International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Loan Approval Prediction Based on a Hybrid Approach of Dynamin Thresholding Genetic Algorithm and Support Vector Machine

Ahmad Abdullah Mohammed Al-Mafrji[1*] , Ahmed M. Fakhrudeen[2] , Lotfi Chaari[3]

[1] Institute of ENET'COM, University of Sfax, Sfax 82713, Tunisia
[2] Software Department, College of Computer Science and Information Technology, Kirkuk University, Kirkuk 36001, Iraq
[3] Institute of IRIT-ENSEEIHT, University of Toulouse, Toulouse 31000, France

Corresponding Author Email: ahabm4040@uokirkuk.edu.iq

ABSTRACT

The loan eligibility prediction model utilizes an analytical approach that adjusts previous and current credit user data to provide forecasts. An important challenge in predicting loan eligibility is accurately forecasting loan outcomes via risk assessment and evaluation analysis. In the nowadays the Predictions of loan approvals now require the utilization of machine learning. Financial institutions are seeking methods to automate the loan approval process while minimizing risk in response to the rising demand for credit. This article introduces a novel application of machine learning methods to predict loan approval. The research centers on various algorithmic architectures, neural networks, support vector machines, decision trees, and random forests. Furthermore, the paper addresses the obstacles encountered in applying Artificial Intelligence (machine learning (ML) algorithms) for predicting loan approval. Moreover, for feature selection, the article proposes a dynamic thresholding genetic algorithm (DTGA) based on loan approval prediction. Besides, we emphasize the importance of data quality and feature selection in designing an effective ML model for loan approval prediction. Compared to conventional approaches, performance evaluation of the DTGA demonstrates the GA feature selection can substantially enhance the accuracy of loan approval prediction. Therefore, this article contributes to utilizing ML models for predicting loan approvals and the potential ramifications. Consequently, it enhances the decision-making procedures of financial institutions.

1. INTRODUCTION

For a long time, the banks have carried the country's economy by providing investments and lending for citizens. Although lending brings revenues to the banks, the banks suffer from repaying from customers or companies [1]. Therefore, the banks utilize the new advances in technologies (particularly Artificial Intelligence (AI) to accept or reject lending (also called loan approval). The main operation of the AI models depends on past customer performance for repaying and several other obstacles [2]. Specifically, it provides a more accurate evaluation of the borrower's creditworthiness. Accordingly, the number of highly risky borrowers is reduced by identifying their repaying history [3]. Consequently, these loan approval models can be argued as tools that aim to improve the decision-making of loan approval. For instance, the authors [4] demonstrated that the ML models in predicting loan approval can provide an accuracy of 85%. Nevertheless, obstacles are still presented in financial institutions. The challenges lie in the fairness of these models concerning every borrower and guaranteeing impartiality, irrespective of their ethnicity and gender. The challenges lie in guaranteeing the explicability and transparency of the models [5].

Recently, the scholars have paid a large attention oof utilizing Machine Learning (ML) algorithms to automatate the procedure of lending (i.e., loan approval process) [6]. Since the loan processing suffer from an increase expense and loan delay. On the other hands, the algorithms can perform the following procedures fastly:
- Analyzing a large data.
- Recognizing the patterns.
- Predict an accurate decision.

Consequently, they accelerate the procedure of approving any loan. The manner in which we approach complex problems in numerous industries, including finance, has been revolutionized by ML algorithms. Loan approval is a highly consequential implementation of ML algorithms within finance. By analyzing large amounts of data, they can predict whether the loan will be repaid [6]. Therefore, they can reduce and mitigate the lending risk and provide an efficient loan approval process. Consequently, professionals and scholars have been paying much attention to developing accurate ML-based loan approval models.

The authors [7] showed that ML algorithms contribute to enhancing decision-making, leading to higher revenues with lower default rates. Additionally, Alonso Robisco

demonstrated [8] that the ML algorithm outperformed the traditional methods. But there are several concerns related to the possibility of lying in ML models. These issues lie in the historical information utilized to train the models that contain impacts related to specific individuals or companies. For instance, low-income individuals and the attitude of the models. Therefore, the issues must be addressed through trusted data selection and validation.

