第 48 卷 第 10 期
2021 年 10 月

湖南大学学报（自然科学版）
Journal of Hunan University（Natural Sciences）

Vol. 48 No. 10
October 2021

Open Access Article

# Machine Learning-Based Prediction Model for Loan Status Approval

**Suliman Mohamed Fati**

College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

**Abstract:** Loan approval in financial organizations is one of the challenges that affect the operational financial process due to the inaccurate estimation or the lack of information. Thus, the banks aim to minimize the credit risks by assessing the loan status through an intensive evaluation process to avoid unforeseen issues. Therefore, loan prediction based on the given and collected information is very important in this regard. Data mining, particularly Machine learning, is a promising direction to give accurate and on-time decisions to approve/disapprove the loans. The main goal of this work is to investigate the loan prediction process by applying different machine learning algorithms. The proposed methodology starts with data pre-processing to clean the data, remove outliers, and find the correlation between the features to find the most noteworthy feature. Then, three machine-learning algorithms will be trained and tested: Logistic Regression, Decision Tree, and Random Forest. The novelty of this research can be represented by comparing three machine-learning algorithms to find the most accurate prediction. The experimental results showed the superiority of Logistic Regression on the other two algorithms in terms of accuracy precision, Recall, F1, and Area under the curve (AUC). The decision tree algorithms also underwent Receiver operating characteristic (ROC), which demonstrated the ability of Logistic Regression to predict the loan status under different thresholds.

**Keywords:** loan approval, machine learning algorithm, logistic regression, data mining, prediction model.

## 基於機器學習的貸款狀態批准預測模型

**摘要:** 由於估計不准確或信息缺乏，金融機構的貸款審批是影響整個財務運作流程的挑戰之一。因此，銀行的目標是通過密集的評估過程評估貸款狀況，以避免出現任何不可預見的問題，從而將信用風險降至最低。因此，基於給定和收集信息的貸款預測在這方面非常重要。數據挖掘，尤其是機器學習，是一個很有前途的方向，可以準確及時地決定批准/不批准貸款。這項工作的主要目標是通過應用不同的機器學習算法來研究貸款預測過程。所提出的方法從數據預處理開始，以清理數據、去除異常值，然後找到特徵之間的相關性以找到最具影響力的特徵。然後，將訓練和測試三種機器學習算法，它們是邏輯回歸、決策樹和隨機森林。這項研究的新穎性可以通過比較三種機器學習算法來找到最準確的預測來體現。實驗結果表明，邏輯回歸在準確率、記起、1 和曲線下面積（澳大利亚大学）等方面均優於其他兩種算法。決策樹算法還通過了接收者操作特徵（鵬），這證明了邏輯回歸預測不同閾值下的貸款狀態的能力。

**关键词：** 貸款審批, 機器學習算法, 邏輯回歸, 數據挖掘, 預測模型。

## 1. Introduction

In the banking industry, loans represent important financial transactions that contribute to the banks' success. In addition, the banks earn their main profit from the distribution of loans, which take the shape of the bank's assets. The primary objective of every bank is to invest its assets in safe hands. Thus, the banks aim to minimize the credit risks by assessing the loan status

Received: June 25, 2021 / Revised: August 26, 2021 / Accepted: September 28, 2021 / Published: October 30, 2021

About the author: Suliman Mohamed Fati, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia

Corresponding author Suliman Mohamed Fati, smfati@yahoo.com

through an intensive evaluation process to avoid any illegal or unexpected events that prevent the lender from fulfilling their obligations. Although banks run a validation and verification mechanism for the approval of a loan, there is no surety that the selected applicant deserves right or not. As one of the options to overcome this issue, the banks may collaborate to share the needed information that helps in deciding the borrower's eligibility for a loan. To achieve this goal, collecting and analyzing customers' historical data from different financial entities is a good way to get accurate results and minimize credit risk.

Therefore, data mining techniques are very useful in this field. Data mining is a process of extracting meaningful knowledge from a large and unstructured amount of data [1]. Such extracted information can help in decision-making through performing important operations. In literature, there are many traditional data mining methods for loan prediction.

