

Machine Learning-Based Credit Default Prediction: Challenges, Concerns, and Model Comparisons

*¹Joshua Bature Hassan, ¹ John Awoyemi Oloruntoba, ¹Daniel Dauda Wisdom, ¹Taiwo David Ajayi,

^{1,2}Federal University Oye Ekiti, Ekiti State, Nigeria

^{3,4}Federal University of Agriculture Abeokuta, Ogun State, Nigeria

*joshua.hassan@fuoye.edu.ng | john.awoyemi@fuoye.edu.ng | danieldw@funaab.edu.ng | ajayitd@funaab.edu.ng

Received: 23-JAN-2025; Reviewed: 20-FEB-2025; Accepted: 06-APRIL-2025

<https://dx.doi.org/10.4314/fuoyejet.v10i2.12>

ORIGINAL RESEARCH

Abstract— the digital era keeps leveraging machine learning (ML) techniques to enhance profitability within the financial sector. With the integration of cutting-edge technologies and services enabling financial institutions to extend microcredit to individuals with lower incomes. This digitization has streamlined various aspects of the loan application process, such as underwriting and electronic signatures, resulting in quicker loan disbursement while upholding conventional underwriting and compliance procedures. Traditional credit scoring models face constraints when adapting to big data technology for risk modeling. Consequently, ML-based credit risk models have emerged as a more efficient approach for forecasting defaults. Thus, this paper introduces ML model for predicting possible credit failure challenges and evolving issues using ML tools. The developed model used kaggle dataset for the study. The model employed decision Tree and Random Forest algorithm to efficiently develop a classification model to anticipate loan defaults before granting the loan.

Keywords— Financial Institutions, Credit, Failure, Prediction, Challenges & Machine-Learning.

1 INTRODUCTION

Digital lending, also known as social or online lending is a swiftly expanding phenomenon in emerging markets. As it encompasses credit products delivered exclusively through digital channels, such as mobile applications and the internet. These loans are typically unsecured cash loans acquired through digital channels without the involvement of a salesperson (Hassan et al., 2023). Digital lending harnesses digital data to make lending determinations through automated procedures. This has provided millions of low-income consumers with the opportunity to access loans from the convenience of their location. Nevertheless, digital lending carries higher risks compared to traditional banking due to inadequate credit checks, limited intermediation, a lack of transparency, and the financial status of online borrowers (Cinca, et al., 2019; Hassan et al., 2023). Credit risk prediction and management assume a critical role in this context. Digital lending has disrupted the conventional consumer banking sector, offering more efficient loan processing, with loan application decisions automatically determined using electronic data-driven algorithms. This has enabled online lenders to offer small loans with short-term maturities, facilitating access to credit for individuals excluded from traditional banking systems (Chang, et al.2022, Wisdom et al. 2020). However, credit risk poses a significant concern for commercial organizations and may lead to a dire situation known as credit default. To mitigate credit risk, lenders conduct comprehensive assessments and verification of a borrower's capacity to fulfill their loan repayment obligations (Hassan et al., 2023).

These platforms employ predictive analytics to establish creditworthiness scores and limits, taking into account factors such as past business records, call logs, text messages, contact lists, age, education, and income (Kisutsa, 2021). Recently, researchers and lending institutions have turned to machine learning algorithms and neural networks for predicting an individual's credit score based on their historical data. Through this approach, the system can also identify potential credit default candidates and reject their loan applications. Despite the associated risks, lending remains a fundamental aspect of financial organizations and lenders strive to minimize credit risk to ensure the smooth operation of their businesses.

This paper examined the specific characteristics of financial data that influence the possible likelihood of loan defaults, using comprehensive data analysis, the study explored the relationships between different attributes and loan defaults to identify the most suitable features for training a predictive model.

2 REVIEW OF RELATED STUDIES

This section presents a review of related studies, credit predicting failures, emerging credit concerns and ml tools, traditional credit risk assessment, discriminant analysis and judgment-based models as follows.

