

Risk Assessment Approaches in Banking Sector –A Survey

Mona Sharaf

Faculty of Commerce and Business Administration, Future University in Egypt, Egypt,
Mona.Aly21@commerce.helwan.edu.eg

shimaa mohamed ouf

BIS helwan university, shimaaouf@yahoo.com

Amira M. Idrees AMI

Future University in Egypt, amira.mohamed@fue.edu.eg

Follow this and additional works at: <https://digitalcommons.aaru.edu.jo/fcij>



Part of the [Computer and Systems Architecture Commons](#), [Data Storage Systems Commons](#), and the [Technology and Innovation Commons](#)

Recommended Citation

Sharaf, Mona; ouf, shimaa mohamed; and Idrees, Amira M. AMI () "Risk Assessment Approaches in Banking Sector –A Survey," *Future Computing and Informatics Journal*: Vol. 8: Iss. 1, Article 3. Available at: <https://digitalcommons.aaru.edu.jo/fcij/vol8/iss1/3>

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Future Computing and Informatics Journal by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact rakan@aarj.edu.jo, marah@aarj.edu.jo, u.murad@aarj.edu.jo.



Risk Assessment Approaches in Banking Sector –A Survey

Mona Sharaf Eldin

Faculty of Commerce and Business Administration, Future University in Egypt, Egypt

Mona.Aly21@commerce.helwan.edu.eg

Shimaa M. Ouf

Faculty of Commerce and Business Administration, Future University in Egypt, Egypt

shimaaouf@yahoo.com

Amira M. Idrees

Faculty of Computers and Information Technology, Future University in Egypt, Egypt

amira.mohamed@fue.edu.eg

Abstract: – Prediction analysis is a method that makes predictions based on the data currently available. Bank loans come with a lot of risks to both the bank and the borrowers. One of the most exciting and important areas of research is data mining, which aims to extract information from vast amounts of accumulated data sets. The loan process is one of the key processes for the banking industry, and this paper examines various prior studies that used data mining techniques to extract all served entities and attributes necessary for analytical purposes, categorize these attributes, and forecast the future of their business using historical data, using a model, banks' business, and strategic goals.

Keywords: Prediction , Classification , Data Mining , Banking Sector

1. Introduction

Data mining is used to collect and prioritize valuable information [1] [2], data mining

involves combining modeling techniques with statistical, artificial intelligence, and machine learning techniques [3] [4]. The combination of these tools ensures that banks can handle a

variety of situations using unsupervised and supervised learning and decision-making expertise. In connection with this, learning and data processing algorithms were developed to achieve the goals of the economic organization, especially in the collection and collection of information. Data mining is now used in financial risk management to provide profiles, trends and real-time insights [5] [6]. Data mining techniques help distinguish between borrowers who repay their loans on time and those who do not. It also helps predict when a borrower will default and whether lending to a particular customer will result in bad credit. Data mining can be used to analyze all banking processes to identify customer behavior. It also helps to determine whether a customer pays on time or late [7] [8], data mining uses many useful techniques to predict statistical data in various business applications. One of the most used techniques with different algorithms is classification. This paper reviews different classification algorithms in terms of accuracy in different areas of data mining applications [9].

After 20 delegated readings of the literature, a comprehensive analysis is prepared. The purpose of this paper is to help data analysts choose the best classification algorithm for various business applications, such as credit forecasting, bankruptcy forecasting, and credit card fraud detection.

According to the results, the most accurate classification algorithms are Support Vector Machine and Random Forest. The Naive Bayes algorithm is the most accurate for agricultural classification.

2. Literature Review

There are several articles on various applications of data mining in banking. In order to diversify the topics of DM applications [10], we selected articles that dealt with various aspects of data mining in banking. These are related to the detection and identification of credit risk assessment to

reduce financial risks in financial services banking caused by financial fraud such as credit cards, insurance fraud and SME bankruptcy forecasting models, loan failure forecasts, customer analysis

2.1- Loan forecasting and credit risk assessment in financial institutions

Mehul et. al. (2021) [11], the proposed results show that banks can improve default prediction models by applying machine learning techniques to a database of 124624 consumer loans with 2 -month maturities from a large Brazilian financial institution. A delay in loan repayment of more than two months indicates a failure to measure efficiency; the performance measure based on AUC was AdaBoost followed by Random Forest and SVM-RBF.