Out of the data-mining field, machine-learning algorithms are available to automate the loan status prediction system. For instance, to predict the right person who deserves the loan, many machine learning algorithms, such as decision tree, gradient boosting, random forest, logistic regression, and many more, can predict the right person who deserves the loan. Therefore, this paper aims at minimizing the credit risk by predicting the loan status based on different loan factors or features.

Particularly, a dataset with 13 features was used in this research to predict whether the applicant deserves the loan or not. This loan prediction problem is a binary classification to get the Loan_Status value as yes or no. Three machine algorithms were trained and tested (i.e., Logistic Regression, Decision Tree, and Random Forest) on the dataset. The pre-processing tasks were performed, including exploratory data analysis to understand the attributes, relationship between attributes, find missing values and remove them, outlier detection, and removal.

The rest of the paper is structured as follows. Section 2 discusses the background and related works; section three demonstrates the dataset used and the implementation details, and section four records and discusses the results.

## 2. Literature Review

Data mining techniques are becoming very popular today and applied in various fields, including bioinformatics, retail industry, intrusion detection, logical data analysis, and banking. Particularly, using data mining techniques in banking helps meet the competition's needs by enhancing the data analysis to make an accurate decision. Thus, authors in [2] have discussed this issue and introduced the usage of different data mining techniques, such as Random Forest, Decision Tree, Bayes classification, Bagging, and Boosting, in financial data analysis and decision-making. The authors [2] proposed loan default risk analysis based on scoring that enables the bank to decrease manual errors.

According to [2], the banks prefer decision trees in credit risk analysis as it is a white-box model showing the information transparency and is easy to understand and trace. In addition, the efficiency of the decision tree can be improved using Boosting. Although the authors in [2] introduced a good explanation for the used algorithms, the work lacks the experimental results, in-depth analysis, or model evaluation. To evaluate the used classification models, the authors in [3] proposed a detailed evaluation framework for the loan classification models used by banks. The proposed evaluation model is based on multiple criteria decision-making (MCDM), whereby the performance of the different models will be evaluated against a set of performance metrics to select the high-rank prediction model.

Furthermore, many case studies were discussed in the literature on applying the data mining-based loan prediction in different countries. For example, the authors in [4] discuss the housing loan approval of a private bank in Turkey. They used a logistic regression model; the authors utilized uncorrelated components rather than original correlated variables to revert the binary response variable. Another work [5] used a decision tree model to calculate the rating for bank customers to reduce the credit risks in Mellat Bank of Iran. The authors [5] introduced a genetic algorithm (GA) for better feature selection.

One of the issues in the proposed models is the performance. Thus, the work [8] focuses on comparing the performance of different prediction models using a data-driven approach. The authors conducted a comparison for three algorithms named Decision Tree (DT), Artificial Neural Network (ANN), and support vector machine (SVM) using 10-fold cross-validation. The results showed that SVM provided higher accuracy (72.05%) than other models with very slight improvement. However, the performance still needs improvement, as per their discussion. Similar work is proposed in [19], whereby the decision tree and k-fold were used to split the loan applicants into different groups based on the most differentiator variables.

Another work in [9] proposed a prediction model by comparing j48, bayesNet, and Naive Bayes algorithms. The highest accuracy that they achieve is 78%. Moreover, the authors in [10] evaluated around 15 machine-learning models to find the most suitable algorithm for loan prediction. Their experimental results showed that logistic regression provided the highest accuracy among its counterparts with 81% in a dataset with three selected features.

Similar work was found in [15], whereby the logistic regression was used and tested on independent (non-correlated) features. Therefore, the authors in [7] used Fuzzy Logic to automate the loan approval process, and they found that Fuzzy logic is more efficient as it reduces the processing time, which improves the satisfaction for both bank staff and clients.

Alternatively, in [6], the authors combined different homogeneous big data models in a smart ubiquitous data mining (UDM) methodology whereby various rule extraction algorithms were used to design a meta-model. The authors in [6] claimed that such integration is superior to the individual models. Another framework was proposed in [11] whereby machine learning was used to validate loan applicants' eligibility to minimize the credit risk factors. The risk factor's weight was calculated and tuned to show better performance. However, for both works, there are no published results that support the claims. Likewise, the authors in [12] proposed a prediction model to decide the loan approval based on some factors such as occupation, financial position, and family background.

