ISSN 2085-4552

Data Mining Motorbike Sales Classification Using a Combination of K-Means and Naïve Bayes Algorithms

Eka Sofiati[1], Dori Gusti Alex Candra[2], Irzon Meiditra[3]
[1,2] Dept. of Information System, Mitra Gama Institute of Technology, Duri, Indonesia
[3] Dept. of Information Technology, Rokan Hilir Institute of Technology, Rokan Hilir, Indonesia
[1]ekasfti@gmail.com, [2]dorigustialexcandra@gmail.com, [3]meiditrairzon@gmail.com

Accepted 04 October 2024
Approved 08 January 2025

***Abstract*— The showroom, which has been established since 2006 and is located in Lapai, Padang City, has problems, namely the difficulty of analyzing consumer demand and the amount of sales data that accumulates. In addition, many stock items are not available when consumer demand is high. From these problems, a data mining application system is needed to improve sales patterns and process sales data to determine items that are often purchased and not using data mining methods, namely K-Means and Naïve Bayes. The data is obtained directly from CV Unique Motor in the form of motorcycle sales data and motorcycle inventory data. The stages of the research include several stages, namely problem identification, data collection, data analysis, clustering stage analysis, classification stage analysis, system analysis, and system implementation. At the system analysis stage, a design will be carried out using data mining by applying the K-Means and Naive Bayes algorithms, where the program will be executed in the PHP programming language and MySQL database. The calculation results show that the highest probability value is in the many variable, namely 0.0607, so it can be concluded that the sales level is a lot. With this system, it can speed up the showroom in making decisions from the data taken so that the showroom can increase the amount of stock that has a best-selling classification so that the showroom does not run out of stock and can help the showroom classify motorcycle sales and can harmonize the availability and inventory of existing motorbikes by classifying sales volume.**

***Index Terms*— Data Mining; Classification; Sales; K-Means; Naïve Bayes**.

I. INTRODUCTION

CV. Unique Motor is a motorcycle buying and selling showroom, which is the object of the author's research this time. This showroom, which has been established since 2006 and is located in Lapai, Padang City, has a problem, namely the difficulty in analyzing consumer demand and a lot of accumulation of sales data based on motorcycle inventory stock sales transaction data from October 2021 to March 2022, totaling 37 motorcycle data. In addition, many stock items are not available when consumer demand is high. From these problems, a data mining application system is needed to improve sales patterns and process sales data to determine what is often purchased and not using data mining methods, namely K-Means and Naïve Bayes.

Data mining is a step in performing knowledge discovery in databases. Knowledge discovery as a process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and knowledge presentation [1]. The KMNB approach is made using a combination of clustering and classification techniques. For data classification based on categories using the Naïve Bayes method. And clustering helps to identify groups that have the same characteristics or show characteristics at the beginning [2]. Clustering is the classification of unsupervised patterns such as observations, data items, and vectors into groups called clusters [3]. The goal is to minimize the objective function set by the clustering process [4]. The basic concept of clustering is to group a large number of objects into a cluster. A good cluster is a cluster that has a high degree of similarity between objects in the cluster [5]. Clustering is a data grouping method that starts by grouping two or more of the most similar objects [6]. K-Means is a non-hierarchical data clustering method that attempts to divide existing data into two or more groups, where data with similar characteristics [2].

Classification is a technique used to predict the class or property of each data instance and classify the data itself based on the topic [7]. Classification performs the process of finding a model or function that describes and characterizes a class of data for a specific purpose [8]. Each classification algorithm used produces a best-fit model that relates the input data to a known classification class [9]. Classification consists of three steps, namely model building, model application, and evaluation [10]. Naïve Bayes is a classification

proposed by British scientist Thomas Bayes and is known as Bayes' theorem because it uses probalistic and statistical methods [11]. Naïve Bayes is a simple algorithm and is used to predict a probability in the membership of a class [12].

Some of the research conducted previously includes the Comparison of Naive Bayes and C4.5 Algorithms in Determining the Sales Level of Honda Motorbikes This research was conducted to determine the sales level of Honda motorbikes using motorcycle sales datasets using data mining [13]. Another study on Product Sales Level Analysis Using K-Nearest Neighbor (KNN) and K-Means explains that high sales require information and identify which products are the most potential and to sell, so that goods in the warehouse do not accumulate as needed [14].

