Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

Analysis of Product Sales in the Application of Data Mining with Naive Bayes Classification

M. Hannata Zahri[1)*], Aswan Sunge[2)], Ahmad Turmudzi Zy[3)]
[1)*2)3)] Faculty of Information Technology, Universitas Pelita Bangsa, Kabupaten Bekasi, West Java, Indonesia.
[1)]hannata23zahri@mhs.pelitabangsa.ac.id , [2)]aswan.sunge@pelitabangsa.ac.id , [3)]turmudi@pelitabangsa.ac.id

ABSTRACT

H&F Shoe Store is a privately owned Micro, Small, and Medium Enterprises retail store that sells merchandise. The owner serves customers directly and also acts as a cashier. In this store, the business owner is less aware of what types or categories of products are most in demand by customers, making sales operations less than optimal. Because of this, special expertise is needed to handle the problems in the retail store, namely data mining or Data Mining with the aim of digging up information related to sales problems, the author will apply the Naive Bayes algorithm in conjunction with the Classification approach in this instance. In this study, the author uses secondary data obtained from sales notebooks and re-collected into Microsoft Excel according to research needs. The data that has been collected on the software is 121 data which have 10 attributes, namely "Nama Produk", "Size Produk", "Kategori Produk", "Jenis Produk", "Gender Produk", "Merek Produk", "Stok Awal", "Stok Terjual", "Stok Sisa", and "Penjualan". The Naive Bayes Classifier method has successfully produced good results in classifying sales on a type or category of marketed products, the results obtained are in the form of product sales analysis and Naive Bayes model evaluation values. The results of the model evaluation values on the Confusion Matrix obtained are accuracy of 86.11%, recall of 84.62% and precision of 84.62%.

Keywords: Data Mining; Classification; Naive Bayes; Product Sales; Retail Stores

INTRODUCTION

In the ever-changing and competitive business world, businesses must constantly think about how to survive and possibly expand their business. To achieve this, business operations need to increase product types and capacity, reduce company operating costs, and increase marketing effectiveness and profits. A product can be defined as anything that can be offered to the market to meet consumer needs or wants, including physical goods, services, individuals, locations, organizations, and ideas. In marketing, a product is defined as something that has value that can provide benefits and satisfaction to consumers (Hartanto & Subagyo, 2019). There are many ways that can be used to meet the above needs. One of them is through company data analysis (Harahap et al., 2023).

Nowadays, competition in the retail and technology industry is growing rapidly and is very competitive. As a result, retailers are competing to use existing technology to grow their business. Retail allows products to meet directly with their consumers, making it the last link in the distribution of goods (Soliha, 2008). The development of retail businesses that are increasingly numerous and mushrooming in an area makes consumer needs increase (Wardani et al., 2023). H&F Shoe Shop is a retail business that has been operating for quite a long time in the Cikarang area, this retail store provides a variety of products in it such as shoes, sandals, and others with certain brands, one of which is products from the Carvil brand. In the Cikarang area itself, this shop is famous for producing more goods from the Carvil brand than from other brands, around 50% of the goods in this shop are Carvil products. Because of this, the business actor does not understand what types or categories of products are most in demand or most needed by customers of this brand's products, so the business actor only does Replenishment or buys the latest products from this brand without knowing what types or categories of products are in great demand by most customers. And because this can also cause the purchase of goods to be Overstocking so that business operations become less optimal.

In dealing with the problems that occur, the business actor must find information related to the number of enthusiasts of the brand's products, so that the business actor can understand what potential customers need and are interested in in the future, so that in the future they can reach a wider range of customers in meeting customer needs. Considering the perceived usefulness of a product is something that companies need to pay attention to in determining the purchase decision of a product (Purba & Syaifullah, 2020).

