

Machine Learning for Credit Scoring Evaluation: A Survey

Baidyanath Sou

Assistant Professor, Jagannath Kishore College, Purulia

Abstract

Financial Risk Management (FRM) is a critical component of any organization's financial success. It helps to protect financial health and long-term growth by identifying and mitigating financial risk. As the risk analysis heavily depends on information-deriving decision making, Machine learning is a promising field for new methods and technologies. In recent decades we have seen increasing adoption of Machine Learning methods for various risk management tasks. Credit Scoring is one of the risk factors involving creditors inability to perform their contractual obligations. This research article examines several scientific literature articles and conference proceedings from reputed databased and found out machine learning methods are applied to predict credit scoring gave promising results than traditional statistical methods.

Keywords: Financial Risk, Credit Scoring, Machine Learning.

Introduction

One of the most challenging subjects in the financial industry is credit risk assessment, which determines the status of a potential borrower. The banking industry is exposed to a variety of threats that could have an impact on both their business and their clients. However, it is a difficult task to achieve precision in risk assessment. Credit scoring is one of the important tools which is used to inform crucial decisions about whether to grant a loan to an applicant. Therefore, creating a reliable credit scoring model has emerged as a crucial tool for academics and the financial sector to accurately identify high-risk and low-risk consumers [1].

The credit score analysis is sometimes viewed as a problem of binary classification, as it compares the socioeconomic characteristics of new credit applicants to decide whether they are "trustworthy" or "untrustworthy" candidates. The models were initially created using statistical techniques, the most popular techniques in this category being discriminant analysis and logistic regression (LR)[2]. In recent decades advancement of machine learning gain attentions in many researchers in financial sectors to shift their research from traditional statistical methods to machine learning based approaches. Machine learning based approaches are more common than traditional statistical models because they can handle nonlinear classification issues and high-dimensional datasets with greater accuracy. The credit scoring models are not completely reliant on machines, though. Banks undergo two processes to process applications before approving loans for the applicant. The applications must first be authorized by a financial analyst (or other professionals), after which they must be processed using computational models. However, in the situation of automated credit scoring systems, the decisions made by the machine learning algorithms are used to determine whether to approve the applications. Compared to statistical methods, traditional machine

learning techniques like Artificial neural network (ANN) [9], Random Forest [4], and Support Vector Machines [5] are more efficient and adaptable

A systemic review of credit scoring evaluation techniques is presented in this study. It critically examines the main statistical methods as well as machine learning based methodologies. The goal is to give a thorough overview of the most advanced credit risk estimation technologies while justifying and connecting earlier and later efforts. There are some research constraints are also mentioned: data imbalance, dataset inconsistency, model transparency.

The rest of this article is organized as follows. The conceptual classification scheme for the systematic literature review is presented in Section 2. Section 3 provides a quick overview of the primary credit rating methods. We give the systematic review's findings for the eligible reviewed papers in Section 4. Section 5 and section 6 contain the paper's concluding remarks and future work.

2 Survey Methodology

2.1 search strategy and study selection

In this paper, scientific articles and conference proceedings were searched in ACM Digital Library, IEEEExplore, Springerlink, and ScienceDirect databases to collect relevant literature. These databases were selected because they include the majority of significant publications and conferences. In this search, the terms used as “credit scoring” with “AI”, “artificial intelligence”, “ML”, “machine learning”, “supervised”, “unsupervised”, “ensemble learning”. The following search was created to find relevant studies from the databases using the terms:

(credit scoring) AND (“AI” OR “artificial intelligence” OR “ML” OR “machine learning” OR “supervised” OR “unsupervised” OR “ensemble learning”)

2.2 Inclusion and exclusion criteria

Here, the inclusion, exclusion, and quality evaluation standards are outlined to filter the search results and find the pertinent papers. The literature is included if:

1. Peer-reviewed paper that was published in English and addressed the relevance of research topic qualified as eligible content.
2. It is published between 2000 to 2019.
3. Limited the study to solely full research publications to retain its scope.