Other research using the K-Means and Naïve Bayes algorithms has also been done before. Research on Determining Livestock Aid Recipients Using K-Means & Naive Bayes Algorithms explains that the combination of classification and clustering is believed to be able to provide accurate classification results when determining data [2]. Another study on Literacy Classification Using K-Means-Naïve Bayes Algorithm explains that to get the optimal number of classes through the clustering process and find out the results of the classification process [15]. The combination of data mining methods in this study hopes to produce accurate classification by combining clustering and classification

II. METHOD

This research proposes a combination of clustering and classification methods. The clustering method is used to obtain several groups of motors. The clustering process is carried out using the K-Means algorithm. After the motorcycle groups are obtained, the classification method is used to obtain the sales classification. The classification process is carried out using the Naïve Bayes algorithm. The research methodology proposed in this study can be seen in Figure 1.

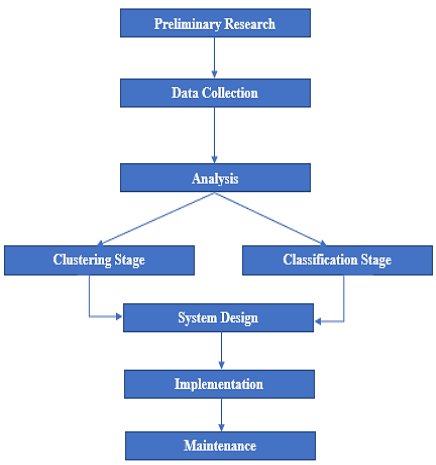


Fig. 1. Research Framework

A. *Preliminary Research*

First, confirm that you have the correct template for your paper size. This template is for ULTIMATICS. It has been tailored for output on the A4 paper size.

B. *Data Collection*

Make direct observations and conduct interviews with motorcycle showrooms and collect literature related to this research.

C. *Analysis*

Analyze the data obtained from the showroom and determine what data mining techniques and methods are used, then analyze the data mining application to be built.

D. *Clustering Stage*

In this clustering stage is the first stage of this research, the clustering process is carried out using the K-Means algorithm. The clustering process begins by determining the number of clusters to be formed, where the motorcycle data will be grouped into clusters, namely “Little Sold” and “Much Sold”. Then determine the initial centroid randomly. Then calculate the distance of each data to each centroid using the Euclidean Distance formula as shown in equation (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}2 \qquad (1)$$

E. *Classification Stage*

The classification stage is the second stage of this research; the classification process is carried out using the Naïve Bayes algorithm. Motorcycle data that has been clustered from the first stage is used as input data for the second stage. The data used is the “sold a lot” result data. The classification process is grouped into several criteria, where these criteria will be the determinant in decision-making. using the Naïve Bayes formula as shown in equation (2).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad (2)$$

F. *Implementation*

To build data mining applications in this study, using the PHP programming language and supported by the MySQL database.

III. RESULTS AND DISCUSSION

The data used in this study is the total inventory data and motorcycle sales for October 2021-March 2022. The sample of motorcycle data that is the object of research is as in table 1 below:

TABLE I. MOTORCYCLE INVENTORY AND SALES DATA

No	Motor Brand	Total Inventor	Total Sales
1	Beat	51	29
2	CB 150 R	9	4
3	PCX	1	1
4	Supra X	10	9
5	Vario 125	12	11
6	Vario CW	6	1
7	Vario fi	7	4
8	Satria FU	10	2
9	Mio	4	3
10	Mio J	5	4
….	….	….	….
…	…	….	….
37	Mio CW	1	0

A. K-Means

The first step is to determine the cluster center or centroid which is obtained randomly or randomly for the initial determination of the cluster assumed:

TABLE II. INITIAL CLUSTER CENTER

Motor Brand	Total Inventor	Total Sales	Claster
NMax	6	5	C1
Revo	3	2	C2

Next, the distance of each data with the cluster center is calculated using the euclidien distance formula (equation 1) from all data to each first center point:

$d_{11} = \sqrt{(51-6)^2} + \sqrt{(29-5)^2} = 51$
$d_{12} = \sqrt{(9-6)^2} + \sqrt{(4-5)^2} = 3{,}1623$
$d_{13} = \sqrt{(1-6)^2} + \sqrt{(1-5)^2} = 6{,}4031$
$d_{14} = \sqrt{(10-6)^2} + \sqrt{(9-5)^2} = 5{,}6569$
$d_{15} = \sqrt{(12-6)^2} + \sqrt{(11-5)^2} = 8{,}4853$
……..
……..
$d_{137} = \sqrt{(1-6)^2} + \sqrt{(0-5)^2} = 7{,}0711$

In the same way calculate the distance of each 2nd center point and we will get:

$d_{21} = \sqrt{(51-3)^2} + \sqrt{(29-2)^2} = 55{,}0727$
$d_{22} = \sqrt{(9-3)^2} + \sqrt{(4-2)^2} = 6{,}3246$
$d_{23} = \sqrt{(1-3)^2} + \sqrt{(1-2)^2} = 2{,}2361$
$d_{24} = \sqrt{(10-3)^2} + \sqrt{(9-2)^2} = 9{,}8995$
$d_{25} = \sqrt{(12-3)^2} + \sqrt{(11-2)^2} = 12{,}7279$
……
……...
$d_{237} = \sqrt{(1-3)^2} + \sqrt{(0-2)^2} = 2{,}8284$

The second is to calculate the new center point of the 1st iteration. By calculating the mean of cluster member data from each cluster, the result of the cluster 2 center point value is:

TABLE III. NEW CENTROID CENTER

C1	12,9091	8,2727
C2	1,9615	1,3462

If in the second iteration no cluster member moves to another cluster then the iteration stops but if any cluster member moves to another cluster then proceed to the next iteration. In this study the iteration stops until the 8th iteration. Here are the final results of the cluster calculation at the 8th iteration:

TABLE IV. NEW CENTROID CENTER

No	Motor Brand	C1	C2	Cluster
1	Beat	0	53,9093	C1
2	CB 150 R	48,8774	5,2214	C2
3	PCX	57,3062	3,3972	C2
4	Supra X	45,6180	8,7424	C2
5	Vario 125	42,9535	11,5704	C2
6	Vario CW	35	2,6639	C2
7	Vario fi	50,6063	3,3228	C2
8	Satria FU	49,0918	6,0952	C2
9	Mio	53,7122	0,3106	C2
10	Mio J	52,3546	1,6789	C2
….	….	….	….	….
…	…	….	….	….
37	Mio CW	57,8014	3,9912	C2

Furthermore, the data that will be forwarded to continue the Naive Bayes classification is C2, namely “Many Sold”.

B. Naïve Bayes

The decision-making process in the classification of motorcycle sales will be grouped into 5 criteria, where these criteria will determine decision-making in the motorcycle showroom. These criteria are:

1) Motor Type

TABLE V. CLASSIFICATION OF MOTOR TYPES

Motor Type	Honda
	Yamaha
	Suzuki
	Kawasaki

2) Color

TABLE VI. COLOR CLASSIFICATION

Monoton	Hitam, Putih, Biru Merah
Variasi	Putih, Biru, Merah Putih
	Hitam, Putih, Biru Merah

3) Year

TABLE VII. YEAR CLASSIFICATION

2021 Down	2010-2021
2006 Above	2006-2017

4) Category

TABLE VIII. YEAR CLASSIFICATION

Category	Matic
	Manual

5) Sales

Sedikit	<=2
Banyak	>=3

The training data used for classification amounted to 36, where the training data will then be processed using the Naïve Bayes method. The training power used can be seen in the table:

