Financial Risk and Management:
An International Journal
Volume 2, Number 1, 2023

Artificial Intelligence Driven Credit Scoring: Opportunities and Risk Implications for Financial Institutions

Syakira Gavrila Haerus[1]

[1] Telkom University, Bandung, Indonesia

Abstract

Article history:
Received: January 7, 2023
Revised: February 21, 2023
Accepted: April 8, 2023
Published: June 30, 2023

Keywords:
Algorithmic Fairness,
Artificial Intelligence,
Credit Scoring,
Financial Inclusion,
Machine Learning.

Identifier:
Zera Open
Page: 29-42
https://zeraopen.com/journal/frmij

This study employs a literature review to examine how artificial intelligence and machine learning are transforming credit scoring and credit risk management in financial institutions. It synthesizes evidence on artificial intelligence model performance, the role of alternative data for "thin file" and unbanked borrowers, and implications for explainability, fairness, and risk governance. The findings show that neural networks, gradient boosting, random forests, and other techniques consistently outperform traditional logistic regression scorecards in predicting default and loss, while alternative data such as digital footprints, transactional records, and platform activity help expand access to credit and support more inclusive lending. At the same time, high-dimensional "black box" models raise concerns around model opacity, privacy, and data governance, and recent work documents "predictably unequal" outcomes across demographic groups. The review concludes that artificial intelligence-driven credit scoring generates an efficiency inclusion risk trade-off and highlights the need for explainable artificial intelligence tools, fairness-aware modelling, and robust regulatory and governance frameworks to ensure that benefits do not come at the expense of consumer protection and prudential stability.

1. Introduction

Artificial intelligence (AI) and machine learning (ML) are transforming credit scoring by enabling financial institutions to process high dimensional data, capture nonlinear relationships, and update risk assessments in near real time. Compared with traditional logistic regression scorecards, ML based models such as gradient boosting, random forests, and neural networks have been shown to deliver significantly higher predictive accuracy in default prediction and loss estimation, especially when combined with alternative data sources such as digital footprints and transactional behavior (Bazarbash, 2019; Berg et al., 2020; Breeden, 2021). These performance gains promise tangible benefits for financial institutions, including improved portfolio quality, more granular risk based pricing, and lower operational costs in credit underwriting.

AI driven credit scoring also opens new avenues for financial inclusion. By leveraging non-traditional data and advanced pattern recognition techniques, lenders can evaluate "thin file" or previously unbanked customers who lack formal credit histories, expanding access to credit in both advanced and emerging markets (Bazarbash, 2019). Evidence from recent applications in consumer and SME lending suggests that AI-enabled models can increase approval rates while maintaining or even reducing default rates, thereby supporting more inclusive yet profitable lending strategies (Breeden, 2021). At the same time, supervisors and industry bodies increasingly view AI as a strategic tool for strengthening credit risk management and stress testing frameworks, provided that appropriate governance and validation mechanisms are in place (Bholat & Susskind, 2021).

However, the deployment of AI-driven credit scoring also introduces new and complex risk implications for financial institutions. High dimensional, nonlinear models often operate as "black boxes," making it difficult for risk managers, auditors, and supervisors to understand drivers of model outputs and to challenge them effectively. This opacity has spurred growing interest in explainable AI (XAI) techniques that decompose model predictions at the level of features and individual borrowers, aiming to reconcile predictive performance with interpretability in credit risk management (Bussmann et al., 2021). At the same time, the use of granular personal data raises concerns around privacy, cyber-security, and data governance, with regulators emphasizing the need for robust controls over data lineage, model risk, and operational resilience (Truby, 2020; Bholat & Susskind, 2021).

A further source of concern is algorithmic fairness. Empirical evidence indicates that while AI based credit scoring can improve overall accuracy, it may also amplify existing disparities across demographic groups if historical biases embedded in data are not addressed (Bono et al., 2021). This has led policymakers and scholars to argue for proactive regulatory approaches that treat AI in credit decisioning as a high-risk application, requiring explicit fairness metrics, bias mitigation strategies, and enhanced accountability from financial institutions. Against this backdrop, this study examines how AI reshapes credit risk assessment, the benefits it offers for efficiency and inclusion, and the emerging challenges it poses for model risk, fairness, and prudential regulation.