2.1 PREDICTING CREDIT FAILURE CHALLENGES

Predicting credit failure challenges, often referred to as credit risk assessment or credit scoring, is a crucial task in the financial industry. The challenges associated with predicting credit failure are multifaceted and involve a combination of statistical, financial, and economic factors. This section presents some key challenges in the area as follows: Reliable data is the cornerstone of credit risk assessment. The accuracy, completeness, and timeliness of data can be a major challenge. Inaccurate or outdated information can lead to incorrect credit risk predictions, more especially with the emergences of machine learning

*Corresponding Author

Section B- ELECTRICAL/COMPUTER ENGINEERING & COMPUTING SCIENCES
Can be cited as:

Hassan, J.B., Awoyemi, J. O., Wisdom, D.D. and Ajayi, T.D (2025) Machine Learning-Based Credit Default Prediction: Challenges, Concerns, and Model Comparisons. FUOYE Journal of Engineering and Technology (FUOYEJET), 10(2), 264-270. <https://dx.doi.org/10.4314/fuoyejet.v10i2.12>

(ML) and Artificial Intelligences (AI) tools employed to ensure precision, efficiency with accuracy (Devi, & Radhika 2019; Wisdom et al., 2023). Credit failure events are usually rare compared to successful credit outcomes. This data imbalance can make it challenging to build robust models that accurately predict credit failures (Eneh, et al. 2022). Developing accurate credit risk models often involves dealing with complex statistical and machine learning techniques. Ensuring that these models are interpretable and explainable can be a challenge. Credit risk is influenced by broader economic conditions. Economic downturns can significantly impact the ability to predict credit failures accurately, as past data may not be representative of future risks. Financial institutions must adhere to various regulations and compliance standards when assessing credit risk. These regulations can vary from one region to another and may affect the data and methodologies used (Ajayi et al. 2020). Selecting the most relevant features (variables) for credit risk assessment can be challenging. There is often a vast pool of potential variables to choose from, and selecting the right ones is critical. Credit risk models need to be rigorously validated to ensure they perform as expected. This includes assessing model performance over time and addressing issues such as model drift. Credit risk is not a static concept. Borrowers' creditworthiness can change over time due to various factors, including changes in their financial situation and life events. Models must be able to adapt to these changes (Garzon, et al. 2020). Ensuring that credit risk models are fair and unbiased is crucial. There is an ongoing challenge in mitigating biases in models, especially regarding race, gender, and other protected attributes. The rise of complex machine learning models like neural networks makes it difficult to explain and interpret decisions. In highly regulated industries like finance, understanding and justifying model predictions is essential. As credit risk assessments increasingly rely on digital data and technology, protecting sensitive information from cybersecurity threats is vital (Wisdom, et al. 2024). Data breaches can have serious implications for credit risk prediction. Addressing these challenges requires a combination of sound data governance, model development practices, regulatory compliance, and ongoing monitoring and validation. Financial institutions and other organizations involved in lending and credit risk assessment must continually adapt their methods to stay current with evolving economic conditions and technological advancements.

2.2 EMERGING CREDIT CONCERNS AND ML TOOLS

Emerging credit concerns refer to the issues and challenges that arise in the world of finance and lending when it comes to assessing the creditworthiness of individuals and businesses. These concerns are particularly significant in the context of financial institutions, as they must strike a balance between offering access to credit and managing the risks associated with lending. (Hassan, et. al. 2023, Wisdom et al. 2021.) Machine learning tools have emerged as a powerful and effective way to address and mitigate these emerging credit concerns.

These challenges have greatly impacted lending

institutions such as banks and other Procure-to-Pay (P2P) lending organizations, resulting in low annual turnover, excessive depths and cash trap (Wisdom et al., 2024b). Such that, despite having sufficient cash, organizations are unable to generate enough revenue or profits to sustain themselves and grow, leading to inefficient operations or poor cash management, with a cycle of stagnation and decline.