Despite these concerns, there are potential benefits of utilizing ML algorithms in loan approval procedures [9]. This article evaluates the pros and cons of different ML-based loan approval methods. Furthermore, to emphasize the best strategies for practical lending cases, we analyze the existing important ML-based loan approval models. Accordingly, we propose a dynamic thresholding genetic algorithm (DTGA) based on loan approval prediction. Then, show how our proposal may outperform the counterparts.

The rest of this article is structured as follows. Section 2 reviews some important related works. Section 3 presents the methodology for implementing the DTGA model. Section 4 evaluates and demonstrates the superiority of the proposal. Finally, Section 5 provides a summary and conclusion of this work.

2. RELATED WORK

Recently, significant interest has been paid to utilize ML models to predict loan approval by accurately evaluating the borrowers' possibility of repaying their loans [10]. This paper presents a novel approach: the application of data mining to classify the risk associated with bank loans. The model comprises three ML algorithms: Naïve Bayes, bayesNet and j48 [11]. However, before delving into the methodology of our proposal, we review and provide a critical analysis of some important previous works.

The authors in the study [12] developed the approval status of a loan application. In their design, they combined a Support Vector Machine (SVM) algorithm and chi-square feature selection. The extensive experiments demonstrated that the model is efficient in predicting the approval status of loan applications with 92.8% accuracy. In developing loan application approval, the authors in the research [13, 14] utilized different ML algorithms in their models: Random Forest (RF), Decision Tree (DT), Logistic Regression (LR) and SVM. The evaluation showed the superiority of SVM over its counterparts in achieving higher accuracy.

Similarly, to develop a credit loan application, the authors [15, 16] utilized a combination model between AI capability and random forests in their models. Performance evaluation experiments showed that their models achieved satisfactory accuracy in loan approval. Moreover, the SVM algorithm has been utilized [17, 18]. Similarly, testing the algorithm showed that SVM can achieve better accurate decision-making status. Finally, Table 1 summarizes some important recent efforts in the loan approval context.

3. METHODS AND MODELLING

After analyzing the related works, we found that genetic algorithm via dynamic thresholding for feature selection could contribute to accurate loan approval prediction. It is important to mention that computational complexity is related to the data

and time utilized in the evaluation. Consequently, more accurate decision-making about loan approval. Therefore, we propose utilizing dynamic thresholding as a dynamic fitness threshold changed according to population fitness distribution. The name of the proposal is dynamic thresholding genetic algorithm (DTGA). Thus, by avoiding premature convergence, it is expected that this combination will enhance the accuracy of loan approval decisions.

Furthermore, DTGA can reduce the complexity of feature selection. Accordinglly, we can achieve further efficient feature selection procedures (i.e., redicate the features that could be redundant. Additionally, our proposal can enhance loan approval steps, making it consistent with the financial institution's main aim. Very briefly, the main aim is increasing profitability and reducing the default risk through managing (i.e., evaluating and controlling) lending issues. Ultimately allowing institutions to identify influential features that may mitigate the potential risk from lending and undrestnat loan approval impact.

Steps to implement the model

As shown in Figure 1, we proposed our model (DTGA), which can predict loan approval for borrowers according to their capability to repay. Through a dynamic threshold, DTGA evaluates the gathered history data of the borrower as an input. Next, the model's output is a response (decision) to the borrower's request.
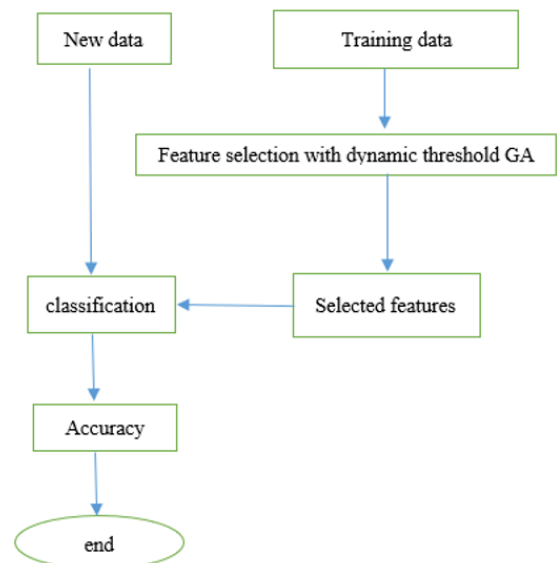