* Corresponding author

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

Therefore, a special ability is needed to deal with the problems that occur in these retail stores. Thus, the use of data mining can be the right choice to predict customer habits and needs related to the products they will buy. In addition, data mining can provide knowledge that can identify unique sales patterns for each product (Pramana et al., 2023). Data Mining is one of the techniques of Machine Learning that was developed to be able to learn by itself without direction from its users. Based on other disciplines such as data mining, mathematics, and statistics, machine learning allows machines to learn through data analysis without the need to be reprogrammed or instructed (Sihombing & Arsani, 2021). There are several techniques owned by Machine Learning, but in this research the author will use Supervised Learning techniques to perform classification models on data. Supervised learning is a machine learning technique where labeled data is used to train the model. Through the use of labeled training data examples, the model gains the ability to map inputs to outputs. The objective is to use patterns discovered from training data to predict or categorize new data (Mahesh, 2020). Data mining is a set of procedures used to extract information from data sources that has additional value but cannot be identified by hand. Extracting information in data mining is very useful for increasing profits and helping in making marketing strategies (Al Faridzi et al., 2024). In this study, data mining will be implemented using the Classification method using the Naïve Bayes algorithm to determine product sales patterns in retail stores. Naive Bayes is a useful classification method according to Bayes' rule used to perform probabilistic calculations in the classification process, the Naive Bayes model presumes that an attribute's input value within a class is independent of the values of other attributes. (Sulastiyono et al., 2023). To estimate an object's unknown class, classification is the process of identifying functions and models that may differentiate or explain ideas or data classes. (Apriliah et al., 2021). In using the techniques mentioned above, it is hoped that it can contribute to the research location in an effort to optimize product sales by finding sales-related information on the implementation of Data Mining, and also as a study material for sales data analysis in future studies.

LITERATURE REVIEW

Some related research on product analysis for sales optimization has been carried out previously by several researchers, including product analysis of medical device sales using the Naive Bayes algorithm. The research aims to improve customer relationship management (CRM) to maintain relationships with customers in an effort to improve product quality, increase the variety of products in demand and determine the level of ownership of company products in customers by understanding the needs and desires of customers in the future, (Ramadhani et al., 2023). The same research has been conducted on analyzing data on interest in products using Naive Bayes Classifier to determine customer interest in choosing existing products, so that companies can find out which products are most in demand by customers, and companies can provide direction to potential customers to determine which products to choose in the future, based on data that has been tested using the Rapidminer application shows an accuracy level of 72.73% which means "interest" in products at PT Jellyfish Education Indonesia. (Susanto et al., 2023). Furthermore, there is research that applies Data Mining to predict the number of best-selling products using Naive Bayes, this research has the same objectives as previous studies, namely to increase sales of UD stores. The results obtained from this research are in the form of a web-based system for predicting best-selling products at Prapti Stores which was developed using the PHP programming language. The results of the system accuracy using confusion matrix with an accuracy value of 83.3%, precision of 84.2% and recall of 88.9% (Abdullah et al., 2022).

Another research on the use of the Naïve Bayes method in data analysis analysis is the classification of office stationery sales data using the Naive Bayes method. In this study, ATK sales data processing was carried out to classify ATK data and find the accuracy value as a determination of what items were most in demand by customers. The purpose of this research is as a marketing strategy in order to optimize store work in processing stock data. The accuracy results obtained from the Naive Bayes method with the help of RapidMiner tools obtained very good results, namely high accuracy of 99.26% (Widodo E, 2019). In other research in the use of Naive Bayes, namely the application of Naive Bayes to predict the level of car sales, this study aims to apply the Naive Bayes algorithm in classifying various categories of car sales in 2022 and to understand consumer buying interest in certain car brands. The results obtained are information that Toyota brand product sales are the most sold based on the Density value obtained, and also the test results show that the algorithm used produces a high accuracy value of 95.38% with a precision of 94.96% and a recall of 90.21% (Hasyim et al., 2022). Based on the literature review that has been done, it can be concluded that most of the research analyzes sales in large or global markets, but few focus on retail stores with more specific products, for example, such as local products. Therefore, this research wants to help small business owners understand the sales patterns of Carvil local products for better sales optimization by focusing only on processing data on Carvil

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

brand products. Based on the description above, the author can conclude that the algorithm used can help solve problems in the classification process for analysis and prediction which will be discussed in this study.