On the other hand, Deniz, Begu (2019) [12] proposed how to build a loan forecast and credit risk assessment model using two machine learning techniques Random Forest and Decision Trees. To better understand the accuracy and other scores of the two models, the results of both models are shown below, along with their classification report and confusion matrix. While the accuracy of the decision tree technique was 73%, the accuracy of the random forest classifier was 80%. For this type of data, the Random Forest model seems to be the best choice.

Princess , Angelo (2020) [13] used odds ratio values to calculate the probability that people with certain characteristics will default on loan payments after these two analyses, pensioners had a 1.2381 times higher probability of avoiding the risk of default than other work groups. Additionally, job seekers are 0.7632 times more likely to become insolvent than other job groups; those in good health are 1.0533 times more likely to avoid default than those who are not; and those who own their home are 1.1008 times more likely to avoid default than those who rent.

Saeed Tariq (2020) [14] focuses on one algorithm in the rating classification (good or bad) prediction model, which reveals the credit history of current borrowers, which can be

used for possible comparisons of loan requests, highlighting the characteristics of good or bad. a loan history Based on their credit history and a certain demographic profile, J8 Algorithm 96.3647%.

A studies that [15] they used data mining to identify and reduce the financial risks associated with financial fraud, and the results show. , the best result was achieved in identifying credible consumers of banks. Despite AUC being 75.8%, F-Measure being 75.3%, Accuracy being 75.3% and Recall being 75.4 %.

Other research was suggested by [16] to compare the outcomes of two classification techniques, Naïve Bayes and Support Vector Machine, in order to forecast the condition of loans. The accuracy results obtained are as follows: Naive Bayes accuracy is 77 percent and Support Vector Machine accuracy is 79 percent, despite the combined execution time of NB and SVM being less than two techniques.

The study proposed by [17], Random Forest technique, is used to develop a model to predict loan payment defaults in loan club, and then compared with decision tree, logistic regression and SVM, random forest still performs best. , outperforms decision trees with 98% accuracy.

On the other hand, in this study [18], they used a random sampling method to compare the ability of machine learning algorithms to distinguish and classify fraudulent and non-fraudulent transactions in a credit card dataset (RUS). Compared to Naive Bayes (NB) and K-nearest neighbor (NN), logistic regression (LR) performed best for all data ratios (KNN). Compared to Naive Bayes and KNN, LR proved to be more accurate. LR achieved 95 percent, NB 91 percent and KNN 75 percent.

In addition, guided methods [19] outperform the unguided K-NN methodology in predicting the future in any situation. They establish a construction process of Artificial Neural Networks (ANNs) through WEKA using multilayer perceptron (MLP) as a data mining forecasting method to choose an appropriate method for forecasting the production and import volume required for

food security. To cover the strategic farming needs of Egyptian citizens due to population growth. According to the findings, Egypt faces a serious problem of availability of wheat and beans, as production and imports are insufficient to meet demand. [20]. They decided on a suitable quantitative model to use loan-related financial information and customer behavior on a mobile network to predict loan performance using credit rating methods to create a machine learning model suitable for payday loans, Random Forest. better for logistic regression, decision trees and other models. If the delay is around 2%, it is worth offering a loan to everyone. The model increases the acceptable default percentage up to four times, from 8% to 32%. Otherwise, the study proposed by [21] presented a comparison between the Naive Bayes (NB) and Support Vector Machine (SVM) classification methods of the machine learning library (MLlib) of the Apache Spark computing system. After applying two classifiers to a dataset consisting of personal and behavioral data of customers from Santander, Spain. According to experimental results, Nave Bayes outperforms Support Vector Machine (% precision, 9% recall and 7.3% F-measure). [22] They used existing customer data from a bank's credit assessment procedure to provide an objective risk measurement method in the lending process of SME customers (Turkish banks). They were able to create a risk metric by looking at the risks identified by bank analysts for current customers.

Their scores are used as target variables in the classification process. Using Weka software, they ran six different algorithms and compared the results for test accuracy, number of rules, recall, precision and Kappa statistics. As a result, the fuzzy accuracy of risk prediction is 78.6 %. [23] They used a machine learning algorithm to detect credit card fraud. Based on a customer's transaction, the accuracy of the system is predicted using machine classification algorithms such as logistic regression and KNN based on credit card data. The KNN algorithm performed best with statistical measures such as precision 0.95,

recall 0.72, and f1 score 0.82 for fraud. This study used other [24] ANFIS, fuzzy clustering, and uncertainty studies to develop an accurate, adaptive, and dynamic model for customer credit risk assessment. The dynamic credit risk assessment model can be adapted for two types of loans: credit card loans and mortgages. Consumers are divided into three categories based on how long they are in arrears: low risk (LR), which means less than two months; medium risk (MR), which means two to six months; and high risk (HR), which means more than six months. In another study proposed by [25], they used a technique based on classification of random forests and deep neural networks to identify customers interested in credit products. The results of empirical studies show that the system is able to extract important patterns from customer history transmission and transaction data and predict the probability of credit purchase, but these patterns may change as the political and economic environment changes. Also another study proposed by [26]. They presented experimental results on real data, which show that the ensemble GLM DT model and the hybrid model based on features selected by probabilistic voting are more effective than all other ensembles and individual classifiers. (GLM DT) Accuracy 87.686928 is the performance measure of the model application.

