Knowledge Engineering and Data Science (KEDS)
Vol 7, No 2, December 2024, pp. 128–138

pISSN 2597-4602
eISSN 2597-4637

Debtor Eligibility Prediction Using Deep Learning with Chatbot-Based Testing

Reski Noviania [a,1], Enny Itje Sela [a,2,*] , Luther Alexander Latumakulita [b,3], Steven R. Sentinuwo [b,4]

[a] *Universitas Teknologi Yogyakarta*
Jalan Siliwangi, Yogyakarta 55285, Indonesia
[b] *Universitas Sam Ratulangi*
Jalan Kampus, Manado 95115, Indonesia
[1] *reskinoviana11@gmail.com;* [2] *ennysela@uty.ac.id*;* [3] *latumakulitala@unsrat.ac.id;* [4] *steven@unsrat.ac.id;*
**corresponding author*

ARTICLE INFO

Article history:
Received 04 September 2024
Revised 13 October 2024
Accepted 11 December 2024
Published online 23 December 2024

Keywords:
Debtor eligibility
Chatbot
Recursive feature elimination
Deep learning
Credit scoring

ABSTRACT

Predicting debtor eligibility is essential for effective risk management and minimizing lousy credit risks. However, financial institutions face challenges such as imbalanced data, inefficient feature selection, and limited user accessibility. This study combines Recursive Feature Elimination (RFE) and Deep Learning (DL) to improve prediction accuracy. It integrates a chatbot interface for user-friendly testing. RFE effectively identifies critical features, while the DL model achieves a validation accuracy of 97.62%, surpassing previous studies with less comprehensive methodologies. The chatbot's novel design not only ensures accessibility but also enhances user engagement through flexible input options, such as approximate values, enabling non-experts to interact seamlessly with the system. For financial institutions, this chatbot-based testing approach offers practical benefits by streamlining debtor evaluation processes, reducing dependency on manual assessments, and providing consistent, scalable, and efficient solutions for credit risk management. It allows institutions to handle inquiries outside business hours, ensuring a continuous service flow. Furthermore, the system's flexibility supports better customer interaction, increasing trust and transparency. By combining advanced machine learning with accessible interfaces, this study offers a scalable solution to improve the precision and practicality of debtor eligibility assessments, making it a valuable tool for modern financial institutions.

I. Introduction

Credit is the most significant activity in the banking sector [1]. A major issue for banks and financial institutions, in general, is bad credit, which can increase allowance costs on the profit/loss statement. While credit contributes significantly to profits, it can also lead to instability in banks or financial institutions [2][3]. Effective management of credit risk is essential, as poor management can result in a higher proportion of problem loans, negatively impacting the financial health of banks [4].

The urgent need to predict debtor assessment is critical in the financial services sector [5], particularly in lending transactions. Accurate assessment of potential debtors is essential to prevent the occurrence of bad credit, which can have widespread adverse effects on the economy [6]. Inaccurate predictions can lead to approving loans for high-risk debtors, resulting in increased default rates and significant financial losses. Furthermore, such errors can erode institutional trust and damage reputations, compounding the financial impact. Therefore, exploring efficient and accurate methods for predicting debtor eligibility is crucial, as errors in approving loans for ineligible debtors can lead to an accumulation of bad credit [7][8]. Predicting debtor eligibility at scale can streamline lending processes, reduce default risks, and increase financial accessibility, leading to more substantial economic stability and equitable credit opportunities.

Currently, information technology plays a significant role in assessing debtor eligibility, especially using machine learning [7][9][10] and chatbots [11]. Machine learning models have been widely adopted for classification tasks and have generally enhanced prediction accuracy [12].

https://doi.org/10.17977/um018v7i22024p128-138

Research on chatbots in finance has gained popularity due to their ability to provide fast and efficient services across various domains, including customer service [13], personal management [14][15], and financial [11][16]. A relevant application of chatbots in this study is their use in debtor services. Chatbots offer several features for debtors, such as 24/7 accessibility, information on financial products and services, predictions of debtor eligibility, and support for customer complaints. Rule-based chatbots are created explicitly with fixed conversation pathways guided by predefined rules. They address queries by comparing the input to predefined keywords [15][17].

Based on this background, there is a need for a model tool that can predict debtor eligibility accurately and flexibly. We will utilize machine learning to address accuracy challenges, while chatbots will enhance flexibility. This research will conduct a debtor feasibility study employing Recursive Feature Elimination (RFE) for feature selection, Deep Learning (DL) classification methods, and chatbot-based testing. RFE works by iteratively removing features and building a model based on the remaining ones. Some advantages of the RFE method include improved model accuracy and flexibility, as it can be applied with various machine learning algorithms, such as linear regression, decision trees, and kernel-based models like SVM [18][19].

The central concept of DL involves gradually learning complex data representations through layers of neural networks. Each layer extracts features from raw data, with deeper layers capable of learning more complex features due to their ability to recognise data patterns. Chatbots can be accessed at any time, day or night, providing users with help or information without being constrained by business hours [14]. Additionally, chatbots deliver almost instant responses to user inquiries, reducing waiting time and improving efficiency [20].

