Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 113-124
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Selecting Best Factors Using Information Gain To Improve Merchant Eligibility Classification

Muhamad Irvan Dandung[1], Bagus Priambodo[2]
[1,2] Informatics Engineering, Universitas Mercu Buana, Jakarta, Indonesia
Email: 41519110136@mercubuana.ac.id[1], bagus.priambodo@mercubuana.ac.id[2]

Abstract

Currently, the financial technology sector growing rapidly. Digital payments significantly drive the digital economy. Payment Service Providers (PSPs) play a crucial role in offering digital payment services and identifying eligible merchants for loan recommendations. The data need further analysis to get accurate load recommendations. This study focuses on supporting the development of merchant businesses, particularly Micro, Small, and Medium Enterprises (MSMEs), by enhancing loan eligibility analysis through the use of machine learning. Information Gain applied to find the best factors to classify merchant eligibility. The factors obtained were then classified using two machine learning algorithms that are Naïve Bayes and K-Nearest Neighbors (KNN). The experiment result shows that classifying merchant eligibility using naïve Bayes and KNN based on selected factors obtained from information gain has better accuracy than based on all factors. These findings demonstrate the importance of selecting relevant variables for more accurate analysis. The results of this study contribute to the understanding of machine learning applications in financial decision-making, offering real solutions for merchant eligibility analysis and supporting the growth of the digital economy

***Keywords**: Financial Technology, Loan Eligibility Analysis, Machine Learning, Payment Service Providers (PSPs), Information Gain.*

1. INTRODUCTION

In the rapid development of financial technology in Indonesia, digital payments have played a crucial role in driving the growth of the digital economy [1]. A survey conducted by Kadence International Indonesia revealed that 44% of users utilize digital payment methods up to four times a week. This finding aligns with a report from Bank Indonesia (BI), which recorded a 49.06% year-on-year growth in electronic money transaction value during 2024 [2]. Several financial technology companies, such as Payment Service Providers (PJP), play a role in offering digital payment services to the Indonesian public [3]. One of the services provided is a payment acceptance system for merchants or Micro, Small, and Medium Enterprises (MSMEs) [4].

To support the business growth of merchants or MSMEs, a Payment Service Provider (PJP) has partnered with a peer-to-peer lending company to offer loan services. These loans are available to merchants or MSMEs that have been recommended as eligible by the PJP, which currently relies on manual analysis by its business team to obtain merchant or MSME eligibility data. With the anticipated growth in the number of merchants or MSMEs over time, the PJP needs to leverage information technology to increase efficiency in analyzing merchants or MSMEs for loan eligibility recommendations. This is crucial to reduce the risk of human error in the analysis process, especially with the increasing volume of data and errors in reading the determining factor parameters for whether a merchant qualifies for a recommendation. One approach that can be used to achieve this is by utilizing data analysis through machine learning, employing classification methods such as Naïve Bayes and K-Nearest Neighbors (KNN).

The selection of machine learning algorithms in a particular case study should first consider the accuracy of the method. Second, we must understand the quality of the results obtained from the method. Lastly, the efficiency of the algorithm is also an important factor to consider [5]. According to research conducted by Zairi Saputra and colleagues in 2024 on the comparison of K-Nearest Neighbor, Naïve Bayes, and Support Vector Machine methods in classifying loan approvals, it was found that both the K-Nearest Neighbor and Naïve Bayes algorithms had the same accuracy, at 77%. Meanwhile, the accuracy of the Support Vector Machine algorithm was 50% [5]. Another study by Priyanto et al., also in 2024, on the application of the Naïve Bayes method for determining PIP (Smart Indonesia Program) aid recipients, showed that this method produced more accurate predictions compared to decisions made by schools. The method achieved an accuracy of 88.89% and a recall of 85.71% from a dataset of 100 students, with 9 students used as test data [6]. In addition to Naïve Bayes, which demonstrated high accuracy in both studies, research conducted by Mardiyyah and her team showed that the K-Nearest Neighbor algorithm also achieved a high accuracy rate, reaching 86.15%. Furthermore, this algorithm had a precision of 90.74% and a recall of 92.45% when applied to data from www.kaggle.com titled "Finance Loan Approval Prediction Data." [7]. Although in the research conducted by Martha and her team in 2024 comparing the K-Nearest Neighbor, Binary Logistic Regression, and Classification Tree methods in credit eligibility analysis, the Classification Tree method demonstrated higher accuracy than the K-Nearest Neighbor method, K-Nearest Neighbor still achieved over 80% accuracy, specifically 83.84%, for the training data. In this case, the K-Nearest Neighbor method ranked second among the three methods in terms of training data accuracy [8].