The literature is excluded if:

1. The study does not include a thesis or dissertation, a short paper, a doctorate symposium paper, or a poster.
2. There is no access or availability to the complete text.
3. The articles that do not written in English.

3. Computing Approaches

This section provides a quick introduction to the primary computing methods—statistical learning, and machine learning—that are used for credit analysis. Each of these methods has unique features and shares common theoretical basis. However, since artificial intelligence advanced quickly, statistical analysis was gradually replaced by machine learning and deep learning [6].

3.1 Statistical techniques

Discriminant analysis, logistic regression, and Naïve Bayes related model are three statistical techniques widely used to evaluate credit scoring.

Linear discriminant analysis (LDA) is a well-known method for forecasting groups of samples [2]. To decide whether to offer a loan to a borrower, it divides borrowers with strong solvency and borrowers with bad solvency.

Logistic regression (LR) is one of the most popular statistical methods for creating credit scoring. To overcome data anomalies, the logistic regression model is an extension of linear discriminant analysis. In order to reduce the output of the linear function into the range (0, 1) and interpret that value as a probability, logistic regression uses the logistic sigmoid function [6].

Nave Bayes methods are statistical learning algorithms that employ Bayes' theorem with the naive assumption that conditional independence exists between each pair of attributes. When the class variable is provided [7]. A probabilistic model built on graphs is known as a Bayesian network. It evaluates the conditional dependence structure of a number of Bayesian-compliant random variables [8].

3.2 Machine learning techniques

Several machine learning techniques have been examined that work well in the field of credit scoring.

Artificial neural network (ANN) is a mathematical model that can adapt to new input and is inspired by the workings principal of the human brain. According to [9], ANN frequently refers to a non-linear optimization method based on the neural design of intelligent species that learn through experience. An ANN goes through two stages: weights, which are typically brief random variables, and a neural network learning optimization strategy. Data distribution assumptions are not necessary for neural networks [10].

Support vector machine (SVM) is a machine learning algorithm that may be used to tackle regression as well as classification problems [11]. It is a non-parametric approach. In a high dimensional feature space, the Support Vector Machine [5] provides a decision boundary that can divide classes.

Tree-related methods are non-parametric technique frequently applied to classification and regression [12]. This approach makes use of a decision model that resembles a tree to enable choice visualization, greatly aiding in the decision-making process [13]. The conditional probabilities that support Decision trees (DT) can be shown as a tree with nodes, branches, and leaves [14]. The two root nodes of the Decision trees (DT) algorithm comprise borrowers with good and bad solvency. According to its overall error rate or the lowest cost of misclassification, the classification rules' goal is to identify the best decision [15] [11].

Another technique, Random Forest (RF) consists of many individual decision trees working together. Each tree has a forecast, and the class with the greatest number of instances properly identified becomes the solution [16]. The technique builds many uncorrelated decision trees and applies them to categorization or forecasting.

The K-nearest neighbor classification (KNN) is a non-parametric approach that examines if the components and their relationships in the test model are identical. The model's subsequent findings are all categorized into the class that the majority of their "neighbors" are in [3][17]. Because the k-nearest neighbor classification is intuitive, business executives should be able to understand it and support its use. The primary issue with this approach is that, even with the right k-factor selection, discrete-value forecasts with relative probabilities without a suitable probability interpretation are produced. The Bayes technique,

which incorporates the k-factor selection into the model, could address this drawback. This method also makes it possible to ascertain the true probabilities, which have an economic significance [18].

4. Research results and discussion

Several scientific literature articles were analyzed, and each model was analysis based on three criteria-quality of the model, interpretability or transparency and efficiency or performance of models.