Several studies focused on improving credit scoring or debtor eligibility and predicting financial behaviours using machine learning techniques. Safarkhani and Moro [21] aimed to enhance the accuracy of predicting bank depositors' behavior through the Decision Tree J48 model, achieving a 94.39% prediction accuracy with a sensitivity of 0.975 and specificity of 0.709. Zhirov [22] utilized a Neural Network Backpropagation approach, achieving an accuracy of 95%. Yiheng Li and Weidong Chen [23] employed methods such as Logistic Regression (LR), Decision Trees (DT), and Naïve Bayes (NB), with LR achieving an accuracy of 81.05%. Meanwhile, Yuelin Wang [24] applied multiple algorithms, including the NB, LR, Random Forest (RF), DT, and K-Nearest Neighbor Classifier, with RF obtaining a high accuracy of 96.53%. Chang Yu [25] focused on a dataset of over 40,000 records provided by a commercial bank, incorporating techniques like PCA, T-SNE, LightGBM, XGBoost, and SMOTE, with the best result achieving an accuracy of 99.89%. These studies highlight the effectiveness of combining feature selection and advanced machine learning methods to optimize credit scoring accuracy. Nallakaruppan et al. [26] examined a loan dataset using an RF model integrated with Explainable AI (XAI). They achieved impressive accuracy, sensitivity, and specificity scores of 0.998, 0.998, and 0.997, respectively. Rabihah et al. [5] analyzed data from Malaysian financial institutions using linear regression, multilayer perceptron, and SVM, with accuracy results of 61.7% for multilayer perceptron, 61.5% for linear regression, and 61.0% for SVM. Xueming et al. [27] explored the effects of chatbots in a financial services company in Asia, finding that chatbot usage led to a 79.7% decrease in purchase rates.

Additionally, Muslim et al. [28] achieved 90.63% accuracy with the LightGBM model. Together, these studies highlight advancements in machine learning applications in financial services, especially in prediction accuracy and technology adoption. Compared to existing methods like Decision Tree or Logistic Regression, which can struggle with feature selection and require extensive preprocessing, this study employs RFE to streamline the dataset and enhance efficiency. RFE is a feature selection technique that iteratively removes less important features to enhance model efficiency. DL, on the other hand, is a machine learning approach that utilizes artificial neural networks to identify complex patterns in data, thereby achieving higher prediction accuracy. Furthermore, by integrating a chatbot interface, the proposed method significantly improves operational scalability, enabling real-time assessments without human intervention. This dual emphasis on algorithmic efficiency and practical scalability makes the approach more adaptable for large-scale implementations in financial institutions.

II. Methods

This research is conducted in a structured manner according to the established stages. The proposed method for this research is illustrated in Figure 1.

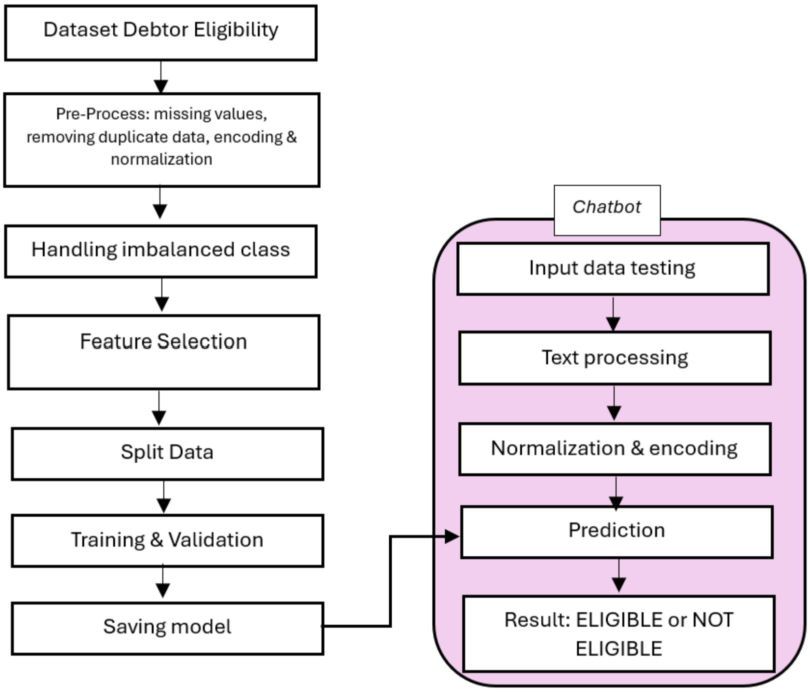

Fig. 1. Proposed method

The data used in this study was sourced from Koperasi Simpan Pinjam XYZ, located in Yogyakarta, Indonesia. The dataset encompasses customer transaction records spanning from 2020 to 2023. Table 1 outlines the attributes and types of data included in the analysis. Over the specified period, 165 records were collected for this research.

Table 1. Attributes and types of data

Variable	Data Type
Age	Integer
Employee occupation	Categorical
Tenure	Integer
Marital status	Categorical
Education	Categorical
Income Amount	Integer
Number of dependents	Integer
Assets	Integer
Loan amount	Integer
Loan term	Integer
Current debt	Integer
Never late	Categorical
Guarantor occupation	Categorical
Debt requirements	Categorical
Class label	Categorical

The preprocessing activities conducted in this research included checking for missing values, removing duplicate data, encoding, and normalization. Upon checking for missing values in the dataset, it was found that there were no missing values. However, the duplicate data check revealed 12 duplicate rows, which were subsequently removed. The initial dataset contained 165 rows; after removing the duplicates, it was reduced to 153.

