Review of Machine Learning models for Credit Scoring Analysis

2 authors:

Dr Rudra Kumar
Annamacharya Institute of Technology & Sciences
22 PUBLICATIONS   **287** CITATIONS

SEE PROFILE

Vinit Kumar Gunjan
CMR Institute of Technology Hyderabad India
168 PUBLICATIONS   **1,600** CITATIONS

SEE PROFILE

Review of Machine Learning models for Credit Scoring Analysis

Revisión del aprendizaje automático modelos para puntuación de análisis de crédito

Revisão de aprendizado de máquina modelos de pontuação de análise de crédito

Madapuri Rudra Kumar[1]
Vinit Kumar Gunjan[2]

Ingeniería Solidaria

Abstract

Introduction: Increase in computing power and the deeper usage of the robust computing systems in the financial system is propelling the business growth, improving the operational efficiency of the financial institutions, and increasing the effectiveness of the transaction processing solutions used by the organizations.

Problem: Despite that the financial institutions are relying on the credit scoring patterns for analyzing the credit worthiness of the clients, still there are many factors that are imminent for improvement in the credit score evaluation patterns. There is need for improving the pattern to enhance the quality of analysis.

Objective: Machine learning is offering immense potential in Fintech space and determining a personal credit score. Organizations by applying deep learning and machine learning techniques can tap individuals who are not being serviced by traditional financial institutions.

Methodology: One of the major insights into the system is that the traditional models of banking intelligence solutions are predominantly the programmed models that can align with the information and banking systems that are used by the banks. But in the case of the machine-learning models that rely on algorithmic systems require more integral computation which is intrinsic. Hence, it can be advocated that the models usually need to have some decision lines wherein the dynamic calibration model must be streamlined. Such structure demands the dynamic calibration to have a decision tree system to empower with more integrated model changes.

Results: The test analysis of the proposed machine learning model indicates effective and enhanced analysis process compared to the non-machine learning solutions. The model in terms of using various classifiers indicate potential ways in which the solution can be significant.

Conclusion: If the systems can be developed to align with more pragmatic terms for analysis, it can help in improving the process conditions of customer profile analysis, wherein the process models have to be developed for comprehensive analysis and the ones that can make a sustainable solution for the credit system management.

Originality: The proposed solution is effective and the one conceptualized to improve the credit scoring system patterns. If the model can be improved with more effective parameters and learning metrics, it can be sustainable outcome.

Limitations: The model is tested in isolation and not in comparison to any of the existing credit scoring patterns. Only the inputs in terms of shortcomings from the existing models are taken in to account and accordingly the proposed solution is developed.

Keywords: Creditworthiness Evaluation, Credit Score Evaluation, Machine Learning for A Credit Score, Solutions for Credit Score Models, Information and Communication Technologies.

Resumen

El aumento de la potencia informática y el uso más profundo de los sistemas informáticos robustos en el sistema financiero impulsa el crecimiento del negocio, mejora la eficiencia operativa de las instituciones financieras y aumenta la efectividad de las soluciones de procesamiento de transacciones utilizadas por las organizaciones. El aprendizaje automático está ofreciendo un inmenso potencial en el espacio Fintech para determinar un puntaje de crédito personal. Organizaciones, mediante la aplicación profunda de técnicas de aprendizaje y aprendizaje automático pueden acceder a las personas que no están siendo atendidas por instituciones financieras. Una de las principales ideas sobre el sistema es que los modelos tradicionales de inteligencia bancaria dan soluciones predominantemente por los modelos programados que pueden alinearse con la información y sistemas que utilizan los bancos. Sin embargo, en el caso de los modelos de aprendizaje automático, que se basan en algoritmos de sistemas, se requiere un cálculo más integral. Por lo tanto, se puede defender que los

modelos usualmente necesitan tener algunas líneas de decisión en las que el modelo de calibración dinámica se debe simplificar. Tal estructura exige que la calibración dinámica tenga un sistema de árbol de decisión para potenciar cambios de modelo más integrados. Si los sistemas se pueden desarrollar para alinearse con términos más pragmáticos para el análisis, pueden ayudar a mejorar las condiciones de proceso del análisis del perfil del cliente, en el que los modelos de proceso tienen que ser desarrollados para un análisis exhaustivo y pueden brindar una solución sostenible para la gestión del sistema de crédito.