The authors used two algorithms named Random Forest (RF) and linear regression (LR). The published results showed that the RF algorithm provided accuracy with 0.79 during testing (0.79), which is slightly higher than LR accuracy (i.e., 0.76).

Another work presented in [16] used the random forest to build a loan prediction model whereby the features were analyzed to prove that credit history is the most important feature that influences the loan approval. The achieved accuracy for the work in [16] was 81%. Another work in [21] used a gradient boosting algorithm to build a loan prediction model. The proposed model achieved an accuracy of 0.79 with two discriminating features: geographic location and applicant age.

On the other hand, the authors in [18] used fuzzy rules to analyze the loan factors such as job status, weight, and income source on a realistic dataset obtained from the bank of England. The authors compared their works with different works and achieved an accuracy of 83% on the given dataset. [17] followed another direction whereby the authors analyzed the descriptive text of the loan to extract some of the delicate features (e.g., Part-of-Speech features, sentiment features, and social relationship information) that can be integrated with the standard features to improve the prediction accuracy. The delicate features were extracted using Latent Dirichlet Allocation (LDA).

In contrast, the authors in [13] used the Gradient Boost Ensemble algorithm to improve the accuracy, and they showed a better accuracy with 81% for the German language-based dataset. Likewise, the authors in [14] used different machine learning algorithms to forecast mortgage loans. The used algorithms are Support-Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Factorization Machines (FM). The last one showed better performance as a nonparametric algorithm.

Besides, the authors in [20] applied LightGBM, XGBoost, Random Forest, and Logistic Regression algorithms to build a binary prediction model that classifies the applicants to good payers and bad payers. In this model, Random Forest showed better performance with an AUC value of 0.89 and an accuracy value of 0.88.

# 3. Proposed Loan Prediction Methodology

As depicted in Figure 1, the proposed mythology starts with the collection, whereby we have used a public dataset for loan prediction from Kaggle. Then, in the data exploration phase, we tried to understand the data and relationships between features through different types of tables and figures such as box plots, bar plots, histograms, heat maps, etc. The third phase is to train three classification algorithms on the explored data based on the best-found variables that may be helpful in the prediction process.

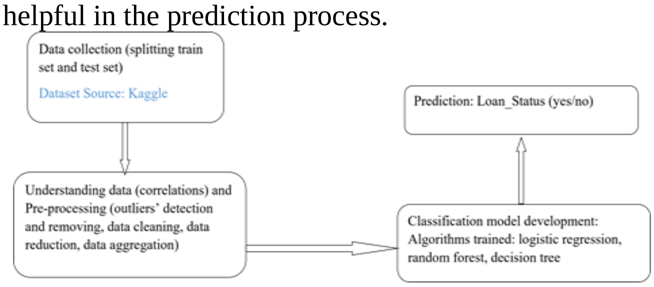


Fig. 1 The proposed loan prediction methodology

## 3.1. Dataset Description and Preprocessing

In order to validate the proposed model, we have used a public dataset for loan prediction from Kaggle, as explained earlier, to benchmark the results with other models. The used dataset consists of 13 features with two sets, a training set with 615 rows, and the reaming 368 rows for the testing purpose. We have loaded the data of the train and test part into the Google collab notebook by mounting the drive. Table.1 depicts the features of the dataset associated with their proper descriptions.

Table 1 Dataset introduction

| Variable | Data-type | Description |
|---|---|---|
| Loan_ID | Categorical (non-numeric) | Unique loan ID |
| Gender | Categorical (non-numeric) | Male/female |
| Married | Categorical (numeric) | Yes/no |
| Dependents | Categorical (numeric) | No. of dependents |
| Education | Categorical (non-numeric) | Graduate/not-graduate |
| Self_employed | Categorical (non-numeric) | Yes/no |
| Applicant Income | Numeric feature | Applicant income |
| CoapplicantInc | Numeric feature | Co-applicant |

| | | |
|---|---|---|
| ome | | income |
| LoanAmount | Numeric feature | Loan amount in thousands |
| Loan_Amount _Term | Numeric feature | Term of the loan in months |
| Credit_History | Categorical (numeric) | 0/1 |
| Property_Area | Categorical (non-numeric) | Urban/semi-urban/rural |
| Loan_Status | Categorical (non-numeric) | Yes/no |

### 3.2. Data Pre-Processing

The first step in data pre-processing is to check if any missing values may affect the prediction, so the heat map was used. As depicted in Figure.2, the heat map shows no multiple missing values in any features. Thus, there is no need to drop any feature except Loan_ID, as it has no impact on the target variable. It is worth mentioning that a feature is dropped if it has a lot of missing values in it compared to its length, as it does not add any valuable information to the dataset.