The encoding process is necessary to convert categorical data into numerical data, ensuring compatibility with the system being developed. In this research, we employed an ordinal encoder. The encoding process was applied to the following attributes: employee occupation, marital status, education, never late, guarantor occupation, debt requirement, and class label. Table 2 displays the results of the attribute encoding.

Table 2. Attributes encoding

Variable	Description
Employee occupation	0 = government employee (ASN); 1 = non-ASN employee; 2 = entrepreneur; 3 = retired
Marital status	0 = married; 1 = divorced; 2 = single
Education	0 = elementary school; 1 = junior school; 2 = high school; 3 = university, academy
Never late	0 = yes; 1 = no
Guarantor occupation	0 = government employee (ASN); 1 = non-ASN employee; 2 = entrepreneur; 3 = retired
Debt requirements	0 = investment; 1 = consumptive; 2 = others
Class label	0 = Not Eligible; 1 = Eligible

Data normalization is essential because the variables in the dataset often have different scales or ranges of values. Normalization helps standardize these variables, enhancing the performance of specific machine learning models and statistical analyses. In this study, the attributes of age, Income amount, Number of dependents, Assets, Current debt amount, Loan amount, length of employment, amount of income, number of dependents, amount of assets, amount of debt, amount of loans, and loan duration will be adjusted using Min-Max normalization as in (1).

$$X_{new} = \frac{x - min_A}{max_A - min_A}(maxn_A - minn_A) + minn_A \tag{1}$$

Let X represent the data to be normalized while min_A and max_A denote the minimum and maximum values for the attribute whose data is to be normalized. Additionally, $minn_A$ and $maxn_A$ represent the new minimum and maximum values for the same attribute after normalization.

Data in the 'eligible' class totals 105 rows, and data in the 'not eligible' class totals 58 rows. This dataset imbalance will be resolved using the SMOTE method [29]. This process produces a new dataset almost twice as large as the previous dataset (duplicate data). The SMOTE algorithm is an oversampling technique that increases the amount of data in the minority class by randomly replicating the amount of minority class data so that the amount is the same or close to the majority class data [30]. The SMOTE algorithm looks for K-Nearest Neighbors, grouping data based on the nearest neighbors. The selection of the closest neighbor is based on the Euclidean distance between a data pair. Given data with p variable, namely $xT = [x1, x2, \dots, xn]$ and $zT = [\mathrm{z1}, \mathrm{z2}, \dots, \mathrm{zn}]$, up to the Euclidean distance $d(x, z)$ calculated as in (2), and for the synthetic data is generated using as in (3).

$$d(x, z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2 + \cdots (x_n - z_n)^2} \tag{2}$$

$$x_{syn} = x_1 + (x_{knn} - x_i) * \gamma \tag{3}$$

Where x_{syn} is data resulting from replication, x_i is the i'th data from the minor class, x_{knn} is data from the minor class with the closest distance from the class x_i, and γ is a random number between 0 and 1. After using the SMOTE method as was done in the research, the distribution of feasible and classes are balanced, with 105 rows each. Next, the SMOTE dataset will undergo a feature selection process.

RFE algorithm is a feature selection technique that aims to identify the most important features in a dataset based on their relevance to the target variable. The main steps in this algorithm are as follows [19]. Begin with a dataset with n features and define the desired number of features *m* to select. A regression model will train the dataset with all available features. The model learns the relationship between the features and the target variable at this step. After training, calculate each feature's importance based on the change in Mean Squared Error (MSE) when that feature is removed from the model. The importance of a feature f_i is calculated as in (4).

$$\Delta MSE(f_i) = MSE_{with\, f_i} - MSE_{without f_i} \tag{4}$$

The magnitude of the change in MSE indicates how critical a feature is to the model's predictions. After calculating each feature's importance, sort them based on the absolute values of their importance scores or model coefficients. This vector of feature importance is denoted as $\mathrm{w} = [\mathrm{w1}, \mathrm{w2} \ldots, \mathrm{wn}]$. Identify and remove the feature with the lowest importance score from the dataset. The feature with index k and the smallest absolute importance value wk was eliminated. Repeat training the model, calculating feature importance, ranking, and removing the least important feature until only mmm desired features remain.

RFE helps retain only the most relevant features by iteratively removing less significant features, improving the model's efficiency and performance. The result of this stage is a dataset with relevant features. The final features selected through RFE include age, number of dependents, assets, the current debt amount, employee occupation, income amount, loan amount, loan term, never late, and guarantor occupation.

Then, the data is divided into training data and validation data. Several scenarios exist for the training and validation data splitting ratio: 60:40, 70:30, 80:20, and 90:10. Each scenario is applied to the DL method by displaying performance results as a confusion matrix.

DL algorithm trains an artificial neural network with L layers using a supervised learning approach. Given a dataset (X, Y), where X is the input data and Y is the target or label, the goal is to train the network to minimize prediction errors. The algorithm requires an activation function (such as ReLU or Sigmoid), a loss function (e.g., Cross-Entropy Loss), the Adam optimizer, and specific hyperparameters like the learning rate η and number of epochs N [31].