This study aims to investigate merchant eligibility data as a determining factor in loan recommendations using machine learning techniques. The research focuses on supporting the growth of merchant businesses, particularly Micro, Small, and Medium Enterprises (MSMEs), by enhancing the accuracy of loan eligibility analysis. Information Gain is applied to identify the best factors influencing merchant eligibility. These factors are classified using two machine learning algorithms, namely Naïve Bayes and K-Nearest Neighbors (KNN).

2. RESEARCH METHODOLOGY

Machine learning, at the intersection of computer science and statistics, is at the core of artificial intelligence and data science. Its growth has been driven by new algorithms, the availability of data, and affordable computing power. This data-driven approach has been widely adopted across various fields, such as healthcare and marketing, to support evidence-based decision-making [9]. In the context of this research, machine learning is applied to selecting best factors using information gain to improve merchant eligibility classification. In this study, the research flow is divided into several stages as shown in Figure 1, where the research begins with problem formulation to establish the boundaries and final objectives. This is followed by data collection, selecting factors, classifying eligible merchants and evaluating the result.

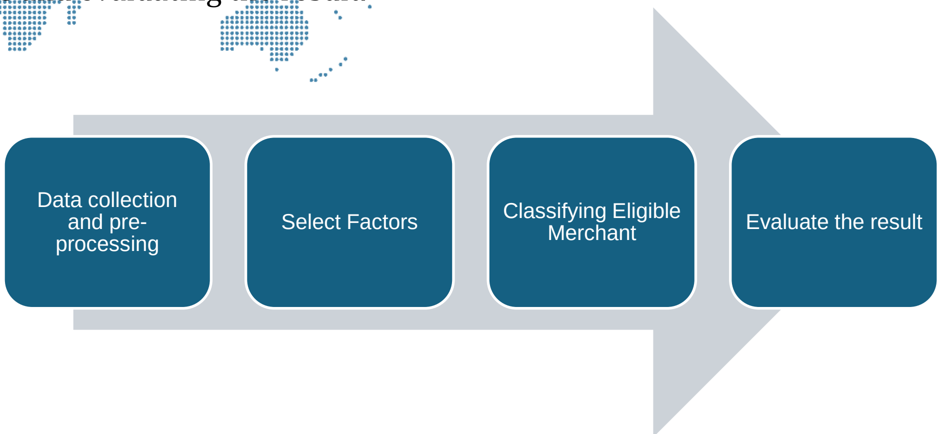

Figure 1. Research Flow

2.1. Data Collection and Pre-Processing

The data is taken from PT Mareco Prima Mandiri as the primary material for analysis. There are 180 rows with 10 factors in the dataset. The example of the data is presented in Table 1

Table 1. The example of merchant data

freqTrx	avgAmountTrx	totalAmountTrx	totalMdrAmountTrx	totalBusinessReceive	acqBusinessName	acqBusinessCity	acqBusinessProvince	acqBusinessType	acqBusinessCriteria	eligible
124	22653,22581	2809000	1164	2807836	Ayam Geprek Papiluk	Wil. Jakarta Timur	Daerah Khusus Ibukota Jakarta	FOOD AND DRINK	UMI	TRUE
15	60666,66667	910000	6370	903630	Aldien Kitchen	Depok	Jawa Barat	FOOD AND DRINK	UKE	FALSE
15	55080,6	826209	5783,46	820425,54	Apotek Indobat Ground Zero	Bandung	Bali	HEALTH AND BEAUTY	UKE	FALSE
302	18854,30464	5694000	315	5693685	Kedai Pelangi Ibu Santi	Wil. Jakarta Selatan	Daerah Khusus Ibukota Jakarta	FOOD AND DRINK	UMI	TRUE
1216	21218,33882	25801500	180610,5	25620889,5	Warung Makan Mahardika	Wil. Jakarta Selatan	Daerah Khusus Ibukota Jakarta	FOOD AND DRINK	UKE	TRUE