METHOD

Research Stages

Before carrying out research, the thing that must be done by researchers is to determine the steps of the research stages to be carried out. The following is a picture of the research stage plan:

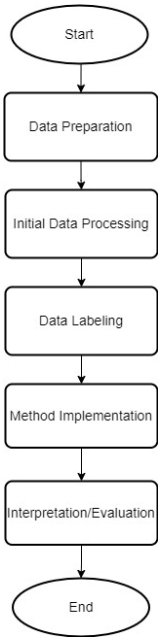


Fig 1. The research stage flow

In Figure 1 describes the research stage flow, the stages of the research carried out are to prepare data by collecting information or data available on the object of research, then the initial data processing is carried out after collecting data, the data that has been processed will be labeled with integrated attributes at the data labeling stage for the implementation of the classification stage, the implementation of the Classification method using the Naive Bayes Classifier algorithm and the final stage, namely Interpretation/Evaluation of the model.

Data Preparation

The initial stage in this research is preparing the data. The data used in this research is secondary data. Secondary data is data that has been collected previously in the form of records or records of information from individuals or groups (Rusdi, 2019). The data that will be used in this study is sales data from the H&F Shoe Store, which will be processed using several methods to produce good data and used in finding information from the shoe store's sales. The data obtained comes from the sales logbook at the H&F Shoe Store which will later be collected back into the Excel software to find out the results of the information obtained. The sales data collected will then be processed using a predetermined calculation algorithm.

Initial Data Processing

In the initial stage of data preparation, the data that has been collected from recording store sales into Excel software is clean data, which means that the data can be directly used in Data Mining implementation. The data that has been prepared at the data research stage will be divided into two for the implementation of the Classification process, namely training data (70%) and testing data (30%). The process of separating training data and testing data is intended to produce good correlation in performing Classification (Haryadi & Mandala, 2019). This data division is done automatically using RapidMiner tools.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

Data Labeling

The data source used in this research comes from the sales record book of the H&F Shoe Shop where the data in the sales record will be entered into an Excel table for the implementation of the Data Mining process. The data that has been collected into the Excel table amounts to 121 data where the data collection process only selects products from the Carvil brand. The data will then be processed using Data Mining with the Naive Bayes Classifier Method. After the data collection process, several attributes will be given to the data table to continue the Data Mining implementation process. The following is an explanation of the attributes given in the table below:

Table 1.
Types of Atributes

No.	Attribute	Description
1.	Product Name	Markers in the form of product names that have been given from the brand.
2.	Product Size	The size of the product that has been grouped.
3.	Product Category	Product grouping that has been given from the brand.
4.	Product Type	Product grouping as a marker from the store.
5.	Product Gender	Targeted products according to gender.
6.	Product Brand	Identification of the origin of a product.
7.	Initial Stock	Number of products or remaining products in the store.
8.	Sold Stock	The amount of a product that has been sold.
9.	End Stock	The remaining amount of a product that has been sold from the store.
10.	Sales	Grouping of sales based on items sold.

In table 1 above on line 10, the "penjualan" grouping is determined based on the percentage of sales of the number of items in each product, this grouping will be used as a class/label for the classification process in the Data Mining implementation. The following is a table explaining the percentage calculation used in this research:

Table 2.
Explanation of the class/label attributes

Class value “Penjualan”	Description
Few	Sold less than 40% of the total items.
Medium	Sold between 40% and 80% of the total items.
Many	Sold more than 80% of the total items.

Method Implementation

The method used in this research is the Classification method with the Naive Bayes Classifier algortima. The process of modeling machine learning technology with supervised teaching methods is called classification, and the goal is to identify models or functions that may be used to separate or explain concepts or data classes. Estimating the class of an object whose label is unknown is the aim of categorization (Novianti, 2019). Classification methods have several popular techniques to be used as case solvers, one of which is the technique used in this research, namely Naive Bayes. Based on the Bayes Theorem, the Naive Bayes algorithm is a straightforward probabilistic classification technique that effectively classifies data using a training set. Naive Bayes makes the assumption that an input attribute's value within a class is unrelated to the values of other attributes (Prast et al., 2022). Naive Bayes is among the top ten most effective algorithms in data mining and is another widely used classification technique (Siradjuddin, 2015). The following is the formula for calculating the Naive Bayes algorithm:

$$P(C|X) = \frac{P(C|X)P(C)}{P(X)} \quad (1)$$

Description:
X : Data with unknown class/label.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

