# APPLICATION OF NAIVE BAYES ALGORITHM FOR SALES ANALYSIS AT ERIGO STORE

Maria Sitanggang[1], Rivandu Ambarita[2], Cantika Marpaung[3], Delima Sitanggang[4*]

[1,2,3,4]Universitas Prima Indonesia, Medan-Indonesia

email: mariasitanggang052020@gmail.com, rivandugomgom125@gmail.com,
cantikamrp24@gmail.com, delimasitanggang@unprimdn.ac.id

## ABSTRACT

The purpose of this study is to research and compare the accuracy of the previous research algorithm,  namely the KNN algorithm with the Naive Bayes algorithm, for the evaluation of Erigo Store sales. Given the increasingly fierce market competition, it is very necessary to formulate a marketing strategy to analyze and predict products using data mining processing methods. Data mining is the introduction of patterns, machine learning techniques, statistics, and visualization techniques that aim to provide information to make better decisions and improve prediction accuracy through the process of analyzing data based on the Knowledge Discovery  in Database (KDD) procedure. The research dataset was taken from shopee Toko Erigo e-commerce sales data using web scraping techniques, starting from January 2021 to June 2023 consisting of 5 categories of Erigo Store products, namely Shirts, T-Shirt, Outwear, Jacket and Pants. The overall accuracy of the previous research product using the KNN algorithm was 83.62% while the study using the application of the Naive Bayes algorithm for sales analysis in Erigo stores achieved an accuracy of 98.3% by using Matlab to analyze the data. The accuracy of the T-shirt category reached 98.6%, the shirt category reached 98.4%, the pants category reached 98.1%, the outwear category reached 98.7% and the accuracy of the jacket category reached 97.6%.

**Keywords:** Forecasting, Naïve Bayes, Data Mining, Sale.

## INTRODUCTION

Data Mining is patten recognition, machine learning techniques, statistics, and visualization techniques to find useful information from database through certain processes and methods [1], the types of data mining techniques, methods or algorithms are very diverse [2]. Data mining aims to provide information to make better decisions and serve to improve prediction accuracy through the process of analyzing previous data in a certain period of time, the output of the analyzed data provides accurate information about the comparison of rising, falling or staying tendencies of products [3][4], data mining helps in terms of planning to maximize the provision of stock of the most popular products to be able to minimize residual products, so that sales are maximized, maximum sales are sales that are able to meet consumer demand [5]. Erigo Store is a successful and inspiring clothing company engaged in fashion, Erigo store focuses production on men's and women's fashion, ranging from T-Shirt, Pants, Jacket,

Outwear, and shirt. Basic human needs such a clothing, food, shelter present business opportunities such as culinary, fashion, and property that continue to grow and become the most popular business fields among business people [6]. Erigo Store was established in 2010, under the name Selected & co, erigo store provides fashion with quality and attractive design according to the latest trendy. Given the increasingly fierce market competition, and in terms of increasing revenue, this company must formulate its marketing strategy to analyze or predict the products its sells, in order to identify the items that costumers are most interested in by utilizing data mining methods [7]. Based on the results of previous research on erigo store sales predictions, namely in a journal entitled "Use Of Data Mining To Predict Sales At Erigo Stores With K-Nearest Neighbor" produces total result prediction of sales categories, with the results of overall product sales classification accuracy of 83.62% where thw jacket category has 95% accuracy, the shirt category has 100% accuracy, the pants category has 92.31% accuracy, the outwear category has 89.47% accuracy, and the T-Shirt category has 60% accuracy [1]. Based on previous research, further research will be carried out with the same case but with different methods. In this study, the method used is The Naive Bayes Algorithm to analyze the accuracy of erigo store sales. The Naive Bayes Algorithm is a classification algorithm based on the Bayesian theorem in statistics [8]. The Naive Bayes method is a machine learning technique that can provide predictions quickly and efficiently in processing large data and calculating the probability value of each attribute examined [9], the selection of the Naive Bayes methods because it has simple algorithm with a high level of accuracy [10]. The application of this algorithm is very suitable for predicting future opportunities and helping companies prdeict customer interest in fashion products sold [7]. Based on the discussion that has been described, research will be carried out on predicting product sales at Erigo Store, this research is entitled Application Of Naive Bayes Algorithm For Sales Analysis At Erigo Store. The author hopes that results of this research will be used in predictive analysis of Erigo Store Sales.