Further explanation about the variable in the dataset is explained in Table 2.

Table 2. Dataset variables

Variable	Description
freqTrx	Merchant transaction frequency
avgAmountTrx	Average merchant transaction amount
totalAmountTrx	Total merchant transaction amount
totalMdrAmountTrx	Total revenue amount of merchant and service provider transactions

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 113-124
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Variable	Description
totalBusinessReceive	The total profit that the merchant will receive from the transaction
acqBusinessCity	Merchant locations by city
acqBusinessProvince	Merchant locations by province
\acqBusinessType	Merchant categorization based on type
acqBusinessCriteria	Merchant categorization based on category
eligible	Whether or not the merchant is eligible to receive a loan

For data preprocessing, this study utilizes Google Colab and Python as the programming language. Google Colab is a hosted Jupyter Notebook service that can be used without setup and provides free access to computing resources, including GPUs [10]. Considering that the language used is Python, it is highly suitable for coding in this research because Python is a flexible and interpreted programming language with a design philosophy that emphasizes code readability. Python is known for combining power and capability with very clear syntax, as well as having a large and comprehensive standard library [11]. Table 3 shows that eight independent variables in the data frame have been selected, and the variables acqBussinessName and acqBusinessProvince have been removed. The acqBussinessName variable was removed because the data in each row is different, and in the case of acqBussinessProvince, it has been omitted since its information is already represented by the acqBusinessCity data, which correlates with the location data in the dataset. Apart from that, table 3 also shows the variables acqBusinessCity, acqBusinessType, and acqBusinessCriteria, which initially had text data types in each column value, adjusted to numeric data types that have been mapped for processing in the classification machine learning algorithm. According to Table 3, which represents the 5 displayed data rows, the value 395 in the first row of the acqBusinessCity variable is the result of mapping from the value Wil. Jakarta Timur is shown in table 3. Therefore, in the acqBusinessCity column, every row with the value Wil. Jakarta Timur will be mapped to the value 395. For the acqBusinessType variable, the value 1 in the first row results from mapping the value FOOD AND DRINK as shown in table 3. Consequently, in the acqBusinessType column, every row with the value FOOD AND DRINK will be mapped to 1. For the acqBusinessCriteria variable, the value 1 in the first row is the result of mapping the value UMI indicated in Table 3, so in the acqBusinessCriteria column, every row with the value UMI will be mapped to 1.

Table 3. The 5 rows of merchant eligibility data that have been mapped

freqTrx	avgAmountTrx	totalAmountTrx	totalMdrAmountTrx	totalBusinessReceive	acqBusinessCity	acqBusinessType	acqBusinessCriteria	eligible
124	22653,22581	2809000	1164	2807836	395	1	1	TRUE
15	60666,66667	910000	6370	903630	197	1	3	FALSE
15	55080,6	826209	5783,46	820425,54	7204	3	3	FALSE
302	18854,30464	5694000	315	5693685	394	1	1	TRUE

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 113-124
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

freqTrx	avgAmountTrx	totalAmountTrx	totalMdrAmountTrx	totalBusinessReceive	acqBusinessCity	acqBusinessType	acqBusinessCriteria	eligible
1216	21218,33882	25801500	180610,5	25620889,5	394	1	3	TRUE

