JIKO (Jurnal Informatika dan Komputer) Accredited KEMENDIKBUD RISTEK, No.105/E/KPT/2022
Vol. 7, No. 3, December 2024, pp. 217-221 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v7i3.9064 e-ISSN: 2656-1948

IMPLEMENTATION OF MSME CREDIT LOAN DETERMINATION USING MACHINE LEARNING TECHNOLOGY WITH K-NN ALGORITHM

Muchamad Taufik Nawawi[1*], Agus Suhendar[2]

[1,2]Universitas Teknologi Yogyakarta
Email: *[1]nawawisk2@gmail.com, [2]agus.suhendar@staff.uty.ac.id

(Received: 6 November 2024, Revised: 12 December 2024, Accepted: 18 December 2024)

Abstract

This research aims to develop a loan eligibility prediction model for Micro, Small, and Medium Enterprises (MSMEs) using the K-Nearest Neighbors (KNN) algorithm. The dataset utilized includes variables such as the length of business operation, number of workers, assets, and net turnover of MSMEs. The data is split into training and test sets with a 70:30 ratio. The KNN model is trained using the training data to classify loan eligibility based on a specified k value. The model predictions include whether a loan is accepted and the probability associated with each decision. The results indicate that the KNN model achieved an accuracy rate of 83.939% in predicting loan eligibility. Based on the predictions, 929 MSMEs were deemed eligible to receive loans according to the KNN model recommendations, while 170 MSMEs were classified as ineligible. These findings contribute significantly to the development of decision support systems in the banking and finance sectors, particularly in evaluating MSME loan eligibility.

Keywords: *K-Nearest Neighbors, MSME, loan eligibility, prediction model, KNN algorithm*

**Corresponding Author: Muchamad Taufik Nawawi*

1. INTRODUCTION

Micro, Small and Medium Enterprises (MSMEs) play a very important role in economic development and growth, not only in developing countries such as Indonesia, but also in developed countries [1]. In Indonesia, in addition to their role in economic development and growth, MSMEs also play a very important role in overcoming unemployment [2]. Therefore, MSMEs need to be developed because they can drive the growth rate of the Indonesian economy[3].

Current business developments have led to a renewal with the presence of Micro, Small and Medium Enterprises (MSMEs). Based on data from the Central Statistics Agency (BPS), the percentage of MSMEs in Indonesia is 98.68% of the total business sector in Indonesia [4]. There are several things that underlie the formation of MSMEs in Indonesia, namely the needs that must be met by the Indonesian people due to the large demand from the community. MSMEs that develop in the community itself are able to increase economic growth 25% of people believe that this technology will result in innovation [4][5][6].

The growth of Micro, Small and Medium Enterprises (MSMEs) in Indonesia has become a major focus in efforts to improve the economy in an inclusive manner [7]. MSMEs play an important role in creating jobs and improving income distribution in various regions [7]. However, the main challenge faced by MSMEs is access to adequate financial resources. The implementation of *Machine Learning* technology *in* this sector has great potential to improve efficiency and effectiveness in fund management [8].

Complicated and inefficient credit application processes are often a major obstacle for MSMEs to get the funds they need. The lack of complete and accurate historical data often results in subjective and sub-optimal risk assessments. This can hinder the growth of MSME businesses and reduce their opportunity to expand [9].

Limitations in risk assessment often discourage savings and loan cooperatives from extending credit to MSMEs, especially those without a strong credit history [10]. Inappropriate risk assessment can also result in misallocated funds, reducing financial sector efficiency and overall economic growth. This is a

crucial issue that needs to be addressed to support the sustainable growth of MSMEs [11].

The application of data mining and the *K-Nearest Neighbors* (KNN) algorithm is in line with Azizah's research (2022) showing that the KKN model succeeded in classifying aid recipients with a high accuracy of 98.46%. The results show that 339 MSME participants are eligible to receive assistance according to the KNN model recommendations, while 42 participants are not eligible [12].

Machine learning models, such as Gradient K Nearest Neighbours (KNN) have proven effective in predicting and optimizing [13] the power of machine learning algorithms becomes apparent [14]. Trained on comprehensive data sets that include damage assessments for various structural models, these algorithms can learn to accurately predict structural damage. This not only simplifies the task of performing accurate structural analysis but also provides invaluable insights that can significantly [14]

Previous research by Elisa (2022) also showed the successful use of the KNN model where out of 26 training data tested, 20 data have been classified correctly, while 6 data have not been classified accurately, resulting in a percentage of Correctly Classified Instances of 77.00%. The results of this analysis can be used effectively to determine future sales targets, which will have an impact on the level of sales of MSMEs in the future [15].