## METHODS

This research stage serves as a process or procedure for studying and analyzing in detail about the Application Of The Naive Bayes Algorithm For Sales Analysis At Erigo Store. The research stages of the application of the naive bayes algorithm for sales analysis at erigo store can be described as follows:
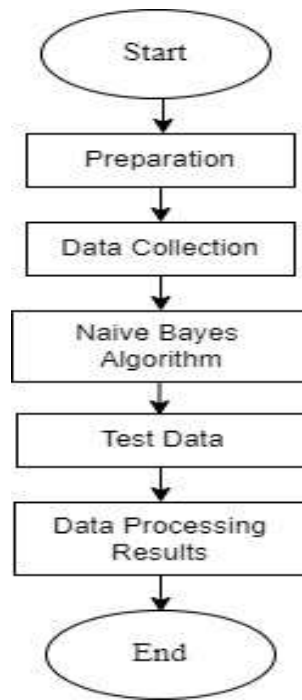
**Figure 1**. Research Framework

B. **Object of Research**

The object of this research was conducted at the Erigo Store clothing company, this company has an online store and offline store, so that many consumers can easily reach this store, so a sales analysis is needed to assist the company's marketing in planning product stock inventory which will be evaluated based on sales data fromthe previous three years.

C. **Research Type**

The type of research used is quantitative research using the Naive Bayes algorithm method. Quantitative method is a method based on, the data can be calculated, the data is in the form of numbers with the aim of obtaining information orinformation from a study [11].

D. **Data Collection Methods**

The data collection methods used are:

- Online Review of Products

Online product reviews are testimonials that contain evaluations from customers and personal experiences made by customers online from purchased products, testimonials are used as information about products and companies [12]. This research data is obtained from Erigo Store customer reviews on the *Shopee e-commerce* platform.

- Documentation

The data collection method with documentation is a way to obtain data and information both in the form of documents, writings, archives, numbers and images that aim to support research [13]. The documentation data in this study is Erigo Store product sales data from January 2021 to June 2023.

- Web Scraping

Web Scraping is a method that aims to collect a lot of data and extract data from websites that are processed into data or information that can be analyzed and used for various purposes [14]. Data was collected from January 2021 to June 2023 throughthe *shopee e-commerce* platform at the Erigo Store.

**E. Data Analysis**

Data analysis is a way to process data systematically to draw research conclusions [15]. The following is a description of data analysis:

- Data Selection. Data selection is one of the stages of KDD or *Knowledge Discover in Database,* data selection is the process of selecting data according to research objectives [16]. The data selected for this research comes from the *e- commerce* sales word *Shopee* Erigo Store from January 2021 to June 2023 and is processed to predict Erigo Store sales. Product category, quantity, and month and year are the attributes selected and processed to predict sales.
- Prepocessing. Prepocessing is also one of the stages of KDD, this stage has two processes, namely cleaning and enrichment, this stage is carried out before the data mining process where cleaning or cleaning is carried out for the data mining process. eliminating unnecessary data such as duplicate data, and irrelevant data. Meanwhile, enrichment is a process to add or enrich data with relevant data [17]. Preprocessing functions to facilitate prediction calculations, preprocessing is used to classify the types of products sold based on the category and quantity sold eachmonth and year. After the data is classified, the data is collected and summed to calculate the total sales for that product category.
- Transformation. Transformation is the process of creating attributes with structured procedures to obtain data or documents that are suitable for certain purposes [18].

**RESULTS**

After making the research design stages of applying data mining using the naive bayes algorithm method to analyze Erigo Store product sales in shopee e-commerce, the researcher implements the naive bayes algorithm method on sales data from shopee e-commerce to analyze erigo store sales. Researchers extracted sales data using Matlab.

## DISCUSSION

### 1. Web Scraping

Web scraping is the process of extracting data from a particular website using specific software or a particular bot. In this way, users can retrieve data from various website and store it in a processable format, so that the data can be analyzed and used for various purposes.

### 2. Data Selection

Data selection is data that is used or selected in accordance with certain research objectives. Data taken from Erigo store e-commerce sales from January 2021 to June 2023, to analyze erigo store sales.