2.2. Select Factors

For factor selection in this research, information gain will be used. Information Gain plays a very important role in feature selection to enhance the results of classification calculations, as its advantage lies in its simplicity compared to other feature selection methods. This makes it widely used in application development [12]. In this phase, we filter the factors that had a high influence on the eligibility of merchants that get loan approval from PJP. Information gain is used to select best factors that influence eligible of the merchant. The formula of information gain is explained in (1).

$$Gain(S,A) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Explanation:

S: The set of cases
A: The attribute
n: The number of partitions of attribute A
|Si|: The number of cases in the i-th partition (Si)
|S|: The total number of cases in S

2.3. Classifiying Eligible Merchant

In machine learning, classification aims to discover a model or function that differentiates data into classes to reveal meaningful patterns within large datasets. Typically, one attribute serves as the target outcome, while the other attributes act as predictors, with the target value indicating the class predicted based on the predictor values [13]. When classifying, data is typically divided into two categories: training data and testing data. Specifically, the testing data serves as the dataset that researchers examine to obtain classification results [14].

The first classification algorithm used is Naive Bayes. Naive Bayes is a probabilistic algorithm that operates directly based on Bayes' theorem and assumes independence among features. This algorithm is commonly used for classification tasks and has proven effective, especially when handling large datasets. Therefore, it is an ideal choice for predicting loan approvals [15]. The Naive Bayes Classifier is a generative model that creates a model for each possible class based on training examples for each class [16]. The formula of Naïve Bayes is explained in (2).

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

Explanation:

X: Data with an unknown class
H: Hypothesis that the data belongs to a specific class (X)
P(H): Probability of hypothesis H (prior probability)
P(X): Probability of the observed sample data (X)

P(H|X): Probability of hypothesis H given the condition X (posterior probability)

P(X|H): Probability of X given the condition of hypothesis H

The second algorithm is K-NN (K-Nearest Neighbors). K-NN is an algorithm that can identify patterns and trends from historical data, including customer history [17]. The KNN method is divided into two stages: the training stage and the classification or testing stage. The KNN algorithm is easy to implement as it operates by searching for the nearest training samples from the instance being tested to determine its nearest neighbours. The distance between neighbours, whether close or far, is measured using Euclidean Distance [18]. The formula of KNN is explained in (3).

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (3)$$

Explanation:

d : Euclidean distance between two points

$x_i y_i$: Feature values of the i^{th} Dimension for points x and y

n: Number of features or dimensions

$\sqrt{\ldots}$: Square root of the sum of squared differences for each feature

2.4. Evaluate The Result

Confusion matrix is used to evaluate the result of classification result using both algorithm. A Confusion Matrix is a calculation used to describe the accuracy level of classification. The basis of the confusion matrix calculation involves comparing the number of data points correctly classified into the appropriate classes with the total number of data points in the matrix [19]. This study use metrics accuracy to see the performance of both algorithm using two conditions. With all factors and with selected factors. The metrics accuracy is explained in result.

3. RESULT AND DISCUSSION

In this research, the main objective is is to develop a machine learning model that can assist in predicting merchant eligibility for loan recommendations. These loans can help merchants grow their businesses. To achieve this goal, classification algorithms such as Naïve Bayes Classifier and K-Nearest Neighbor (K-NN) will be used. These algorithms will learn patterns from training data that involve various relevant variables. In this context, the variables considered include merchant transaction frequency, average transaction amount, total transaction amount, total transaction revenue, total profit the merchant will receive from the transactions, merchant location, type of business, and merchant criteria. The classification will be conducted in two stages: using all factors and using selected factors filtered through Information Gain. This approach will compare the effectiveness of incorporating all available variables versus prioritizing those most influential in predicting merchant eligibility.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 113-124
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

3.1. Classification Using All Factors

The result of classification based on all factors using KNN is present in figure 2.