The key points to consider when discussing the use of machine learning (ML) tools in addressing emerging credit concerns are namely:

1. **Data-Driven Decision Making:** One of the primary benefits of machine learning in credit assessment is its ability to process and analyze vast amounts of data. Traditional credit scoring models often rely on limited financial information, such as credit scores and income. Machine learning algorithms, on the other hand, can consider a wide range of alternative data sources, including social media activity, transaction history, and more. This helps financial institutions make more informed and accurate lending decisions (James, 2017).

2. **Risk Assessment:** Machine learning models can be used to assess credit risk more accurately. These models can identify subtle patterns and trends in historical data that might be missed by traditional methods. By analyzing factors like payment history, loan duration, and external economic indicators, machine learning algorithms can provide a more comprehensive understanding of a borrower's credit risk.

3. **Fraud Detection:** Machine learning can play a vital role in fraud detection and prevention. It can detect anomalies and suspicious activities in real-time, which is crucial in mitigating emerging credit concerns related to fraudulent applications or transactions. Advanced fraud detection models can adapt to evolving fraud techniques and protect lenders from financial losses (Wisdom et al. 2021).

4. **Credit Scoring Models:** Machine learning enables the development of custom credit scoring models that can be tailored to specific populations or industries. These models can help address emerging credit concerns by being more adaptable and responsive to changing economic conditions or market dynamics.

5. **Financial Inclusion:** Machine learning tools can promote financial inclusion by evaluating the creditworthiness of individuals with limited or no traditional credit history. By analyzing alternative data sources, such as utility payments or rental history, machine learning can extend credit access to underserved populations (Wisdom et al. 2021)

6. **Dynamic Decision-Making:** Credit assessment is not a static process, and emerging credit concerns may arise due to economic downturns or other external factors. Machine learning models can continuously learn and adapt to changing conditions, allowing financial institutions to make dynamic lending decisions and adjust credit policies in real-time (Isaac et al. 2020).

7. **Regulatory Compliance:** Financial institutions must comply with various regulations related to lending practices. Machine learning models can help automate and streamline compliance processes by ensuring that lending decisions adhere to regulatory guidelines, which can help mitigate potential concerns related to regulatory violations (Wisdom et al. 2021).

8. Interpretable Models: While machine learning can significantly enhance credit assessment, it's essential to use interpretable models to ensure that lending decisions can be explained and justified. Explainable AI (XAI) methods can help provide transparency and accountability, reducing concerns about the "black box" nature of some machine learning algorithms (Li, & Li 2022). Finally, emerging credit concerns in the financial industry can be effectively addressed by leveraging machine learning tools. These technologies provide more accurate risk assessment, better fraud detection, and improved financial inclusion while allowing institutions to adapt to changing economic conditions and regulatory requirements. However, it's crucial to strike a balance between innovation and transparency to build trust and confidence in machine learning-powered credit assessments.

2.3 TRADITIONAL CREDIT RISK ASSESSMENT

The conventional credit scoring method relies on historical data, comprising bank transactional data like previous credit transactions, records of delayed payments, credit bureau inquiries, and commercial data like financial statements and duration of credit history (Madaan, et al. 2021).

Table 1: Executive Summary of Related Studies

2.4 DISCRIMINANT ANALYSIS

Discriminant analysis, initially introduced by (Audu, et al.2020), is a statistical method that continues to be widely employed for segregating two groups, particularly in the context of categorizing customers as either having good or bad credit. This technique is extensively utilized in credit scoring to categorize group variables into distinct classes.

2.5 JUDGMENT-BASED MODELS

Diverse techniques, such as the analytic hierarchy process (AHP), are utilized in the creation of models that rely on human judgment. AHP is a structured methodology that simplifies intricate decision-making by dissecting it into more manageable sub problems (Zhang, et al. 2019). In AHP, human assessments are integrated with data to carry out the evaluations.

2.6 OVERVIEW OF MACHINE LEARNING

Arthur Samuel coined the term "Machine Learning (ML)," as the field that focuses on enabling computers to acquire knowledge without explicit programming. Machine Learning operates as a subset of AI, aiming to comprehend the structure of data and integrate it into models that can be applied by humans (Tagliaferri 2017; Setiawana et al. 2019; Wisdom et al., 2025). Machine learning implementations are categorized into three primary types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning entails the process of learning a function that links input to output, guided by labeled data, with the algorithm adjusting the model according to errors. Unsupervised learning, on the other hand, involves deducing patterns from input data lacking labels and revealing hidden structures within a dataset (Alhasan et al., 2020). Reinforcement learning, the third category,

centers on the maximization of rewards by instructing an agent to interact with the environment.