Weight Initialization: Randomly initialize the network’s weights and biases for each layer. Training Process: The training proceeds over N epochs, where each epoch involves iterating through mini-batches in the dataset to adjust weights based on the error. Forward Pass: For each batch, input X passes through the network from the input layer to the output layer. Each layer’s output, a(l), is calculated based on the activation function as in (5) and (6).

$$z^{[l]} = W^{[l]}a^{[l-1]} + b^{[l]} \tag{5}$$

$$a^{[l]} = p(z^{[l]}) \tag{6}$$

Where $W^{[l]}$ and $b^{[l]}$ represent the weights and biases, respectively, and σ is the activation function, such as ReLU, defined as in (7).

$$ReLu(X) = \max(0, x) \tag{7}$$

Loss Calculation: Calculate the network’s loss using the loss function L based on the predicted output $\hat{Y}$, and accurate labels Y as in (8):

$$loss = L(\hat{Y}, Y) \tag{8}$$

Backward Pass (Backpropagation): Compute gradients of the loss to each layer's output. This is done by propagating the gradients backward through the network, as in (9) and (10).

$$\delta^{[l]} = \nabla_a L(\hat{Y}, Y) \odot \sigma'(z^{[L]})) \tag{9}$$

$$\partial^{[l]} = (W^{[l+1]})^T \partial^{[l+1]} \odot \sigma'(z^{[l]}) \tag{10}$$

Where $\delta^{[l]}$ represents the loss gradient for layer l.

Weight Update: Using the gradients calculated, update weights as in (11) and biases as in (12) using the Adam optimizer.

$$W^{[l]} = W^{[l]} - \eta \nabla_W L(\hat{Y}, Y) \tag{11}$$

$$b^{[l]} = b^{[l]} - \eta \nabla_b L(\hat{Y}, Y) \tag{12}$$

After all epochs, the neural network has been trained, and its weights are adjusted to minimize the error between predicted and actual labels.

The training results with the highest accuracy are saved as the reference model, designated as the "best model." This model, built on a rule-based approach, is the foundation for testing new data within a rule-based chatbot framework. The chatbot utilizes optimized knowledge derived from classification learning. In the developed application, the rule-based chatbot is implemented using arrays. It is coded in Python with the NLTK library. This allows users to provide responses flexibly based on features prompted by the system. User input undergoes a preprocessing pipeline that includes tokenization, stop word removal, filtering, word distribution analysis, and feature extraction. The processed data is then normalized and encoded. Finally, the system predicts whether the input corresponds to the "eligible" or "not eligible" class.

III. Results and Discussion

Training is conducted with data splitting ratios: 60:40, 70:30, 80:20, and 90:10. Validation results using RFE and DL can be seen in Table 3. The best results achieved an accuracy value of 97.62% with precision, recall, specificity, and F1 score, each valued at 100.00%, 94.74%, 100.00%, and 97.30% in the 80%:20% data split scheme. The 80:20 split scheme outperformed others because it best balanced training and testing data. With 80% of the data used for training, the model effectively learned patterns without overfitting. In comparison, 20% of testing data provided reliable validation. This split resulted in the highest accuracy (97.62%), perfect precision (100%), and specificity (100%), indicating zero false positives and false negatives. Additionally, the F1 Score (97.30%) and recall (94.74%) demonstrated an optimal balance between precision and recall. Compared to other splits, the 80:20 scheme offered the most robust performance across all metrics, avoiding the overfitting seen in the 90:10 split while achieving superior generalization.

Table 3. Validation results

Split ratio	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	F1 Score (%)
60:40	94.49	92.06	97.35	91.63	94.63
70:30	94.70	92.09	97.79	91.63	94.85
80:20	97.62	100.00	94.74	100.00	97.30
90:10	90.48	90.91	90.91	90.00	90.91

Figure 2 shows the training and validation confusion matrices. The training accuracy reached 100%, meaning all 168 training samples were successfully classified into their respective classes. Meanwhile, the validation accuracy was 97.62%, indicating that out of 42 validation samples, 41 were correctly classified, with 23 True Positives (TP), 18 True Negatives (TN), 1 False Positive (FP), and 0 False Negatives (FN).