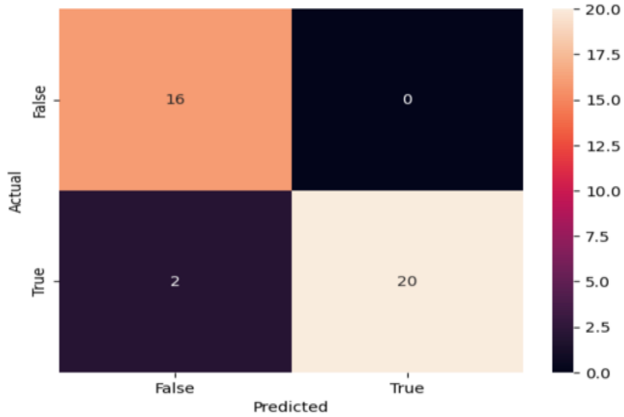

Figure 2. Confusion matrix graph results from K-Nearest Neighbors (K-NN)

Figure 2 shows the results of the confusion matrix from the data that has been processed in Figure 5 and passed through the K-Nearest Neighbors (K-NN) algorithm. It includes 20 True Positives, 16 True Negatives, 2 False Negatives, and 0 False Positives. Let's discuss each evaluation metric presented in the graph above:

a. True Positive (TP) = 20
The model successfully identified 20 positive cases correctly. This means that in 20 cases where the actual label is positive, the model also predicted positive.

b. True Negative (TN) = 16
The model successfully predicted negative for 16 cases that were actually negative. This means that in 16 cases where the actual label is negative, the model also predicted negative.

c. False Negative (FN) = 2
The model missed 2 positive cases that should have been detected. This means that in 2 cases where the actual label is positive, the model predicted negative.

d. False Positive (FP) = 0
There are no false positive cases, meaning the model did not incorrectly detect negative cases as positive.

The accuracy score of K-Nearest Neighbors (K-NN) is 0.9473684210526315, indicating a fairly high accuracy in line with the confusion matrix results. Overall, the K-NN model demonstrates excellent performance, particularly because FP = 0, which means there is no risk of misclassifying negative cases as positive at all. The result of classification based on all factors using Naïve Bayes is present in figure 3.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 10, Nomor 1, Februari 2025, pp 113-124
Terakreditasi Nomor 204/E/KPT/2022 | ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

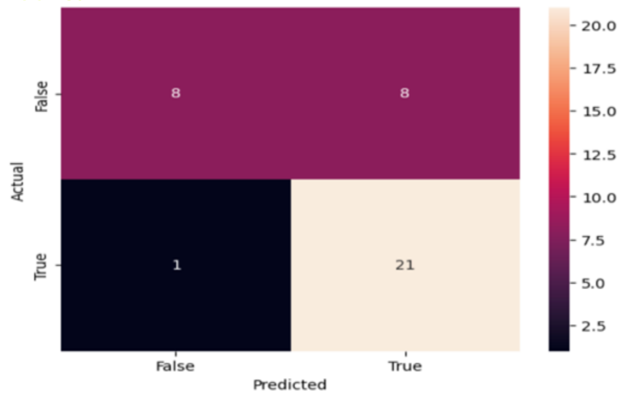

Figure 3. The confusion matrix graph resulting from the Naïve Bayes algorithm

a. True Positive (TP) = 20
   The model successfully detected 21 positive cases correctly. In 21 cases where the actual label was positive, the model also predicted positive.
b. True Negative (TN) = 8
   The model correctly predicted negative for 8 cases. Thus, in 8 cases where the actual label was negative, the model also predicted negative.
c. False Negative (FN) = 1
   The model had only 1 positive case that was not detected (identified as negative).
d. False Positive (FP) = 8
   The model made mistakes by predicting positive for 8 cases that were actually negative.

This indicates that the model often incorrectly identifies negatives as positives, which can lower the accuracy score. Based on the confusion matrix in Figure 3 the Naive Bayes accuracy only achieved 0.7631578947368421.

3.2. Classification with Selected Factors Using Information Gain

The result of filtered factors using information gain is presented in figure 4. Figure 4 shows that acqBusinessType ranks lowest in the results of the second, third, and fourth executions and second lowest in the first execution. This indicates that the acqBusinessType variable has the least influence compared to the other independent variables. In contrast, the totalAmountTrx variable ranks highest in the results of the second, third, and fourth executions and second highest in the first execution, confirming that the totalAmountTrx variable is the most influential independent variable compared to the others. Based on this result acqBusinessType factors need to be removed from the independent variable for classifying eligible merchants.