2. Literature Review

Recent empirical work on AI-based credit risk modelling shows that machine and deep learning techniques can substantially outperform traditional scorecards in predicting default probabilities and loss rates. Using bank portfolio data, Addo et al. (2018) demonstrate that neural networks and gradient-boosting models generate more accurate probability of default estimates than logistic regression, particularly when non-linear interactions and higher order effects are important. In the context of peer-to-peer lending, Ariza-Garzón et al. (2020) find that boosted trees and other non-linear algorithms not only improve classification accuracy but also capture structural breaks and dispersion in borrower risk, suggesting that AI models are better suited to dynamic credit markets than static scorecards. Building on these results, Tyagi (2022) compares several machine learning algorithms for credit scoring and reports that ensembles such as XGBoost and random forests consistently deliver higher discriminatory power and more stable risk rankings across different market conditions.

A second strand of research focuses on alternative data and financial inclusion. Using proprietary data from a large fintech lender in India, prior research shows that mobile phone digital footprints such as app usage, social connections, and communication patterns can substitute for traditional bureau scores and enable profitable lending to borrowers with limited formal credit histories. Complementing this micro evidence, a World Bank ICCR study documents how transactional, utility, and platform data are increasingly integrated into credit risk assessment frameworks worldwide, expanding access for unbanked and underbanked segments while raising new challenges around data quality and consumer protection. At a more macro level, Philippon (2019) argues that fintech and big data credit scoring can reduce intermediation costs and narrow financial access gaps, but warns that market power and opaque algorithms may offset inclusion gains if regulatory oversight is weak.

Given the opacity of high-dimensional models, a growing body of work investigates explainable AI (XAI) in credit scoring. Gramegna and Giudici (2021) evaluate SHAP and LIME explanations for SME credit risk models and show that these tools can meaningfully decompose complex predictions into feature level contributions, helping lenders validate whether AI models rely on economically sensible drivers. De Lange et al. (2022) develop an XAI framework for bank credit assessment and report that combining gradient boosting with SHAP based explanations achieves a favourable trade off between predictive accuracy and interpretability, sufficiently transparent for use in regulated environments. Davis et al. (2022) reach similar conclusions for home equity lending, illustrating how rule based models, tree ensembles, and post-hoc explanation methods can be tailored to the information needs of lenders, regulators, and borrowers. More broadly, the financial risk literature emphasises that XAI should be embedded within risk based governance and model validation frameworks, rather than treated as a purely technical add on.

At the same time, distributional and fairness implications of AI driven credit scoring have become a central concern. Fuster et al. (2022) show that machine learning based mortgage models can increase overall predictive accuracy but also generate "predictably unequal" outcomes across demographic groups, as historical disadvantages encoded in data are propagated and sometimes amplified in credit allocations. Various studies propose statistical tests and diagnostic tools for assessing the fairness of credit scoring models, offering guidance on how lenders and supervisors can identify variables that drive disparate impacts. Complementing these contributions, Szepannek (2021) reviews alternative fairness definitions and develops a counterfactual-based approach for constructing risk scores that satisfy explicit fairness constraints while preserving as much predictive power as possible.

Overall, the literature indicates that AI-driven credit scoring can enhance predictive performance and support more inclusive lending through the use of alternative data, but only when accompanied by robust explain ability, fairness safeguards, and risk governance arrangements that address model risk, privacy, and regulatory compliance.

3. Methods

This study employs a systematic literature review (SLR) to synthesize existing evidence on artificial intelligence driven credit scoring and its implications for financial institutions. The review begins with the development of a clear research protocol specifying the main questions related to model performance, use of alternative data, financial inclusion, explain ability, fairness, and risk governance. A structured search strategy is then applied across major academic databases such as Google Scholar, Scopus, Web of Science, and SSRN, using combinations of keywords including “artificial intelligence,” “machine learning,” “credit scoring,” “credit risk,” “financial inclusion,” “explainable AI,” and “algorithmic fairness.”

Inclusion criteria focus on scholarly articles that examine AI or machine learning models in the context of credit scoring or credit risk assessment for financial institutions, covering both consumer and SME lending and encompassing empirical, conceptual, and methodological contributions. Studies that concentrate solely on technical algorithm development without clear financial or credit risk applications, non-financial domains, non-bank contexts unrelated to credit decisions, or duplicate publications are excluded.