C : Hypothesize that data X belongs to a certain class/label.
P(C|X) : The value of hypothesis H depending on the condition X (Posterior).
P(C) : Probability value of hypothesis H (Prior).
P(X|C) : The value of hypothesis X depending on condition H (Likelihood)..
P(X) : Probability value of hypothesis X

Interpretation/Evaluation

Following the acquisition of the Data Mining process, the Knowledge Discovery in Database (KDD) process is executed at the Interpretation/Evaluation stage. This stage is the process of the last KDD which covers the pattern of examining information that contradicts the previous hypothesis. In this method, the data mining process yields patterns or information in the form of rules that are derived from the Naive Bayes Classifier algorithm's computation. At this stage the interpretation/evaluation results are obtained from the Confusion Matrix calculation using RapidMiner tools. The RapidMiner software platform, created by the same-named firm, offers an integrated environment for text mining, predictive analysis, machine learning, deep learning, and data preparation. Analytics for data mining, text mining, and prediction are provided by RapidMiner open source software. Many descriptive and predictive methods help people make better decisions.Less than half of the data mining operators available in rapidminer are devoted to data input, output, pre-processing, and visualization (Dewi et al., 2019).Confusion Matrix is a method that is usually used to calculate accuracy in data mining concepts. The Confusion Matrix, which has been predicted by the classification system, includes details on the actual classification outcomes.Confusion Matrix is a tool used to assess an algorithm model's accuracy, precision, and recall values (Fansyuri, 2020).The Confusion Matrix formula or formulation is as follows:

1. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$ (2)
2. $Precision = \frac{TP}{TP+FP}$ (3)
3. $Recall = \frac{TP}{TP+FP}$ (4)

Description:
1) The number of positive records that are considered positive is called TP (True Positive).
2) The number of negative records that are considered positive is called FP (False Positive).
3) The number of positive records that are considered negative is called FN (False Negative).
4) The number of negative records that are considered negative is called TP (True Negative).

RESULT

After analyzing the design at the Data Mining stage to predict product sales to the sales conditions of the H&F Shoe Store using the Naive Bayes Classifier method, this analysis ends by carrying out the actual Data Mining process. So the result that the author wants to achieve is the analysis of product sales to explore information about products that will be marketed at the H&F Shoe Store. researchers use RapidMiner tools to facilitate the Data Mining implementation process which will then produce accurate data information.

Method Implementation on RapidMiner

In the implementation of the method, the author will directly implement the Naive Bayes Classifier method in the RapidMiner tool. At this stage, the RapidMiner tools are arranged to perform the data classification process using the Naive Bayes algorithm. The following is the arrangement of operators in RapidMiner:

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

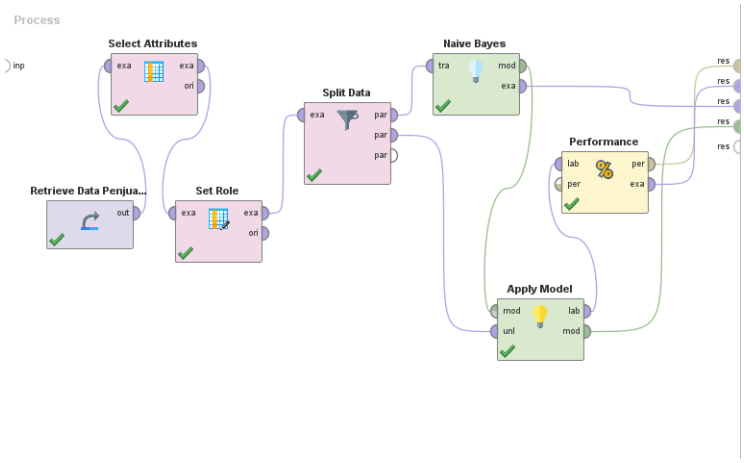


Fig. 2 Operator Setup on RapidMiner

Figure 2 shows the arrangement of operators to perform all analysis processes using the Classification method with the Naive Bayes algorithm. The above processes will produce a predictive value obtained from classifying the Training Data by testing the Testing Data and also produce an evaluation value to measure the performance results performed on the model used.