The quality of a model that represent key characteristics that are considered when model is built. The superiority of non-parametric approaches is that they do not have such tight assumptions and can handle a wide range of functions and relationships, but they are prone to overfitting. In contrast to classic statistical methods, ANN does not have strict assumptions and may model highly complex functions [19]. When there is a dearth of understanding regarding the relationship between the dependent and independent variables, SVM performs well. The SVM has no constraints except that the data variables must be independent and uniformly distributed [20]. Random Forest (RF) is seen as an appropriate tool for assessing borrowers' solvency because it is not sensitive to multicollinearity. The results are resistant to missing and imbalanced data [21]. RF is also appropriate for huge amounts of data with sufficient noise, as it can avert overfitting and separate crucial features in classification [22]. The Decision Tree (DT) approach finds non-linear relationships with a high degree of accuracy, and it is especially well suited for data mining tasks, where there is frequently little prior knowledge or any assumptions about which variables are related to [23]. K-Nearest Neighbours (KNN), on the other hand, considers that similar items are close to each other, and this method is based on this assumption, hence smooth data with no outliers is required. KNN is sensitive to data scale and irrelevant attributes [24].

In contrast with Non-Parametric, Parametric approaches, in general, have limitations, tight assumptions, and a limited ability to handle complicated and non-linear data. There are various assumptions and limits to Logistic Regression (LR), for example, the dependent variable should be discrete and primarily binary. There should not be any multicollinearity between the independent variables because all independent variables should be independent of each other [25]. Although the Linear Discriminant Analysis (LDA) produces a binary output, the connection between the independent and dependent variables must be linear [26]. There should not be outliers and multicollinearity. For smooth running of Naïve Bayes algorithm and obtaining precise results requires elimination of multicollinearity meaning all variables are independent of each other and avoiding null observations [27][28]. Thereby, in terms of quality of a model, machine learning methods are more reliable then statistical methods.

Transparency is another important parameter to examine a model. Thus, transparency of the model examines if it can demonstrate the primary elements influencing the borrower's solvency and determine what the decision consists of, thereby determining the main reasons for denying or approving the loan. However, much research overlook interpretation since common credit scoring methods adopt a black-box approach, example ANN, SVM, KNN and developing comprehensive credit scoring models is complicated by these and many other ways [29]. Statistical methods such as LR and LDA are transparent and can be easily understood by humans. Although the LDA method frequently results in a satisfactory solution, it does not include a risk assessment for the borrower's insolvency [26].

Among several Black-box methods (ANN, KNN, SVM) the non-parametric structure of the ANN approach is a fundamental deficiency in explaining the importance of credit scoring elements and their relationship to excellent or poor borrower solvency, making it difficult to justify a loan decision based on these methodologies [30]. The use of the SVM approach to separate non-linear credit scoring data makes

the findings harder to comprehend [31]. Similarly, KNN does not produce easily understood findings due to its non-parametric character [32]. As a result, it cannot discern the most crucial aspects of the borrower's solvency or why the loan was rejected. Another method that is difficult to interpret is the Random Forest (RF) strategy, which is based on the usage of many decision trees and does not produce simple answers [33]. To make the findings more comprehensible, try to limit the complexity of this method's aggregation rule by selecting only a few decision trees that will work together to produce the outcome.

The DT approach makes the model transparent. A decision tree gives managers explicit and broad standards for deciding whether to lend to a borrower [34]. The interpretation of the data in the tree is simple and allows for the rapid and easy classification of new observations, as well as a simple explanation of why the observations are classified in this manner [35]. Therefore, Decision Tree (DT) is most efficient in this category.

