International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

Credit risk modeling in Nigerian banking sector: A comparative study of machine learning algorithms

Ucheoma Austie Ehimare *

Michael Okpara University of Agriculture, Statistics, Umudike, Abia State, Nigeria.

International Journal of Science and Research Archive, 2020, 01(01), 231-235

Publication history: Received on 22 October 2020; revised on 25 November 2020; accepted on 29 November 2020

Article DOI: https://doi.org/10.30574/ijsra.2020.1.1.0046

Abstract

In Nigeria, economic instability, a high unemployment rate and banking rules mean that credit risk continues to be Very significant and could result in loan defaults. When traditional credit risk models, like logistic regression and credit scoring, are used, the data's complex structure usually cannot be understood well which keeps the models from being very accurate. This work cheques how using machine learning algorithms of different types can lead to improvement in credit risk modelling in Nigerian banks. The research combines specific borrowers' details with economic information to compare both ML models and standard methods, looking closely at prediction accuracy, what the algorithms reveal and if they are in line with regulation.

The report shows how ML works better than traditional methods at predicting credit risk globally, showing clear advantages during volatility and also notes where local studies are insufficient. Important matters related to the quality of data, the transparency of models and harmonisation with local laws are sharply examined. The study assesses different ML algorithms by comparing them with Nigerian loan data and analysing how accurate, useful for testing performance and how quickly they run. With the findings, banks will learn how to use ML in a way that supports compliance without missing out on the benefits of new risk management tools. Recommendations include boosting data systems, making AI understandable, to preparing custom rules for using ML in Nigerian banking.

Keywords: Credit Risk; Machine Learning; Economic Instability; Risk Management; Predictive Analysis; Algorithm

1. Introduction

Credit risk has long been regarded as one of the most critical challenges facing the banking sector, not only globally but particularly within developing economies such as Nigeria (Adeusi, Akeke, Adebisi, & Oladunjoye, 2014). When the person borrowing money is unlikely to meet their commitments according to the terms set, Nigerian banks depend a lot on lending, so strong credit risk management is crucial for their financial health, safety of deposits and the stability of the bank industry (Okafor, 2016). Over the years, non-performing loans and loan defaults have often been linked to financial problems and sector weaknesses, yet good, data-supported credit risk evaluation practises have not always been used (Sanusi, 2011). In the past few years, ML techniques have given banks new opportunities to make traditional credit assessment models more advanced. ML detects complex, non-linear details in large data sets, unlike standard statistics that base predication on pre-supposed values and linear connexions. This is why ML excels in forecasting and controlling risks in the credit market (Lessmann, Baesens, Seow, & Thomas, 2015). The use of ML with customer and transactional data now held by Nigerian banks provides an exciting new way to improve credit risk modelling accuracy, efficiency and timeliness.

* Corresponding author: Ucheoma Austie Ehimare

This conceptual review aims to explain the theory, important concepts and methods linked to using machine learning in modelling credit risk, as applied to the Nigerian banking sector. The review covers the historical way credit risk was measured, explains important concepts tied to using ML in finance and considers how different ML models perform in risk evaluation. Confirmation of the approach allows for further investigation where various ML methods are examined to predict credit risks in Nigerian banks.

2. Conceptual framework

To understand this study, there is need to understand the concepts around credit risk, credit risk modelling, machine learning in finance and comparative analysis of algorithms

When a loan borrower cannot make timely repayments and complies with agreements, credit risk leads to losses for the lender as indicated by Saunders and Cornett (2018). The sources of trade credit include retail loans, mortgages, loans to corporations and exposures to other banks. Nigeria's credit risk situation is made challenging by unstable economic conditions, large unemployment, rising and falling commodity costs and limited regulation (Adegbite, 2015; CBN, 2020). If credit risk is not properly handled, the results can be reduced earnings, shortages of funds, wearing away of the bank's protection and, sometimes, failure of the bank.