3. METHODOLOGY

The emergent digital era is advancing in academic development with Artificial Intelligence (AI) and Machine Learning (ML) at the four front. This paper developed Machine Learning-Based Credit Default Prediction: Challenges, Concerns, and Model Comparisons. The research employed Random forest and Decision Tree for the study. The model used kaggle dataset for the training and preprocessing. And used decision Tree and Random Forest algorithm for the analysis and prediction, efficiently developing a classification model anticipating loan defaults before granting the loan respectively.

3.1 DECISION TREE

The Decision Tree is a supervised learning method primarily employed for classification, although it can also be used for regression tasks. This classifier consists of a tree structure, with internal nodes corresponding to dataset features, decision rules represented by branches, and outcomes at the leaf nodes. It is composed of two fundamental elements: Decision Nodes and Leaf Nodes. Decision nodes, which have multiple branches, make decisions by evaluating dataset features, while leaf nodes provide outputs without further branches (Java Point, 2021). This tree-like structure represents all potential solutions to a problem or decision based on specified conditions, starting with the root node and branching out. The CART algorithm is responsible for constructing this tree, and the decision tree works by posing a question and dividing the tree into sub-trees based on the response. The structure of a decision tree is visually depicted in Figure 2.1

3.2 DECISION TREE TERMINOLOGIES

- I. The root node initiates the decision tree, representing the entire dataset and dividing it into subsets.
- II. Leaf nodes are the final output nodes that cannot be divided further.
- III. Splitting involves dividing a decision or root node into smaller nodes based on specific conditions.
- IV. A branch or sub-tree is created by splitting a node.
- V. Pruning refers to the removal of unnecessary branches from the tree.
- VI. The root node is the parent node, while other nodes are considered child nodes.

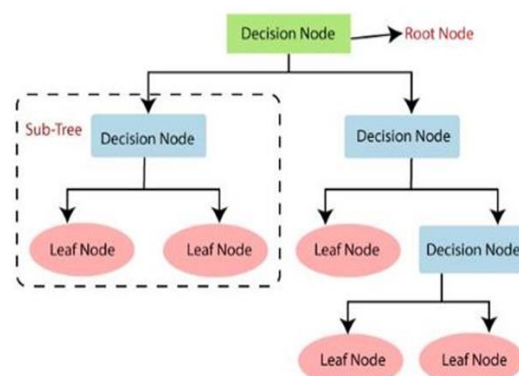


Fig 1. Decision Tree

3.3 How the Decision Tree algorithm Works

- i. When employing a decision tree for classifying a particular dataset, the process commences by positioning itself at the tree's root node. It then proceeds to evaluate the attribute value at the root in relation to the dataset record's attribute, and based on this evaluation, it selects a branch and advances to the subsequent node.
- ii. As the algorithm advances to the subsequent node, it iterates the attribute comparison within the sub-nodes, continuing this procedure until it eventually reaches the final leaf node of the tree.

ADVANTAGES

- i. Decision Trees have the ability to imitate human reasoning when making a decision, which makes them easy to comprehend.
- ii. The decision tree's rationale can be easily comprehended due to its tree-like structure.

3.4 RANDOM FOREST

The Random Forest is a supervised machine learning technique that utilizes ensemble learning to address complex problems and enhance model performance. Its approach involves generating multiple decision trees using various subsets of the dataset and then merging their results to enhance the dataset's predictive accuracy (JavaPoint, 2021). This algorithm aggregates predictions from each individual tree and arrives at its final prediction by tallying the most frequently occurring prediction among these forecasts. Increasing the number of trees in the forest improves accuracy and serves as a defense against overfitting problems, as illustrated in Figure 2.2, which depicts the Random Forest.