Efficiency measures how soon the model can provide the desired results or whether the model must be modified, and the data analyst involved before the results are obtained. This evaluation criterion examines if expanding the data set has a substantial impact on the speed of model training, learning, and development. SVM is a competitive technique to assessing borrowers' solvency and provides a quality edge over other conventional approaches, but it comes at a cost: the substantially longer computing time necessary to discover the ideal core function parameters [36]. Although ANN is widely used for determining a borrower's solvency, the training procedure is lengthy for large amounts of data. However, if the amount of data is not too vast, ANN model training and decision making can deliver great efficiency quickly. Also, the main drawback of the KNN approach is that it gets substantially slower as the number of data increases, hence in the case of credit score modeling (for huge volumes of data), the method may be impracticable if projections must be made rapidly [37]. The computational complexity of the DT technique in the event of big data sets is the most important disadvantage, as all properties must be examined and evaluated at each node. Because the generated tree is frequently huge, the process of learning the model becomes prohibitively time-consuming [38].

The NB algorithm creates and evaluates models quickly and efficiently. It can be used to handle both binary and multiclass classification issues, therefore it is computationally efficient - it analyzes multidimensional and voluminous data fast without impacting the dimensions themselves (Taylor, 2008). Other parametric algorithms, like as LR and LDA, can also get results fast. Thereby, they work well in this category.

The cost and complexity of the calculations must also be considered while evaluating approaches [39]. It might be assessed by assessing the model's simplicity and potential cost. The procedures determined the sum of all the basic principles and their percentage value from the greatest number of points achievable, considering the weights of the basic principles. The findings suggest that ANN and decision trees are the best strategies for creating a credit scoring model. The key benefits of these methods are their high quality, as evidenced by classification accuracy, and their efficiency, which is especially significant for businesses seeking straightforward and unambiguous findings. The SVM approach, which has the highest classification accuracy in scientific literature, has taken the second position but is still far behind the best methods in this examination. It is primarily owing to the method's black-box nature. The following figure shows the overall performance based on averaging three categories after analysis of research articles and conference proceedings.

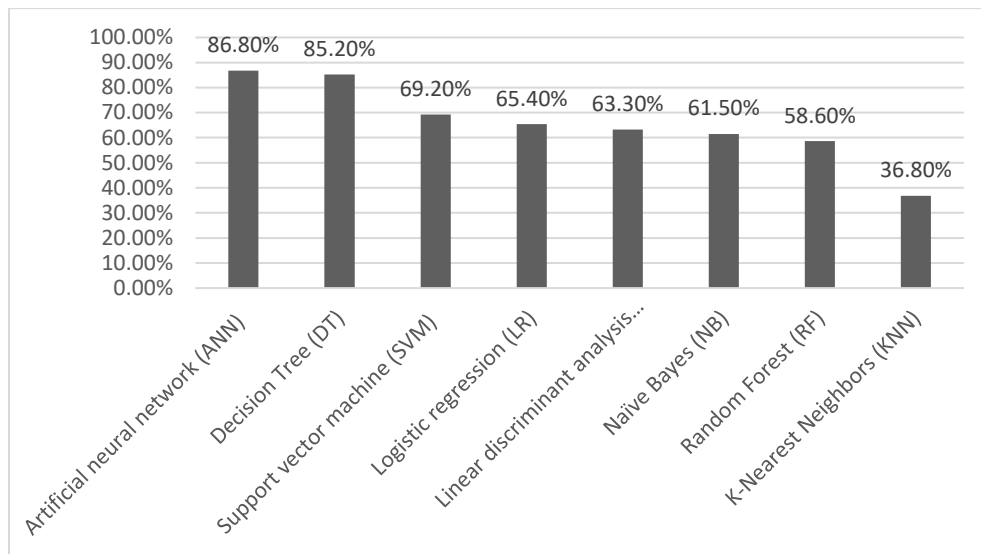


Fig: Graphical representation of the observation.

However, the classification accuracy is dependent on the size and properties of the data set used to build the model [40]. As a result, it is not possible to say that ANN and decision trees are always the optimal methods. Experts and researchers are recommended to utilize a variety of approaches to evaluate which method is best suited to a certain data set, but this assessment gives a full systematic methodology based on four fundamental principles.