Figure 1. Flowchart of DTGA implementation

It is possible to predict the probability and severity of future loan defaults through various data analytics tools that identify the most influential determinants of loan approval. By utilizing the DTGA, financial institutions can enhance loan approval decisions with greater precision and productivity, diminish default risk, and improve profitability. At this time, the training data set is supplied by the ML model, which employs the data for its training. The information provided by every new applicant on the application form functions as a test dataset. Once the model has completed the testing phase, it uses the inferences it derives from the training data sets to forecast whether a new loan applicant is qualified for approval. Financial institutions can improve risk management and profitability by utilizing a GA with dynamic thresholding for feature selection in loan acceptance prediction and reducing the time and resources required to evaluate loan applications.

The steps of implementing DTGA can be summarized as follows:

First step: Dataset utilization stage

The data utilized by our proposed model comprised training and testing sets and was obtained from Kaggle (a widely recognized source of educational data). Typically, the training data was divided in half again, with the proportions being 70:30 or 80:20. Both the major and minor datasets were employed to train and evaluate the accuracy of the models.

Second step: Preprocessing stages

To address missing values, obtain the most suitable data imputation technique, which may involve employing a next or previous value.

Specify appropriate actions to address anomalies present in the dataset, including their removal, transformation, or imputing.

The process of dividing continuous or numerical data into smaller intervals, or bins, simplifying and facilitating the interpretation of information.

Convert categorical data to numeric data using one hot encoding technique and create binary columns for each category present in the categorical variable.

Table 1. Summary of literature review algorithms for loan acceptance decision support system

Method	Usage	Advantage	Disadvantage
Chi-square feature selection [19, 20]	A statistical method selects a dataset's most important features, particularly for categorical features.	Simple and easy to use, computationally efficient for large datasets.	Requires data to be frequency data and a sufficient sample size for the chi-square approximation to be valid.
SVM [21, 22]	For regression, classification and analysis.	Effective in handling high-dimensional data, memory efficient, versatile, and not sensitive to outliers.	Computationally expensive for large datasets, choice of the kernel can greatly affect performance, long training time for large datasets, and difficulty in interpreting the final model.
Logistic regression [23, 24]	A statistical method used for predicting the probability of a binary outcome.	Simple and interpretable, it provides probabilities for outcomes.	Assuming a linear relationship between the log odds of the dependent variable and the independent variables may not perform well with non-linear relationships.
Decision tree [25, 26]	For regression, classification and analysis.	1. It can be used for both classification and regression problems. 2. It can handle both categorical and numerical data. 3. It can be used for feature selection. 4. Easy to understand and interpret. 5. It can handle missing values.	1. Unstable in nature and affected by small changes in the data. 2. Time-consuming to build, especially with large datasets. 3. Prone to overfitting, especially with deep and complex trees.
Random forest [27]	1. At training time, it operates by constructing multiple decision trees. 2. use for regression and classification.	1. Robust and less likely to overfit the data. 2. High accuracy. 3. Fast and can handle large datasets. 3. Provides feature importance measures for feature selection and data understanding.	1. There are inferior gradient-boost tees on complex problems, impacting prediction accuracy. 2. Less interpretable than a single decision tree. 3. Suffers from overfitting in cases where the number of trees is very deep or very high.
Genetic algorithm feature selection [28]	Genetic algorithms can choose the right features to make a model more.	Capable of global optimization, even in complex.	The complexity and computational cost of genetic algorithms can be high.

3.1 Model description

For feature selection-based minimization problems, a binary representation has been achieved by the DTGA [29]. Through the implementation of this binary representation, any linear position can be encoded as a bit, which serves as the smallest unit of information recorded in a two-state computer. The potential position of the bit is located between "1" and "0" [30]. Since it can evolve and adjust in response to the fitness distribution of the feature subsets, the DTGA is regarded as the most effective method for selecting features in loan acceptance prediction. By dynamically adjusting the fitness threshold, it effectively demonstrates its capability to identify the most influential features and eliminate unnecessary ones.

Consequently, the computational complexity diminishes, and the feature selection process is optimized. Preventing premature convergence is facilitated by adaptability, which enables the identification of critical characteristics for loan approval decisions with greater precision and efficiency. In line with enhancing risk management procedures, the DTGA's dynamic characteristics also facilitate a deeper awareness of the variables influencing loan approval results. Consequently, the algorithm enables the companies to make an accurate and efficient loan decision [31].