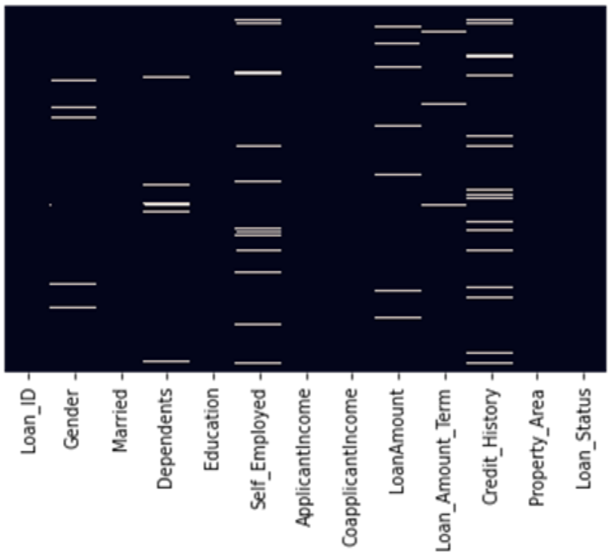


Fig. 2 Heatmap technique for missing feature values discovery

For the few missing values, we will use the Mean Imputation technique to estimate the missing values. For example, in the gender feature. As depicted in Figure3, there are 489 males and 112 females in the dataset. In addition, we found 13 missing values in the train part and 11 missing values in the test part.

By calculating the correlation between gender and Loan_status, we find that Males have a high correlation, as shown in Table 2. Thus, the missing values were filled with the male category. The gender values were converted to numeric so that the model works properly.

Table 2 Correlation between gender and loan-status

| | Gender | Loan-status |
|---|---|---|
| 0 | Female | 0.669643 |
| 1 | Male | 0.693252 |

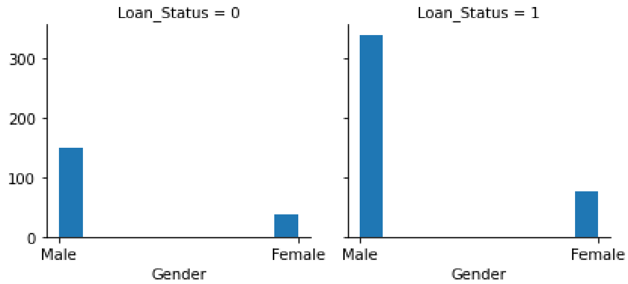


Fig. 3 Relation between gender and loan-status

Likewise, as depicted in Table 3 and Figure 4, 398 persons are married, 213 persons are unmarried, and only three are missing values. Married values have a high correlation in the married feature. According to Figure 4, if anyone is married, he/she would be more likely to have a loan. Thus, the missing values were filled with "married" values.

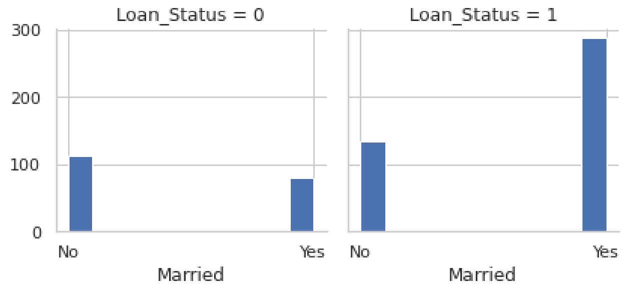


Fig. 4 Relation between married and loan-status

Table 3 Relation between married and loan-status

| | Married | Loan_Status |
|---|---|---|
| 0 | No | 0.629108 |
| 1 | Yes | 0.718204 |

The other features, as education (graduated or non-graduated), several dependents, and property areas, were tested for the missing values and treated the same way. According to figure 5, if a female is married and has not graduated, she would be more likely to loan. If any female is not married, she still would be more likely to have a loan than a male (zero is for female and one is for male).