Attribute Selection

The first stage that must be done in this implementation is to enter the previously created dataset into the RapidMiner tool, then the data that has been entered will be selected attributes as needed to extract data. The attribute columns that are not used are "Merek Produk" because the contents of the value in it are all the same value, and "Nama Produk" because the attribute does not provide data distribution to a class. Figure 3 shows the attributes that will be used for the Naive Bayes testing process.

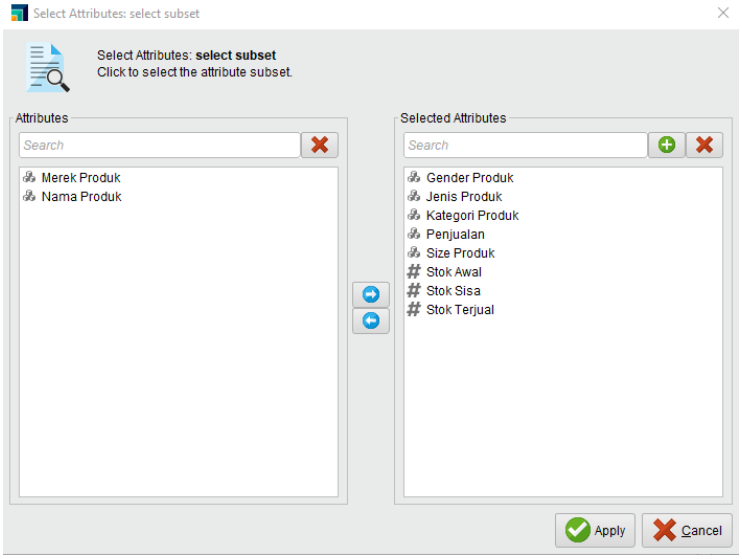


Fig. 3 Attribute Selection

Role Selection on Attributes

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

The next stage is the selection of the role of an attribute in a dataset, at this stage the Set Role operator is used to mark the "Penjualan" attribute into a label that will later be used to classify data. Figure 4 shows how to use the Set Role operator.

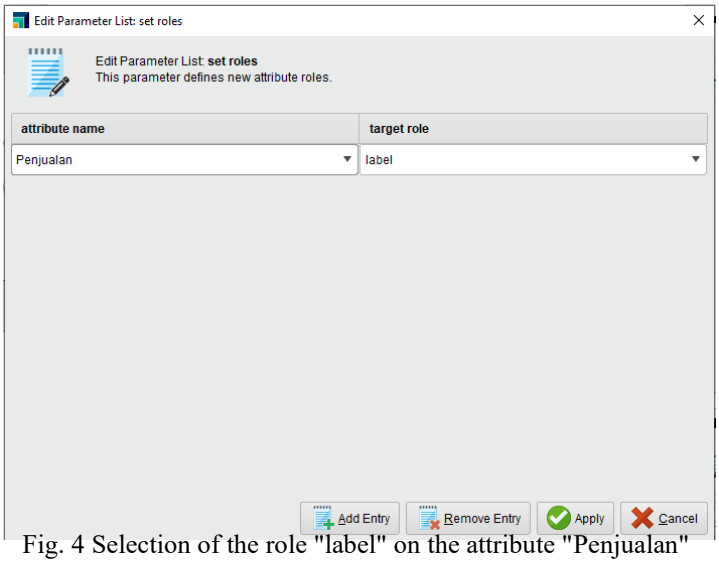


Fig. 4 Selection of the role "label" on the attribute "Penjualan"

Sharing of Training Data and Testing Data

The next stage is to divide the data, in this process it is done to apply the model that has been trained on the Testing Data or on new data that will be predicted. At this stage the data used amounted to 121 rows of data which were divided into 7:3, 70% as much as 85 rows of data were used to become Training Data and 30% as much as 36 rows of data were used to become Testing Data, this data division was carried out with automatic or automatic sampling type in RapidMiner which has been adjusted based on the distribution of the dataset we have.

