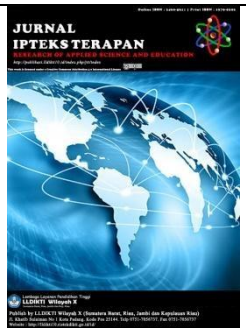

Vol. 17 No.2 2023

COMPARATIVE ANALYSIS OF K-NEAREST NEIGHBOR AND NAÏVE BAYES CLASSIFIER ALGORITHMS FOR CLASSIFICATION OF FEASIBILITY OF ACCEPTANCE OF PAWN GOODS

Mely Syafira*[1)], Mariza Devega[2)]
[12]Universitas Lancang Kuning Pekanbaru, Indonesia
*Email: melysfr@gmail.com

Submit :
06/12/2022

Accept :
28/06/2023

Publish :
30/06/2023

Abstract

One of the problems faced by companies engaged in pawn services is that they must always pay attention to the appropriateness of the value of the loan based on the goods being pawned. Sometimes the pawnbrokers do not come back after depositing their goods with the company and do not pay them off. If the value of the goods does not match and the amount of money lent is too large compared to the value of the pawned goods, the company will lose money. There needs to be a way to determine whether or not it is appropriate for a pawned item to be accepted by the company and the value of the loan given to the pawnbroker. The way to find out can be classified by data mining classification techniques using the K-Nearest Neighbor method and the Naïve Bayes method. The analysis carried out by manual calculations and testing with Rapidminer tools resulted in the accuracy of the comparison of the two methods of the K-Nearest Neighbor and Naïve Bayes algorithms in predicting the classification of the feasibility of receiving pawned goods using the Confusion Matrix formula which resulted in an accuracy rate of the K-Nearest Neighbor method of 40%. while the Naïve Bayes method has an Accuracy Value of 20%, so we can determine that the accuracy of the prediction value using the K-Nearest Neighbor method is better than the Naïve Bayes method.

Keywords: Data Mining, Classification, K-Nearest Neighbor, Naïve Bayes, Rapidminer

ISSN : 1979-9292 / 2460-5611 © 2023 Higher Education Service Institute Region X

JIT is Sinta 3 Journal (https://sinta3.kemdikbud.go.id/journals/profile/2143) accredited by the Ministry of Research & Technology, Republic of Indonesia

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V17.i2 (401-406)

ISSN : 1979-9292

E-ISSN : 2460-5611

INTRODUCTION

CV. Malau Jaya Gadai is one of the pawn services in Pekanbaru City. The company in the field of loan services runs with a pawn scheme. The types of goods that can be mortgaged are cellphones, laptops, motorcycles, moto bpkb, car bpkb, house certificates, gold and others. The loan price list is determined by the company for mortgagers who want to pawn their goods, of course, the goods to be mortgaged by the provisions of their value. One of the problems faced by companies engaged in pawn services is that they must always pay attention to the feasibility of borrowing value based on the goods being mortgaged. Sometimes, the mortgagers do not return after entrusting the goods to the company and not paying them off. If the value of the goods does not match and the amount of money lent is too large when compared to the value of the pawn goods, the company will lose money. There needs to be a way to find out whether or not a lien is accepted by the company and the value of the loan that will be given to the mortgagee.

The way to find out can be classified with data mining classification techniques. Data mining is a science used to mine from several data to produce useful information [1]. This research uses the Naïve Bayes method and also the K-Nearest Neighbor method to dig up hidden information from existing data and analyze it to get information.

Naïve Bayes is a simple probabilistic classification method [2]. This method will calculate a set of probabilities by summing the frequency and combination of values from a given dataset [3]. The naïve Bayes method assumes all attributes in each category have no dependencies on each other (independent) [4]. The advantage of using Naive Bayes is that it only requires a small amount of training data to determine the mean parameters and variance of the variables required for classification [5]. Naive Bayes is a supervised document classification method, that requires training data before the classification process [6].