**Table 1**. Erigo Store Data Selection

| No. | Nama Produk | Kuant | Bulan |
|---|---|---|---|
| 1 | Erigo T Shirt Project Summer | 159 | JANUARI |
| 2 | Erigo T Shirt Longboard Green Unisex | 98 | JANUARI |
| 3 | Erigo Chino Short Pants Light Color Unisex | 50 | JANUARI |
| ...... | ........................................... | ..... | ............ |
| 1167 | Erigo Flannel Vilga Brown Unisex | 26 | JUNI |

### 3. Preprocessing

The purpose of the preprocessing stage is to group or categorize erigo store sales data according to the number of sales in three years. After data grouping, all data is summed according to the erigo store product sales cayegory.

### 4. Transformation

The transformations stage is a stage to convert data into data that is suitable for the data mining test process. The transformation stage can also create training data based on data that will be selected first wich aims to see the attributes that affect Erigo Store sales, where from this training data can be determined relevant target data for the data mining test

process. When data has been properly processed, data mining algorithms can find more relevant patterns and informations, wich in turn increases the ability to make more informed decisions and more accurate predictions.

### 5. Naive Bayes

$$P(H|X) = P(H|X)P(H)/P(X)$$

**Description :**

X        = Unknown class data

H        = Hypothesized class specific X data

P(H|X) = Probability of hypothesis H based on condition X

P(H)     = Probability of hypothesis H

P(X|H)  = Probability of X based on the condition H

P (X)     = Probability of X

### 6. Matlab Test

After the selection data is obtained, the selection data is input into matlab. The data analyzed is 1167 data, starting from January 2021 to June 2023. Data to predict sales is in excel format so that it can be accessed and analyzed using matlab with the Naive Bayes Algorithm.
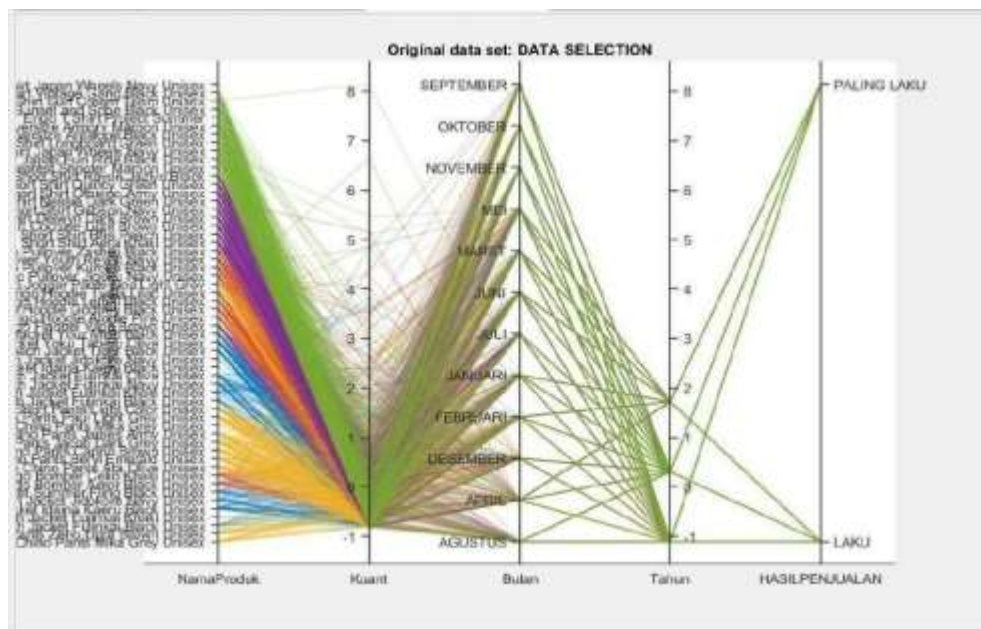


**Figure 2.** Graph of Original Set Sales Predictions

In figure 1. Above, it can be seen that the accuracy rate is 98.3%. Predictors or variables used to analyze Erigo Store sales consist of product name, quantity, month, year and sales results and also consist of 5 Erigo Store product categories, namely Shirt, T-Shirt, Outwear, Jacket and Pants using sales data from January 2021 to June 2023.