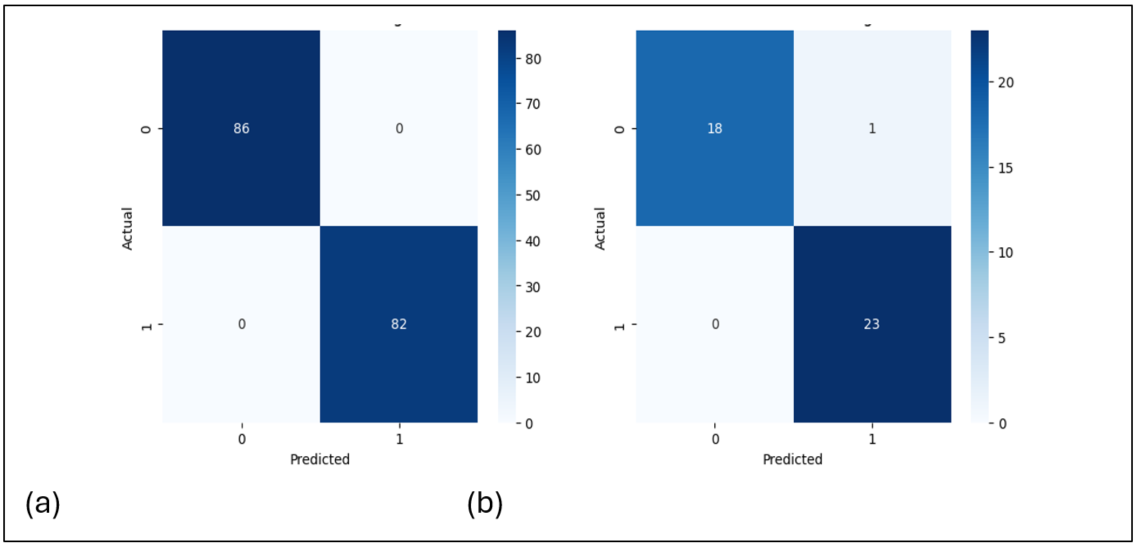

Fig. 2. Validation confusion matrix

The optimal DL architecture consists of an input layer with 10 nodes, two dense layers with 64 neurons each, and a final dense layer with a single neuron. The deep learning parameter tuning was conducted with epoch values of 400, 500, and 600, respectively. The learning rates tested were 0.0001, 0.001, and 0.01. The best results were obtained with 500 epochs and a learning rate 0.001. Batch size was set up to 32 (default). Meanwhile, Figure 3 shows a comparison between accuracy values and loss values.

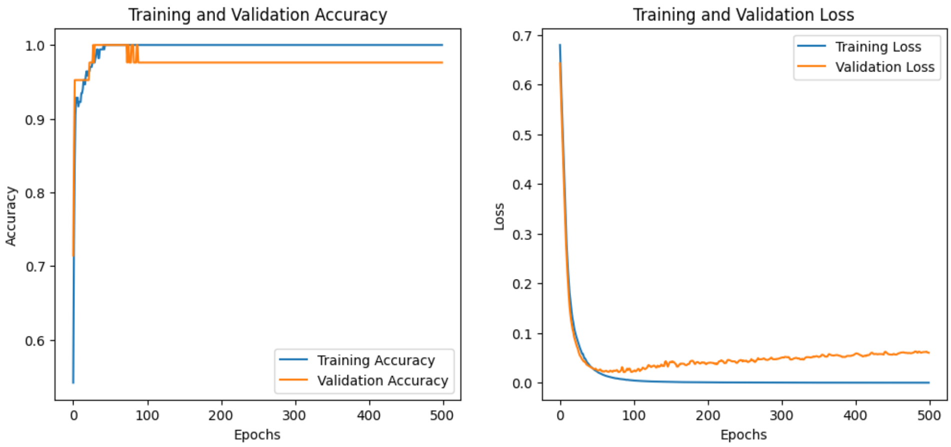

Fig. 3. Comparison of training and validation accuracy and loss

Next, the data from the selected features is stored and used as input for chatbot-based testing. In this study, a rule-based chatbot was developed to evaluate new data from prospective customers using the optimal combination of RFE and DL. The results of the chatbot's evaluation are presented in Figure 4. The chatbot allows users to include related terms, particularly for integer numeric attributes such as "around," "approximately," "about," "+/-," or "maximum" to enhance user flexibility when providing input. This flexibility addresses the challenge that users may not always have precise information regarding age, total assets, debts, or income. This feature simplifies user interactions, making sharing their data easier for non-expert users. As a result, it enhances user engagement and improves the system's overall accuracy in assessing debtors.