K-Nearest Neighbor is a method of finding cases by calculating the proximity between new cases and old cases based on the weight match of several existing features [7]. To define the distance between two points in the training data (x) and points in the testing data (y) the formula is used euclidean distance [8].

By conducting data mining analysis using the two methods above to classify whether or not a pawn item is accepted by the company. And also aims to be able to help the company in the decision-making process related to mortgage issues.

RESEARCH METHODS

This study used two algorithms, namely Naïve Bayes and K-Nearest Neighbor. The Naive Bayes algorithm is a classification with probability and statistical methods while the K-Nearest Neighbor algorithm is a classification method based on attributes and training samples. Here is the general equation of the Naïve Bayes algorithm [9] :

$$(Ci|X) = P(X|Ci)\ P(Ci)\ P(X) \qquad (1)$$

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V17.i2 (401-406)

ISSN : 1979-9292

E-ISSN : 2460-5611

Here is the general equation of the K-Nearest Neighbor algorithm [10] :

$$d(x,y) = \sqrt{\sum_{i=1}^{r}(Xik - Yik)^2} \qquad (2)$$

Where :
d = is the euclidean distance between points in the training data
X = data training
Y = data testing
R = number of variables

RESULTS AND DISCUSSION

In preprocessing this data is done by taking raw data derived from the CV. Malau Jaya Gadai is then recapitulated in the form of data stored in Excel (.xls) format, which will later be used as new data consisting of item type data (Var. A), Loan Category (var. B), and Timeframe (Var. C) as a calculation material.

No	Item	Loan	Tenor	Rate
1	SMARTPHONE	2.000.000	30	260.000
	SMARTPHONE	1.000.000	15	130.000
	SMARTPHONE	1.200.000	15	156.000
2	LAPTOP/NOTEBOOK	1.500.000	15	195.000
	LAPTOP/NOTEBOOK	2.000.000	30	260.000
	BARANG ELEKTRONIK	800.000	15	104.000
..	…	…	…	…
424	SMARTPHONE	1.200.000	30	156.000
	SMARTPHONE	4.000.000	30	520.000
	SMARTPHONE	1.300.000	30	169.000

Table 1. Preliminary Data

Testing

The testing process is a very important process to find out the extent to which Data Mining designs can be tested using hardware or software.

1. Rapidminer's Initial Look

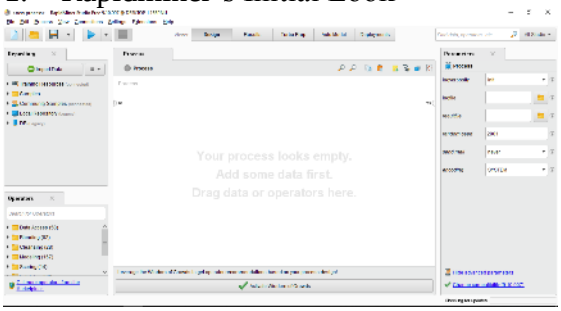

Figure 1. Initial Look

2. Import Database

Data import is done to enter the data to be tested in Excel format.

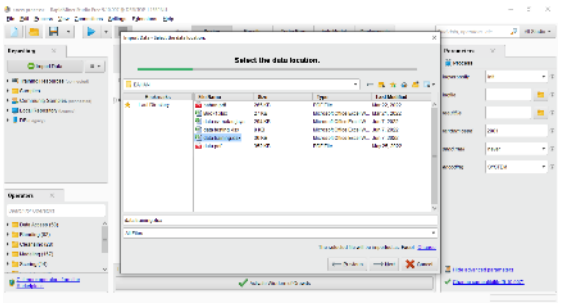

Figure 2. Import Database

3. Import Data Training

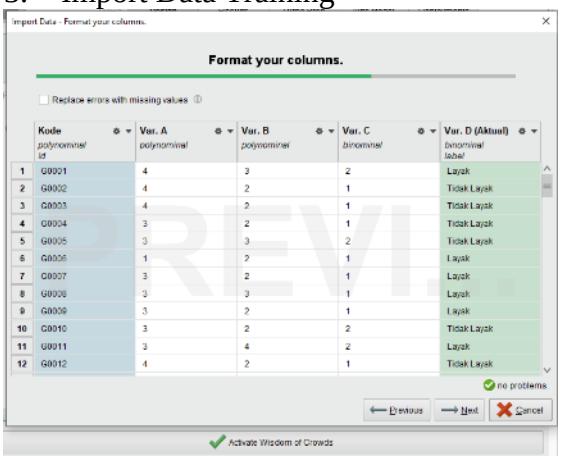

Figure 3. Import Data Training

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V17.i2 (401-406)