TABLE IX. TRAINING DATA

Brand	Type	Color	Year	Category	Sales
CB 150 R	Honda	Variasi	2021 Kebawah	Manual	Banyak
PCX	Honda	Monoton	2006 Keatas	Matic	Sedikit
Supra X	Honda	Variasi	2006 Keatas	Manual	Banyak
Vario 125	Honda	Variasi	2021 Kebawah	Matic	Banyak
Vario CW	Honda	Monoton	2006 Keatas	Matic	Sedikit
Vario fi	Honda	Varasi	2021 Kebawah	Matic	Banyak
Satria FU	Suzuki	Variasi	2006 Keatas	Manual	Sedikit
Mio	Yamaha	Monoton	2006 Keatas	Matic	Banyak
Mio J	Yamaha	Monoton	2021 Kebawah	Matic	Bannyak
Mio S	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
NMAX	Yamaha	Monoton	2021 Kebawah	Matic	Banyak
Vixion	Yamaha	Monoton	2006 Keatas	Manual	Banyak
Kharisma	Honda	Variasi	2006 Keatas	Manual	Sedikit
Supra X	Honda	Variasi	2006 Keatas	Manual	Sedikit
Vario 150	Yamaha	Variasi	2006 Keatas	Matic	Banyak
Spin	Suzuki	Variasi	2006 Keatas	Matic	Sedikit
Freego	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Mio GT	Yamaha	Variasi	2006	Matic	Sedikit

Xeon	Yamaha	Variasi	2006 Keatas	Matic	Banyak
Beat Street	Honda	Monoton	2021 Kebawah	Manual	Sedikit
GSX	Suzuki	Variasi	2021 Kebawah	Manual	Sedikit
R15	Yamaha	Monoton	2006 Keatas	Manual	Sedikit
Soul GT	Yamaha	Monoton	2006 Keatas	Matic	Sedikit
Revo	Honda	Monoton	2006 Keatas	Manual	Sedikit
Fino	Yamaha	Variasi	2006 Keatas	Matic	Sedikit
Mio Soul	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Scoopy	Honda	Variasi	2021 Kebawah	Matic	Banyak
Ninja SS	Kawasaki	Variasi	2006 Keatas	Manual	Sedikit
Ninja RR	Kawasaki	Variasi	2006 Keatas	Manual	Sedikit
MX King	Yamaha	Variasi	2006 Keatas	Manual	Sedikit
Mega Pro	Honda	Monoton	2006 Keatas	Manual	Sedikit
Supra FIT	Honda	Variasi	2006 Keatas	Manual	Sedikit
Mio Z	Yamaha	Monoton	2006 Keatas	Matic	Sedikit
Genio	Honda	Monoton	2021 Kebawah	Matic	Sedikit
Mioo M3	Yamaha	Monoton	2021 Kebawah	Matic	Sedikit
Mio CW	Yamaha	Variasi	2006 Keatas	Matic	Sedikit
Vario 150	Yamaha	Variasi	2006 Keatas	Matic	Banyak

In solving using the Naïve Bayes method if the sales level is not known as the testing data below:

TABLE X. DATA TESTING

Merek	Jenis	Warna	Tahun	Kategori	Penjualan
Mio	Yamaha	Variasi	2021 Kebawah	Matic	???

The first is to calculate the probability value of each class:

P(Penjualan = “Sedikit”) = 24/36 = 0,6667

P(Penjualan = “banyak”) = 12/36 = 0,3333

P(Jenis = Honda | Penjualan = “Sedikit”)

= 8/24 = 0,6667

P(Jenis = Honda | Penjualan = “Banyak”)

= 7/12 = 0,375

P(Warna = Variasi | Penjualan = “Sedikit”)

= 12/24 = 0,5417

P(Warna = Variasi | Penjualan = “Banyak”)

= 6/12 = 0,5833

P(Tahun = 2021 Kebawah | Penjualan = “Sedikit”)

= 6/24 = 0,2917

P(Tahun = 2021 Kebawah | Penjualan = “Banyak”)

= 8/12 = 0,75

P(Kategori = Matic | Penjualan = “Sedikit”)

= 9/12 = 0,8333

P(Kategori = Matic | Penjualan = “Banyak”)

= 13/24 = 0,5833

Selanjutnya kalikan semua variable Sedikit dan Banyak

dengan kelas yang sama:

P(Penjualan = “Sedikit”) X P(Penjualan = “Sedikit”)

= 0,6667 x 0,6667 x 0.5417 x 0,2917 x 0,8333

= 0,0307

P(Penjualan = “Banyak”) X P(Penjualan = “Banyak”)

= 0,3333 x 0,375 x 0.5833 x 0,75 x 0,5833

= 0,0607

From the results of the above calculations, it can be seen that the highest probability value is in the many variable so it can be concluded that the sales level is a lot. System Interface Testing Before entering the system, the admin must input the username and password in the database. The system then validates the username and password entered by the admin.

