matter of ROC Class 1 = 69%, and a value of ROC class 2 = 83%.

Split Data Train Test Results

The best results may be obtained on the 90:10 split data with the most significant evaluation metric value on each method from the outcomes of the 70:30, 80:20, and 90:10 data splits.

Table 3. Train Test Split

Model	Split	Score				
		Accuracy	Precision	Recall	F1-Score	Train/ Test
			94% = 0	98% = 0	96% = 0	763 Train
KNN	70: 30	96%	95% = 1	91% = 1	93% = 1	327 Test
			100% = 2	98% = 2	99% = 2	
			55% = 0	35% = 0	42% = 0	763 Train
NB	70: 30	50%	0% = 1	0% = 1	0% = 1	327 Test
			48% = 2	93% = 2	63 = 2	
			93% = 0	98% = 0	95% = 0	872 Train
KNN	80: 20	96%	95% = 1	90% = 1	92% = 1	218 Test
			100% = 2	96% = 2	98% = 2	
			57% = 0	38% = 0	45% = 0	872 Train
NB	80: 20	52%	0% = 1	0% = 1	0% = 1	218 Test
			50% = 2	93% = 2	65% = 2	
			98% = 0	100% = 0	99% = 0	981 Train
KNN	90: 10	99%	100% = 1	96% = 1	98% = 1	109 Test
			100% = 2	100% = 2	100% = 2	
			55% = 0	40% = 0	44% = 0	981 Train
NB	90: 10	54%	0% = 1	0% = 1	0% = 1	109 Test
			56% = 2	93% = 2	70% = 2	

As a result, the ultimate conclusion was a 90:10 data split, in which the K-Nearest Neighbor model achieved outstanding classification and successfully predicted multiclasses (excellent classification), whereas the Naïve Bayes model failed in classification (failure classification) and failed to predict multiclasses.

CONCLUSION

The Results of Product Sales Classification Using K-Nearest Neighbor and Naïve Bayes Algorithms with a total of 1,090 sales data from 2020 to 2022. 90% of the data, which is 981 data, was used as training data, and 10%, which is 109 data, was used as testing data. The K-Nearest Neighbor algorithm achieved an accuracy of 99% with Precision scores for Class 0 at 98%, Class 1 at 100%, and Class 2 at 100%. The Recall scores for Class 0 were 100%, Class 1 96%, and Class 2 100%. The F1-Score for Class 0 was 99%, Class 1 98%, and Class 2 100%. It was categorized as an Excellent Classification. On the other hand, the Naïve Bayes algorithm achieved an accuracy of only 54% with Precision scores for Class 0 at 55%, Class 1 at 0%, and Class 2 at 56%. The Recall scores for Class 0 were 40%, Class 1 0%, and Class 2 93%. The F1-Score for Class 0 was 44%, Class 1 0%, and Class 2 70%. It was categorized as a Failure Classification. The performance evaluation using ROC Curve showed that the K-Nearest Neighbor algorithm had a higher value compared to the Naïve Bayes algorithm, with a micro-average AUC of 99% for KNN, which was close to 1. In comparison, the Naïve Bayes algorithm had a micro-average AUC value of only 74%. Based on the evaluation of both algorithms, K-Nearest Neighbor achieved the highest accuracy, while the Naïve Bayes algorithm failed in the classification.

At the end, where one algorithm fails to classify, it can be compared to the other algorithm. And could be further developed by making a decision support system to make predicting the product sales rate easier.

PT Kotamas Bali can make this research a consideration in the sale of products so it can determine what products should be multiplied.

REFERENCE

Alfani W.P.R., A., 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

Arhami, M., & Muhammad, N. (2020). *Data Mining Algoritma Dan Implementasi*. Penerbit Andi, 2020. https://books.google.co.id/books?id=AtcCEAAAQBAJ&lpg=PP1&hl=id&pg=PP1#v=onepage&q&f=false

Astuti, N. K. M., Utami, N. W., & Juliharta, I. G. P. K. (2022). Classification of Blood Donor Data Using C4.5 and K-Nearest Neighbor Methods (Case Study: Utd Pmi Bali Province). *Jurnal Pilar Nusa Mandiri*, *18*(1), 9–16. https://doi.org/10.33480/pilar.v18i1.2790

Budiyanto, A., & Dwiasnati, S. (2018). The Prediction of Best-Selling Product Using Naïve Bayes Algorithm (A Case Study at PT Putradabo Perkasa). *Ijctjournal.Org*, *5*(6), 68–74. http://www.ijctjournal.org/Volume5/Issue6/IJCT-V5I6P10.pdf

Fayyad, U., Piatetsky-shapiro, G., & Smyth, P. (1996). *From Data Mining to Knowledge Discovery in*. *17*(3), 37–54. http://www.aaai.org/

Kotamas Makmur - Houseware & Lifestyle Products. (2022). Diakses dari https://ptkotamas.com/

Muqorobin, M., Kusrini, K., Rokhmah, S., & Muslihah, I. (2020). Estimation System for Late Payment of School Tuition Fees. *International Journal of Computer and Information System (IJCIS)*, *1*(1), 1–6. https://doi.org/10.29040/ijcis.v1i1.5

Rahmad, F. Suryanto, Y. Ramli, K. (2020). Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification. *IOP Conference Series: Materials Science and Engineering*, 1–11. https://doi.org/10.1088/1757-899X/879/1/012076

Romadhon, M. R., & Kurniawan, F. (2021). A Comparison of Naive Bayes Methods, Logistic Regression, and KNN for Predicting Healing of Covid-19 Patients in Indonesia. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIConCIT 2021*, 41–44. https://doi.org/10.1109/EIConCIT50028.2021.9431845

Saputra, I., & Ajeng Kristiyanti, D. (2022). *Machine Learning Untuk Pemula*. Informatika, Bandung, Indonesia.

Soepriyanto, B. (2021). Comparative Analysis of K-NN and Naïve Bayes Methods to Predict Stock Prices. *International Journal of Computer and Information System (IJCIS) Peer Reviewed-International Journal*, *02*(02), 2745–9659. https://ijcis.net/index.php/ijcis/index

Suyanto. (2017). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Informatika, Bandung, Indonesia. https://suyanto.staff.telkomuniversity.ac.id/textbook-data-mining/

Swamynathan, M. (2019). *Mastering Machine Learning With Python In Six Steps*. Bangalore, Karnataka, India. https://books.google.co.id/books?id=dEqzDwAAQBAJ&lpg=PR3&dq=python programming concept and implementation&lr&hl=id&pg=PA3#v=onepage&q=python programming concept and implementation&f=false

Tarmizi. (2021). *Marketing Strategy to Increase Sales Volume*. *2*(2), 322–328. https://doi.org/10.51612/teunuleh.v2i2.49

Yulianto, T. (2019). Prediksi Penjualan Produk Menggunakan Algoritma Naive Bayes. *Journal Teknologi Yogyakarta*, *1*(2), 3–10. http://eprints.uty.ac.id

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 19, No. 2 September 2023 | DOI: 10.33480/pilar.v19i2.4420

This sheet was deliberately left blank.