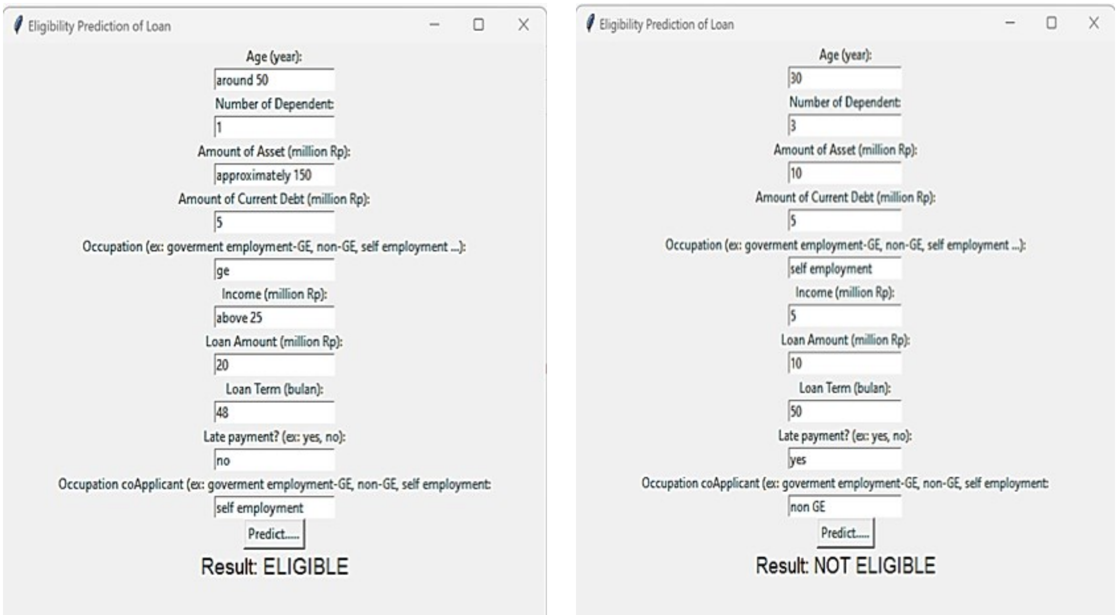

Fig. 4. Chatbot for testing new data

In recent years, the rise of financial technologies (fintech) has transformed the roles of traditional intermediaries and created new opportunities for consumers and investors. In the context of credit scoring, fintech has enabled more efficient, data-driven approaches to assessing creditworthiness, offering faster and more accessible solutions for borrowers and lenders [32]. This study has developed a debtor assessment model using a combination of RFE for feature selection and DL for the learning process. The results of 3 scenarios are presented: scenario 1, dataset learning with RFE combined with four machine learning (ML) methods: DL, NB, RF, and LR. In scenario 2, dataset learning with Forward Selection combined with the same ML methods (NB, RF, DL, and LR). Scenario 3, dataset learning without applying any feature selection. The results for scenario 1, using the four machine learning methods, are summarized in Table 4. These results indicate that the combination of RFE and DL achieves the highest accuracy.

Table 4. RFE Validation results (Scenario 1)

ML	Acc	Pre	Rec	Spec	F1Sc
NB	82.54	90.91	85.11	75.00	87.91
DL	97.62	100.00	94.74	100.00	97.30
RF	90.48	90.91	90.91	90.00	90.91
LR	80.95	81.82	81.82	80.00	81.82

Scenario 2 tests the dataset using four machine learning methods and the forward selection feature selection method. The test results can be seen in Table 5. The performance value for scenario 2 is the best accuracy value achieved using the DL method with an accuracy value of 94,87% with precision, recall, specificity, and F1 score, each worth 94.74; 94.74%, 95.00%, and 94.74%. Based on this test scenario, select as many as 11 features when using Forward Selection: age, employee occupation, marital status, education, number of dependents, assets, current debt amount, income amount, loan amount, loan term, and guarantor occupation.

Table 5. Forward selection validation results (Scenario 2)

ML	Acc	Pre	Rec	Spec	F1Sc
NB	76.19	80.00	72.73	80.00	76.19
DL	94.87	94.74	94.74	95.00	94.74
RF	85.71	83.33	90.91	80.00	86.96
LR	76.19	75.00	81.82	70.00	78.26

In scenario 3, a classification trial was performed using NB, DL, RF, and LR without feature selection. The following are the results of implementing the classification (Table 6).

Table 6. Validation results without selection feature (Scenario 3)

ML	Acc	Pre	Rec	Spec	F1Sc
NB	70.24	77.78	70.00	70.59	73.68
DL	76.19	75.00	81.82	70.00	78.26
RF	80.95	81.82	81.82	80.00	81.82
LR	76.19	75.00	81.82	70.00	78.26

Table 6 shows that the RFE method produces the highest accuracy, 80.95%, with precision, recall, specificity, and F1 score, each worth 81.82%, 81.82%, 80.00%, and 81.82%. Figure 5 shows a graph of the accuracy values of the three scenarios worked on, with the highest accuracy value obtained in scenario 1, namely a combination of RFE and DL. The limitations of the RFE method in selecting features for highly imbalanced datasets are primarily due to its reliance on the model's performance to evaluate feature importance. In imbalanced datasets, the model tends to be biased toward the majority class, affecting the feature selection process. RFE may prioritize features that are more predictive of the majority class, neglecting features important for detecting the minority class. This bias can lead to suboptimal feature selection, reducing the model's ability to generalize and accurately classify the minority class. Additionally, RFE does not account for class imbalance directly, which can further exacerbate this issue.

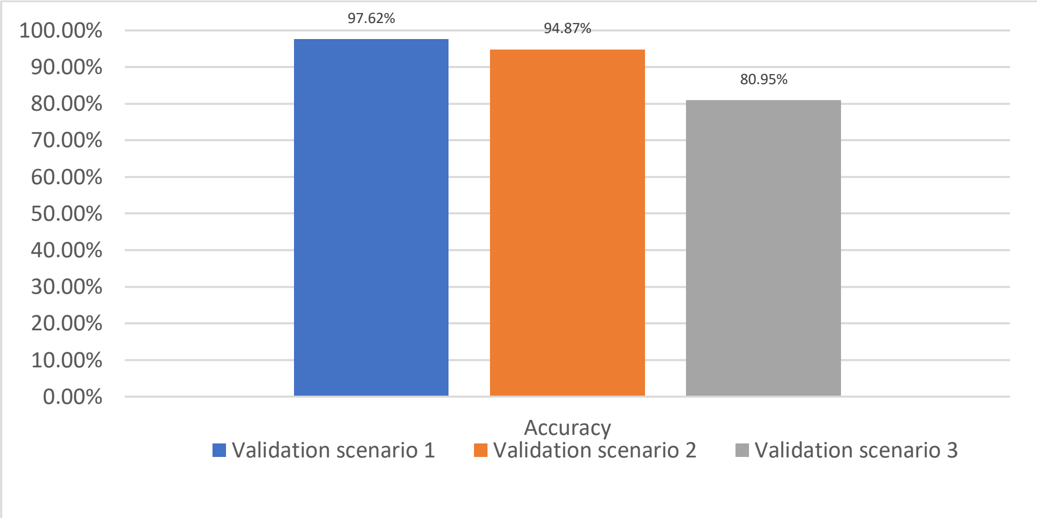

Fig. 5. Comparison of accuracy values

The accuracy of the research has better results when compared with research [5][21][22][23][24][33][34]. This difference in performance results is thought to occur because the research did not carry out complete preprocessing as in this study. However, this study's accuracy results are slightly lower than other studies [25][26]. The analysis can be written because the data processed has different features and quantities. Research [25] showed that over 40,000 records and 16 features were processed data. Meanwhile, research [26] consists of the variables gender, character, Col, financing, type of financing, and customer eligibility, and the amount of data is 430 records. This has an impact on differences in performance.