Credit risk modelling is the practise of using different approaches from finance to predict if a borrower cannot meet their payments, to estimate upcoming losses and to set safety limits for credit (Bluhm, Overbeck, & Wagner, 2016). For years, risk assessment in banking has been built on traditional credit risk models such as the Altman Z-score, credit scoring and logistic regression (Altman, 1968; Crook, 2002). They typically use past data, information about borrowers and major economic indicator to find out how likely it is that someone will default. Still, they are frequently held back by strict rules, unavoidable linearity and little ability to change when dealing with flexible data patterns.

Risk evaluation and prediction have made a major leap forward in finance because of machine learning (ML) (Mullainathan & Spiess, 2017). ML deals with a collection of methods that help algorithms uncover patterns from data, keep improving their accuracy and perform most work without needing to write code for every job (Goodfellow, Bengio, & Courville, 2016). When dealing with credit risk, ML models are able to work with diverse data, including accounts, social media behaviour and broader economic situations, to point out patterns that can mean higher possible risks (Lessmann et al., 2015). Among the methods used in credit risk modelling are decision trees, random forests, support vector machines (SVM), gradient boosting and neural networks (Bishop, 2006). Worldwide research finds that these models provide higher prediction accuracy than the usual statistical approaches (Barboza, Kimura, & Altman, 2017). But there are issues in explaining these models, checking and cleaning data and ensuring compliance which remain especially important in a growing financial sector such as Nigeria's.

By performing comparative algorithm analysis, we assess and compare ML algorithms in terms of their predictions, stability and appropriateness for a given situation (Provost & Fawcett, 2013). While doing comparative analysis for credit risk modelling, people cheque the algorithms against important metrics such as accuracy, AUC-ROC, precision, recall, F1-score and running speed (Lessmann et al., 2015). When comparing algorithms, researchers and practitioners see which of them provide the strongest outcomes for being useful, easy to understand and workable in practise (Khandani, Kim, & Lo, 2010). Since Nigerian banks need to follow local laws closely such insights give them a useful guide in choosing technologies, interacting with regulators and staying aligned with both the local market and risk measures.

The areas of this review have been precisely defined to suit the Nigerian banking sector. Special attention is given to retail and corporate lending actions by commercial banks under the supervision of the Central Bank of Nigeria (CBN). The review analyses information from the past, including borrowers' identities, details of their loans, their timely payments and other important economic aspects. Banks' followance of the CBN Prudential Guidelines and Basel rules is also examined, given that they specify the types of risk management approaches and reports allowed for Nigerian banks. Importantly, reviews consider that gathering and managing data locally is a challenge, requiring careful balancing of model accuracy with what consumers (and regulators) can see and analyse.

3. Theoretical Background

For many years, the approach to credit risk modelling has moved from using statistics to depending on advanced machine learning. To set this study in context, we need to identify ways that methods have evolved.

Over the years, traditional credit scoring models depended mainly on logistic regression, discriminant analysis and credit scoring methods (Crook, 2002). Within banking, using logistic regression to estimate a borrower's likelihood of default from a list of explanatory variables is widespread (Hosmer & Lemeshow, 2013). People find linear generalised equations attractive because they are easy to use, interpret and provide estimates that matter for both setting premiums and dividing capital (Altman & Saunders, 1998). Credit reporting systems, by contrast, attach numbers to a borrower's qualities and combine them into a rating which indicates their creditworthiness (Hand, 2005). Having these systems ensures that criteria are the same for all the large numbers of credit applications banks consider. Even so, traditional models are unable to deal with situations where there are many highly interconnected variables, cases involve multiple linear relations or there is a lot of complex data (West, 2000).

Machine learning techniques, however, pave the way for a new system of evaluating and anticipating credit risk (Mullainathan & Spiess, 2017). A range of ML algorithms helps systems uncover patterns from data and make decisions independently, without having to be told specifically (Goodfellow, Bengio, & Courville, 2016). Among the models used most often in credit risk modelling are decision trees, random forests, gradient boosting machines, support vector machines and neural networks (Lessmann et al., 2015). Data is arranged in decision trees as branches, determined by splits in features, so the classification process is easy to understand (Breiman, Friedman, Olshen, & Stone, 1984). This class uses a combination of decision trees which reduces both overfitting and improves how stable the predictions are (Liaw & Wiener, 2002). Ensembles of models are improved further by gradient boosting which adjusts past mistakes one by one, enabling them to frequently perform well on structured data tasks (Friedman, 2001). Since SVMs have strong classifying abilities, they work in high-dimensional areas by recognising the best separating hyperplanes (Cortes & Vapnik, 1995). Because they are modelled on the brain, neural networks do well at finding connexions between many variables.