**Figure 3.** Classification of True Positive & False Negative Data Matlab

Based on figure 3, the results of Matlab classification modeling obtained the overall accuracy of products for 5 categories is 98.3% with different accuracy results for each product. Accuracy for the T-Shirt category is 98.6% with a false negative rate (FNR) of 1.4%, the accuracy of the Shirt category is 98.4% while the FNR rate is 1.6%, the accuracy for the Pants category is 98.1% with FNR rate of 1.9%, the accuracy of the Outwear category is 98.7% with an FNR of 1.3% and the accuracy for the Jacket category is 97.6% with an FNR of 2.4%.
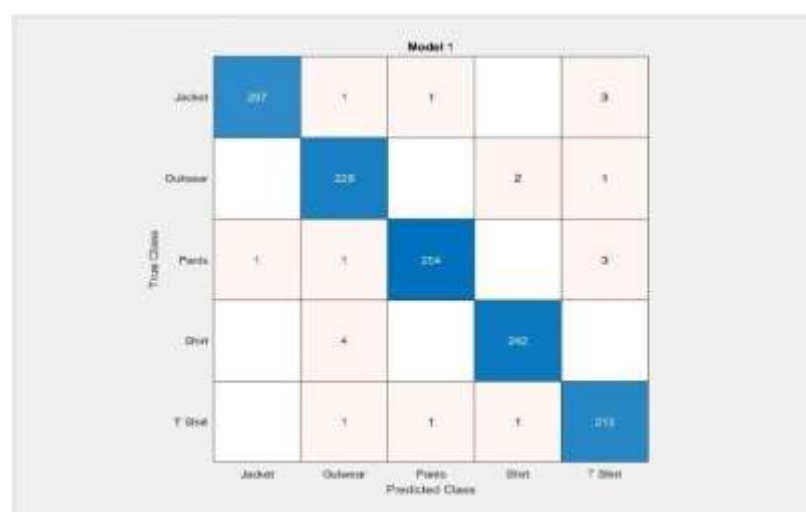


**Figure 4.** Classification Model Number of Observations

Figure 4 is the number of observations model classification modeling, where the number on the graph is the number of categories that are considered the most salable and salable. Before the data in training is processed using matlab, the author determines the average value of the data first using excel software to determine the quantity that is categorized as most salable adan salable. In this study there is no category that is not practible, because there is no category quantity that is 0. The average value obtained is 122, so it determined; >122 = most practicable; <122 = practicable.

## CONCLUSION

Based on the results of previous research entitled Application of Data Mining To Predict Sales at Erigo Store with K-Nearest Neighbor, it is noted that the overall accuracy of product category sales classification is 83.62%. Accuracy for the Jacket category reached 95%, Shirt reached 100%, Pants reached 92.31%, Outwear reached 89.47% and T-Shirt reached 60%, where this previous research used rapid miner tools.

Whereas in the research Application of Naive Bayes Algorithm for Sales Analysis at Erigo Store has an overall accuracy of 98.3% with different accuracy results for each category. Accuracy for the T-Shirt category is 98.6% with a False Negative Rate (FNR) of 1.4%, Shirt accuracy is 98.4% and for an FNR level of 1.6%, accuracy for the Pants category is 98.1% with an FNR of 1.9%, Outwear category accuracy is 98.7% with an FNR of 1,3%, and accuracy for the Jacket category is 97.6% with an FNR of 2.4%, using matlab to determine sales of each product. The data set used for this research starts from January 2021 to June 2023.

## References

[1]  R. H. Saragih, W. A. Pamungkas, F. Yumna, D. Sitanggang, dan S. Wardani, "Penerapan Data Mining Untuk Memprediksi Penjualan Di Erigo Store Dengan K-Nearest Neighbor," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, vol. 8, no. 2, hlm. 649–661, 2023, [Daring]. Tersedia pada: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

[2]  I. R. Pratama, Maimunah, dan E. R. Arumi, "Sistem Klasifikasi Penjualan Produk Alat Listrik Terlaris Untuk Optimasi Pengadaan Stok Menggunakan Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, hlm. 2135–2139, Okt 2022, doi:

10.30865/mib.v6i4.4418.

[3]    F. Widiastuti, W. Murniati, dan Salkin, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI PENJUALAN KAIN TENUN MNGGUNAKAN REGRESI LINEAR STUDI KASUS: UD.BINTANG REMAWE SUKARARE," *JURNAL ILMIAH TEKNIK MESIN, ELEKTRO DAN KOMPUTER (JURITEK)*, vol. 2, no. 1, hlm. 27–39, 2022.