This study leverages advanced chatbot features that enable users to independently assess debtors, a capability not found in other studies. The chatbot can process approximate inputs using terms such as "approximately," "about," "around," and the "+/-" symbol for numeric data. However, compared to more sophisticated chatbots discussed in [14][35]. The chatbot in this study is relatively more straightforward, relying on rule-based processing to manage the sequence and number of features. While effective for debtor assessment, rule-based chatbots have certain limitations. They operate within predefined rules and logic, making them unable to address queries or symptoms outside their programmed scope. This restricts their adaptability and limits their ability to provide personalized responses in more complex or ambiguous scenarios.

Furthermore, unlike machine learning-based chatbots [36], rule-based systems cannot learn and evolve, reducing their capacity to handle nuanced or dynamic knowledge, thereby impacting their long-term effectiveness. This study enhances the chatbot's ability to provide personalized and dynamic responses based on user inputs, adapting to changing contexts and preferences. By utilizing machine learning algorithms, the chatbot can predict user needs, intent, and behavior more accurately, improving the overall interaction. This connection between prediction and real-time interaction ensures that the chatbot can respond intelligently, providing a seamless and efficient user experience. Additionally, the study may explore techniques for improving the chatbot’s learning process from user feedback, further closing the gap between static predictions and evolving user demands.

IV. Conclusions

Relevant features were obtained using RFE: age, number of dependents, assets, current debt amount, employee occupation, income amount, loan amount, loan term, never late, and guarantor occupation. The proposed combination of RFE and DL methods can be used as an alternative to assess debtor eligibility with validation accuracy values reaching 97.62%. Chatbot-based testing that has been implemented can fill in test data more flexibly and can provide feasibility assessment results. This research has the advantage of a new chatbot-based data testing feature that can provide predictive results. This feature has not been implemented in previous research. This feature helps users to provide more flexible input. This research needs to be developed with a more flexible chatbot feature where users can provide self-descriptions freely. The model can be adapted for industries or regions by incorporating region-specific financial data, consumer behaviour, and local economic conditions to improve its accuracy and relevance. The model can be adapted for industries or regions by incorporating region-specific financial data, consumer behaviors, and local economic conditions to improve its accuracy and relevance. Additionally, industry-specific variables, such as unique risk metrics in banking or consumer preferences in retail, can further tailor the model's predictions. Future improvements could involve integrating advanced machine learning techniques, sentiment analysis, and multi-turn conversation capabilities, enhancing the chatbot's flexibility and intelligence and allowing it to understand better and respond to user needs in real time.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Funding Statement

The authors would like to thank the Ministry of Research, technology, and Higher Education of the Republic of Indonesia for funding this research in the scheme of Penelitian Tesis Magister in 2024.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] N. Kozodoi, J. Jacob, and S. Lessmann, "Fairness in credit scoring: Assessment, implementation and profit implications," Eur. J. Oper. Res., vol. 297, no. 3, pp. 1083–1094, Mar. 2022.

[2] S. Filomeni, G. F. Udell, and A. Zazzaro, "Communication frictions in banking organizations: Evidence from credit score lending," Econ. Lett., vol. 195, no. 109412, p. 109412, Oct. 2020.

[3] Y. Tan, M. C. K. Lau, and G. Gozgor, "Competition and Profitability: Impacts on Stability in Chinese Banking," Int. J. Econ. Bus., vol. 28, no. 2, pp. 197–220, May 2021.

[4] A. Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, and B. Tarantino, "Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses," Socioecon. Plann. Sci., vol. 90, no. 101746, p. 101746, Dec. 2023.

[5] R. M. Sum, W. Ismail, Z. H. Abdullah, N. F. M. N. Shah, and R. Hendradi, "A new efficient credit scoring model for personal loan using data mining technique for sustainability management," J. Sustain. Sci. Manag., vol. 17, no. 5, pp. 2672–7226, 2022.

[6] D. Kumar, R. Maddikunta, A. Chella, and R. Myadaraboina, "Machine learning approaches for predicting loan approval using chat bots," Int. J. Progress. Res. Eng. Manag. Sci., vol. 04, no. 05, pp. 288–297, 2024.

[7] B. Dushimimana, Y. Wambui, T. Lubega, and P. E. McSharry, "Use of machine learning techniques to create a credit score model for airtime loans," J. Risk Financ. Manag., vol. 13, no. 8, 2020.

[8] Y. Wu and Y. Pan, "Application analysis of credit scoring of financial institutions," Complexity, vol. 2021, no. 9222617, 2021.

[9] C. Wang, D. Han, Q. Liu, and S. Luo, "A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM," IEEE Access, vol. 7, pp. 2161–2168, 2019.

[10] D. West, "Neural network credit scoring models," Comput. Oper. Res., vol. 27, no. 11–12, pp. 1131–1152, Sep. 2000.

[11] S. Prathipa, S. Haroon Prakash, D. Dinesh Kumar, and G. Tejesh Kumar, "Loan Management System Using Chatbot," in 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Apr. 2024, pp. 1–6.