ML models offer significant advantages over the traditional models used before. An important point is that ML models can find connexions between variables that are not linear, whereas old linear models would miss them (Bishop, 2006). As a result, the system recognises unexpected relationships that typically improve the accuracy of forecasts (Ngai, Hu, Wong, Chen, & Sun, 2011). Also, ML models can process many kinds of data, like the financial information you see in spreadsheets, text written expressively and behavioural or social media information (Provost & Fawcett, 2013). More importantly, using ensemble and boosting practises gives better reliability to the predictions made on unseen data sets (Dietterich, 2000). It is possible to regularly build the models with new data, so they follow new patterns among borrowers and changes in the overall economy (Khandani, Kim, & Lo, 2010). Yet, this technology is not without problems, mainly because it is often difficult to explain, run on big data and fit within needed regulatory frameworks. Consequently, their ability to make accurate predictions gives them appeal among Nigerian banks trying to update their credit risk reviewing systems. By examining these advanced algorithms in comparison to earlier models, this research seeks to discover which techniques make the best, most accurate and practical forecasts of credit risk for Nigerian banks.

4. Review of Relevant Literature

An abundance of research has applied machine learning to credit risk models and conclusively showed it outperforms usual methods. In 2020, Wang et al. looked at five classifiers along with data from bank loans. The team found that Random Forest worked best among the models, showing 97.16% precision and 96.53% accuracy. Decision Trees are at risk for overfitting, according to Wang et al. and KNN can be affected by outliers and when working with high-dimensional data (Wang et al., 2020).

Many studies reported facing the problems of quality information, clear model development and overfitting. Wang et al. (2020) pointed out how data-oriented tasks play an important role and Vuppala (2020) discussed the connexon between how advanced your model is and whether it is easy to understand. Experience in Nigeria and Ghana suggested that scaling up cheap ML solutions is especially required for working with limited or noisy data.

4.1. Knowledge Gaps and Justification for Study

ML technology is well studied globally, but its application to credit risk is still lacking in Nigeria. Most related studies such as those by Wang et al. (2020) and Vuppala (2020), deal with countries with strong, complete financial programmes and good data. In Nigeria, however, data is usually less complete and reliable. A lot of local research attempts simples classifiers such as and does not compare them to advanced techniques like SVM, GBM or deep learning. There is also a large gap in linking local macroeconomic conditions such as changes in inflation, currency rates, risk and prices, with credit risk assessments. Such variables have a big impact on how borrowers act in Nigeria's unstable economy, yet they are often missed in current research. In addition, little is known about the hands-on and regulatory

problems associated with ML use in Nigerian banks. Many articles deal with explainable AI and model governance, but only a limited number analyze how advanced ML models follow the standards from the CBN and Central Bank standards.

This research uses a comparison of various ML methods to fill these gaps using mixed borrower and macroeconomic data for Nigeria. Additionally, the study cheques if each model is easy to understand, has efficient computations and meets regulatory requirements. By providing evidence for decision-making, the research will improve credit risk prediction, guide portfolio adjustments for banks and help regulators see the balance between new approaches, steady compliance and the safety of bank clients.

5. Conclusion

This review has formed a basis for investigating ML tools in credit risk modelling among Nigerian banks. Although logistic regression and credit scoring have been handy before, they are challenged now by the complex and curved patterns of risk. Compared to others, these techniques—decision trees, random forests, gradient boosting, SVM and neural networks—have much better accuracy and can handle more situations. Around the world, research confirms that ML outperforms more traditional approaches when markets are unpredictable, yet Nigerian research does not yet match this level. Removing these gaps is important to prevent mismanaging credit risk which jeopardizes bank stability. This investigation will evaluate various ML algorithms using Nigerian lever information and economic data, paying attention to both correctness and obstacles in practice. The research hopes to increase banks' resilience and inclusion by supporting responsible use of ML and also by assisting regulators.