Resumo

1. Introduction

Information and Communication Technologies (ICT) have changed the way in which banking and transaction processes are handled globally. Unlike the conventional banking system, wherein people must visit bank branches and other financial institutions for handling their financial transactions, in the current scenario, millions of transactions are taking place every hour over smartphone applications, online banking systems and even social banking models[1].

There is a paradigm shift in the way banking processes are currently handled when compared to earlier conditions. For instance, a few decades ago, if a customer required a loan, many sets of documentation formalities and personal verification, etc. had to be completed to ensure clearances. One of the key elements in the case of the banking system is the credit score of the individuals who wish to avail credit from the financial institutions [2].

In the early days of credit score usage, the scores were pertinent only to the services like unsecured loan issuance. Today, across various financial services, credit score-based analyses and criteria fulfillment have become mandatory. Globally, across countries, there was a distinct set of credit score management platforms and models, wherein the score would be analyzed by the financial institutions and banking systems [3] [4]. However, in the recent past, there is a phenomenal change in the way financial institutions are focusing on understanding the creditworthiness records of individuals [3].

Organizations no longer rely exclusively on one kind of credit score pattern, and there are multiple dimensions in which credit evaluation teams understand and

analyze the models wherein the credit evaluation pattern could be improved significantly. With evolving financial systems and the increasing set of instant loan solutions, peer to peer lending companies and pay-later solutions are emerging as significant models. It is essential that within the limited set of data available, the organizations have to take to holistic solutions, which can lead to a more emphatic set of creditworthiness evaluation systems [5].

In other dimensions, the critical aspect is about the lending business for the case of both small- and large-scale companies. The competitive business environment is pushing organizations to focus on multiple facets that can support having an overall credibility analysis in the case of lending models. The application of technology has been on the rise in day to day activities. Emerging technologies like Artificial intelligence, Machine learning, Blockchain have created a serious impact at both the business and personal level [6].

2. Purpose of Research

Increases in computing power and the deeper usage of robust computing systems in the financial system is propelling business growth, improving operational efficiency of the financial institutions, and increasing the effectiveness of the transaction processing solutions used by organizations. Due to this, there is an opportunity to process huge amounts of historical data. In a recent development, the application of emerging technologies like Machine learning is seen in Personal Finance. Though the application of machine learning, there is in all aspects of Finance, the recent application of Machine learning in assessing personal finance scores or creditworthiness [3].

Globally, one of the key concerns for financial institutions is about the responsible and effective lending system, integral to their business operations. All the financial institutions that lend money use the Personal Finance or credit score to analyze the creditworthiness of an individual or business. The application of Machine learning can be termed as a breakthrough technology, that helps to get accurate Personal Finance or credit score, which can be used by financial institutions [6].

The Indian Bank regulatory authority, Reserve Bank of India, have found that nearly 25% of non-performing assets or NPAs cannot be recovered. This is a very big number and will create a dent in bank balance sheets. The impact will be higher if these loans are big-ticket loans [5].

In the United States, banks and financial institutions use the FICO score to understand creditworthiness and gives a score between 300 and 850. The FICO score is based on past credit history, credit utilization and credit repayment. It is estimated

that around 100 million Americans cannot access credit from traditional sources as their credit score (considered subprime) is between 300 and 670. This has made many individuals lose out on different loans.