3.5 ADVANTAGES AND DISADVANTAGES

- i. in comparison to other algorithms, the training time required is shorter.
- ii. Even when handling large datasets, it can efficiently predict outputs with high accuracy.
- iii. It is capable of maintaining accuracy even when a significant amount of data is missing.
- iv. Random Forests are resilient and can perform well in both regression and classification tasks.
- v. RF algorithms are effective with large datasets and various data types, including numerical, binary, and categorical.
- vi. However, when the number of trees is high, the complexity and computational time are relatively increased, leading to a longer training time. Additionally, the sampling of subsets may introduce bias.

4 APPLICATIONS

Random Forests can be utilized in virtually any classification or regression scenario, although they are frequently employed in areas such as Remote Sensing, predicting stock market trends, identifying fraudulent activity, analyzing sentiment, and recommending products.

4.1 EXTRA TREES CLASSIFIER

The Extra Trees Classifier, akin to the Random Forest Classifier, is an ensemble machine learning technique, but it distinguishes itself through its approach to constructing decision trees. It generates several unpruned decision trees from the training dataset, where each tree employs

a random subset of k features at every testing node to divide the data (Gupta, 2020). Subsequently, the algorithm combines the predictions from these trees to produce the ultimate classification outcome, relying on majority voting for classification tasks or an arithmetic average for regression tasks.

4.2 Extra Trees has two main advantages:

- it reduces bias by sampling from the entire dataset during the construction of the trees, which prevents the introduction of different biases that may arise when different subsets of the data are used.
- it reduces variance by randomizing the splitting of nodes within the decision trees, thereby preventing the algorithm from being heavily influenced by specific features or patterns in the dataset.

4.3 Applications

Extra Trees, much like Random Forests, find application in both classification and regression tasks. Furthermore, in specific scenarios, Extra Trees are employed for the purpose of feature selection as well.

4.4 Extra Trees versus Random Forest

The two ensembles, Random Forest and Extra Trees, share many similarities. They both consist of a significant number of decision trees, and finally the overall decision is made according to the predictions of every tree. In classification problems, this is achieved through majority voting, while in regression problems, it is done through the arithmetic mean.

4.5 DIFFERENCES BETWEEN EXTRA TREES AND RANDOM FOREST

1. In Random Forest, bootstrap replicas are employed, indicating that it subsamples the input data with replacement, whereas Extra Trees utilizes the entire original dataset.

2. These two algorithms diverge in their methods for selecting cut points to split nodes. Random Forest opts for the optimal split, whereas Extra Trees opts for random selection. Nevertheless, both algorithms choose the best split point from subsets of features.

Consequently, Extra Trees introduces randomness to mitigate both bias and variance while still optimizing the outcome.

3. Using the complete original sample, as opposed to a bootstrap replica, diminishes bias, while the random selection of split points for each node diminishes variance.

4. When it comes to computational efficiency, the Extra Trees algorithm is quicker because it randomly selects split points instead of computing the optimal ones, leading to time savings during the process.

4.6 COMPARATIVE ANALYSIS OF LOAN DEFAULT PREDICTION FOR DECISION TREES AND RANDOM FOREST

The primary objective of this study was to create machine learning models capable of forecasting whether an individual should be approved for a loan, streamlining

the loan selection process for banking authorities. The research involved a performance evaluation of the Random Forest and Decision Trees algorithms using the same dataset, and the results indicated that Random Forest exhibited greater accuracy in predicting loan eligibility (Madaan, et al. 2021). Figure 2 below illustrates a typical loan default prediction. The researchers underscored the escalating rate of loan defaults in the banking sector, which presents challenges in evaluating loan applications and mitigating the risk of loan default.