This research proposed by [27] develops an effective forecasting model for Jordanian commercial banks using Artificial Neural Networks (ANN). The results show that the multi-layer feed forward (MLFF) neural network successfully classified loan requests with an accuracy of 91.3 using data from Jordanian banks from 2010 to 2018.

Another study proposed by [28]. Financial institutions can use advanced machine learning techniques to reduce credit risk. They used 10-fold cross-validation for model selection before comparing several machine learning classification methods using the AUROC methodology. The results show that the Lasso regression with an AUROC of 0.808 provides the best guess, followed by the

Random Forest model with an AUROC of 0.7869.

On the other hand, this study proposed by [29] creates a pioneering set of deep genetic SVM classifier (16-layer system) that enables an efficient binary classification of accepted or rejected borrowers. They developed a new approach based on the synthesis of several approaches, including: two types of SVM classifiers, two types of normalization, two types of feature extraction, three types of core features, three types of parameter optimization, stratified 10-fold CV, ensemble learning, deep learning, layer learning, supervised training, feature selection using a genetic algorithm (attributes), parameter optimization using a genetic algorithm, and training a new genetic layer.

Anil Kuma et.al. (2021) [30] The primary goal of this study is to present different machine learning algorithms that researchers previously identified as potentially applicable to rural borrowers, especially those with poor credit history, or not at all to assess the credit rating of borrowers. It would be interesting to learn more about how financial institutions can successfully combine traditional and modern approaches without violating moral principles. With AIML-based credit scoring algorithms, credit checks can be run in the shortest possible time and potentially with greater accuracy. The literature recommends the use of hybrid or AI-ML based credit scoring methods, but the real challenge for financial institutions is to implement this on the ground using an appropriate combination of traditional and digital methods. The two most important aspects of the success of any banking institution are speed. and accuracy in lending decisions Their work includes changing the structure of the system using individual machine learning methods (classifiers, KNN, fuzzy system), Bayesian decision tree (NBT) and sequential minimum optimization (SMO), and creating frameworks for ensemble classification. in single and multilayer assemblies with different aggregates. Deep Genetic Cascade Ensemble of SVM Classifiers (DGCEC) achieved 97.39 percent accuracy in binary classification

3. Conclusion

From the literature, it appears that various practical methods are used in banking in the mining industry to anticipate loan defaults and credit risk. We found that the most commonly used loan and credit risk forecasting techniques are support vector machines, Bayesian classifiers, and logistic regression and decision trees. In particular, the SVM algorithm is the most accurate to classify the highest yield level to identify risky loans. It also suggested that hybrid methods are better than individual classifiers in improving the predictive ability of credit loan forecasting models.

Supervised Learning techniques are more accurate and reliable but requires more computing time to efficiently acquire and label input in the training phase. Unsupervised learning is the best method to determine unknown data patterns and characteristics, but imprecise, especially in the financial field, describes the good results of ensemble learning using Random Forest to identify credible customers in banks, and its frequency of use is higher than others. Techniques For most of the study, Random Forest was the best classifier,

with accuracy showing that nonlinearity and the ensemble approach were superior.

Despite the fact that recent predictive analytics research is based on deep learning, the performance compared to supervised learning is still under investigation. It also noted that the credit risk performance study gave positive results for sensitivity, precision, accuracy, AUC, recall, F1-I-score and F-measure values, and suggested that performance could be improved by increasing the amount of data.

Previous research focused on customer information and details and found that the primary characteristic was useful in predicting customers' total income, household, gender, and credit history. They also focused on special number types, such as float or integer, because numeric values were easier to categorize. Several factors affect the data structure correct use of loan maturity and customer identity are critical to credit rating to avoid excessive and unreasonably accurate ratings. It must be remembered how important it is for consumers to be aware of their previous loans, and to obtain reliable results, the appropriate amount of information must be used.