Some Fintech startups, by using custom-built machine learning algorithms, are trying to find a solution to access individuals who have a subprime score. They use machine learning and deep learning techniques to determine creditworthiness. These companies are using data of rejected applications who requested a loan earlier and are applying ML for data analysis. Earlier the rejection data was not analyzed to determine creditworthiness [6].

A US Fintech startup is using machine learning on 10,000 different variables, such as how a customer fills his online application form, time spent on the website or app and other parameters often ignored by traditional financial services companies to determine creditworthiness. By using machine learning and different variables, it has been able to provide loans to many customers who have subprime scores according to FICO scores. With a simplified loan process, which determines loan eligibility, the entity was able to provide loans to around 600,000 customers and the model of evaluation is evolving as an ongoing process. With regular refinement taking place in the algorithm models and the historic set of data that is being garnered by the institution, the volume of business that is being signed to the system is on the rise [7].

Machine learning is offering immense potential in Fintech space and determining a personal credit score. Organizations by applying deep learning and machine learning techniques can access individuals who are not being serviced by traditional financial institutions [8].

There are challenges, as the machine learning system is still in the nascent stage and will evolve over a period. In the initial stages, there could be bad debts. However, with constant evolution, the percentage of bad debts can be controlled. In this review paper, the emphasis is on understanding the distinct set of machine learning models that were proposed in the past, and the gaps in the existing system of machine learning models that were proposed and the contemporary models that can be used in terms of improving the patterns and the ease with which the credit scoring models, can be improved for overall system development.

3. Related Work

In a Research paper published in 2016 by Regina ESI Turkson et al. titled "A machine learning approach for predicting Bank creditworthiness", the authors examined bank creditworthiness data and applied various machine learning and deep learning

techniques for determining credit score using comparative analysis. By identifying 23 of the most critical features, this ML technique can determine customer paying ability. Authors have been able to authenticate their algorithm with other techniques. The authors have also designed a predictive model based on a machine learning algorithm using critical features that determine and help in predicting the creditworthiness of different individuals. The authors of the research paper entitled "A combination of genetic programming and deep learning" published in 2016, try to identify the relationship between input data and results generated by the input data, which the traditional models ignore as they use a black-box model [9]. In this paper, the authors intend to identify a model that combines genetic programming and deep learning techniques. Genetic programming was used as it is flexible and robust, while deep learning and Machine learning techniques are emerging technological areas. The authors review past papers and propose a hybrid model that combines the best functional and induction model. The authors conclude that there is scope for improvement and model will evolve with more research [10].

A review paper published in 2018, tries to understand and explore different models of credit rating using machine learning. Most of them use neural networks and SVM models. Due to the presence of relatively large amounts of data and some redundant features present in credit data, accurate predictions are not always possible. In this paper, the authors propose a credit model using NCSM which reduces the amount of irrelevant and redundant data using Optimized Random Forest. The authors propose a smaller number of features which would reduce the work of the credit evaluators. The authors intend to continue the research on credit rating data structures and the use of clustering approach with Random Forest Method [11].

In a research study published in the year 2002, the authors did a literature survey of previous papers that used machine learning and neural networks to understand its influence on credit card fraud detection, mortgage underwriting, forecasting bankruptcy, among others. For the research, by using a multi-layer perceptron model, the authors wanted to investigate companies that are listed in China. They have classified the companies as Good and Bad based on their credit history. The multi-layer perceptron model is successful in resolving complex and diverse problems. The authors conclude that the model can predict accurately and with a strong suitability ratio. This model can be further developed to understand companies in distress; the same can also be applied for individuals [12].

In the paper submitted in the year 2018, the authors try to find a relevant accurate credit model for peer to peer lending. The authors propose the model of peer to peer lending, which is growing at a faster pace in China due to the simple process of

lending. Though the lending process is easy, risks associated for lenders are high. The authors use borrowers' online behaviors data and have proposed a credit score that uses machine learning. They use the Long Short - Term Memory (LSTM) technique. The authors state that the research done of data sets using LSTM showed improved results. In future research, authors intend to add the time interval of each event [13].