The screening process is conducted in multiple stages, starting with title and abstract screening followed by full text review to ensure that only studies directly relevant to AI-based credit scoring and its risk implications are retained. For each selected study, key information is systematically extracted, including data sources, AI/ML techniques used, performance metrics, treatment of alternative data, approaches to explain ability and fairness, and discussion of governance, regulatory, and operational risk issues. The quality of the evidence is assessed using a structured checklist that considers clarity of research design, transparency of methods, robustness of analysis, and relevance to the research questions. The extracted data are then synthesized using a narrative and thematic approach, allowing the review to map the evolution of AI driven credit scoring, identify converging and diverging findings across studies, and highlight gaps and future research directions related to efficiency, inclusion, fairness, and prudential oversight.

4. Results and Discussion

The systematic review shows strong and consistent evidence that AI based credit scoring models outperform traditional logistic regression scorecards in predicting default and loss outcomes. Across bank portfolios and digital lending platforms, studies by Addo et al. (2018), Ariza-Garzón et al. (2020), and Tyagi (2022) demonstrate that neural networks, gradient boosting, random forests, and other ensemble methods deliver higher discriminatory power and more stable risk rankings than conventional models, especially in the presence of nonlinear interactions and complex borrower profiles. These empirical findings are in line with earlier work highlighting the superior predictive performance of ML based models when they are fed with high dimensional inputs and alternative data sources (Bazarbash, 2019; Berg et al., 2020; Breeden, 2021). Together, this body of evidence supports the view that AI driven credit scoring can improve portfolio quality, enable more granular risk-based pricing, and reduce underwriting costs for financial institutions.

At the same time, the results underscore that the performance gains of AI are closely tied to the use of alternative data and have important implications for financial inclusion. Micro level evidence from a large fintech lender shows that mobile-phone digital footprints such as app usage and social connections can substitute for traditional bureau scores and support profitable lending to "thin file" borrowers, echoing earlier findings that AI models can expand access for previously unbanked and underbanked segments (Bazarbash, 2019). This is complemented by global evidence from the World Bank ICCR, which documents how transactional, utility, and platform data are increasingly integrated into credit risk frameworks worldwide, and by Philippon (2019), who argues that big data credit scoring can reduce intermediation costs and narrow access gaps. However, these inclusion benefits come with trade-offs: the same expansion of data sources raises concerns around data quality, consumer protection, and the concentration of data and algorithmic power in a few large providers, suggesting that unregulated use of alternative data may erode some of the social gains from AI enabled inclusion (Philippon, 2019).

A third key result concerns model opacity and the growing role of explainable AI in credit risk management. High dimensional AI models often behave as "black boxes," creating challenges for model validation, internal risk governance, and supervisory scrutiny, as emphasized by both regulatory and academic work (Bholat & Susskind, 2021; Bussmann et al., 2021). In response, several studies evaluate XAI tools such as SHAP and LIME in real credit settings. Gramegna and Giudici (2021) show that these techniques can decompose complex SME risk models into intuitive feature level contributions, while de Lange et al. (2022) find that combining gradient boosting with SHAP explanations yields a favourable balance between accuracy and interpretability for bank credit assessment. Davis et al. (2022) reaches similar conclusions in home equity lending, demonstrating that post hoc explanations and rule-based summaries can be tailored to the information needs of lenders, regulators, and borrowers. These results indicate that XAI can partially mitigate black box concerns, but the literature also stresses that explanation tools must be embedded in broader, risk based governance and model validation frameworks rather than treated as a cosmetic add on.

The review also highlights that fairness and distributional impacts are now central to the debate on AI-driven credit scoring. Fuster et al. (2022) provide evidence that machine learning mortgage models can increase overall predictive accuracy while still generating "predictably unequal" outcomes across demographic groups, as historical disadvantages encoded in data are reproduced or amplified in credit allocations. Complementary work reviews alternative fairness definitions and develops counterfactual techniques for constructing risk scores that satisfy explicit fairness constraints while preserving as much predictive power as possible (Szepannek, 2021). In parallel, policy oriented studies show that algorithmic credit scoring can exacerbate existing disparities if biases in training data are not explicitly addressed, leading to calls for fairness metrics, bias mitigation procedures, and heightened accountability for financial institutions (Bono et al., 2021). Various contributions also propose statistical tests and diagnostic tools that help lenders and supervisors identify which variables drive disparate impacts, providing a practical foundation for fair-lending oversight in an AI environment.