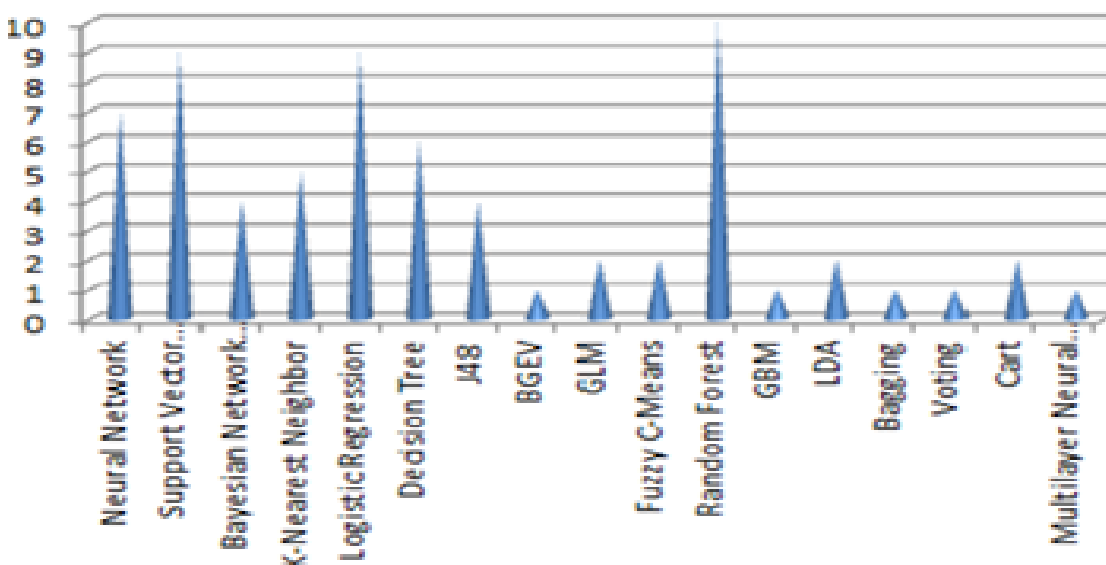


Fig. 1 The average usage of different algorithm

Most previous studies did not consider special loans as cattle loans, and this is addressed in this study, which uses data mining techniques to predict non-performing loans to improve forecasting accuracy and prevent default. to customers in the repayment and schedule of loans from distressed customers., which helps the banking sector to better manage its lending, operations, carefully select customers, minimize losses, increase returns and reduce risks, reduce the amount of loan defaults for special loans, especially livestock loans, credit loan default forecasting, multivariable considering externalities to improve bankruptcy forecasting models.

References

- [1] A. E. Khedr, A. I. Elseddawy and A. M. Idrees, "Performance Tuning of K-Mean Clustering Algorithm a Step towards Efficient DSS," *International Journal of Innovative Research in Computer Science & Technology*, vol. 2, no. 6, 2016.
- [2] A. E. Khedr, A. M. Idrees and A. Elseddawy, "Adaptive Classification Method Based on Data Decomposition," *Journal of Computer Science*, vol. 12, no. 1, pp. 31-38, 2016.
- [3] A. E. Khedr, A. M. Idrees, A. E.-F. Hegazy and S. El-Shewy, "A proposed configurable approach for recommendation systems via data mining techniques," *Enterprise Information Systems*, vol. 12, no. 2, 2018.
- [4] A. E. Khedr, A. M. Idrees and F. K. Alsheref, "A Proposed Framework to Explore Semantic Relations for Learning Process Management," *International Journal of e-Collaboration*, vol. 15, no. 4, 2019.
- [5] A. E. Khedr and A. M. Idrees, "Adapting Load Balancing Techniques for Improving the Performance of e-Learning Educational Process," *Journal of Computers*, vol. 12, no. 3, pp. 250-257, 2017.
- [6] N. Sultan, A. E. Khedr, A. M. Idrees and S. Kholeif, "Data Mining Approach for Detecting Key Performance Indicators," *Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 59-65, 2017.
- [7] A. Al Mazroi, A. E. Khedr and A. M. Idrees, "A Proposed Customer Relationship Framework based on Information Retrieval for Effective Firms' Competitiveness," *Expert Systems With Applications*, vol. 176, 2021.
- [8] A. M. Mostafa, Y. M. Helmy, A. E. Khedr and A. M. Idrees, "A PROPOSED ARCHITECTURAL FRAMEWORK FOR GENERATING PERSONALIZED USERS' QUERY RESPONSE," *JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY*, vol. 55, no. 5, 2020.
- [9] N. A. Bayomy, L. A. Abd-Elmegid, A. E. Khedr and A. M. Idrees, "A Literature Review for Contributing Mining Approaches for Business Process Reengineering," *Future Computing and Informatics Journal*, vol. 5, no. 2, 2021.
- [10] A. M. Idrees, A. E. Khedr and A. A. Almazroi, "Utilizing Data Mining Techniques for Attributes' Intra-Relationship Detection in a Higher Collaborative Environment," *International Journal of Human-Computer Interaction*, 2022.
- [11] Mehul ,Aniket , Chirag and Preeti, "Loan default prediction using decision trees and