Recommendations

From the outcomes, below are some key suggestions to Nigerian banks and regulators looking to use machine learning (ML) in credit risk management.

Banks should start by strengthening their data processes so they gather strong data on both borrowers and the wider economy in order to make the most out of ML techniques. Care must be taken to clean, unite and manage data to deal with the common problems of missing, imbalanced or noisy data Nigeria faces. Secondly, banks need to apply several types of algorithms together, so they can achieve both accurate predictions and clarity in understanding their effects. Even though gradient boosting and neural networks give more accurate results, decision trees or random forests are often better for understanding and show more transparency which matters greatly for following the rules and gaining people's trust. The fourth issue concerns the Central Bank of Nigeria (CBN) which should create clear guidelines for using ML in credit risk assessment that strongly emphasize matters of fairness, accountability, explain ability and data privacy. If you promote the use of clear AI models and write up their explanations, ML-based decisions can always be understood and verified. All involved should strengthen capacity by delivering knowledge about ML to risk managers, data scientists and compliance officers. With internal expertise, these institutions will be able to start using ML confidently so it is both useful locally and up to international standards.

References

[1] Adegbite, E. (2015). The role of regulation in bank risk-taking in Nigeria. Journal of Financial Regulation and Compliance, 23(3), 318-336.

[2] Adeusi, S. O., Akeke, N. I., Adebisi, O. S., & Oladunjoye, O. (2014). Risk management and financial performance of banks in Nigeria. Journal of Business and Retail Management Research, 8(2), 93-103.

[3] Altman, E. I. (1968). Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589-609.

[4] Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. Journal of Banking & Finance, 21(11-12), 1721-1742.

[5] Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405-417.

[6] Basel Committee on Banking Supervision (BCBS). (2019). Principles for the sound management of credit risk. Bank for International Settlements.

[7] Bessis, J. (2015). Risk management in banking (4th ed.). Wiley.

[8] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

[9] Bluhm, C., Overbeck, L., & Wagner, C. (2016). Introduction to credit risk modeling (2nd ed.). Chapman & Hall/CRC.

[10] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth.

[11] Central Bank of Nigeria (CBN). (2020). Financial stability report.

[12] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

[13] Crook, J. (2002). Credit scoring and its applications. Journal of the Operational Research Society, 52, 997-1006.

[14] Dietterich, T. G. (2000). Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems, 1-15.

[15] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.

[16] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

[17] Hand, D. J. (2005). Good practice in retail credit scorecard assessment. Journal of the Operational Research Society, 56(9), 1109-1117.

[18] Hosmer, D. W., & Lemeshow, S. (2013). Applied logistic regression (3rd ed.). Wiley.

[19] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), 2767-2787.

[20] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[21] Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124-136.

[22] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R News, 2(3), 18-22.

[23] Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106.

[24] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection. Decision Support Systems, 50(3), 559-569.

[25] Okafor, F. O. (2016). Credit risk management and bank performance in Nigeria. Journal of Economics and Sustainable Development, 7(4), 85-94.

[26] Provost, F., & Fawcett, T. (2013). Data science for business. O'Reilly Media.

[27] Sanusi, L. S. (2011). Global financial meltdown and the reforms in the Nigerian banking sector. Central Bank of Nigeria Bulletin, 34(3), 1-12.

[28] Sanusi, L. S. (2012). Banking reform and its impact on the Nigerian economy. CBN Journal of Applied Statistics, 3(2), 115-122.

[29] Vuppala, S. K. (2020). Impact of global economic shocks on credit risk assessment models in the banking industry. International Journal of Leading Research Publication, 2(5), 1–16.

[30] Wang, Y., Zhang, Y., Lu, Y., & Yu, X. (2020). A comparative assessment of credit risk model based on machine learning—A case study of bank loan data. Procedia Computer Science, 174, 141–149.

[31] West, D. (2000). Neural network credit scoring models. Computers & Operations Research, 27(11-12), 1131-1152.