To overcome this problem, the implementation of *Machine Learning* technology with the *K-Nearest Neighbors* (KNN) algorithm can be an effective solution [16]. By utilizing historical data and relevant variables, the system can generate predictive models that are able to provide a more accurate and objective risk assessment [17]. This not only makes it easier for financial institutions to make lending decisions, but also provides greater opportunities for MSMEs to gain access to the funding they need to grow their businesses.

2. RESEARCH METHOD

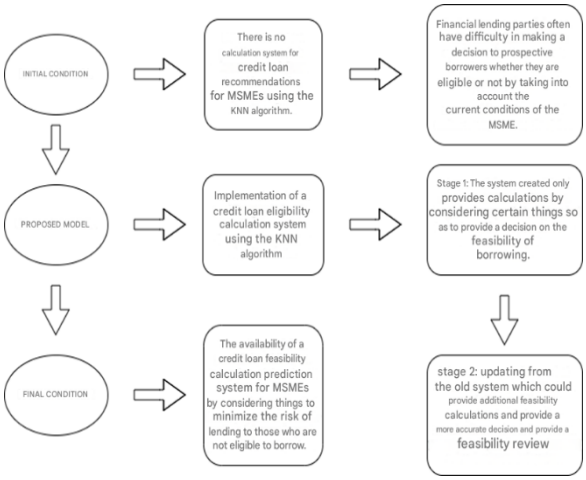


Figure 1. Research Freamwork

This research uses Primary data where the dataset is taken from the Savings and Loan Cooperation which focuses on MSME data, covering various statistical aspects that are relevant for credit risk assessment analysis. The data used includes length of business establishment in years, business assets, number of employees, income in months, etc. Using such data, this research aims to provide a more accurate risk assessment and support better lending decisions. Data collection for this study was conducted from the beginning of 2024 until May 2024. The data collected included relevant MSME statistics such as length of business establishment, assets, number of employees, and net turnover, which were obtained from Koprasi and MSME managers over the last 3-year period.

2.1 Model architecture

The admin inputs information related to the applicant/customer's business, such as length of business, number of employees, net turnover, and assets which are then sent via HTTP request to the web server. There, the data is processed using the K-Nearest Neighbors (KNN) module to predict loan eligibility. The results of this prediction are then conveyed to the user through a web interface, providing the admin with the information needed to make an informed decision on the loan decision to be made. The architecture of the system built is shown in Figure 2.

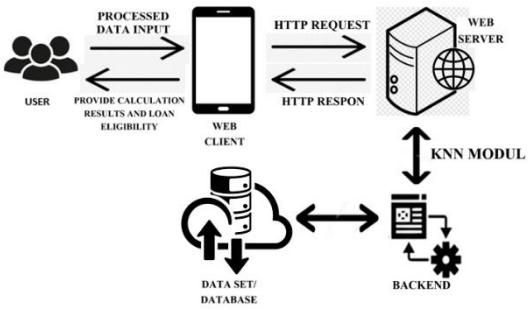


Figure 2. Model Architecture

2.2 K-Nearest Neighbors (KKN) Model

The calculation of the neighboring distance uses the euclidien algorithm as shown in the equation 1 [16].

euc=√((a_1−b_1)^2+⋯+(a_n−b_n)^2)

Where a = a1, a2, ..., an, and b = b1, b2, ..., bn represent n attribute values from two records. For attributes with category values. The formulation is as given

$$euc = \sqrt{(\sum_{i=1}^{n}(pi - qi)^2)} \quad (1)$$

Description

pi = sample data / training data

qi = test data / testing data

i = data variable
n = data dimension

The accuracy of the KNN model is calculated by comparing the prediction results with the actual data on the test dataset. For obtain the accuracy value used Equation 2.