C. Design Implementation

at this stage is the result of the design

1) Alternative Page View

The alternative page functions to add motorcycle brands that will be clustered. On this page, the admin can view and edit motorcycle brand data.

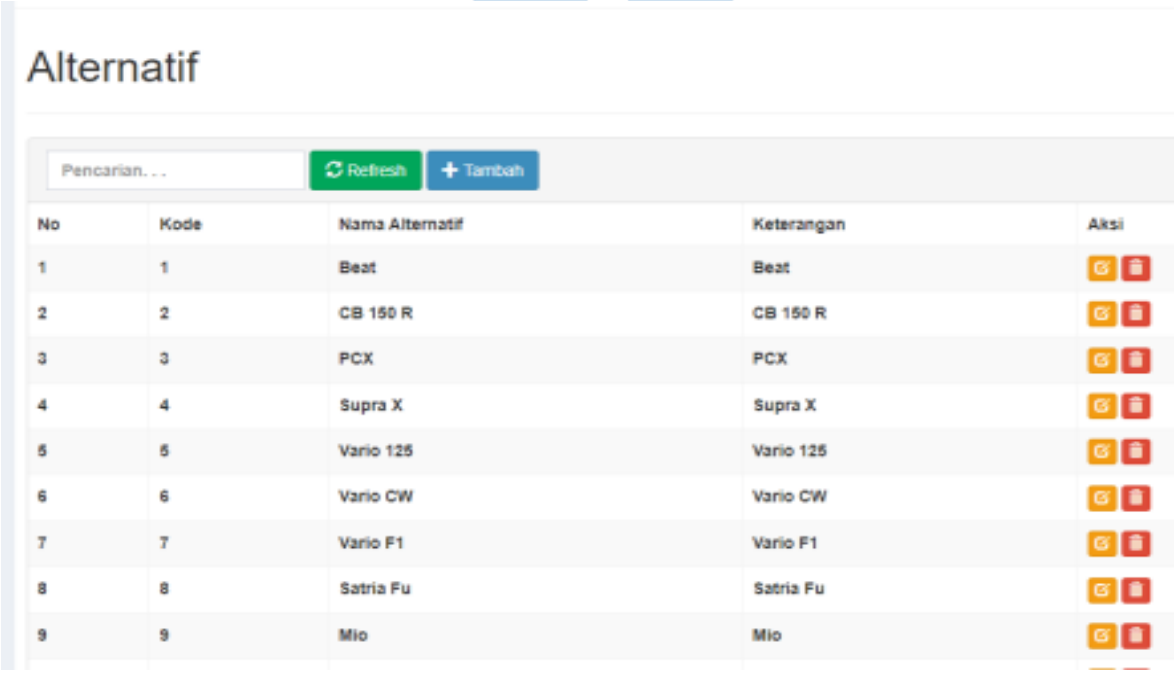
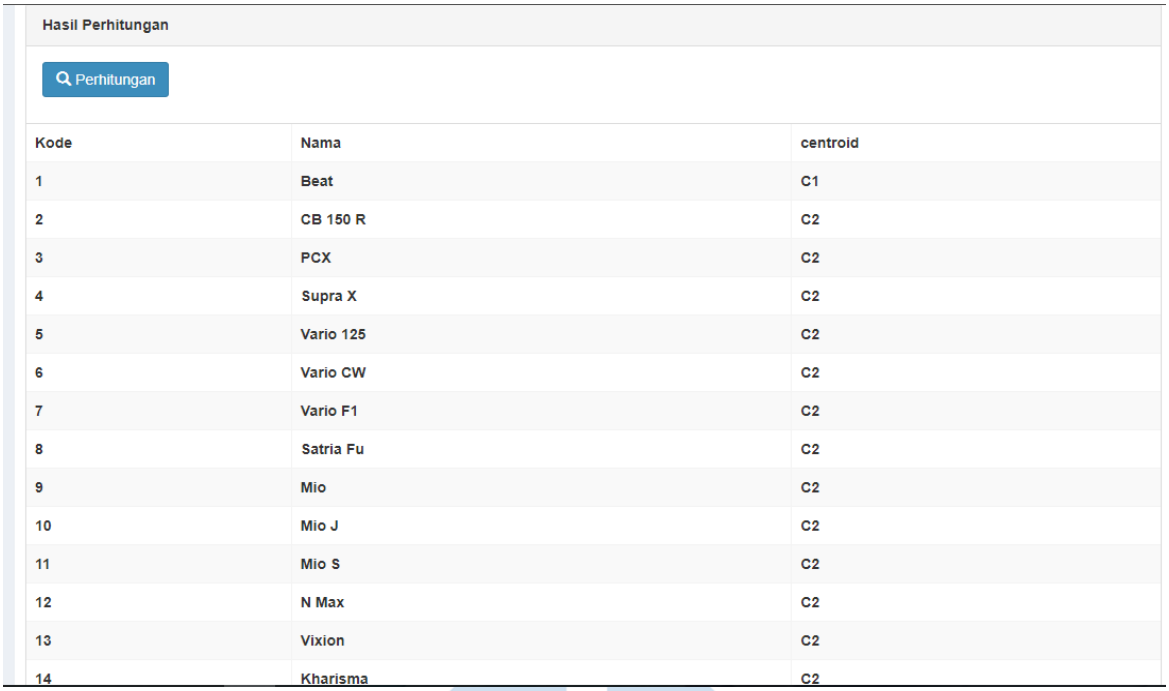