5. Conclusion

Credit rating is crucial for identifying credit defaulters, for this reason precise data for forecasting is needed. After analyzing the papers, I have learned that there are several difficulties in the subject of credit scoring evaluation. Each model has its own pitfalls and difficulties, therefore they cannot all be relied upon for evaluation. The problem of credit scoring cannot be solved by a single complicated classifier. The dataset will differ greatly because different financial institutions from various geographical regions, or even the same location, would have different laws and regulations. Therefore, we will lose accuracy if we train the model on a dataset from one domain and test it on a dataset from another domain. Ensemble approaches are being used by researchers to investigate this issue [41]. Single classifiers have been observed to perform poorer than ensemble approaches [42-44]. A significant disadvantage of ensemble learning is the output' interpretability or readability. As a result, increasing the interpretability of ensemble models is another crucial study field that requires more investigation.

6. Discussion and future Work

Applying feature selection techniques is an essential stage in addressing the dimensionality problem. To improve the model's accuracy, some algorithms like GAs [45][46] with classifiers such as SVM can be implemented. This type of hybridized model is growing in popularity as more and more researchers create them. Their use has allowed academics to explore a brand-new field.

Data preprocessing of datasets is another area that can be improved. A dataset may have recurrent or repetitive features. Low precision and needless computation may result from this. Therefore, pre-processing data is a crucial step in enhancing a model's performance. There are few techniques that data

pre-processing can boost classifier performance. However, there are ample scopes to improve in real world scenarios.

References

1. Yoichi Hayashi (2016), Application of a rule extraction algorithm family based on the Re-RX algorithm to financial credit risk assessment from a Pareto optimal perspective, *Operations Research Perspectives*, vol. 3: [32-42]
2. M. Mihalovič (2016), Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction, *Economics & Sociology*, vol. 9 (4): 101–118.
3. T Cover, P Hart (1967), Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, vol. 13(1): 21-27.
4. Breiman L (2001), Random forests. *J Mach Learn*, vol. 45(1):5–32.
5. Cortes C, Vapnik V (1995), Support-vector networks. *Mach Learn* 20(3):273–297.
6. Goodfellow I, Bengio Y, Courville A (2016) *Deep Learn*. MIT press, Cambridge.
7. Buitinck L, Louppe G, Blondel M, et al (2013) API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: languages for data mining and machine learning*, pp 108–122
8. Liu S, McGree J, Ge Z et al (2015) *Computational and statistical methods for analyzing big data with applications*. Academic Press
9. Fausett, L. V. (1993), *Fundamentals of Neural Networks: Architectures, Algorithms And Applications*. 1st ed. London: Pearson.
10. Noh, H. J., Roh, T. H., Han, I. (2005), Prognostic Personal Credit Risk Model Considering Censored Information. *Expert Systems with Applications*, 28(4), 753-762.
11. Wah, Y. B., Ibrahim, I. R. (2010), Using Data Mining Predictive Models to Classify Credit card Applicants. 6th International Conference on Advanced Information Management and Service (IMS) Seoul, IEEE Xplore.
12. Pedregosa F. et al. (2011), Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, Vol. 12:[2825-2830].
13. P. Gupta. (2019), Decision Trees in Machine Learning. Towards Data Science. Available online at: <https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>.
14. Abellán, J. Castellano, J. G.,(2017), A Comparative Study on Base Classifiers in Ensemble Methods for Credit Scoring. *Expert Systems with Applications*, vol. 73(1) :[1-10].
15. Eddy, Y. L., Bakar, E. E.(2017), Credit Scoring Models: Techniques and Issue. *Journal of Advanced Research in Business and Management Studies*, vol. 7(2): 29-41.
16. Yiu, T. (2019), Understanding Random Forest. Available online at: <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>.
17. Bolton, C. (2009), *Logistic Regression and its Application in Credit Scoring*. Pretoria: University of Pretoria.
18. Vojtek, M., Kocenda, E., 2006. Credit Scoring Methods. *Czech Journal of Economics and Finance*, vol. 56(3-4): 152-167.
19. Abdou, H. A., Pointon, J. El-Masry, A. A., 2008. Neural Nets Versus Conventional Techniques in Credit Scoring in Egyptian Banking. *Expert Systems with Applications*, vol. 35(3): [1275-1292].
20. Wilson, M. D. (2008), *Support Vector Machines*, Encyclopedia of Ecology. Amsterdam: Elsevier.