The dynamic threshold of the proposal will increase the population diversity and exploration space of the genetic algorithm. A threshold is established for every individual in the population using this approach; the threshold is subsequently dynamically adjusted based on the individual's performance. The threshold can be increased if he/she (i.e., individual) exceeds its threshold. Consequently, this leads to exploring a larger portion of the search space. Otherwise, the

threshold of the individual will be decreased and limit the capability of exploring. DTGA maintains a diverse population and allows individuals to explore various parts of the search space through the dynamic adjustment of thresholds. This leads to improved performance and the generation of superior solutions.

In DTGA, we represent each gene by a single byte, making this technique highly adaptable and simple to comprehend. We represent the state of the operator by utilizing bits as follows.

$$q_j^t = [\beta_1^t | \beta_2^t | \ldots \ldots | \beta_m^t |] \tag{1}$$

where, m represents the gen index number of q_j^t which defines the chromosome of the jth individual and the tth generation. Instantly, a single individual can represent all states via bit encoding. Making DTGA to be more diverse than traditional genetic algorithms. Consequently, convergence can be achieved using the bit statement as well. The bit chromosome joins to one single state as β attitudes to 1 or 0 [31]. Figure 2 illustrates the pseudocode of DTGA, and Table 2 lists the parameters of the DTGA.

Pseudo Code

Input: Dataset T; No. of Generation r; No. of individuals j; Initial Populations Pop.

```
1-      while t < MAX_GENS do
2-        t ← t+1
3-        Pops ← bit_Encoding(Pops)
4-        Fitness_Vals=Fitness_Evaluation(Pops)
5-        Pop ← Selection_Best (Pops, Fitness_Vals)
6-        Update POPs using DT_Function
7-        If Termination_Condition is False, then
8-          for i ← 0 to (POP_SIZE − 1) do
9-            New_Pops (i) ←Crossover(Pop)
10-           New_Pops (i) ← Mutation(New_Pops)
11-         end
12-        Pops= New_Pops
13-     end
14-     If Termination_Condition is True, then
15-        Return Pops
16-     end
17-   end
18-     Best individual= Pops
```

Figure 2. DTGA pseudo code

Following this, each of the updated populations is measured. Subsequently, assess the adequacy of the solution set in conjunction with its binary solution. P(t) is previously designated and entrusted to subsequent generations following the classification accuracy of the present population.

By employing dynamic thresholding, the exploration space and diversity of a Classical Genetic Algorithm (CGA) can be significantly expanded. This method assigns a threshold value, represented by T(i), to each population member. The threshold value is determined through an analysis of the performance of each individual in the population using fitness or correlation coefficient values. In the case of rapid population evolution, the threshold value is raised by one through the augmentation of the population's one-gene count. If the fitness value remains constant or the population remains stable, the threshold value is reduced by one gene within the population. Cross-sectional and mutational genetic operators are employed to process further solutions that meet or surpass the threshold value.

Investigating different regions of the search space and enhancing diversity are the main aims of this process. The DTGA algorithm returns high-quality solutions in exchange for the following parameters: a dataset, dynamic thresholding function, the initial population size, the number of individuals and the number of generations.

Table 2. GA parameter examples

Parameter	Default Value
Threshold	[1, 2, 3, …., 11]
Mutation Ratio	0.3
Crossover Ratio	0.7
Population Size	50
Generation	10

3.2 Approach overview

DTGA is proposed to be utilized in loan decision-making to find the most informative features for predicting loan defaulters. This can aid in developing effective and more accurate algorithms for assessing creditworthiness and reducing the risk of default. The DTGA methodology involves multiple stages. In the beginning, we generate a random population of candidate feature subsets. Next, a fitness function is used to evaluate each subset. It uses accuracy as the fitness value that gauges its ability to predict loan defaulters in the database accurately.

Important loan characteristics include loan amount, employment status, debt-to-income ratio, income, and credit score. These are crucial in defining a loan application. Here, we explain the steps of our methodology.

B.1 Data description

This research uses loan application data; it contains information about borrowers' creditworthiness and ability to repay loans. The loan application data includes credit scores, income, debt-to-income ratios, employment status, loan amounts, and other relevant features. In the database, each is marked up by trained financial analysts, who have assessed the borrowers' creditworthiness and likelihood of defaulting on their loans. Various types of loan applications are listed in the database. In the loan application data, each record has a set of features that can be used to predict loan defaulters.