Hasil Interpretasi/Evaluasi pada RapidMiner

accuracy: 86.11%

	true Sedikit	true Sedang	true Banyak	class precision
pred. Sedikit	7	0	0	100.00%
pred. Sedang	1	13	2	81.25%
pred. Banyak	0	2	11	84.62%
class recall	87.50%	86.67%	84.62%	

Fig. 5 Testing the model with Confusion Matrix

The testing process in Figure 5 above uses Confusion Matrix to produce accuracy, recall, and Precision values. The values obtained from the test results are accuracy of 86.11%, recall of 84.62% and precision of 84.62%. The following is the calculation of the accuracy, recall, and Precision values:

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

a. Accuracy
By knowing the amount of data that is classified correctly, the prediction accuracy value can be obtained, which is 86.11%. Because this research has 3 classes/labels, the author will make a formula in table 3 as follows:

Table 3.
Explanation of Confusion Matrix calculation with 3 classes/labels

Process		Predicted Class		
		Class (A)	**Class (B)**	**Class (C)**
Real Class	**Class (A)**	TP (A)	FP (B|A)	FP (C|A)
	Class (B)	FP (A|B)	TP (B)	FP (C|B)
	Class (C)	FP (A|C)	FP (B|C)	TP (C)

$$= \frac{TP_A + TP_B + TP_C}{TP_A + TP_B + TP_C + FP_{A|C} + FP_{B|C} + FP_{A|B} + FP_{C|B} + FP_{B|A} + FP_{C|A}}$$
$$= \frac{11 + 13 + 7}{11 + 13 + 7 + 0 + 2 + 1 + 2 + 0 + 0}$$
$$= \frac{31}{36}$$
$$= 0{,}8611 * 100\%$$
$$= 86{,}11\%$$

b. Recall
Recall is the amount of True Positive data divided by the amount of data that is actually Positive (True Positive + False Negative), while True Negative is divided by the amount of data that is actually Negative (True Positive + False Positive). The following is the calculation of the Recall value:

$$= \frac{TP_C}{TP_C + TP_{C|A} + TP_{C|B}}$$
$$= \frac{11}{11 + 0 + 2}$$
$$= 0{,}8462 * 100\%$$
$$= 84{,}62\%$$

c. Precision
Precision is the number of true "Many" data divided by the number of data recognized as "Many", while the number of true "Few" data is divided by the data recognized as "Few", and so on. Here is the calculation form for Precision:

$$= \frac{TP_C}{TP_C + FP_{A|C} + FP_{B|C}}$$
$$= \frac{11}{11 + 0 + 2}$$
$$= 0{,}8462 * 100\%$$
$$= 84{,}62\%$$

Information obtained from SimpleDistribution

SimpleDistribution Model is used to analyze data tables to predict sales in a type/category of branded products using the Naive Bayes Classifier method so as to produce 3 main classes, namely "Few (Sedikit)", "Medium (Sedang)" and "Many (Banyak)". In SimpleDistribution, the author gets information about sales analysis at the H&F Shoe Store which is obtained from the results of the Density value for each instance that belongs to a certain class. In this analysis the author will focus on four attributes, namely the attributes "Size Produk", “Kategori Produk", "Jenis Produk" and "Gender Produk". The following is a detailed explanation:

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

1. Sales Class “Many”
   - “Product Size” = “Adult Sandals”
   - “Product Category” = “Flip-flop Sandals”
   - “Product Type” = “Flip-flops”
   - “Product Gender” = “Male”
2. Sales Class “Medium”
   - “Product Size” = “Kids Sandals”
   - “Product Category” = “Flip-flop Sandals”
   - “Product Type” = “Flip-flops”
   - “Product Gender” = “Male”
3. Sales Class “Few”
   - “Product Size” = “Adult Sandals”
   - “Product Category” = “Flip-flop Sandals”
   - “Product Type” = “Flip-flop Sandals”
   - “Product Gender” = “Male”

In the details of the explanation above, a decision can be made to regulate the production of Carvil brand product stock at the H&F Shoe Shop. If a type/category of product has "Many (Banyak)" sales, it can be done to add more stock, if a type/category of product has "Medium (Sedang)" sales, adjustments can be made to replenishment or restock on the product, and if a type/category of product has "Few (Sedikit)" sales, it can be rearranged in reducing restock on the product for the efficiency of capital spent and minimizing losses that might be obtained.