- random forest," *A comparative study, IOP Conf. Series: Materials Science and Engineering*, 2021, Vol 1022.
- [12] Deniz, Begu , "Comparison of Data Mining Classification Algorithms Determining the Default Risk," *Hindawi ,Scientific Programming ,Volume 2019, Article ID 8706505.*, 2019.
- [13] Princess , Angelo, "Sample Model for the Prediction of Default Risk of Loan Applications Using Data Mining," *International Journal Of Scientific & Technology Research*, 2020, VOL 9, Issue6.
- [14] Saeed Tariq, "The Application of Data Mining Techniques for Financial Risk Management," *International Journal of Computer Secince and Network Securiry*, 2020, VOL.20.
- [15] Vimala and Sharmili, " Prediction of loan risk using naive bayes and support vector machine," *International Conference on Advancements in Computing Technologies* ,2018, Vol.4 ,Issue: 2, pp. pp. 110-113.
- [16] Lin, Dafeng , Daji and Cai, "study on predicting loan default based on the random forest algorithm,," *Science Direct Procardia Computer Science*, 2019, pp. 503-513.
- [17] Fayaz, Meenakshi and Satwinder, "Comparison and analysis of logistic regression, Naive Bayes and KNN machine learning algorithms for credit card fraud detection," 2020.
- [18] A. M. Idrees and A. E. Khedr, "A Collaborative Mining-Based Decision Support Model for Granting Personal Loans in the Banking Sector," *International Journal of E-Services and Mobile Applications (IJESMA)*,2022, vol. 14, no. 1, pp. 1-23.
- [19] Bernard, Yvonne,Timothy and Patrick, "Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans," *Journal of Risk and Financial Management*, 2020.
- [20] Wael, Mariam and Ghazi Naymat, "Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System,," *The 4th International Symposium on Emerging Information, Communication and Networks*, 2017, p. 559–565.
- [21] S. K. Nihan Gulsoy, "A data mining application in credit scoring processes of smalland medium enterprises commercial corporate customers," *WIREs Data Mining Knowl Discovery*, 2018, pp. 1-12.
- [22] K. Vengatesan , A. Kumar and S. Sabni, "Credit Card Fraud Detection Using Data Analytic Techniques," *Advances in Mathematics: Scientific Journal* 2020, VOL3, p. 1185–1196.
- [23] Farimah and Somayeh, "Dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks, Moradi and Mokhatab Rafiei Financial Innovation," *Moradi and Mokhatab Rafiei Financial Innovation*, 2019.
- [24] Piotr and Kamil, "Direct marketing campaigns in retail banking with the use of deep learning and random forests, .," *Expert Systems With Applications,Elsevier*, 2019, vol134, pp. 28-35.

- [25] Jasmina , Goran and Drago, "New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers.," *Advanced Engineering Informatics, Elsevier, 2020, Vol 45.*
- [26] Khaled Alzeaideen, "Credit risk management and business intelligence approach of the banking sector in Jordan," *Cogent Business & Management, 2019 , Vol 6 ,Issue 1.*
- [27] E. E. Nyoni1, "Credit scoring using machine learning algorithms,Vol.13," *Zimbabwe Journal of Science & Technology , 2018, pp. 26 – 34.*
- [28] Ying , Mohammad and Kunpeng, " Feature selection in credit risk modeling," *an international evidenceEconomic Research-Ekonomiska, Economics and Management Dalian University of Technology, Dalian, 2020.*
- [29] R. a. M. A. Pławiak, "Application of new deep genetic cascade ensemble of SVM classifiers," *Applied Soft Computing Journal, 2019.*
- [30] Anil Kumar, Suneel Sharma and Mehregan Mahdavi, "Machine Learning (ML) Technologies for Digital Credit Scoring for Digital Credit Scoring: A Literature Review," *MDPI -Risks, 2021, pp. 1-15.*