In another study, authors try to understand the existing models and propose a different model based on CF − MF − D − IE. As the first step, the authors try to understand existing models based on SVM, Neural networks and others. They analyze that these models are not accurate due to mass data and are not a cost-effective solution. The authors have built a simple credit scoring model using collaborative filtering based on Matrix Factorization of data sets. At the end of the research, the authors conclude that the CF − MF − D − IE model is cost-effective and gives relatively accurate results. In the future, the author wants to integrate algorithms with classification accuracy for better results [14].

In a research study published in 2017, authors wanted to understand and predict loan status in commercial banks using machine learning techniques. The authors want to build a credit scoring model that can be used by banks to provide credit. From the previous research papers, authors find that there is no one model that is sacrosanct. The authors use an analysis model for credit data. This is done with a combination of K − NN classifiers and the final objective is derived from R. The authors believe this model can help to build a tool that can predict loan status in commercial banks. The authors conclude that the proposed model provides an accuracy of 75% and can be improved with further study [15].

In a study carried out in 2009, the authors propose the use of ensemble machine learning technique to give accurate credit scores. The authors discuss the earlier models for credit score proposed by different researchers using Linear Discrimination Analysis and Logistic regression. After the research the authors conclude, with the application of ensemble machine learning techniques for finding credit score, using data from Australia, that the results were accurate when compared with CART. Authors state that ensemble learning yields accurate results when compared to CART. The authors conclude that better outcomes of credit score can be obtained with continuously improvising algorithms [16].

In a research paper published in the year 2016, the authors try to understand how machine learning can be implemented effectively when imbalanced data is available in asserting credit risk. The authors have applied the CART technique as it helps when the data sample is less than required, data is imbalanced and has other similar issues. The authors applied different methods, like ROC curves, Classification and

regression trees, logistic regression to build a model. The authors conclude that CART and Logistic regression technique together is the right mode for building machine learning algorithms for credit score evaluation. This can be further optimized to get better results [17].

4. Scope of Existing Models

It is imperative that, in the case of the business solutions, focusing on contemporary solutions and evolving the models in line with new age solutions can increase the operational efficiency in business operations. Categorically in the case of the financial business solutions and the personal finance categories, having a more robust and accurate system of lending models and credit evaluation patterns, holds critical importance for business. There are many systems and solutions, that can be integral to business conditions [18].

Even though there are some comprehensive solutions that are proposed in the recent past, Fintech solutions are still focusing on more integrated decision-making systems that can improve the algorithms that can lead to performance outcomes in their models.

One of the major insights into the system is that the traditional models of banking intelligence solutions are predominantly the programmed models that can align with the information and banking systems that are used by the banks. But in the case of the machine-learning models that rely on algorithmic systems, they require more integral computation which is intrinsic. In many of the proposed machine learning models, the implications of such models are ignored in the development of complex predictive models [19]. The key gap that exists in the system is that the banks' existing information and banking systems fail to support the complex systems that are designed for the organizational process [20].

In many of the models that are proposed as a development system, the latency levels in the models do not comply with the standards. One of the critical success factors for the banking system models is about how the model is able to have validators that can assess the range of model risks that were integrated into the implementation conditions. However, in the case of the credit scoring or credit evaluation kind of machine learning models, the scope and features that must be analyzed have to be dynamic and deeper [21].

It is important that the machine learning models also take into account the estimations pertaining to the volume of data to be analyzed based on the model, the production-system architecture and the runtime that is integral to the process. Unless

such holistic systems are developed as a comprehensive condition, the purpose and the desired outcome from the system might not be achieved in practical terms [19].

4.1 Calibration of Dynamic models

Many of the machine learning models that were proposed earlier could be attributed to a certain level of development in the creditworthiness evaluation pattern. However, there are conditions wherein the models must be more dynamic so as to support detection of creditworthiness of the profiles in high accuracy systems [22].