[12] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," Soft Comput., vol. 25, no. 3, pp. 2277–2293, 2021.

[13] J. Singh, M. H. Joesph, and K. B. A. Jabbar, "Rule-based chabot for student enquiries," J. Phys. Conf. Ser., vol. 1228, no. 1, 2019.

[14] C. Papakostas, C. Troussas, A. Krouska, and C. Sgouropoulou, "A Rule-Based Chatbot Offering Personalized Guidance in Computer Programming Education," in Lecture Notes in Computer Science, vol. 14799, 2024, pp. 253–264.

[15] D.-H. Kim, H.-S. Im, J.-H. Hyeon, and J.-W. Jwa, “Development of the rule-based smart tourism chatbot using Neo4J graph database,” Int. J. Internet, Broadcast. Commun., vol. 13, no. 2, pp. 179–186, 2021.

[16] M. Jang, Y. Jung, and S. Kim, “Investigating managers’ understanding of chatbots in the Korean financial industry,” Comput. Human Behav., vol. 120, p. 106747, Jul. 2021.

[17] S. A. Thorat and V. Jadhav, “A review on implementation issues of rule-based chatbot systems,” SSRN Electron. J., no. Icicc, pp. 1–6, 2020.

[18] Y. Kwon et al., “Osteoporosis pre-screening using ensemble machine learning in postmenopausal Korean women,” Healthc., vol. 10, no. 6, 2022.

[19] W. Lian, G. Nie, B. Jia, D. Shi, Q. Fan, and Y. Liang, “An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning,” Math. Probl. Eng., vol. 2020, 2020.

[20] E. A. Kadasah, “Artificial Intelligence Powered Chatbot for Business,” Int. J. Inf. Technol. Bus., vol. 4, no. 2, pp. 61–66, May 2023.

[21] F. Safarkhani and S. Moro, “Improving the accuracy of predicting bank depositor’s behavior using a decision tree,” Appl. Sci., vol. 11, no. 19, 2021.

[22] V. K. Zhirov, N. A. Staroverova, M. L. Shustrova, and M. N. Tomilova, “Neural network as a tool to solve the problem of credit scoring,” J. Phys. Conf. Ser., vol. 2032, no. 1, 2021.

[23] Y. Li and W. Chen, “A comparative performance assessment of ensemble learning for credit scoring,” Mathematics, vol. 8, no. 10, pp. 1–19, 2020.

[24] Y. Wang, Y. Zhang, Y. Lu, and X. Yu, “A comparative assessment of credit risk model based on machine learning a case study of bank loan data,” Procedia Comput. Sci., vol. 174, pp. 141–149, 2020.

[25] C. Yu, Y. Jin, Q. Xing, Y. Zhang, S. Guo, and S. Meng, “Advanced User Credit Risk Prediction Model Using LightGBM, XGBoost and Tabnet with SMOTEENN,” in 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS), Jul. 2024, pp. 876–883.

[26] M. K. Nallakaruppan, B. Balusamy, M. L. Shri, V. Malathi, and S. Bhattacharyya, “An Explainable AI framework for credit evaluation and analysis,” Appl. Soft Comput., vol. 153, no. 111307, p. 111307, Mar. 2024.

[27] X. Luo, S. Tong, Z. Fang, and Z. Qu, “Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases,” Mark. Sci., vol. 38, no. 6, pp. 937–947, 2019.

[28] M. A. Muslim, Y. Dasril, A. Alamsyah, and T. Mustaqim, “Bank predictions for prospective long-term deposit investors using machine learning LightGBM and SMOTE,” J. Phys. Conf. Ser., vol. 1918, no. 4, 2021.

[29] E. Ileberi, Y. Sun, and Z. Wang, “Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost,” IEEE Access, vol. 9, no. 165286, pp. 165286–165294, 2021.

[30] A. Fernández, S. García, F. Herrera, and N. V. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” J. Artif. Intell. Res., vol. 61, pp. 863–905, 2018.

[31] Y. Li, Y. Fan, S. Wang, J. Bai, and K. Li, “Application of YOLOv5 based on attention mechanism and receptive field in identifying defects of thangka images,” IEEE Access, vol. 10, no. July, pp. 81597–81611, 2022.

[32] P. Giudici, B. Hadji-Misheva, and A. Spelta, “Network based credit risk models,” Qual. Eng., vol. 32, no. 2, pp. 199–211, 2020.

[33] M. A. Muslim, A. Nurzahputra, and B. Prasetiyo, “Improving accuracy of c4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction,” Int. Conf. Inf. Commun. Technol., pp. 141–145, 2018.

[34] E. Priyanto, E. I. Sela, L. A. Latumakulita, and N. Islam, “Decision Tree C4.5 Performance Improvement using Synthetic Minority Oversampling Technique (SMOTE) and K-Nearest Neighbor for Debtor Eligibility Evaluation,” Ilk. J. Ilm., vol. 15, no. 2, pp. 373–381, Aug. 2023.

[35] M. Ghaleb et al., “Mining the chatbot brain to improve COVID-19 bot response accuracy,” Comput. Mater. Contin., vol. 70, no. 2, pp. 2619–2638, 2022.

[36] S. Zhang and J. Song, “A chatbot based question and answer system for the auxiliary diagnosis of chronic diseases based on large language model,” Sci. Rep., vol. 14, no. 1, p. 17118, Jul. 2024.