[4]    F. S. Wahyuni dan H. Z. Zahro', "Penerapan Teknik Data Mining untuk Menentukan Rencana Strategi Penjualan," *JUPITER (Jurnal Pendidikan Teknik Elektro)*, vol. 07, no. 01, hlm. 47–54, 2022, [Daring]. Tersedia pada: https://www.kaggle.com/

[5]    Nosiel, Sriyanto, dan F. Maylani, "Perbandingan Teknik Data Mining Untuk Prediksi Penjualan Pada UMKM Gerabah," *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat*, hlm. 72–86, 2021.

[6]    A. Purwanto, S. I. Isnawati, dan N. L. Ramadhani, "PENGEMBANGAN USAHA BISNIS RETAIL MODERN PADA TOKO PAKAIAN KEDJORA GROSIR," *Bakti Humaniora*, vol. II, no. 1, hlm. 29–37, 2022.

[7]    Q. A'yuniyah, W. Elvira, N. Nazira, I. Ambarani4, S. F. Intan, dan D. Ramadhani, "Naïve Bayes Classifier (NBC) Algorithm Analysis for Prediction Medical Device Sales," *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, vol. 3, no. 2, hlm. 119–126, 2023.

[8]    J. Suntoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT. Elex Media Komputindo, 2019.

[9]    A. A. Ronaldi dan N. Hunafi, "IMPLEMENTASI DATA MINING UNTUK PREDIKSI PENJUALAN PESTISIDA PADA CV MITRA ARTHA SEJATI MENGGUNAKAN ALGORITMA NAIVE BAYES," *eProsiding Teknik Informatika (PROTEKTIF)*, vol. 1, no. 1, hlm. 250–257, 2020, [Daring]. Tersedia pada: http://eprosiding.ars.ac.id/index.php/pti

[10]   N. P. G. Naraswati, D. C. Rosmilda, D. Desinta, F. Khairi, R. Damaiyanti, dan R. Nooraeni, "Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 1, hlm. 222–238, 2021, [Daring]. Tersedia pada: http://sistemasi.ftik.unisi.ac.id

[11]   M. Ali, T. Hariyati, M. Y. Pratiwi, dan S. Afifah, "Metodologi Penelitian Kuantitatif Dan Penerapan Nya Dalam Penelitian," *Education Journal : Penelitian Ibnu Rusyd*

*Kotabumi*, vol. 2, no. 2, 2022.

[12]  Mardiayanti dan A. N. Andriana, "Pengaruh Harga dan Kualitas Produk serta Review Produk terhadap Keputusan Pembelian Produk Scarlett Whitening (Studi Kasus pada Mahasiswa Adminstrasi Bisnis)," *Jurnal Pendidikan dan Kewirausahaan*, vol. 10, no. 3, hlm. 1091–1109, Okt 2022, doi: 10.47668/pkwu.v10i3.560.

[13]  O. N. Zafira, "STUDI LITERATUR TENTANG PENGARUH MODIFIKASI MEDIA PEMBELAJARAN PENDIDIKAN JASMANI TERHADAP HASIL BELAJAR," 2020.

[14]  M. Djufri, "PENERAPAN TEKNIK WEB SCRAPING UNTUK PENGGALIAN POTENSI PAJAK (Studi Kasus pada Online Market Place Tokopedia, Shopee dan Bukalapak)," *Jurnal BPPK*, vol. 13, no. 2, hlm. 65–75, 2020.

[15]  Ahmad dan Muslimah, "Memahami Teknik Pengolahan dan Analisis Data Kualitatif," *Proceedings : Palangkaraya International and National Conference On Islamic Studies (Pincis)*, vol. 1, no. 1, hlm. 173–186, 2021.

[16]  Gustientiedina, M. H. Adiya, dan Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 05, no. 01, hlm. 17–24, Apr 2019, doi: 10.25077/TEKNOSI.v5i1.2019.17-24.

[17]  S. Z. Harahap dan A. Nastuti, "TEKNIK DATA MINING UNTUK PENENTUAN PAKET HEMAT SEMBAKO DAN KEBUTUHAN HARIAN DENGAN MENGGUNAKAN ALGORITMA FP-GROWTH (STUDI KASUS DI ULFAMART LUBUK ALUNG)," *Informatika : Jurnal Ilmiah Fakultas Sains dan Teknologi*, vol. 7, no. 3, hlm. 111–119, 2019.

[18]  D. R. F. Daud, B. Irawan, dan A. Bahtiar, "PENERAPAN METODE NAIVE BAYES PADA ANALISIS SENTIMEN APLIKASI MCDONALDS DI GOOGLE PLAY STORE," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, hlm. 759–766, 2024.