Few of the machine learning models have the capability model of dynamically changing the parameters towards reflecting on distinct conditions of patterns over the data. Such models replace the system of traditional approach towards handling the periodic manual review and refreshing the model. Some of the reinforcement-based learning algorithms or Bayesian methods lead to such dynamic solutions. However, there are certain risks that are highly integral to such models, as they lack sufficient controls in the system, and the other impact factor is the models rely on some short-term pattern emphasis, which might impact the performance quality of the models over time [23].

Hence, it can be advocated that the models usually need to have some decision lines wherein the dynamic calibration model must be streamlined. Such structure demands dynamic calibration to have a decision tree system to empower more integrated model changes. At some point, every dynamic model might turn to the conditions of static solutions and there is a need for mitigating such risks which might emerge in the conditions [24].

Though such changes could result in a change of thresholds that capture the material shifts over the health of the model, a simple out-of-sample performance measure might make an impact in terms of exposure limits and handling the models of pre-defined values which might trigger the need for human intervention systems in the model.

In the process of improving the machine learning models, the model is making sure the inventories constitute more effective scrutinizing of learning-based models. Validation policies and practices are to be modified in terms of addressing the risks that are integral to many of the existing machine learning models, which leads to more refinement of policies and practices that can make a significant outcome [25].

The systems need to be tested under vivid banking transaction conditions, and only upon having the right kind of proofing for the system execution and its accuracy, alongside the alignment of the model to the existing information systems of the banks.

4.2 Deep-Learning

The deep learning models are more prevalent in the case of the application system modeling for creditworthiness solutions. The usage of deep learning models is implicit with hundreds of models that are used in the systems proposed for credit evaluation; lending pattern decisions of the business. The key advantage of the model is about its usage towards managing the unstructured datasets like the images, audio and text conditions. Some of the significant deep learning models used in the system are [26]:

Voice to Text: Off-the-shelf deep learning software must convert customer's audio to text. This can be used in combination with ML methods towards handling the automated systems of customer service conditions and evaluating the necessary factors for analysis.

Social Listening is the other model of unstructured text data-based analysis from the social feed. The patterns of insights in terms of relative discussions, major expenditure related inputs, or the social and demographic facets of the model can be more impacting for the business systems. Unsupervised clustering is phenomenally used for segment and profiling of the customer base in terms of understanding the development strategy towards each of the segments. However, the issue that impacts the model is about the curse of dimensionality [18].

Though some of the models have embraced the PCA application systems, deep learning can be an alternative for handling more advanced levels of the lower set of dimensional features. Contemporary end-to-end big data platforms that are currently available can provide computational power towards training many new-age machine learning models and streamline for deployment [19].

The other critical factors that are integral to the conditions are partial dependence and distance to the decision boundary that is integral to the process. The structure followed by the models is called Model Explain-ability. It is highly important in terms of using the appropriate kind of techniques towards addressing analytical problems and towards addressing complexity.

Certain key parameters to be used in the model analysis are about using the incremental value business, customer experience, the value proposition that could be integral to offering a service to the customer, robustness of the model and other significant factors [18].

4.2.1 Features Selection and Models

Machine learning enables a more insightful model that reflects on various factors based on specific outcomes that are attained. Correlation matrices dismiss correlated

variables and feature selection methods, like phased regression, used in terms of filtering irrelevant predictors. It handles the conditions of best feature selection, by ensuring that the worst feature related words are deleted at every round. Also, the other strategic model followed in the case of machine learning model development is about relying on the iteration using the cross-validation in order to keep the predictor's subset as a contemporary feature selection system [20], [26].

Profoundly, in many of the models that were proposed earlier, logistic regression and decision trees are vividly used in terms of improving the popular classification techniques, towards handling the behavioral scorecards, which are used to analyze datasets for improving the relativity between predictors, independent variables and the response in terms of dependent variable evaluation and assessment [27].