In each class there is a classification result value from the Naive Bayes method on this SimpleDistribution, namely the value of the "Banyak" class = (0.365), the value of the "Sedang" class = (0.412) and the value of the "Sedikit" class = (0.365). Here's a picture of the results on the SimpleDistribution model.

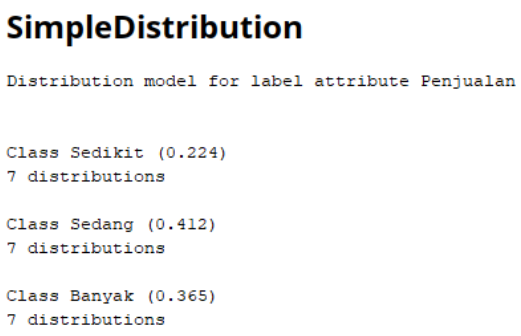


Fig. 6 Calculation results of 3 classes in SimpleDistribution

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

CONCLUSION

Based on the results of the research that has been done, the author has processed 121 product sales data at the H&F Shoe Shop with the aim of extracting pertinent information about sales. The results of the information obtained are product sales with the "Many (Banyak)" class found in [Size Produk= Sandal Dewasa], [Kategori Produk= Sandal Flip-flop], [Jenis Produk= Sandal Jepit], [Gender Produk= Laki-laki], in the "Medium" class product sales found in [Size Produk= Sandal Anak], [Kategori Produk= Sandal Flip-flop], [Jenis Produk= Sandal Jepit], [Gender Produk= Laki-laki], and in product sales with the class "Few (Sedikit)" there are [Size Produk= Sandal Dewasa], [Kategori Produk= Sandal Flip-flop], [Jenis Produk= Sandal Selop], [Gender Produk= Laki-laki]. The results of this study are expected to provide more insight into the sales information of Carvil products at the H&F Shoe Store, and the results of obtaining this information can be applied to the research location in an effort to optimize product sales by adjusting the production of Carvil goods at the retail store. In the results of testing the model using Confusion Matrix, good accuracy results were obtained, namely the accuracy value of 86.11%, recall 84.62% and precision 84.62%. For future research, it is suggested to expand the scope of research in order to get better results in analyzing product sales at this research location because this research only focuses on one local product, and it is also recommended to use other classification algorithms to compare the results of the algorithms used in this study with the algorithms that will be used in future research.

REFERENCES

Abdullah, R. W., Hartanti, D., Permatasari, H., Septyanto, A. W., & Abi Bagaskara, Y. (2022). Penerapan Data Mining untuk Memprediksi Jumlah Produk Terlaris Menggunakan Algoritma Naive Bayes Studi Kasus (Toko Prapti). Jurnal Ilmiah Informatika Global, 13(1).

Al Faridzi, S., Azizah, F. S., Mustafa, F., Putri, A. N., Ramadhika, G., Aditya, F. R., Fadilah, R. S., Habibi, Y., Sutrisno, M., & Jumail, J. (2024). PENGOLAHAN DATA: Pemahaman Gempa Bumi Di Indonesia Melalui Pendekatan Data Mining. Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS, 2(1), 262–270.

Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi kemungkinan diabetes pada tahap awal menggunakan algoritma klasifikasi Random Forest. Sistemasi: Jurnal Sistem Informasi, 10(1), 163–171.

Dewi, S. M., Windarto, A. P., & Hartama, D. (2019). Penerapan Datamining Dengan Metode Klasifikasi Untuk Strategi Penjualan Produk Di Ud. Selamat Selular. KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer), 3(1).

Fansyuri, M. (2020). Analisa algoritma klasifikasi k-nearest neighbor dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus pt. Trigatra komunikatama). Jurnal Ilmiah Humanika, 3(1), 29–33.

Harahap, F., Fahrozi, W., Adawiyah, R., Siregar, E. T., & Harahap, A. Y. N. (2023). Implementasi Data Mining dalam Memprediksi Produk AC Terlaris untuk Meningkatkan Penjualan Menggunakan Metode Naive Bayes. Jurnal Unitek, 16(1), 41–51.