The next step after checking the missing values is to find any outliers in the collected data. Thus, we use the box plots technique to find the outliers, as depicted in Figure 8. The box plots figure shows that there are four features with outliers, which are Applicant Income, Coapplicant Income, Loan Amount, and Loan-Amount-Term. Thus, the detected outliers were removed using the univariate method.

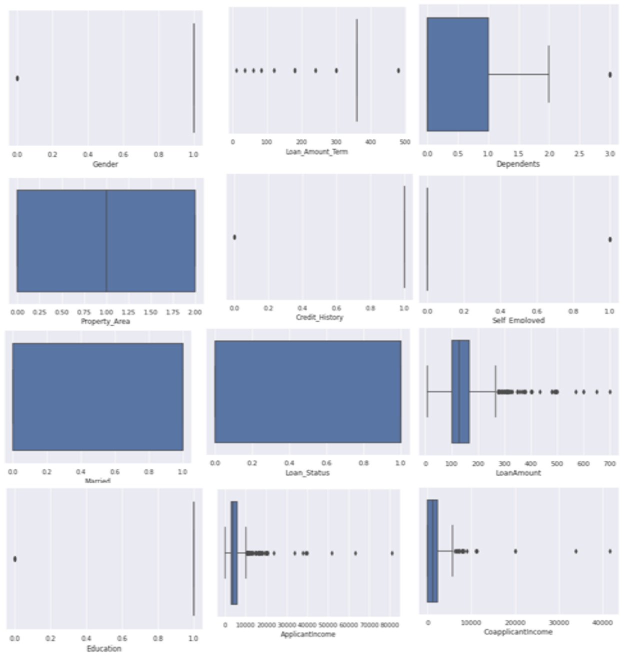

Fig. 5 Outlier detection using box plots techniques

The last step in the pre-processing is to test the correlation between data attributes to find the most noteworthy feature in the prediction process. For this purpose, we use a heat map to visualize the correlation of the variables. Figure 9 depicts the heat map for the collected data attributes. From the heat map in figure 9, we can easily notice the most important feature for loan prediction. Notably, Loan_ID has been removed from the heat map, as it has no impact on the prediction process.

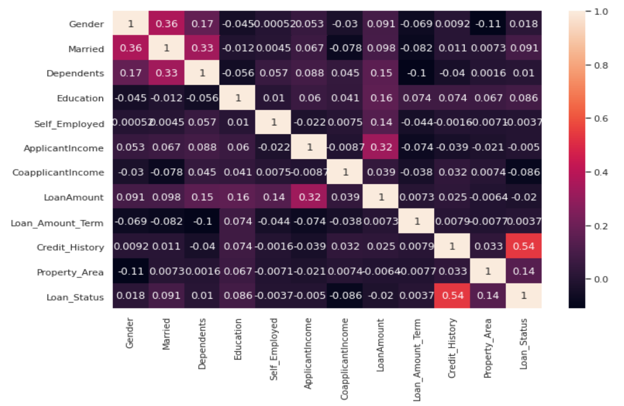


Fig. 6 Representing the correlation between attributes using the heat map

### 3.3. The Proposed Prediction Module

After performing pre-processing, we have trained and tested three algorithms on the data. The first algorithm is Logistic regression. Logistic regression is a statistical technique that is used to predict the probability of binary response variables. It is used when our label(y) is a binary response variable in 1 or 0, yes or no, etc. It is a basic and popular algorithm to solve a classification problem. The second algorithm is a Decision Tree classifier. A decision tree is a popular algorithm that is used to build classification models. The models are built in the form of a tree-like structure. Each node in the tree indicates a test on a variable, and each branch descending from that node indicates one of the possible values for that attribute. The third algorithm is Random Forest Classifier. Random Forest is a classification algorithm that consists of many decision trees. It tries to create an uncorrelated forest of trees by using feature randomness and Bagging. The prediction of an uncorrelated forest of trees is more accurate than that of any individual tree.

## 4. Results and Discussion

This section records the proposed prediction model's experimental results using the three machine algorithms (logistic regression, decision tree, and random forest). Table 4 shows the summary results of the three algorithms.