From the inputs reviewed in the literature about potential machine learning models and their features that are chosen for analysis, what follows are some of the key features that are chosen as integral to the analysis models, which can help in improving the overall system of the machine learning model.

Some of the common classifications of the features are depicted in the Table 1.

Table 1. Classifications of the features

Socio economic	Financial	Occupational	Personal
Location	Occupational Income	Current employment	Credit score value
Age	Additional Sources of Income	Category of the company in which the current employment	Asset/Liabilities liquid ratio
Work practices	Prospective income sources	Professional opportunities	Marital Status
Social network profile	Existing lending models used		Interpersonal relationships

Source: own work

Though the dynamics of features that are used in the analysis might vary based on the dynamics of the changing conditions, it can be stated that the metrics used for the process are usually modeled based on the conditions that are considered more appropriate to the conditions chosen for the machine learning model [25].

In addition to the usage of features, the other important aspect in the case of the machine learning solutions is the facets of using the right kind of classifiers that can make a significant difference to the system. For instance, in the case of the training models, the emphasis is on understanding the model of training classifiers which can support the overall enhancement of the system [3], [28].
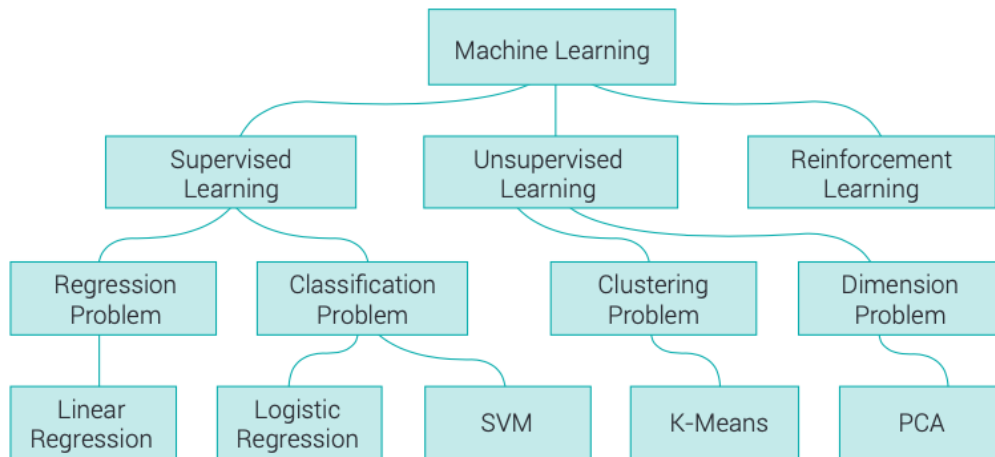
Figure 1. Numerous Sets of Classification Models Integral to Developing
Source: own work

Figure 1 represents numerous sets of classification models integral to developing a comprehensive system that can lead to more indigenous models of analysis that are more related to clustering the profile into various credit patterns.

5. Gaps Analysis

Many of the current machine learning models focus on the creditworthiness models that are integral to a holistic system of evaluation. However, in the case of the contemporary scenarios, where financial applications can take decisions on the credit provided to the clients over the minimal set of information furnished over the apps, there are many distinct ways in which the information can be garnered by the applicant tracking system and even the conditions of pattern analysis [1], [29].

5.1 Illustrative Scenario

In the case of an applicant applying for a certain amount of credit, the current system takes into account the credit score, occupation and income levels, predominantly for the decision on the loan extension. However, there could be many other significant factors that might help in the decision-making process. For instance, in the case of some applicants, it is important to understand the current profile conditions, a pattern of spending, banking transaction conditions of the user, etc. Probably, such models can be highly resourceful in ensuring there is a more emphatic system in place. If the machine learning model can attribute the profile history to various conditions as to

why a specific level of default or the current socio-economic status of the individual is positive or negative for a credit extension, the models can be more integral to the conditions [2], [30].