B.2 Approach description

Step 1: Data Preprocessing

Here, we preprocess the loan application data and extract the pertinent features. Features are normalized, missing values are credited, and outliers are removed to determine the feature set that can be utilized for classification.

Step 2: Population at the Outset

DTGA starts by extracting features set from the loan application data. Every member of the population is associated with a subset of the features. It is represented by a binary string in which each bit signifies whether a specific feature is excluded or contained in the subset. For instance, suppose there are ten features in the dataset. In that case, an individual may be denoted as a binary string of length 10, where a bit value of 0 indicates that the corresponding feature is excluded, and a bit value of 1 indicates that it is included in the subset.

Each subset corresponds to the loan decision-making attributes selected. For instance, if a subset contains the loan amount, credit score, employment status, and debt-to-income ratio, it is represented as a binary string with 1s in the corresponding positions. The population size is proportional to

the number of feature subsets considered. A larger population size permits the exploration of more diverse subsets during the GA search procedure. However, it also increases the algorithm's computational cost.

A random sample of prospective feature subsets is generated to initiate the search procedure of a genetic algorithm. As previously described, each subset is represented as a binary string. Based on past knowledge or experiments, we determine the size of the initial population.

Step 3: Evaluation of Fitness:

We use a fitness function dubbed "accuracy" to asses each feature subset candidate. Using the considered feature subsets, we implement the model (i.e., classifier) to the data results of the fitness function. Therefore, the evaluation determines how the classifier can accurately predict loan defaulters.

Step 4: Best Feature Subsets:

In loan approval prediction, the highest-performing feature subsets can be chosen from the initial population by the GA through tournament selection operators. For loan defaulters, the operators select the subsets with the highest predictive accuracy. Using randomly selecting subsets and rarely allowing weaker ones to prevail, tournament selection promotes diversity. This ensures a thorough exploration of the search space and prevents early convergence. By adjusting the tournament size, tournament selection can be designed to address particular loan decision-making challenges and fitness functions. In other words, this approach is appropriate for situations involving large populations or limited computational resources because it involves evaluating a small subset of the population during each tournament selection iteration. Accordingly, it can facilitate the development of more accurate and efficient algorithms for assessing creditworthiness and reducing default risk.

Step 5: Crossbreeding and Mutation

Additional candidate feature subsets are generated by utilizing the crossover and mutation operations of the genetic operator on the chosen subsets. The algorithm's crossover contains two sites exchanging subsequences between parent chromosomes to generate offspring. This type of crossover investigates and exploits the search space effectively. Each bit (or gene) in the offspring chromosomes has a minuscule chance of being flipped from 0 to 1 or vice versa. This form of mutation permits small modifications to the feature subsets, which can aid in exploring new regions of the search space and improve loan decision-making accuracy.

Step 6: Stopping criterion or accurate Loan decision-making.

We repeat steps 3, 4, and 5 until loan decision-making accuracy is adequate or a stopping criterion is met (e.g., the maximum generation number).

Step 7: Feature selection completion

Increasing population diversity and exploration space are two potential benefits of dynamic thresholding in Classical Genetic Algorithms (CGA), which involve assigning a threshold value to each loan application according to its fitness or correlation coefficient values. By employing crossover and mutational genetic operators to assess and modify high-performance solutions, one can augment the accuracy of loan decision-making. Precise loan decisions are generated as the algorithm concludes its execution or a termination criterion is fulfilled. Through subset selection, the final feature set is determined from the candidate loan decision-making solutions that exhibited the highest accuracy in the past iteration.

By mining loan data, DTGA can potentially improve the

accuracy of loan decisions.

To select loan features, however, challenges must be overcome, such as the requirement for large and diverse datasets, the standardization of data acquisition and processing, and the assurance of clinical relevance. To enhance the precision of loan decisions using DTGA, further examination of diverse parameter configurations and alternative genetic algorithms is required. Therefore, if financial institutions and researchers continue to work together, DTGA can become the superior tool for identifying high-risk loans and providing individualized service.