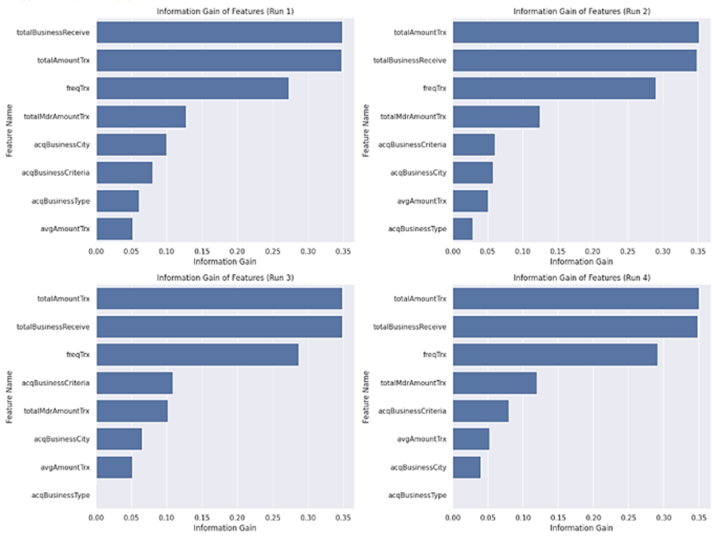

Figure 4. Graph of information gain of independent variables in 4 trials

The confussion matrix of classification based on selected factors using Naïve Bayes is show in figure 5. Figure 5 shows that after the acqBusinessType variable was removed from the dataset, the Naive Bayes model achieved an accuracy score of 0.868421052631579, with the confusion matrix results yielding 19 True Positives, 14 True Negatives, 3 False Negatives, and 2 False Positives.

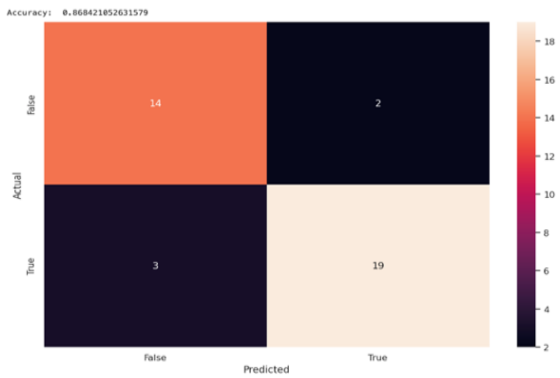

Figure 5. Accuracy score and confusion matrix graph Naïve Bayes after removing the acqBusinessType variable

3.2. Discussion

Table 4 shows that the use of information gain has a positive impact on improving the accuracy of two machine learning models, namely Naive Bayes and K-NN. In the Naive Bayes model, the initial accuracy of 0.7631 increased to 0.8684 after the application of information gain, resulting in an accuracy improvement of 0.1053. This indicates that information gain significantly enhanced the performance of the Naive Bayes model. Meanwhile, the K-NN model showed an initial accuracy of 0.9474, which increased to 0.9737 after the use of information gain, with a difference of 0.0263. Although the improvement in K-NN is not as pronounced as that in Naive Bayes, information gain still contributes positively to the performance of this model. Overall, this table supports the hypothesis that information gain can improve the performance of machine learning models, particularly in classification, with a more notable increase in the Naive Bayes model compared to the K-NN model.