The other dimension in which the model can be emphatic is about minimal criteria for qualifying, followed by emulating certain levels of peer trust scores. For instance, if the social network members of an applicant can endorse the credibility of the individual applicant, it reflects on the trust and the socio status of the individual. This could seriously reflect on the possible scope of the individual in terms of credit conditions, repayment history factors, etc. [3], [4].

Improvements in the system can be attributed to the requirements of all the key stakeholders integral to the business process. In pragmatic considerations, the model is about ensuring that the loan extending institutions have the right kind of profiles for those who wish to extend the loan and information regarding applicants who might have some disturbing history of credit conditions. Taking such factors into account, if there is a comprehensive system which can be used for analysis, it shall help in improving the ways in which credit scoring patterns are evaluated and towards developing a sustainable solution[3], [31].

6. Conclusion

Financial institutions are always in the burden of ensuring that all the loan disbursals to the customers are highly secured. It is very important that the company must develop an effective creditworthiness evaluation system, that can lead to a better kind of classification system. Many AI-based systems are used in credit evaluation patterns. The review of the literature indicates that there are many potential systems that are used in credit evaluation patterns.

However, one of the significant challenges that has been integral to the problem is about the alignment of the model to the current banking transaction and information systems that can support overall system enhancement of the systems. More often, the features that are chosen for analysis are not effective in terms of the collection of relevant data from the information system. However, there is a need for more significant systems wherein the machine learning models can be improved to take into account diverse conditions that can analyze the credit profile of the individuals in more diversified conditions. If the features can be supportive in garnering holistic conditions of creditworthiness analysis, it can be more pragmatic for businesses in tracking a more effective customer base.

The usage of predictive modeling, in terms of selecting non-conventional metrics for credit evaluating models, can provide more in-depth analysis of the customer profile and the conditions that can lead to a more pragmatic evaluation of customer profiles. If such a holistic system can be developed, it can lead to sustainable business practices in credit extension for businesses.

References

[1]   N. C. Hsieh and L. P. Hung, "A data-driven ensemble classifier for credit scoring analysis," *Expert Syst. Appl.*, vol. 37, no. 1, pp. 534–545, Jan. 2010.

[2]   W. Chen, G. Xiang, Y. Liu, and K. Wang, "Credit risk Evaluation by hybrid data mining technique," *Syst. Eng. Procedia*, vol. 3, pp. 194–200, 2012.

[3]   S. Moradi and F. Mokhatab Rafiei, "A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks," *Financ. Innov.*, vol. 5, no. 1, Dec. 2019.

[4]   S. Akkoç, "An empirical comparison of conventional techniques, neural networks and the three-stage hybrid Adaptive Neuro-Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data," *Eur. J. Oper. Res.*, vol. 222, no. 1, pp. 168–178, Oct. 2012.

[5]   R. P. Bunker, M. A. Naeem, and W. Zhang, *Improving a Credit Scoring Model by Incorporating Bank Statement Derived Features.*

[6]   M. B. Waad, "On Feature Selection Methods for Credit Scoring," no. January 2015.

[7]   P. Abdou, H. A. Abdou, J. Pointon, and H. Abdou, "Credit scoring, statistical techniques and evaluation criteria: A review of the literature Title Credit scoring, statistical techniques and evaluation criteria: A review of the literature Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature," *Financ. Manag.*, vol. 18, no. 3, pp. 59–88, 2011.

[8]   M. Schumann Yang Liu, P. Matthias Schumann, and D. Werk, "The evaluation of classification models for credit scoring Institut für Wirtschaftsinformatik."

[9]   R. E. Turkson, E. Y. Baagyere, and G. E. Wenya, "A machine learning approach for predicting bank credit worthiness," in *2016 3rd International Conference on Artificial Intelligence and Pattern Recognition, AIPR,* pp. 81–87.