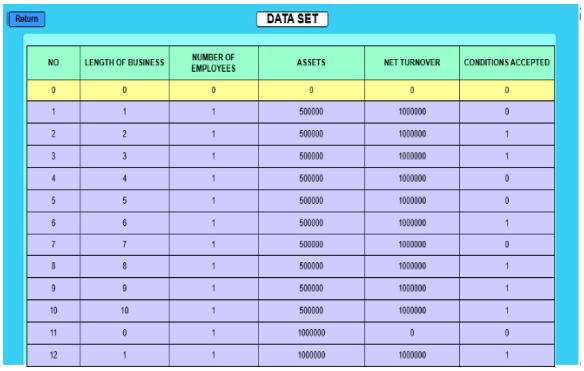
$$\text{Accuracy} = \frac{\text{Number of Correct Accuracy}}{\text{Total Test Data}} \text{ x } 100\% \quad (2)$$

K-Nearest Neighbor (K-NN) belongs to the instance-based learning group. This algorithm is also one of the *lazy* learning techniques. kNN is performed by finding the k groups of objects in the training data that are closest (similar) to objects in the new data or testing data.

3. RESULT AND DISCUSSION

3.1 Dataset

The dataset used is a data set taken from corporation data, the variables taken are length of business, number of workers, assets, net turnover, and also the terms accepted, the data has a range of 1099 contents per variable. Where this dataset is important koprasi data taken from customer data that has performed transaction actions, the data set has vulnerable data values that vary from 0-10 years of business, the number of workers from 0-9 people, the number of assets from vulnerable RP. 0-15000000,00, the amount of net turnover is vulnerable from RP. 0-5500000,00 and the conditions accepted are in the form of yes with variable 1 and not accepted with variable 0.



NO	LENGTH OF BUSINESS	NUMBER OF EMPLOYEES	ASSETS	NET TURNOVER	CONDITIONS ACCEPTED
0	0	0	0	0	0
1	1	1	500000	1000000	0
2	2	1	500000	1000000	1
3	3	1	500000	1000000	1
4	4	1	500000	1000000	0
5	5	1	500000	1000000	0
6	6	1	500000	1000000	1
7	7	1	500000	1000000	0
8	8	1	500000	1000000	1
9	9	1	500000	1000000	1
10	10	1	500000	1000000	1
11	0	1	1000000	0	0
12	1	1	1000000	1000000	1

Figure 3. Dataset Application

3.2 System View

On the input page there are several functions, namely the data input function according to the variables. Where each input variable is input according to the data of each member who applies for a loan, the variables include length of business, number of workers, assets, net turnover, and value as a benchmark for the calculation, and the prediction button to process the data that has been inputted earlier using the knn algorithm and dataset as a benchmark for vulnerable variables. For the input data page is shown in Figure 4.

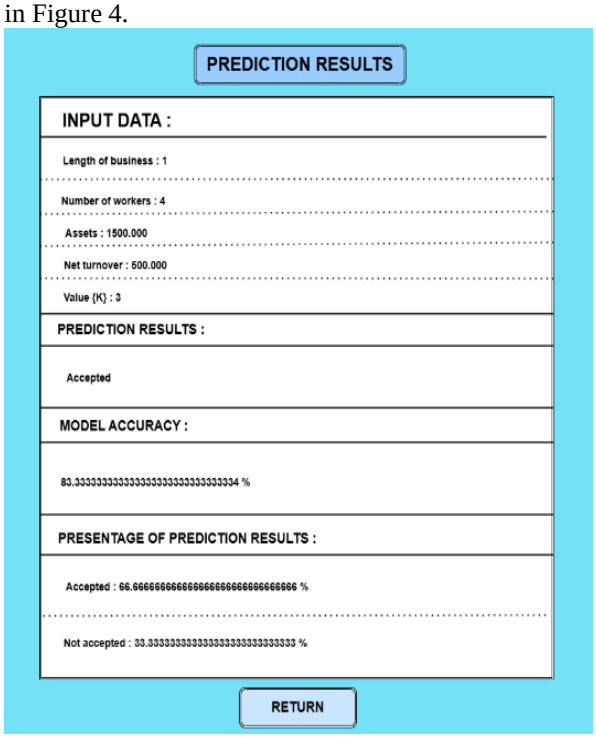


Figure 4. Data Input Page

The results page displayed in Figure 5 is the result of calculations using the K-NN algorithm and displays several results, including: input data according to previously inputted data variables, prediction results, accuracy model with k value and percentage of prediction results that are close to the data.