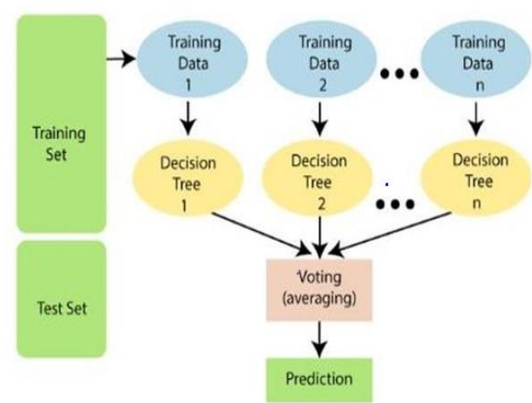


Fig 2. Loan default prediction

5 PROPOSED CREDIT DEFAULT PREDICTION MODEL

This paper proposed a Credit default prediction model depicted in Figure 3 in the proposed model a unique confusion matrix, decision Tree confusion Matrix, Random Forest confusion matrix as well as Extra Tree Confusion Matrix were used to depict the results in Figure 4, 5, and 6. Decision Tree and Random forest algorithm were preferred and used for this study. Decision Trees and Random Forests are ideal for credit default prediction because they handle both numerical and categorical data efficiently while capturing complex decision boundaries. Random Forest, an ensemble of Decision Trees, improves prediction accuracy and reduces overfitting by averaging multiple tree outcomes, making it more robust for financial risk assessment. Additionally, both algorithms provided feature importance rankings, helping financial institutions identify key factors influencing credit default. The accuracy, precision, recall, F1 Scores and support were depicted with their individual percentage (%) in figure 4, 5 and 6 respectively. Figure 4 depicts the Decision Tree Confusion matrix, Figure 5 depicts the Random Forest Confusion matrix and Figure 6 depicts the Extra Tree Classifier Confusion Matrix respectively.

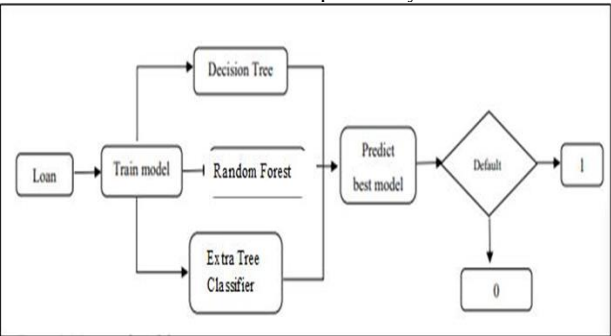


Fig 3. Proposed Credit default prediction Model

Credit default prediction system was proposed in (Hassan, et al. 2023). In the proposed scheme, a confusion matrix displays the ratio of accurately classified and inaccurately classified items in each category, presenting a comprehensive view of the test results.

The results are rounded to the nearest whole number.