21. Breiman, L. (2001), Random Forests. *Machine Learning*, vol. 45(1): [5–32].
22. Deng, Menghua, et al. (2018), Agricultural drought risk evaluation based on an optimized comprehensive index system. *Sustainability* vol. 10.10: [3465].
23. Bolton, C. (2010). Logistic regression and its application in credit scoring (Doctoral dissertation, University of Pretoria).
24. Hastie, et al. (2009). *The elements of statistical learning: data mining, inference, and prediction* New York: springer. Vol. 2: [1-758].
25. Gupta, D. (2018). *Applied analytics through case studies using Sas and R: implementing predictive models and machine learning techniques*. Apress.
26. Lee, T. S., & Chen, I. F. (2005), A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with applications*, vol. 28(4): [743-752].
27. Buitinck L, et al (2013) API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: languages for data mining and machine learning*, [108–122].
28. Liu S, et al (2015) *Computational and statistical methods for analysing big data with applications*. Academic Press.
29. Xia, Yufei, et al. (2017), A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications* vol. 78: [225-241].
30. Eletter, S. F., & Yaseen, S. G. (2017). Loan decision models for the Jordanian commercial banks. *Global Business and Economics Review*, vol. 19(3): [323-338].
31. Le, J. (2018). Support Vector Machines in R. DataCamp. Available online at: <https://www.datacamp.com/community/tutorials/support-vector-machines-r>.
32. Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE access*, vol. 7: [154096-154113].
33. Tomczak, J. M., & Zięba, M. (2015), Classification restricted Boltzmann machine for comprehensible credit scoring model. *Expert Systems with Applications*, vol. 42(4): [1789-1796].
34. Xia, Yufei, et al. (2017), A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications* vol. 78: [225-241].
35. Vidal, M. F., and Barbon, F. (2019). *Credit Scoring in Financial Inclusion. Technical Guide; CGAP: Washington, DC, USA*.
36. Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert systems with applications*, vol. 36(2): [3302-3308].
37. Harrison, O. (2018). *Machine learning basics with the k-nearest neighbors algorithm. Towards data science*, 11.
38. Bastos, J. (2007). *Credit scoring with boosted decision trees. School of Economics and Management (ISEG)*.
39. Li, X. L., & Zhong, Y. (2012). An overview of personal credit scoring: techniques and future work. *International Journal of Intelligence Science*, vol. 2(4): [181-189].
40. Szwabe, A., & Misiorek, P. (2018). Decision trees as interpretable bank credit scoring models. In *Beyond Databases, Architectures and Structures. Facing the Challenges of Data Proliferation and Growing Variety: 14th International Conference, 24th IFIP World Computer Congress*, [207-219]. Springer.

41. Xia Y, Liu C, Li YY, Liu N (2017) A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Syst Appl* 78:02.
42. Hung C, Chen J-H (2009) A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Syst Appl* 36:5297–5303, 04.
43. Yu L, Wang S, Lai KK (2008) Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Syst Appl* 34(2):1434–1444.
44. Wang G, Hao J, Ma J, Jiang H (2011) A comparative assessment of ensemble learning for credit scoring. *Expert Syst Appl* 38(1):223–230.
45. Cheng-Lung H, Mu-Chen C, Chieh-Jen W (2007) Credit scoring with a data mining approach based on support vector machines. *Expert Syst Appl* 33:847–856, 11
46. Zhao Zhenyu, Zhang Wei, Zhou Yayue (2011) National student loans credit risk assessment based on gabp algorithm of neural network. In: 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2196–2199