Fig. 2. Alternate Page View Page

2) Calculation Result Page Display

The page displays the training data page that will be used for the Naïve Bayes calculation process, on this page the admin can edit the categories in the dataset and can delete data in the dataset.



Hasil Perhitungan

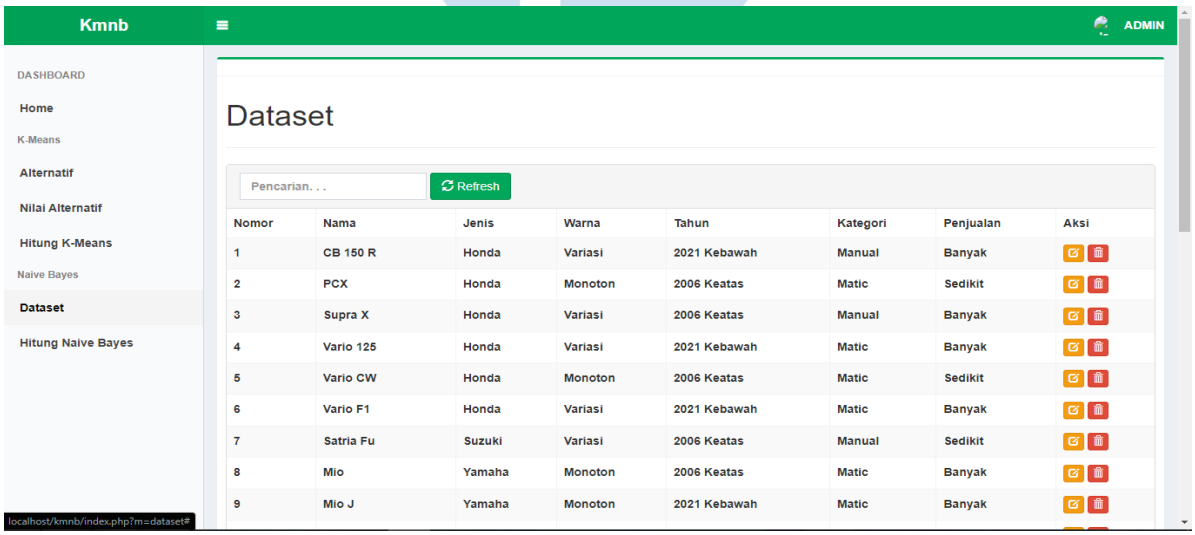
Perhitungan

Kode	Nama	centroid
1	Beat	C1
2	CB 150 R	C2
3	PCX	C2
4	Supra X	C2
5	Vario 125	C2
6	Vario CW	C2
7	Vario F1	C2
8	Satria Fu	C2
9	Mio	C2
10	Mio J	C2
11	Mio S	C2
12	N Max	C2
13	Vixion	C2
14	Kharisma	C2

Fig. 3. *Calculation Result Page Display*

3) Alternative Page View

This page displays the selection of data that will be used as test data for the Naïve Bayes process. displays the selection of data that will be used as test data for the Naïve Bayes process.