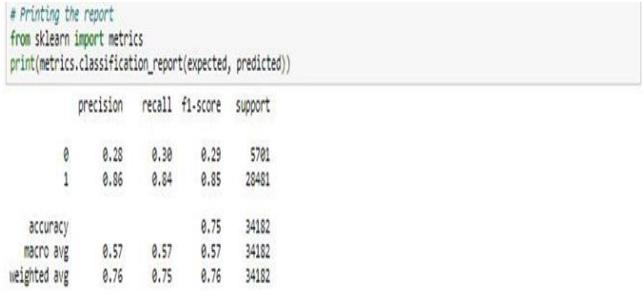


Fig 4: Decision Tree Confusion matrix

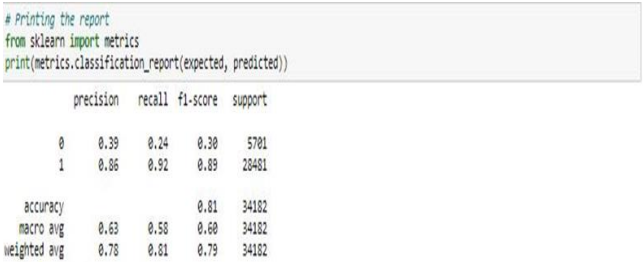


Fig 5: Random Forest Confusion matrix

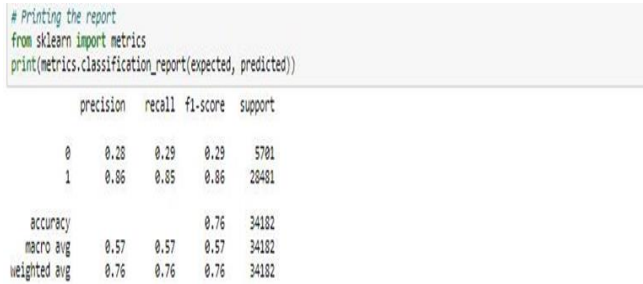


Fig 6: Extra Tree Classifier Confusion Matrix

6 CONCLUSION

In the era of digital technology, there is a growing interest in leveraging machine learning techniques to enhance profitability within the financial sector, particularly in domains such as risk evaluation, credit scoring, and bankruptcy prediction. This paper, designed to uncover the specific characteristics of financial data that influence the likelihood of loan defaults. Through comprehensive data analysis, the study investigates the relationships between different attributes and loan defaults to identify the most appropriate features for training a predictive model. Subsequently predicting credit failure challenges and emerging credit concerns using machine learning tools, the research employed three distinct machine learning algorithms which were evaluated using both training and testing datasets, and crucial performance metrics were utilized to determine which algorithm provides the most precise predictions for loan defaults.

Future directions for improving credit default prediction includes integrating feature engineering techniques, such as incorporating alternative data sources (e.g., transaction history or social behavior), to enhance predictive power. Additionally, hybrid models combining Random Forest with deep learning or gradient boosting algorithms (e.g., XGBoost) can improve accuracy and generalization. Lastly, explainable AI (XAI) techniques may be explored to ensure model transparency and regulatory compliance in financial decision-making.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for the constructive comment which has improved this paper, however this work was not supported by any grant.

REFERENCES

- Alhasan, S. Ajayi, E. A. Wisdom, D.D. (2020) a Comparative Performance Study of Machine Learning Algorithms, for Efficient Data Mining Management of Intrusion Detection Systems, *International Journal of Engineering Applied Sciences and Technology*, vol. 5, issue 6, issn no. 2455-2143, pages 85-110.
- Audu, A., Olawoyin, I. O., Zakari, Y., Wisdom, D. D., Muili, J., & Ndatsu, A. M. (2020). Regression-cum-exponential ratio imputation class of estimators of population mean in the presence of non-response. *Science Forum (Journal of Pure and Applied Sciences)*, 20(1), 58-58.
- Akitunde, A. (2020). Digital Lending in Nigeria. Retrieved from LinkedIn: <https://www.linkedin.com/pulse/digital-lending-nigeria-akitobi-akitunde-sfc-ssyb-acib>.
- Aslam, U., Aziz, H. T., Sohail, A., & Batcha, N. K. (2019). An Empirical Study on Loan Default Prediction Models. *Journal of Computational and Theoretical Nano science*, 16, 3483-3488.
- Hassan, J.B, Wisdom, D. D., Odufuwa, T. T., & Ayetuoma, I. O. (2023). Credit Default Prediction System using Machine Learning. *COMPUTOLOGY: Journal of Applied Computer Science and Intelligent Technologies*, 3(1), 51-60. Journal Press India. DOI: 10.17492/computology.v3i1.2305.
- Chang, A. H., Yang, L. K., Tsaih, R. H., & Lin, S. K. (2022). Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. *Quantitative Finance and Economics*, 6(2), 303-325.
- Wisdom, D. D., Igulu, K., Esther, O. O., Baba, G. A., Ahmad, A., & Sidi, A. (2021). A Review of Best Practices and Methods of Mitigating Cybersecurity Risks in Healthcare Systems and a Newly Proposed Algorithm. *Computology: Journal of Applied Computer Science and Intelligent Technologies*, 1 (2), 27-36. DOI: 10.17492/computology.v1i2.2104
- Cinca, C. S., Begoña, G. N., & Palacios, L. L. (2019). Determinants of Default in P2P Lending. *PLOS ONE*, 10.
- Devi, S., & Radhika, Y. (2019). A Survey on Machine Learning and Statistical Techniques in Bankruptcy Prediction. *International Journal of Machine Learning and Computing*, 8.2019.
- Garzon, M. J., Arroyo, J., Caparrini, A., & Segovia-Vargas, M.-J. (2020). Explainability of Machine Learning Granting Scoring Model in Peer to Peer Lending. *Open Access Journal*, 8.
- Kisutsa, G. T. (2021). Loan Default Prediction Using Machine Learning: A Case of Mobile Lending. University of Nairobi, College of Biological and Physical Sciences, School of Computing and Informatics M.Sc. Thesis
- Wisdom, D. D., & Vincent, O. R. (2024). Cybersecurity concerns and risks in emerging healthcare systems. *Cybersecurity in Emerging Healthcare Systems*, Institute of Engineering and Technology (IET), UK. https://doi.org/10.1049/PBHE064E_ch1
- James, F. (2017). Machine Learning in Credit Risk Modeling: Efficiency should not come at the expense of Explainability, published by Crowd Process Inc, 1177 Avenue of the Americas 10036 NY, USA.
- Java Point. (2021). Decision Tree Classification Algorithm. Retrieved from Java Point: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.
- Li, D., & Li, L. (2022). Research on Efficiency in Credit Risk Prediction Using Logistic SBM Model. *Wireless Communications and Mobile Computing*, 2022, 11.
- Madaan, M., Kumar, A., Keshri, C., Jain, R., & Nagrath, P. (2021). Loan Default Prediction Using Decision Trees and Random Forest: A Comparative Study. *Materials Science and Engineering*, 1022.
- Setiawana, N., Suharjitoa, & Dianab. (2019). A Comparison of Prediction Methods for Credit default on Peer-to-Peer Lending using Machine Learning. 4th International Conference on Computer Science and Computational Intelligence 2019, 157, 38-45.
- Wisdom, D. D., Vincent, O. R., Igulu, K. T., Arowolo, M. O., Hyacinth, E. A., Christian, A. U., Oduntan, O. E., Baba, G. A., (2023). Mitigating Cyber Threats in Healthcare Systems: The Role of Artificial Intelligence and Machine Learning, *Artificial Intelligence and Blockchain Technology in Modern Telehealth Systems*. Institute of Engineering and Technology (IET), UK.
- Shi, S., Tse, R., Luo, W., D'Addona, S., & Pau, G. (2022). Machine Learning-driven Credit Risk Prediction Systems: A Systemic Review. *Neural Computing and Applications* (34), 14327- 14339.