[10] K. Tran, T. Duong, and Q. Ho, "Credit scoring model: A combination of genetic programming and deep learning," in *FTC 2016 - Proceedings of Future Technologies Conference,* 2017, pp. 145–149.

[11] X. Zhang, Y. Yang, and Z. Zhou, "A novel credit scoring model based on optimized random forest," in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC,* vol. 2018-January, pp. 60–65.

[12] Su-Lin Pang, Yan-Ming Wang, and Yuan-Huai Bai, "Credit scoring model based on neural network," in *Proceedings. International Conference on Machine Learning and Cybernetics,* vol. 4, pp. 1742–1746.

[13] C. Wang, D. Han, Q. Liu, and S. Luo, "A Deep Learning Approach for Credit Scoring of Peer-to-Peer Lending Using Attention Mechanism LSTM," *IEEE Access*, vol. 7, pp. 2161–2168, 2019.

[14] X. Zheng, "A credit scoring model based on collaborative filtering," in *Proceedings - 9th International Conference on Computational Intelligence and Security, CIS 2013*, pp. 144–148.

[15] G. Arutjothi and C. Senthamarai, "Prediction of loan status in commercial bank using machine learning classifier," in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, 2018,* pp. 416–419.

[16] P. Yao, "Credit scoring using ensemble machine learning," in *Proceedings - 2009 9th International Conference on Hybrid Intelligent Systems, HIS 2009*, vol. 3, pp. 244–246.

[17] S. Birla, K. Kohli, and A. Dutta, "Machine Learning on imbalanced data in Credit Risk," in *7th IEEE Annual Information Technology, Electronics and Mobile Communication Conference, IEEE IEMCON,* 2016.

[18] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending," *Appl. Econ.*, vol. 47, no. 1, pp. 54–70, Jan. 2015.

[19] H. Ince and B. Aktan, "A comparison of data mining techniques for credit scoring in banking: A managerial perspective," *J. Bus. Econ. Manag.*, vol. 10, no. 3, pp. 233–240, 2009.

[20] J. N. Crook, D. E. Edelman, and L. C. Thomas, "Credit Scoring," *J. Oper. Res. Soc.*, vol. 56, no. 9, pp. 1003–1005, Sep. 2005.

[21] Y. Liu and M. Schumann, "Data mining feature selection for credit scoring models," *J. Oper. Res. Soc.*, vol. 56, no. 9, pp. 1099–1108, Sep. 2005.

[22] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *J. Oper. Res. Soc.*, vol. 54, no. 6, pp. 627–635, Jun. 2003.

[23] F. Reserve Board, *Board of Governors of the Federal Reserve System Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit.*

[24] "FICO Credit Score Algorithm Group 14."

[25] S. Sayad, *Comparing Different Classification Techniques in Credit Scoring.*

[26] S. Bhatia, P. Sharma, R. Burman, S. Hazari, and R. Hande, *Credit Scoring using Machine Learning Techniques*, 2017.

[27] Y. Hou, X. Ma, G. Mei, N. Wang, and W. Xu, "A Trial of Student Self-Sponsored Peer-to-Peer Lending Based on Credit Evaluation Using Big Data Analysis," *Comput. Intell. Neurosci.*, vol. 2019, 2019.

[28] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Trans. Syst. Man Cybern.*, vol. 23, no. 3, pp. 665–685, 1993.

[29] M. Bensic, N. Sarlija, and M. Zekic-Susac, "Modelling small-business credit scoring by using logistic regression, neural networks and decision trees," *Intell. Syst. Accounting, Financ. Manag.*, vol. 13, no. 3, pp. 133–150, Jul. 2005.

[30] M. Pagano and T. Jappelli, "Information Sharing in Credit Markets," *J. Finance,* vol. 48, no. 5, pp. 1693–1718, 1993.

[31] B. Twala, "Multiple classifier application to credit risk assessment," *Expert Syst. Appl.,* vol. 37, no. 4, pp. 3326–3336, Apr. 2010.