Kmnb

ADMIN

DASHBOARD
Home
K-Means
Alternatif
Nilai Alternatif
Hitung K-Means
Naive Bayes
Dataset
Hitung Naive Bayes

Dataset

Pencarian. . .

Refresh

Nomor	Nama	Jenis	Warna	Tahun	Kategori	Penjualan	Aksi
1	CB 150 R	Honda	Variasi	2021 Kebawah	Manual	Banyak	
2	PCX	Honda	Monoton	2006 Keatas	Matic	Sedikit	
3	Supra X	Honda	Variasi	2006 Keatas	Manual	Banyak	
4	Vario 125	Honda	Variasi	2021 Kebawah	Matic	Banyak	
5	Vario CW	Honda	Monoton	2006 Keatas	Matic	Sedikit	
6	Vario F1	Honda	Variasi	2021 Kebawah	Matic	Banyak	
7	Satria Fu	Suzuki	Variasi	2006 Keatas	Manual	Sedikit	
8	Mio	Yamaha	Monoton	2006 Keatas	Matic	Banyak	
9	Mio J	Yamaha	Monoton	2021 Kebawah	Matic	Banyak	

localhost/kmnb/index.php?m=dataset#

Fig. 4. Dataset Page Display Page

4) Display of Naive Bayes Result Page

This page is the result of the Naïve Bayes calculation process. Naïve Bayes result page display

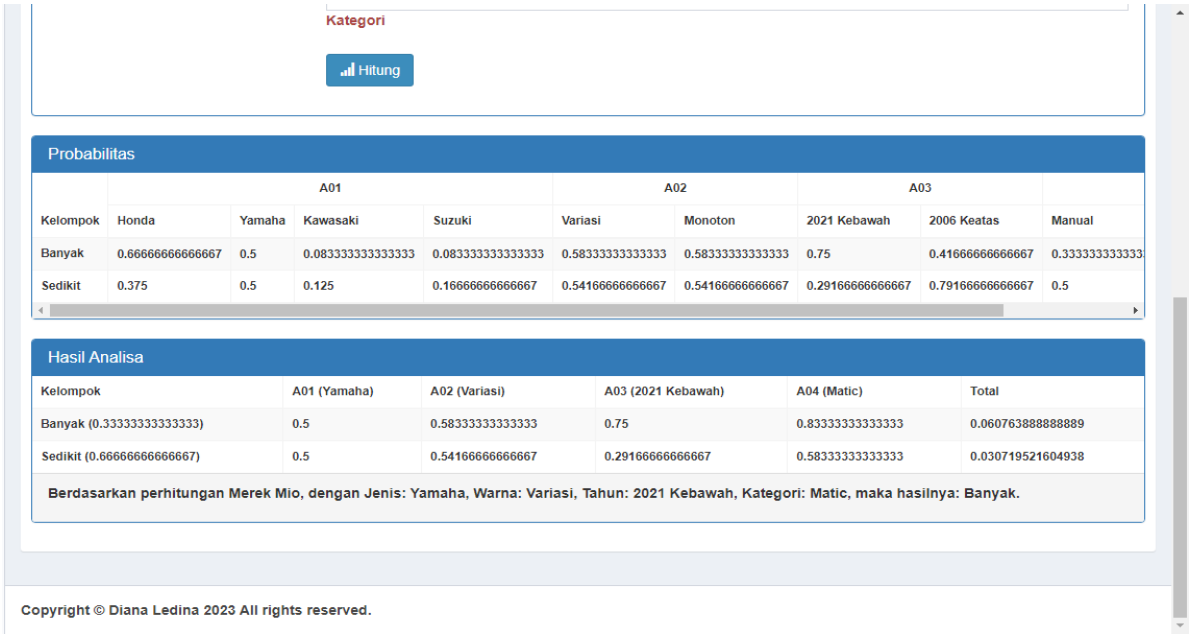


Fig. 5. Naïve Bayes Result Page Display Page

IV. CONCLUSIONS

Based on the total motorcycle inventory and sales data for October 2021 - March 2022 used in this study using the K-means algorithm, it shows that the 8th iteration with the updated cluster center has not changed, so the iteration process is stopped. So the number of clusters to be formed at this stage is two clusters of 37 total inventory and sales data. The clusters to be formed are C1, namely "Little Sold" and C2, namely "Much Sold". Furthermore, the data that will be forwarded to continue the Naive Bayes classification is C2, namely "Sold a lot". From the calculation results it can be seen that the highest probability value is in the many variable which is 0.0607 so it can be concluded that the sales level is a lot. The results of the implementation of the data mining system application are proven to be able to classify sales. This is evidenced by the data on motorcycle inventory and sales that can classify motorcycle sales with the most sales. And the existence of this system is able to solve the problem of empty motorcycle availability when there is a request from the buyer.