ISSN : 1979-9292

E-ISSN : 2460-5611

4. Import Data Testing

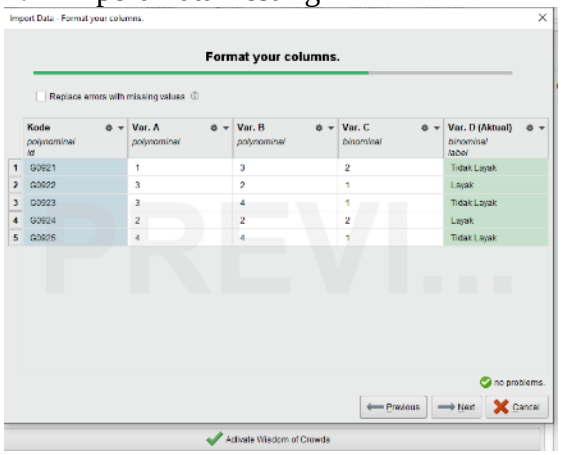

Figure 4. Import Data Testing

5. Repository

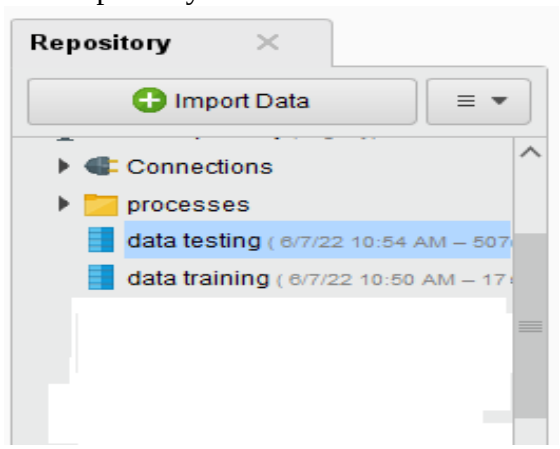

Figure 5. Repository

6. Select Operator

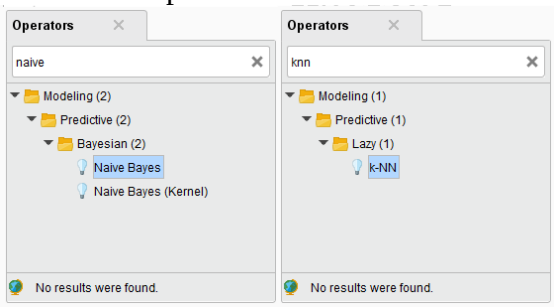

Figure 6. Select Operator

7. Process Method

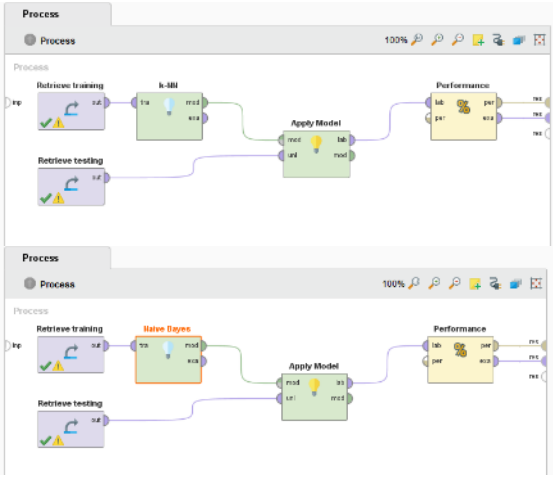

Figure 7. Process Method

8. Setting Parameters

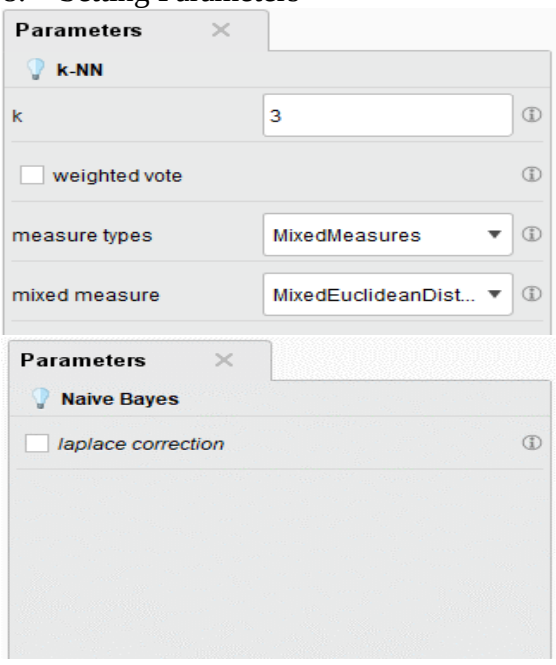

Figure 8. Setting Parameters

9. Example Set K-Nearest Neighbor

JURNAL IPTEKS TERAPAN
Research of Applied Science and Education V17.i2 (401-406)