Table 4 Summary results of the three machine learning algorithms

| Algorithms | Precision | Recall | F1 | Accuracy | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.79 | 0.98 | 0.88 | 0.91 | 0.80 |
| Decision tree | 0.77 | 0.83 | 0.80 | 0.82 | 0.75 |
| Random forest | 0.78 | 0.93 | 0.84 | 0.86 | 0.79 |

The performance metrics used are precision, Recall, F1, Accuracy, and Area under the curve (AUC). AUC (Area under Curve) tells us about the measure of performance across all possible classification thresholds. The higher value of AUC indicates the better model at distinguishing that loan should be approved or not. As shown in Table 4, logistic regression has high results than the other two algorithms.

Furthermore, Receiver operating characteristic (ROC) has been used to show the diagnostic ability of binary classifiers based on AUC values. Particularly, the prediction model that is closer to the top-left corner has better performance than others do. Figure 10 depicts the ROC for the three models and shows that linear regression performs better than the others in different classifications of thresholds do.

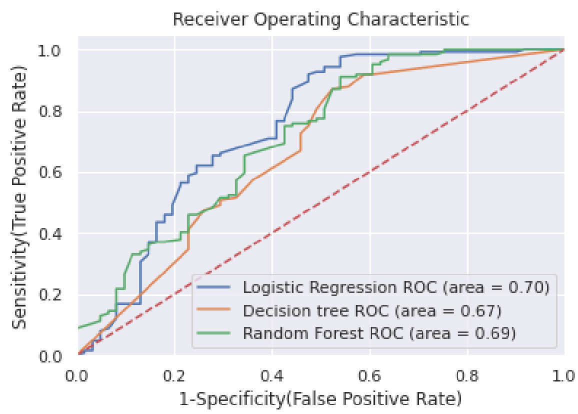


Figure 7 Performance comparison of the different prediction algorithms using ROC

### 4.1. Comparison with Related Works

This section compares the accuracy values and AUC values of different models discussed in the literature review sections. Notably, some authors have not given exact calculations of some metrics.

Table 5 Result comparison

| Models | Accuracy | AUC (area under the curve) |
|---|---|---|
| [3] | 83% with PCA dataset and 78% with ICA dataset | 0.86 with PCA dataset and 0.78 with ICA dataset |
| [4] | 91.1% | - |
| [5] | 88% | - |
| [8] | 72.05% | - |
| [9] | 78% | - |
| [10] | 81% | - |
| [12] | - | 0.79 |
| [13] | 88% on Australian dataset and 81 %in German dataset | - |
| [14] | - | 0.91 |
| [16] | 81% | - |
| [17] | 86% | 0.84 |
| [18] | 83% | - |
| [20] | 88% | - |
| [21] | 79% | - |
| The proposed model | LR:91%<br>DT: 82%<br>RF:86% | LR:0.8<br>DT:0.75<br>RF:0.79 |

[4] have used principal components as the explanatory variables, so they improved accuracy. [5] achieved 88% accuracy because they have used a genetic algorithm to select features on the classification of algorithms. [13] have improved the performance of a single regression classification using boosting method with gradient boosting algorithm. [17] used soft features extracted from descriptive loan text, which are more related to loan prediction, achieving the best accuracy. [20] used sampling scenarios to ensure the data was balanced.

Some of the papers have small accuracy values than the model proposed in this paper. We have analyzed every feature in the dataset used separately, filled missing values, removed outliers, and created relationships between variables. Credit_History was the attribute with the best correlation value with Loan_Status.

## 5. Conclusion

Loan application processing is one of the main tasks for banks. Many approaches are proposed in the literature for loan prediction. Among those approaches, machine-learning algorithms are proposed to predict the loan status based on different factors and parameters. Therefore, in this paper, we have trained and tested a dataset to predict loan approval. There are 13 features in the dataset, and we found that Credit_History is the most important feature for the prediction of loans. The pre-processing starts with data understanding, data cleaning, outlier detection, and removal. Three machine-learning algorithms were trained and tested in the proposed prediction model on the data: linear regression, decision tree, and Random Forest. Logistic regression showed better performance than the others with 81% accuracy, while Decision tree and Random Forest got 72% and 76% accuracy, respectively. The recorded results have been validated using the ROC curve. In addition, the proposed model was compared with the related works. The possible future directions of this work will be to acquire a realistic dataset with more prediction features to improve the prediction. Besides, the prediction accuracy must be improved. Such improvement can be achieved through applying feature extraction applying hybrid machine learning algorithms.