Table 4. Results of the accuracy using both algorithm

Model	Accuracy (all factors)	Accuracy (selected factors)	Score increase
Naive Bayes	0.7631	0.8684	0.1053
K-NN	0.9474	0.9737	0.0263

4. CONCLUSION

Based on the research results conducted on the classification model for predicting the eligibility of merchants to receive loans, it can be concluded that the application of machine learning methods, namely Naïve Bayes and K-Nearest Neighbor (K-NN), using feature selection through information gain, successfully improved prediction accuracy. This model was trained using independent variables such as transaction frequency, average transaction amount, total transaction amount, total revenue, location, business type, and merchant criteria. Information gain proved effective in enhancing model performance by identifying the most relevant variables, with the totalAmountTrx variable having the greatest influence, while the acqBusinessType variable was removed due to its minimal impact. The accuracy results showed that Naïve Bayes improved from 0.7631 to 0.8684, while K-NN increased from 0.9474 to 0.9737 after the application of information gain. The K-NN model demonstrated higher performance, but Naïve Bayes also experienced a significant increase with contributions from feature selection. Evaluation using metrics such as confusion matrices indicated that the K-NN model had no False Positive cases, signifying a high accuracy in predicting merchants that are ineligible for loans. Overall, this study shows that the combination of classification algorithms and feature selection through information gain can assist financial institutions in making more accurate decisions regarding loan recommendations. The developed machine learning model can efficiently predict the eligibility of merchants, thereby supporting the growth of merchant businesses and decision-making in loan disbursement based on new merchant profile data.

REFERENCES

[1] Admin Winpay, "Digitalisasi Pembayaran Sebagai Pendorong Ekonomi Digital Indonesia!," Winpay Id. Accessed: May 11, 2024. [Online]. Available: Https://Www.Winpay.Id/Digitalisasi-Pembayaran-Sebagai-Pendorong-Ekonomi-Digital-Indonesia/

[2] Admin Bri Api, "Menilik Tren Pembayaran Digital Di Indonesia Dan Peran Briapi Di Dalamnya," Bri Api. Accessed: May 11, 2024. [Online]. Available: Https://Developers.Bri.Co.Id/Id/News/Menilik-Tren-Pembayaran-Digital-Di-Indonesia-Dan-Peran-Briapi-Di-Dalamnya