3.3 Simulation result and analysis

In our method, the scale factors procedure is used to normalize the feature values so that they all have equal weight in the classification model. We present experimental evidence supporting the effectiveness of the proposed method concerning the precision of loan decision-making. The output of the DTGA is compared to the output of the prescribed approach for traditional evaluations.

The framework is incorporated into user-specific applications by providing a MATLAB library. The evaluations were performed on a computer with 32GB RAM, two Intel(R) Zeon(R) CPU E5430@2.66GHz, and Microsoft Windows 10-64 bit. Based on the simulation results, it can be concluded that the proposed method can produce a precise and all-encompassing loan decision. 30% of the Kaggle dataset was evaluated by the DTGA algorithm using 30% of the Kaggle data set for testing purposes with a mutation ratio of 0.5, a crossover ratio of 0.5, and a threshold of 11.

By mining loan data, the DTGA algorithm offers an innovative method for identifying high-risk loans. DTGA offers an alternative approach to traditional research methodologies by employing a genetic algorithm to generate model populations via the evolution of random initial models. Table 2 illustrates the results of considering different combinations of GA parameters (e.g., threshold, mutation ratio, and the crossover ratio). Firstly, we calculated the following predicted measurements:

True positive (TP) refers to the corrected prediction of the approval.

True positive (TN) denotes the corrected prediction of the rejection of the approval.

True positive (FP) represents an uncorrected prediction of the approval.

True positive (FN) is the uncorrected prediction of the rejection of the approval.

Then, we determined the evaluation metrics: Accuracy, True Positive Rate (TPR), True Negative Rate (TNR), and F-score:

$$\text{F\_Score} = \frac{2TP}{2TP + FP + FN} \qquad (2)$$

$$\text{PPV} = \frac{TP}{TP + FP} \qquad (3)$$

$$\text{TNR} = \frac{TN}{TN + FP} \qquad (4)$$

$$\text{TPR} = \frac{TP}{TP + FN} \qquad (5)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (6)$$

Table 3. Performance of DTGA algorithm

Gn	Ps	Features Index	TP	TN	FP
110	95	[1,3,9,10,11]	178	89	1
115	10	[2,4,5,10,11]	178	89	1
120	20	[3,5,9,10,11]	178	89	1
125	30	[2,3,10,11]	178	89	1
125	100	[4,5,11]	178	89	1
130	5	[2,3,10,11]	178	89	1
5	20	[1,2,3,5,10,11]	177	88.5	2
5	25	[1,2,5,9,10,11]	177	88.5	2
5	40	[2,4,5,10,11]	177	88.5	2
5	100	[1,3,5,10,11]	177	88.5	2
10	5	[1,5,11]	177	88.5	2
10	20	[4,9,10,11]	177	88.5	2
5	10	[1,5,9,10,11]	176	88	3
5	50	[1,2,9,10,11]	176	88	3
5	55	[1,3,10,11]	176	88	3
5	65	[5,10,11]	176	88	3
5	80	[1,2,4,11]	176	88	3
10	10	[1,2,5,10,11]	176	88	3
10	50	[1,4,5,11]	176	88	3
10	60	[1,2,3,4,9,10,11]	176	88	3
10	85	[1,2,3,4,5,10,11]	176	88	3

Table 4. The accuracy of ML algorithms used in this study

Gn	Acc	TPR	TNR	PPV	F_score
1	99.26	0.99	0.99	0.99	0.99
1	99.26	0.99	0.99	0.99	0.99
1	99.26	0.99	0.99	0.99	0.99
1	99.26	0.99	0.99	0.99	0.99
1	99.26	0.99	0.99	0.99	0.99
1	99.26	0.99	0.99	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
2	98.52	0.99	0.98	0.99	0.99
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98
3	97.78	0.98	0.97	0.98	0.98

According to the GA parameters employed, Table 3 demonstrates the accuracy rates of the DTGA range of 97.78% to 99.26%. The results shown in this table confirm that the proposed method improves the accuracy of loan decision-making and emphasizes the importance of selecting GA parameters with care.