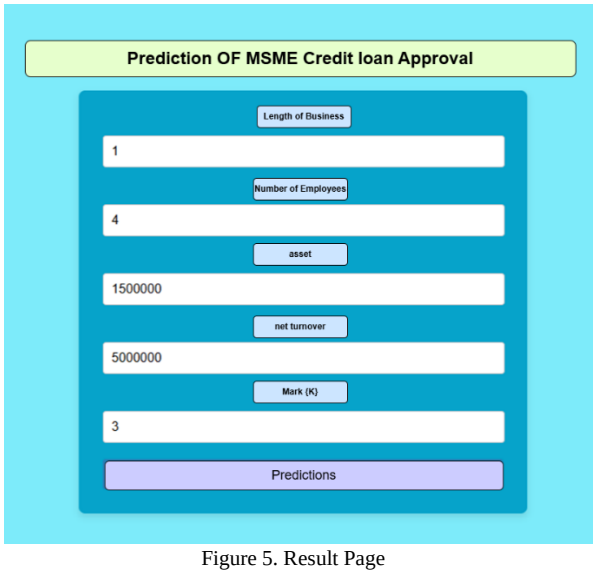


Figure 5. Result Page

3.3 Result KNN Model

Through this implementation, we can see that this application still has some limitations that need to be improved. However, the results achieved show the potential of this application in predicting SME loan approval using relevant data and the KNN algorithm.

Where the total data tested is 1099 data, each variable tested displays a different model accuracy value depending on the K value, where the lower the K value, the higher the percentage value of model accuracy, for example with data on business length 2, number of workers 6, assets 5000000, net turnover 2500000, with a K value of 100, the accuracy of the resulting model is 57. 898% while with the same variables but a smaller K value of 10 then the accuracy of the model will increase to 70.303%, the highest model accuracy is of course with a K value of 1 which is 83.939% but from this data it does not yet have high accuracy due to limited test data and test data so that each different variable data will produce the same percentage of model accuracy with each of the same k values. The results showed that the KNN model achieved an accuracy level of 83.939% in predicting loan eligibility. From the prediction results, 929 MSME data were declared eligible to receive loans according to the KNN model recommendations, while 170 MSME data did not qualify.

4. CONCLUSION

This research aims to develop a loan eligibility prediction model for Micro, Small, and Medium Enterprises (MSMEs) using the K-Nearest Neighbors (KNN) algorithm. The dataset used in this study consists of real data from cooperative customers (KSP) and includes variables such as business duration, number of workers, assets, and net turnover of MSMEs. The KNN algorithm was chosen due to its ease of operation, and the dataset is relatively straightforward to process. The application developed from this research simplifies the process for users to determine the percentage of eligible loans. The resulting model demonstrated high accuracy, achieving 83.939% in predicting loan eligibility. The model classifies MSME data into two categories: eligible or ineligible, while also providing a probability for each decision. Based on the prediction results, 929 MSMEs were classified as eligible to receive loans, while 170 MSMEs were deemed ineligible. These findings represent a significant contribution to the development of decision support systems in the banking and finance sector, particularly in evaluating MSME loan eligibility.

5. REFERENCE

[1] Munizu, M., Tandiono, R., Pravitasari, E., et al. *UMKM (Peran Pemerintah Dalam Daya Saing UMKM Di Indonesia*). PT. Sonpedia Publising Indonesia. 2023.

[2] Azizah, H. W., Nurdiawan, O., Dwilestari, G., Kaslani, K., & Tohidi, E. Klasifikasi Pemberian Bantuan UMKM Cirebon Dengan Menggunakan Algoritma K-Nearest Neighbor. *Journal Of Computer System And Informatics (Josyc)*, *3*(3), 110-115. 2022 Doi: 10.47065/josyc.v3i3.1392

[3] Sedyastuti, K. Analisis Pemberdayaan UMKM Dan Peningkatan Daya Saing Dalam Kancah Pasar Global. *INOBIS: Jurnal Inovasi Bisnis Dan Manajemen Indonesia*, *2*(1), 117 - 127. 2018. Doi: 10.31842/jurnal-inobis.v2i1.65

[4] Mardanugraha, E., & Akhmad, J. Ketahanan UMKM di Indonesia menghadapi Resesi Ekonomi. *Jurnal Ekonomi dan Pembangunan*, *30*(2), 101-114. 2023. Doi: 10.14203/JEP.30.2.2022.101-114

[5] Evi Sirait, Bagus Hari Sugiharto, Jenal Abidin, Nely Salu Padang, & Johni Eka Putra. Peran UMKM dalam Meningkatkan Kesejahteraan Perekonomian di Indonesia. *El-Mal: Jurnal Kajian Ekonomi & Bisnis Islam*, *5*(7), 3816 –. 3822. 2024. Doi: 10.47467/elmal.v5i7.4160