REFERENCES

[1] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.

[2] M. F. Asikin, D. E. Ratnawati, and M. A. Fauzi, "Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means & Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 2, pp. 641–649, 2018, [Online]. Available: http://download.garuda.kemdikbud.go.id/article.php?article=871445%5C&val=10384%5C&title=Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means Nave Bayes

[3] A. Muni, "Analisis Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Penjualan Sepeda Motor Studi Kasus PT. Alfa Scorpii," *Juti Unisi*, vol. 4, no. 1, pp. 1–8, 2020, doi: 10.32520/juti.v4i1.1087.

[4] H. Priyatman, F. Sajid, and D. Haldivany, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.

[5] Zulrahmadi, S. Defit, and Y. Yunus, "Pemetaan Wilayah Potensial Terhadap Penjualan Sepeda Motor Honda Menggunakan K-Means Clustering," *J. Inform. Ekon. Bisnis*, vol. 2, no. 2018, pp. 53–59, 2020, doi: 10.37034/infeb.v2i2.41.

[6] Haditsah Annur, "Penerapan Data Mining Menentukan Strategi Penjualan Variasi Mobil Menggunakan Metode K-Means Clustering (Studi Kasus Toko Luxor Variasi Gorontalo)," *J. Inform. Upgris*, vol. 5, no. 1, pp. 40–45, 2019.

[7] M. P. Tana, F. Marisa, and I. D. Wijaya, "Penerapan Metode Data Mining Market

Basket Analysis Terhadap Data Penjualan Produk Pada Toko Oase Menggunakan Algoritma Apriori," *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 2, pp. 17–22, 2018, doi: 10.37438/jimp.v3i2.167.

[8] P. I. Lestari, D. E. Ratnawati, and L. Muflikhah, "Implementasi Algoritme K-Means Clustering Dan Naive Bayes Classifier Untuk Klasifikasi Diagnosa Penyakit Pada Kucing," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 1, pp. 968–973, 2019.

[9] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[10] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.

[11] D. Rezekika, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penjualan Spare Part Sepeda Motor," *J. Pelita Inform.*, vol. 8, no. 3, pp. 326–329, 2020.

[12] N. E. Akbar, P. P. Adikara, and R. C. Wihandika, "Analisis Sentimen Layanan Astra Honda Motor Menggunakan Metode Naïve Bayes dan Identifikasi Aspek pada Layanan Menggunakan DBSCAN," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, pp. 929–937, 2020, [Online]. Available: http://j-ptiik.ub.ac.id

[13] D. Prabowo, F. Hidayat, G. Gumelar, D. Qintoro, and A. Setiawan, "PERBANDINGAN ALGORITMA NAÏVE BAYES DAN C4.5 DALAM MENENTUKAN TINGKAT PENJUALAN MOTOR HONDA," *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 16, no. 3, pp. 67–76, Nov. 2023, doi: 10.61805/fahma.v16i3.91.

[14] Herianto, N. Syamsiyah, and S. Anggraini, "ANALISA TINGKAT PENJUALAN PRODUK MENGGUNAKAN K-NEAREST NEIGHBOR (K-NN) DAN K-MEANS (STUDI KASUS PERUSAHAAN KAYU ELANG PERKASA)," *J. Sains Teknol.*, vol. XI, no. 1, pp. 8–18, 2021.

[15] M. F. A. Saputra, T. Widiyaningtyas, and A. P. Wibawa, "Illiteracy Classification Using K Means-NaÃ¯ve Bayes Algorithm," *JOIV Int. J. Informatics Vis.*, vol. 2, no. 3, pp. 153–158, May 2018, doi: 10.30630/joiv.2.3.129.