The performance of our proposed DTGA compared to previous studies is illustrated in Table 4. In contrast to the 67.41 to 83.73 percent accuracies of ML algorithms in k-nearest Neighbours, Decision Tree, Naive Bayes and Random Forest), the DTGA method achieves an outstanding 99.26 % accuracy. The outcomes of this analysis demonstrate that DTGA is exceptionally effective at identifying the most influential features that can be used to predict loan acceptance. As a result, decision-making is considerably more precise and accurate. The significant enhancement in accuracy attained through the implementation of DTGA highlights its capacity to transform feature selection and predictive modeling

regarding loan approval fundamentally. This presents a formidable and transformative resolution for financial institutions aiming to optimize lending practices and risk management. In Table 5 we have been compared current work against other ML algorithms

Table 5. Evaluation of the performance of our proposal against other ML algorithms

Reference	Algorithm	Accuracy
[29]	Random Forest	77.23%
[30]	Naive Bayes	83.73%
[31]	k-Nearest Neighbors	77.23%
Proposed Method	DTGA	99.26%

4. CONCLUSIONS

By identifying the most relevant features from a vast dataset, we have presented the DTGA algorithm in this paper as an efficient algorithm for improving the accuracy of loan decision-making. The potential implementation of the proposed method could aid financial institutions in improving the accuracy of their loan decisions, thereby resulting in enhanced risk management and financial performance. The DTGA algorithm achieved remarkable levels of precision, varying between 97.78% and 99.26%, by utilizing the most effective amalgamation of GA parameters. The results that have been presented offer empirical validation for the effectiveness of the suggested approach in detecting high-risk loans. Furthermore, they underscore the importance of accurately determining appropriate parameters for the genetic algorithm. The feasibility of employing DTGA-based feature selection for loan approval prediction cannot be ruled out. By implementing this strategy, substantial improvements can be achieved in prediction accuracy, computational efficiency, prevention of overfitting, and model interpretability.

Accordingly, the findings of this article contribute to the expanding corpus of knowledge concerning the application of DTGA-based feature selection in loan approval forecasting. In the future, one of our primary focuses is determining the applicability of DTGA-based feature selection in the context of loan approval prediction.

REFERENCES

[1] Srinivasa Rao, M., Sekhar, C., Bhattacharyya, D. (2021). Comparative analysis of machine learning models on loan risk analysis. In Machine Intelligence and Soft Computing: Proceedings of ICMISC 2020, pp. 81-90. Springer Singapore. https://doi.org/10.1007/978-981-15-9516-5_7

[2] Khan, A., Bhadola, E., Kumar, A., Singh, N. (2021). Loan approval prediction model: A comparative analysis. Advances and Applications in Mathematical Sciences, 20(3).

[3] Dagar, A. (2021). A comparative study on loan eligibility. International Journal of Scientific Research & Engineering Trends, 7(3): 1646-1649.

[4] Sobana, S., Ebenezer, P.J.L. (2022). A comparative study on machine learning algorithms for loan approval prediction analysis. International Research Journal of Modernization in Engineering Technology and Science.

[5] Supreeth S.A., Thanmai B.K., Varshini, K.S., Sukesh, S.B., Harish, K. (2022). A comparative study of machine learning algorithms for predicting loan default and eligibility. Perspectives in Communication, Embedded-systems and Signal-processing-PiCES, 5(12): 116-118. https://doi.org/10.5281/zenodo.6543983

[6] Spoorthi, B., Kumar, S.S., Rodrigues, A.P., Fernandes, R., Balaji, N. (2021). Comparative analysis of bank loan defaulter prediction using machine learning techniques. In 2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER) Nitte, India, pp. 24-29. https://doi.org/10.1109/DISCOVER52564.2021.9663662

[7] Mhlanga, D. (2021). Financial inclusion in emerging economies: The application of machine learning and artificial intelligence in credit risk assessment. International Journal of Financial Studies, 9(3): 39. https://doi.org/10.1109/AIARS59518.2023.00037

[8] Alonso Robisco, A., Carbo Martinez, J.M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. Financial Innovation, 8(1), 70. https://doi.org/10.1186/s40854-022-00366-1

[9] Tejaswini, J., Kavya, T.M., Ramya, R.D.N., Triveni, P.S., Maddumala, V.R. (2020). Accurate loan approval prediction based on machine learning approach. Journal of Engineering Science, 11(4): 523-532.

[10] Hamid, A.J., Ahmed, T.M. (2016). Developing prediction model of loan risk in banks using data mining. Machine Learning and Applications: An International Journal, 3(1): 1-9. https://doi.org/10.5121/mlaij.2016.3101

[11] Sawant, A.A., Chawan, P.M. (2013). Comparison of data mining techniques used for financial data analysis. International Journal of Emerging Technology and Advanced Engineering, 3(6): 112-116.