## Acknowledgements

The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.


## References


[1] WITTEN, H., and FRANK, E. *Data mining practical machine learning tools and techniques with Java implementations*. Elsevier, 2017.

[2] ZHOU, P.Y., CHAN, K.C., OU, C.X., and CHAWAN, P.M. Corporate communication network and stock price movements: insights from data mining. *IEEE Transactions on Computational Social Systems*, 2018, 5(2): 391-402.

[3] YONGMING, S., and PENG, Y. A MCDM-based evaluation approach for imbalanced classification methods in financial risk prediction. *IEEE Access*, 2019, 7: 84897-84906.

[4] KEMALBAY, G., and KORKMAZOĞLU, Ö.B. Categorical principal component logistic regression: a case study for housing loan approval. *Procedia-Social Behavioral Science*, 2017, 109: 730–736.

[5] ZAMANI, S., and MOGADDAM, A. Natural Customer Ranking of Banks in Terms of Credit Risk by Using Data Mining: A Case Study: Branches of Mellat Bank of Iran. *Journal of UMP Social Science Technology Management*, 2016, 3(2).

[6] BAE, J.K., and KIM, J. A personal credit rating prediction model using data mining in ubiquitous smart environments. *International Journal of Distributed Sensor Networks*, 2018, 11(9): 179060.

[7] KUMAR, B., BAWANE, I., SHIRSATHE, A., and PARDESHI, P. An Expert System Based On Fuzzy Logic for Automated Decision Making For Loan Approval. 2016.

[8] JIN, Y., and ZHU, Y. A data-driven approach to

predicting loan default risk for online peer-to-peer (P2P) lending. *2015 Fifth International Conference on Communication Systems and Network Technologies*, 2017, 609–613.

[9] HAMID, A.J., and AHMED, T.M. Developing prediction model of loan risk in banks using data mining. *Machine Learning Appliances an International Journal*, 2016, 3(1).

[10] TURKSON, R.E., BAAGYERE, E.Y., and WENYA, G.E. A machine learning approach for predicting bank creditworthiness. *2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR)*, 2016, 1–7.

[11] ARUN, K., ISHAN, G., and SANMEET, K. Loan Approval Prediction based on Machine Learning Approach. *IOSR Journal of Computing Engineering*, 2016, 18(3): 18–21.

[12] GAHLAUT, A., and SINGH, P.K. Prediction analysis of risky credit using Data mining classification models. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017, 1–7.

[13] LAWI, A., AZIZ, F., and SYARIF, S. Ensemble GradientBoost for increasing classification accuracy of credit scoring. *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, 2017, 1–4.

[14] BAGHERPOUR, A. *Predicting mortgage loan default with machine learning methods*. University of California/Riverside, 2017.

[15] VAIDYA, A. Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval. *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017, 1–6.

[16] PRIYA, K.U., PUSHPA, S., KALAIVANI, K., and SARTIHA, A. Exploratory Analysis on Prediction of Loan Privilege for Customers using Random Forest. *International Journal of Engineering Technology*, 2018, 7(2.21): 339–341.

[17] JIANG, C., WANG, Z., WANG, R., and DING, Y. Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annual Operation Resources*, 2018, 266(1–2): 511–529.

[18] TANEJA, S., SURI, B., GUPTA, S., NARWAL, H., JAIN, A., and KATHURIA, A. A fuzzy logic-based approach for data classification. *Data engineering and intelligent computing*. Springer, 2018.

[19] YADAV, O., SONI, C., KANDAKATLA, S., and SAWANT, S. *Loan Prediction System Using Decision Tree*.

[20] COŞER, A., MAER-MATEI, M.M., and ALBU, C. Predictive Models for Loan Default Risk Assessment. *Economical Computer: Economic Cybernetics Study Resources*, 2019, 53(2).