ISSN : 1979-9292

E-ISSN : 2460-5611

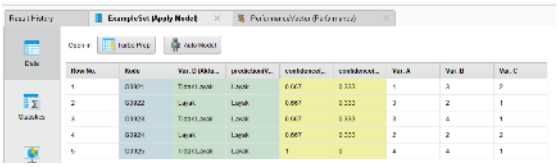

Figure 9. Example Set K-Nearest Neighbor

10. Scatter Plot K-Nearest Neighbor

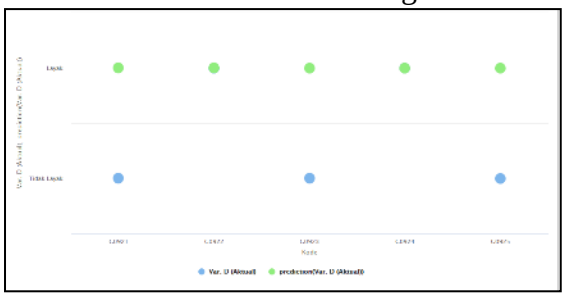

Figure 10. Scatter Plot K-Nearest Neighbor

11. Performance Vector K-Nearest Neighbor

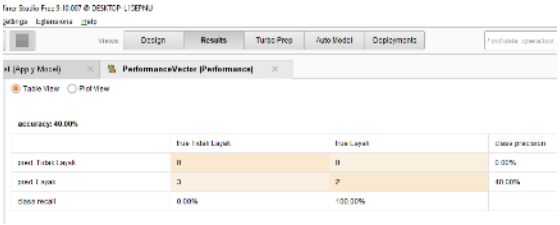

Figure 11. Performance Vector K-Nearest Neighbor

12. Example Set Naïve Bayes

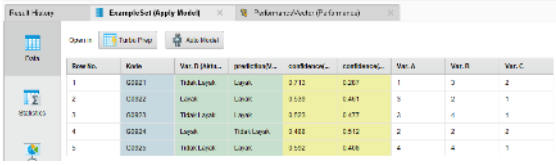

Figure 12. Example Set Naïve Bayes

13. Scatter Plot Naïve Bayes

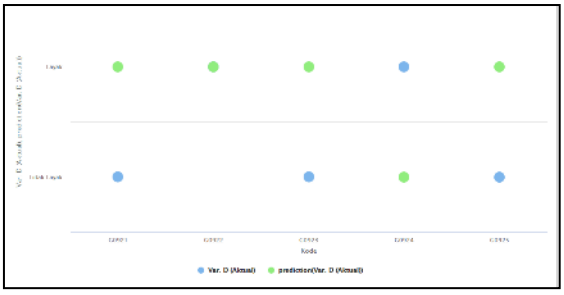

Figure 13. Scatter Plot Naïve Bayes

14. Performance Vector NB

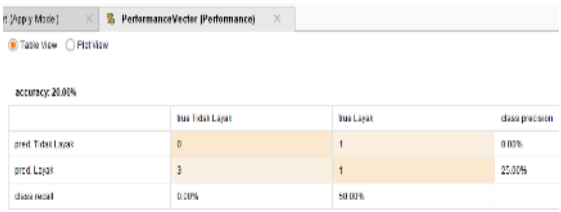

Figure 14. Performance Vector Naïve Bayes

CONCLUSION

Based on the research that has been carried out, the author concludes that to see the level of accuracy of the comparison of the two methods of the K-Nearest Neighbor and Naïve Bayes Algorithms in predicting the classification of the feasibility of receiving pawn goods, the Confusion Matrix formula is used which produces an accuracy rate of the K-Nearest Neighbor method of 40% while the Naïve Bayes method has an Accuracy Value of 20%, so we can determine that the accuracy of the accuracy value from the prediction with the K-Nearest Neighbor method is a better one when compared to the Naïve Bayes method.

BIBLIOGRAPHY

[1] Y. Apridonal M, W. Choiriah, and A. Akmah, "Penerapan Data Mining Menggunakan Metode Association

ISSN : 1979-9292
E-ISSN : 2460-5611

Rule Dengan Algoritma Apriori Untuk Analisa Pola Penjualan Barang," *JURTEKSI*, vol. V, no. 2, pp. 193–198, 2019.