Hartanto, B. W., & Subagyo, S. (2019). Kerangka kerja perencanaan pengembangan produk sebagai peningkatan daya saing industri kecil menengah. Jurnal Teknosains, 8(1), 26–38.

Haryadi, D., & Mandala, R. (2019). Prediksi Harga Minyak Kelapa Sawit Dalam Investasi Dengan Membandingkan Algoritma Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor. IT for Society, 4(1).

Hasyim, A., Fatchan, M., & Hadikristanto, W. (2022). Penerapan Algoritma Naïve Bayes Dalam Memprediksi Tingkat Penjualan Mobil Tahun 2022. Jurnal Ilmiah Intech: Information Technology Journal of UMUS, 4(02), 207–215.

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381–386.

Novianti, D. (2019). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner. Paradig.-J. Komput. Dan Inform, 21(1), 49–54.

Pramana, I. M. A. A., Sudiarsa, I. W., & Nugraha, P. G. S. C. (2023). Penerapan Algoritma Naive Bayes Untuk Prediksi Penjualan Produk Terlaris Pada CV Akusara Jaya Abadi. JATISI (Jurnal Teknik Informatika Dan Sistem Informasi), 10(4).

Prast, T. P., Zakaria, H., & Wiliantoro, P. (2022). Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier. OKTAL: Jurnal Ilmu Komputer Dan Sains, 1(06), 573–582.

Purba, R. S., & Syaifullah, S. (2020). Pengaruh Promosi Penjualan Dan Kualitas Produk Terhadap Keputusan Pembelian Produk Carvil Di Kota Batam. Akrab Juara: Jurnal Ilmu-Ilmu Sosial, 5(3), 228–239.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 3, July 2024
https://doi.org/10.47709/cnapc.v6i3.4255

Submitted : 9 July 2024
Accepted : 21 July 2024
Published : 27 July 2024

Ramadhani, D., A'yuniyah, Q., Elvira, W., Nazira, N., Ambarani, I., & Intan, S. F. (2023). Analisa Algoritma Naïve Bayes Classifier (NBC) Untuk Prediksi Penjualan Alat Kesehatan: Naïve Bayes Classifier (NBC) Algorithm Analysis for Prediction Medical Device Sales. Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), 3(2), 119–126.

Rusdi, M. (2019). Strategi Pemasaran untuk Meningkatkan Volume Penjualan pada Perusahaan Genting UD. Berkah Jaya. Jurnal Studi Manajemen Dan Bisnis, 6(2), 83–88.

Sihombing, P. R., & Arsani, A. M. (2021). Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018. Jurnal Teknik Informatika (Jutif), 2(1), 51–56.

Siradjuddin, H. K. (2015). Penerapan Algoritma Naïve Bayes Untuk Memprediksi Tingkat Kualitas Kesuburan (Fertility). J. Ilm., No. September, 1–14.

Soliha, E. (2008). Analisis industri ritel di Indonesia. Jurnal Bisnis Dan Ekonomi, 15(2), 24251.

Sulastiyono, R., Setiawan, A., & Nugroho, S. (2023). Sentimen Analisis Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naïve Bayes. Journal of Information System Research (JOSH), 4(4), 1387–1394.

Susanto, A. B., Wiharjo, S., & Liyana, T. (2023). ANALISIS DATA MINAT CUSTOMER TERHADAP PRODUK DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER: STUDI KASUS: PT JELLYFISH EDUCATION INDONESIA. Scientia Sacra: Jurnal Sains, Teknologi Dan Masyarakat, 3(2), 234–241.

Wardani, N. W., Nugraha, P. G. S. C., & Mahendra, G. S. (2023). Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel. JST (Jurnal Sains Dan Teknologi), 12(3).

Widodo E, S. P. (2019). KLASIFIKASI DATA PENJUALAN ALAT TULIS KANTOR (ATK) TERLARIS UNTUK OPTIMASI STRATEGI PEMASARAN DI TOKO CITRAMEDIA MENGGUNAKAN METODE NAIVE BAYES. SIGMA – Jurnal Teknologi Pelita Bangsa, 10(2), 138–152.

* Corresponding author

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.