[3] Admin Bi, “Penyelenggara Jasa Sistem Pembayaran,” Bank Indonesia. Accessed: May 12, 2024. [Online]. Available: Https://Www.Bi.Go.Id/Id/Fungsi-Utama/Sistem-Pembayaran/Perizinan/Default.Aspx
[4] Admin Faspay, “Pentingnya Sistem Pembayaran E-Payment Untuk Bisnis,” Faspay. Accessed: May 12, 2024. [Online]. Available: Https://Faspay.Co.Id/Id/Pentingnya-Menggunakan-Sistem-Pembayaran-E-Payment-Di-Indonesia/
[5] Zairi Saputra, H. A. Supahri, R. Ismanizan, And R. Rahmaddeni, “Perbandingan Algoritma Knn (K-Nearest Neighbors), Naïve Bayes, Dan Svm (Support Vector Machine) Untuk Klasifikasi Pemberian Pinjaman Nasabah,” *Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika (Jisti)*, Vol. 7, No. 1, Pp. 67–75, Apr. 2024, Doi: 10.57093/Jisti.V7i1.182.
[6] I. Priyanto *Et Al.*, “Penerapan Algoritma Metode Naïve Bayes Untuk Penentuan Penerimaan Bantuan Program Indonesia Pintar (Pip) (Studi Kasus Smp Pgri 1 Cilacap),” *Jurnal Manajemen Informatika Jayakarta*, Vol. 4, No. 2, Pp. 162–172, 2024, Doi: 10.52362/Jmijayakarta.V4i2.1355.
[7] N. W. Mardiyyah, N. Rahaningsih, And I. Ali, “Penerapan Data Mining Menggunakan Algoritma K-Nearest Neighbor Pada Prediksi Pemberian Kredit Di Sektor Finansial,” 2024. [Online]. Available: Www.Kaggle.Com
[8] S. Martha, W. Andani, And S. W. Rizki, “Perbandingan Metode K-Nearest Neighbor, Regresi Logistik Biner, Dan Pohon Klasifikasi Pada Analisis Kelayakan Pemberian Kredit,” *Euler : Jurnal Ilmiah Matematika, Sains Dan Teknologi*, Vol. 10, No. 2, Pp. 262–273, Nov. 2022, Doi: 10.34312/Euler.V10i2.16751.
[9] B. Mega Pratiwi And N. Q. Nada, “Penerapan Model Machine Learning Dalam Menentukan Rekomendasi Objek Wisata Provinsi Jawa Tengah,” 2022.
[10] G. I. E. Soen, M. Marlina, And R. Renny, “Implementasi Cloud Computing Dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants,” *Jitu : Journal Informatic Technology And Communication*, Vol. 6, No. 1, Pp. 24–30, Jun. 2022, Doi: 10.36596/Jitu.V6i1.781.
[11] R. N. Melinda, L. Meitya Ningrum, I. B. Suryabrata, G. Swarna Bayu, A. Dwipa, And T. P. Sukoco, “Program Perhitungan Rab Pekerjaan Struktur Baja (Wf Beam) Menggunakan Bahasa Python,” *Tiers Information Technology Journal*, Vol. 2, No. 1, Pp. 31–38, 2021, [Online]. Available: Https://Journal.Undiknas.Ac.Id/Index.Php/Tiers
[12] E. Rahmanita, Y. D. P. Negara, Y. Kustiyahningsih, V. Sasmeka, And B. K. Khotimah, “Implementasi Metode Naïve Bayes Dan Information Gain Untuk Klasifikasi Penyakit Dan Hama Tanaman Jagung,” *Teknika*, Vol. 12, No. 3, Pp. 198–204, Oct. 2023, Doi: 10.34148/Teknika.V12i3.684.
[13] E. Rahayu, N. Irawati, And R. Ananda, “Sistemasi: Jurnal Sistem Informasi Klasifikasi Kelayakan Warga Penerima Bpnt Dengan Algoritma K-Nearest Neighbor Classification Of Eligibility For Bpnt Recipients Using The K-Nearest Neighbor Algorithm.” [Online]. Available: Http://Sistemasi.Ftik.Unisi.Ac.Id
[14] R. Widyasari, M. Adi Putri Siregar, S. Dan Teknologi, And K. Kunci-Klasifikasi, “Hybrid Naive Bayes Classifier Dan Markov Chain Untuk Prediksi Tingkat Penggunaan Quick Response Indonesian Standard (Qris),” *Jifotech (Journal Of Information Technology*, Vol. 4, No. 1, 2024.
[15] J. Khatib Sulaiman, A. Syakur, R. Purwandi Putra, C. Juliane, S. Likmi Bandung, And S. Tinggi Ilmu Kesehatan Sukabumi, “Optimalisasi Metode Naive Bayes Classifier Untuk Prediksi Persetujuan Kredit,” *Indonesian Journal Of Computer Science Attribution*, Vol. 13, No. 1, Pp. 2024–1091.

[16] M. Fatchan, I. K. Nanang Tedi, And E. Widodo, "Perbandingan Dalam Memprediksi Penyakit Liver Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor Abstrak Informasi Artikel," *Jurnal Pelita Teknologi*, Vol. 16, No. 1, Pp. 15–21.

[17] D. K. Purba, I. M. Sarkis, And A. Situmorang, "Penentuan Kelayakan Pemberian Kredit Pada Nasabah Cu Merdeka Berastagi Dengan Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," 2024. [Online]. Available: Https://Ejurnal.Methodist.Ac.Id/Index.Php/Methosisfo

[18] B. Aji Santoso And A. Dwi Hartanto, "Comparison Of Accuracy Levels Of Random Forest And K-Nearest Neighbor (Knn) Algorithms For Classifying Smooth Bank Credit Payments," *Jurnal Teknik Informatika (Jutif)*, Vol. 5, No. 1, Pp. 77–87, 2024, Doi: 10.52436/1.Jutif.2024.5.1.1195.

[19] D. Sartika And I. Saluza, "Penerapan Metode Principal Component Analysis (Pca) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang Km 12 Palembang Menggunakan Metode Decision Tree."