[6] Eryc, E. Analysis the Metaverse Adoption of Batam City Micro, Small, Medium Enterprises Using Technology Acceptance Model. *JIKO (Jurnal Informatika Dan Komputer)*, *6*(3), 165–173. 2023. Doi: 10.33387/jiko.v6i3.6719

[7] Dian Sudiantini, Wahyu Untoro, & Teguh. MANAJEMEN OPERASIONAL BAGI UMKM: PENINGKATAN EFISIENSI DAN PRODUKTIVITAS. *Musytari : Neraca Manajemen, Akuntansi, Dan Ekonomi*, *1*(1), 30–40. 2023. Doi: 10.8734/mnmae.v1i1.185

[8] Dauni, P., Pratiwi, P., & Prasetio, R. T. Pemetaan Keberlangsungan Hidup Umkm Guna Optimalisasi Bantuan Kredit Menggunakan Algoritma Fuzzy C-Means. *Jurnal Responsif: Riset Sains Dan Informatika*, *5*(1), 61-69. 2023. Doi: 10.51977/jti.v5i1.1051

[9] Bangsa, R. A. P. S. A. Metode Knn Pada Sentiment Analisis Review Produk Game Android. *Indonesian Journal Of Networking And Security (Ijns)*, *11*(2). Pp. 123-128. 2022. Doi: 10.55181/ijns.v11i2.1769

[10] Hambali, M. Penerapan Manajemen Koperasi Berbasis Syariah (Studi Kasus Di Koprasi Raudlatul Ulum 1 Malang). *Al-Astar*, *1*(2), 39-54. 2022.

[11] Pardede, H. F. Penerapan Pembelajaran Mesin (*Machine Learning*) dan Pembelajaran Dalam (Deep Learning) Berkinerja Tinggi untuk Mendukung Sektor Pertanian di Indonesia. 2023. Doi: 10.55981/brin.872

[12] Faiza, I. M., & Andriani, W. Tinjauan Pustaka Sistematis: Penerapan *Metode Machine Learning* Untuk Deteksi Bencana Banjir. *Jurnal Minfo Polgan*, *11*(2), 59-63. 2022. Doi: 10.33395/jmp.v11i2.11657

[13] Yang, P. B., Chan, Y. J., Yazdi, S. K., & Lim, J. W. Optimisation and economic analysis of industrial-scale anaerobic co-digestion (ACoD) of palm oil mill effluent (POME) and decanter cake (DC) using machine learning models: A comparative study of Gradient Boosting Machines (GBM), K-nearest neighbours (KNN), and random forest (RF). *Journal of Water Process Engineering*, *58*(December 2023), 104752. Doi: 10.1016/j.jwpe.2023.104752

[14] Calofir, V., Munteanu, R. I., Simoiu, M. S., & Lemnaru, K. C. Innovative approach to estimate structural damage using linear regression and K-nearest neighbors machine learning algorithms. *Results in Engineering*, *22*(May), 102250. 2024. Doi: 10.1016/j.rineng.2024.102250

[15] Wahyudi, S. Pengembangan Sistem Informasi Klinik Berbasis Web. *Riau Journal of Computer Science*, *6*(1), 50-58. 2020.

[16] Elisa, E., Tukino, T., & Handoko, K. Penerapan Forecasting Methods Untuk Penjualan Produk Umkm Dengan Algoritma K-Nearest Neighbor. *Jurnal Tekinkom (Teknik Informasi Dan Komputer)*, *5*(2), 455-463. 2022. Doi: 10.37600/tekinkom.v5i2.629

[17] Zalukhu, A., Purba, S., & Darma, D. (2023). Perangkat Lunak Aplikasi Pembelajaran Flowchart. *Jurnal Teknologi, Informasi Dan Industri*, *4*(1), 61-70. 2023. Doi:

[18] Susnaweli, S. (2024). Faktor Faktor Yang Mempengaruhi Pedagang Muslim Melakukan Kredit Pada Peminjaman Modal Perorangan Di Kota Jambi: Studi Kasus Pasar Modern Handil. *Mutiara: Jurnal Penelitian Dan Karya Ilmiah*, *2*(2), 87-102. 2024. Doi: 10.59059/mutiara.v2i2.1067

[19] Admojo, F. T. Klasifikasi Aroma Alkohol Menggunakan Metode KNN. *Indonesian Journal of Data and Science*, *1*(2), 34-38. 2020. Doi: 10.33096/ijodas.v1i2.12