[12] Kala, K. (2018). A customized approach for risk evaluation and prediction based on data mining technique. International Journal of Engineering Research & Technology (IJERT), 3(33): 20-25.

[13] Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. Decision Support Systems, 37(4): 543-558. https://doi.org/10.1016/S0167-9236(03)00086-1

[14] Pandey, T.N., Jagadev, A.K., Mohapatra, S.K., Dehuri, S. (2017). Credit risk analysis using machine learning classifiers. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) Chennai, India, pp. 1850-1854. https://doi.org/10.1109/ICECDS.2017.8389769

[15] Attigeri, G.V., Pai, M.M., Pai, R.M. (2017). Credit risk assessment using machine learning algorithms. Advanced Science Letters, 23(4): 3649-3653. https://doi.org/10.1166/asl.2017.9018

[16] Kalaycı, S., Kamasak, M., Arslan, S. (2018). Credit risk analysis using machine learning algorithms. In 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, pp. 1-4.

[17] Saragih, M.G., Chin, J., Setyawasih, R., Nguyen, P.T., Shankar, K. (2019). Machine learning methods for analysis fraud credit card transaction. International Journal of Engineering and Advanced Technology, 8,

870-874. https://doi.org/10.35940/ijeat.F1164.0886S19

[18] Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M., Anderla, A. (2019). Credit card fraud detection-machine learning methods. In 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), Bosnia and Herzegovina, pp. 1-5. https://doi.org/10.1109/INFOTEH.2019.8717766

[19] Dornadula, V.N., Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. Procedia Computer Science, 165: 631-641. https://doi.org/10.1016/j.procs.2020.01.057

[20] Melendez, R. (2019). Credit Risk analysis applying machine learning classification models. In Intelligent Computing: Proceedings of the 2019 Computing Conference. https://doi.org/10.1007/978-3-030-22871-2_57

[21] Amin, M. M., Hassan, M., Aljomayli, A. (2018). Increasing the lifetime of wireless sensor networking depending on neighbors relationship. Journal of Electronic Systems, 8(3): 103-111. https://doi.org/10.6025/jes/2018/8/3/85-94

[22] Shoumo, S.Z.H., Dhruba, M.I.M., Hossain, S., Ghani, N. H., Arif, H., Islam, S. (2019). Application of machine learning in credit risk assessment: A prelude to smart banking. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON), Kochi, India, pp. 2023-2028. https://doi.org/10.1109/TENCON.2019.8929527

[23] Shaikh, R. (2018). Feature selection techniques in machine learning with python. Towards Data Science. https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e.

[24] Alcaraz, J., Labbé, M., Landete, M. (2022). Support Vector Machine with feature selection: A multiobjective approach. Expert Systems with Applications, 204: 117485. https://doi.org/10.1016/j.eswa.2022.117485

[25] Gárate-Escamila, A.K., El Hassani, A.H., Andrès, E. (2020). Classification models for heart disease prediction using feature selection and PCA. Informatics in Medicine Unlocked, 19: 100330. https://doi.org/10.1016/j.imu.2020.100330

[26] Vikas, P. K., Kaur, P. (2021). Lung cancer detection using chi-square feature selection and support vector machine algorithm. International Journal of Advanced Trends in Computer Science and Engineering, 10(3). https://doi.org/10.30534/ijatcse/2021/801032021

[27] Deng, N., Tian, Y., Zhang, C. (2012). Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions. CRC Press.

[28] Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M. (2002). Logistic regression (536 p.). Springer-Verlag. https://doi.org/10.1161/CIRCULATIONAHA.106.682658

[29] Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6): 275-285. https://doi.org/10.1002/cem.873

[30] Victor, L., Raheem, M. (2021). Loan default prediction using genetic algorithm: A study within peer-to-peer lending communities. International Journal of Innovative Science and Research Technology, 6(3): 2456-2165.

[31] Pernkopf, F., O'Leary, P. (2001). Feature selection for classification using genetic algorithms with a novel

encoding. In Computer Analysis of Images and Patterns: 9th International Conference, CAIP 2001 Warsaw, Poland, pp. 161-168. https://doi.org/10.1007/3-540-44692-3_20