[21] ODEGUA, R. Predicting Bank Loan Default with Extreme Gradient Boosting. 2020, *arXiv Prepr. arXiv2002.02011*.

**参考文:**

[1] WITTEN, H., 和 FRANK, E. 數據挖掘：具有 爪哇實現的實用機器學習工具和技術。愛思唯爾，2017 年。

[2] ZHOU, P.Y., CHAN, K.C., OU, C.X., 和 CHAWAN, P.M. 企業通訊網絡與股價走勢：來自數據挖掘的洞察。計算社會系統彙刊，2018 年，5(2)：391-402。

[3] YONGMING, S., 和 PENG, Y. 基於的金融風險預測中不平衡分類方法的評估方法。訪問，2019，7：84897-84906。

[4] KEMALBAY, G., 和 KORKMAZOĞLU, Ö.B. 分類主成分邏輯回歸：住房貸款審批案例研究。继续-社會行為科學，2017，109：730-736。

[5] ZAMANI, S., 和 MOGADDAM, A. 使用數據挖掘在信用風險方面對銀行的自然客戶排名：案例研究：伊朗梅拉特銀行分行。聯合醫務社科技術管理學報, 2016, 3(2).

[6] BAE, J.K., 和 KIM, J. 在智能泛在環境中使用數據挖掘的個人信用評級預測模型。國際分佈式傳感器網絡雜誌, 2018, 11(9): 179060。

[7] KUMAR, B.、BAWANE, I.、SHIRSSATHE, A., 和 PARDESHI, P. 基於模糊邏輯的貸款審批自動化決策專家系統。2016 年。

[8] JIN, Y., 和 ZHU, Y. 一種預測在線貸款違約風險的數據驅動方法。2015 第五屆通信系統和網絡技術國際會議，2017，609-613。

[9] HAMID, A.J., 和 AHMED, T.M.使用數據挖掘開發銀行貸款風險預測模型。機器學習設備國際期刊，2016 年，3(1)。

[10] TURKSON, R.E.、BAAGYERE, E.Y., 和 WENYA, G.E. 一種預測銀行信用價值的機器學習方法。2016 第三屆人工智能與模式識別國際會議，2016，1-7。

[11] ARUN, K., ISHAN, G., 和 SANMEET, K. 基於機器學習方法的貸款審批預測。計算工程雜誌, 2016, 18(3): 18–21。

[12] GAHLAUT, A., 和 SINGH, P.K. 使用數據挖掘分類模型對風險信用進行預測分析。 2017 第八屆計算、通信和網絡技術國際會議，2017，1-7。

[13] LAWI, A.、AZIZ, F., 和 SYARIF, S. 用於提高信用評分的分類準確性。2017 第四屆計算機應用與信息處理技術國際會議, 2017, 1-4.

[14] BAGHERPOUR, A. 用機器學習方法預測抵押貸款違約。加州大學/河濱分校，2017 年。

[15] VAIDYA, A. 使用邏輯回歸的預測和概率方法：貸款批准預測的應用。 2017 第八屆計算、通信和網絡技術國際會議，2017，1-6。

[16] PRIYA, K.U.、PUSHPA, S.、KALAIVANI, K., 和 SARTIHA, A. 使用隨機森林預測客戶貸款特權的探索性分析。國際工程技術雜誌, 2018, 7(2.21): 339–341。

[17] JIANG, C., WANG, Z., WANG, R., 和 DING, Y. 通過結合從在線借貸中的描述性文本中提取的軟信息進行貸款違約預測。年度運營資源， 2018， 266(1-2)：511-529。

[18] TANEJA, S., SURI, B., GUPTA, S., NARWAL, H., JAIN, A., 和 KATHURIA, A. 基於模糊邏輯的數據分類方法。數據工程與智能計算。斯普林格， 2018 年。

[19] YADAV, O., SONI, C., KANDAKATLA, S., 和 SAWANT, S. 使用決策樹的貸款預測系統。

[20] COŞER, A.、MAER-MATEI, M.M., 和 ALBU, C. 貸款違約風險評估的預測模型。經濟計算機：經濟控制論研究資源， 2019， 53(2)。

[21] ODEGUA, R. 使用極端梯度提升預測銀行貸款違約。 2020 年， 阿尔西准备, 阿尔西 2002.02011。