Overall, the findings from this SLR suggest that AI-driven credit scoring offers a clear efficiency inclusion risk trade off. On the positive side, there is robust evidence that machine and deep learning models can enhance predictive performance, support more inclusive lending through the use of alternative data, and strengthen credit risk management and stress testing (Addo et al., 2018; Berg et al., 2020; Breeden, 2021; Bholat & Susskind, 2021) On the other hand, these benefits are conditional on the presence of strong governance arrangements that address model opacity, data governance, and algorithmic fairness. Without explainability frameworks, bias-control mechanisms, and appropriate regulatory oversight, the same technologies that improve risk measurement can undermine consumer protection, entrench discrimination, and create new forms of model and operational risk (Philippon, 2019; Bussmann et al., 2021; Szepannek, 2021; Bono et al., 2021; Fuster et al., 2022). This tension points to an important agenda for future research and policy: designing AI enabled credit systems that jointly optimize predictive accuracy, financial inclusion, and fairness within a prudent risk-governance framework.

5. Conclusion

The review concludes that artificial intelligence driven credit scoring fundamentally reshapes how financial institutions assess credit risk, combining higher predictive accuracy with the potential to broaden financial inclusion. Machine and deep learning models consistently outperform traditional scorecards in distinguishing between good and bad borrowers, especially when they exploit high dimensional inputs and alternative data such as digital footprints, transactional records, and platform activity. These capabilities allow lenders to refine risk based pricing, improve portfolio quality, and reduce underwriting costs, while also extending credit to "thin file" and previously unbanked customers who lack conventional credit histories. In this sense, AI based credit scoring is not just a technical upgrade but a strategic tool for building more efficient and inclusive credit markets.

At the same time, the findings highlight that these benefits are accompanied by significant challenges related to model opacity, data governance, and fairness. Complex AI models often function as “black boxes,” making it difficult for institutions and supervisors to understand or challenge individual decisions, which drives the need for explainable AI techniques embedded within robust governance and validation frameworks. The expansive use of granular personal data raises concerns over privacy, cybersecurity, and the concentration of informational and algorithmic power, while evidence of “predictably unequal” outcomes across demographic groups underscores the risk that AI may reinforce or amplify existing inequalities if historical biases in data are not actively addressed. Overall, the study emphasizes that realizing the full promise of AI-driven credit scoring requires an integrated approach that balances predictive accuracy and financial inclusion with strong safeguards for consumer protection, fairness, and prudential stability, and calls for future research and policy design focused on governance architectures that can support this balance.

References

Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., & Segovia-Vargas, M. J. (2020). Explainability of a machine learning granting scoring model in peer-to-peer lending. *IEEE Access, 8*, 64873–64890.

Bazarbash, M. (2019). *Fintech in financial inclusion: Machine learning applications in assessing credit risk*. Washington, DC: International Monetary Fund.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies, 33*(7), 2845–2897.

Bholat, D., & Susskind, D. (2021). The assessment: Artificial intelligence and financial services. *Oxford Review of Economic Policy, 37*(3), 417–434.

Bono, T., Croxson, K., & Giles, A. (2021). Algorithmic fairness in credit scoring. *Oxford Review of Economic Policy, 37*(3), 585–617.

Breeden, J. L. (2020). Survey of machine learning in credit risk. *Available at SSRN 3616342.*

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*(1), 203–216.

Davis, R., Lo, A. W., Mishra, S., Nourian, A., Singh, M., Wu, N., & Zhang, R. (2022). Explainable machine learning models of consumer credit risk. *SSRN Electronic Journal.*

De Lange, P. E., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for credit assessment in banks. *Journal of Risk and Financial Management, 15*(12), 556.

Fuster, A., Goldsmith‐Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance, 77*(1), 5–47.

Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk. *Frontiers in Artificial Intelligence, 4*, 752558.

Philippon, T. (2019). *On fintech and financial inclusion* (Working Paper No. 26330). National Bureau of Economic Research.

Szepannek, G., & Lübke, K. (2021). Facing the challenges of developing fair risk scoring models. *Frontiers in Artificial Intelligence, 4*, 681915.

Truby, J., Brown, R., & Dahdal, A. (2020). Banking on AI: Mandating a proactive approach to AI regulation in the financial sector. *Law and Financial Markets Review, 14*(2), 110–120.

Tyagi, S. (2022). Analyzing machine learning models for credit scoring with explainable AI and optimizing investment decisions. *arXiv preprint*, arXiv:2209.09362.