[2] D. Remawati, R. D. Nugroho, and P. Harsadi, "Penerapan Sistem Penunjang Keputusan Menggunakan Algoritma Naive Bayes Pada konsep Human Resource Information System (HRIS) (Studi kasus :Penerusan Kontrak Kerja Karyawan di PT. XYZ)," *J. Ilm. SINUS*, vol. 18, no. 1, pp. 63–74, 2020.

[3] R. M. Sari and Y. Apridonal M, "Data Mining Implementation For Printer Sales Prediction Using Naive Bayes Method," *ICoSSIT.*, vol. 1, no. 1, pp. 215–220, 2020.

[4] R. Yendra, L. Marifni, and I. Suryani, "Klasifikasi Data Mining Untuk Seleksi Penerimaan Calon Pegawai Negeri Sipil Tahun 2017 Menggunakan Metode Naïve Bayes," *J. Sains Mat. dan Stat.*, vol. 6, no. 1, pp. 65–78, 2020.

[5] R. Jayadi, H. M. Firmantyo, M. T. J. Dzaka, M. F. Suaidy, and A. M. Putra, "Employee Performance Prediction Using Naïve Bayes," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 6, pp. 3031–3035, 2019.

[6] C. Fadlan, S. Ningsih, and A. P. Windarto, "Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *J. Tek. Inform. Musirawas*, vol. 3, no. 1, p. 1, 2018.

[7] A. D. Afifaturahman and M. Firmansyah, "Perbandingan Algoritma K-Nearest Neighbour (KNN) dan Naive Bayes pada Intrusion Detection System (IDS)," *Innov. Res. Informatics*, vol. 3, no. 1, pp. 17–25, 2021.

[8] R. N. Devita, H. W. Herwanto, and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa indonesia," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 427–434, 2018.

[9] A. S. Fitrani, F. Fajrillah, and W. Novarika, "Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection in Jemirahan Village , Jabon District," *Int. J. Informatics Comput. Sci.*, vol. 3, no. 2, pp. 66–79, 2019.

[10] A. Indriani, "Analisa Perbandingan Metode Naïve Bayes Classifier Dan K-Nearest Neighbor Terhadap Klasifikasi Data," *Sebatik*, vol. 24, no. 1, pp. 1–7, 2020.