Tagliaferri, L. (2017). An Introduction to Machine Learning. Retrieved from DigitalOcean: <https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning>.

Wisdom, D. D., Vincent O.R, Christian, A. U., Igulu, K., Hyacinth, E. A., Odunayo, O. E. & Umar, B (2024b). Industrial IoT Security Infrastructure and Threats. In A. Prasad, T.P. Singh and S.D. Sharma (Eds.), Communication technologies in IoT and their security challenges: Present and future. Springer Nature.

Zhu, L., Qiu, D., Ergu, D., Ying, C., & Liu, K. (2019). A study on predicting Loan default based on the Random Forest Algorithm. 7th International Conference on Information Technology and Quantitative Management, 162, 503-513

Eneh, A. H., Wisdom, D. D., Wisdom, D. D., Arinze, U. C., Philibus, E., & Philibus, E. (2020). An Optimized Database Management System. In Proceedings of the LASUSTECH 30th Multidisciplinary Innovation Conference, Series 30 Vol-2 (p. 25). Lagos, Nigeria. DOI:10.22624/AIMS/iSTEAMS/LASUSTECH2022V30P25

Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2019). Research on Credit Scoring by fusing social media information in Online Peer to Peer Lending. Information Technology and Quantitative Management, 91, 168-174.

Wisdom, D. D., Vincent, O. R., Igulu, K., Christian, A. U., Hyacinth, E. A., Baba, G. A., Esther, O. O. (2025). The Protection of Industry 4.0 and 5.0: Cybersecurity Strategies and Innovations, CI-Industry-4.0: Computational Intelligence in Industry 4.0 and 5.0 Applications. Taylor and Francis group.