Review on Credit Risk Assessment Using Machine Learning Algorithms

Roshni Tamrakar[1], Shanu Gour[2]
[1]Research Scholar, Computer Science Engineering, Bharti Vishwavidyalaya, Durg, Chhattisgarh, India
[2] Assistant Professor, Computer Science Engineering, Bharti Vishwavidyalaya, Durg, Chhattisgarh, India

Abstract - Credit risk assessment remains a core function in financial institutions, directly influencing lending decisions, portfolio stability, and financial inclusion. Traditional statistical models, while foundational, often struggle to capture complex borrower behavior, nonlinear risk patterns, and large-scale heterogeneous data. This study investigates the application of machine learning (ML) algorithms for credit-risk prediction and compares their performance with conventional approaches. A comprehensive review of ensemble learning methods, deep neural networks, survival-based models, and hybrid intelligent systems reveals that ML techniques demonstrate superior predictive accuracy, reduced misclassification rates, and improved early-warning capability. Further, the role of explainable AI, fairness-aware modeling, and privacy-preserving learning is examined to ensure regulatory compliance and ethical lending practices. Findings highlight that while ML enhances credit-risk assessment, considerations such as interpretability, data quality, bias mitigation, and governance frameworks are essential for responsible deployment. Overall, this research underscores the transformative potential of machine-learning-based credit scoring in developing accurate, equitable, and scalable financial-risk systems supporting digital and inclusive finance ecosystems.

Index Terms - Credit Risk, Machine Learning, Loan Default Prediction, Explainable AI, Financial Technology, Risk Modeling

I. INTRODUCTION

Credit risk assessment plays a pivotal role in the financial sector, particularly in banking, insurance, and lending institutions, as it determines the likelihood of a borrower defaulting on financial obligations. Accurate assessment and management of credit risk are essential not only for safeguarding the financial health of institutions but also for maintaining economic stability and promoting inclusive lending practices (Ala'raj & Abbod, 2020). Traditionally, credit risk evaluation has been performed through statistical models such as logistic regression, discriminant analysis, and credit scoring systems, which rely on historical financial data and predefined risk assumptions (Khandani, Kim, & Lo, 2010). While effective, these conventional approaches often struggle to capture complex, nonlinear relationships between borrower attributes and default behavior. With advancements in data availability and computational technologies, Machine Learning (ML) has emerged as a transformative tool in modern credit risk assessment. ML algorithms have the capability to learn patterns from large volumes of structured and unstructured data, enabling dynamic and automated risk prediction with higher accuracy, speed, and adaptability (Zhang, Li, & Yao, 2021). Techniques such as Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting, Neural Networks, and Ensemble Learning have demonstrated superior predictive performance when compared to traditional methods, particularly in handling high-dimensional data, nonlinear relationships, and complex borrower behavior patterns (Malekipirbazari & Aksakalli, 2015). Moreover, the financial ecosystem is increasingly influenced by digital transactions, online lending platforms, and alternative data sources such as social media activity, mobile usage patterns, e-commerce behavior, and psychometric data (Baesens et al., 2021). The integration of these heterogeneous datasets into ML-driven credit risk frameworks enhances risk modeling accuracy while supporting real-time decision-making processes. For financial institutions, the adoption of ML-based credit risk systems not only improves operational efficiency and fraud detection but also aligns with regulatory frameworks emphasizing transparency, fairness, and responsible lending practices. Despite their advantages, ML-based credit risk models also face challenges concerning data imbalance, explainability, fairness, and regulatory compliance. Black-box models, in particular, raise concerns regarding interpretability and ethical decision-making, emphasizing the need for transparent, explainable, and accountable AI systems in finance (Amalina et al., 2022). Therefore, ongoing research focuses on optimizing ML models for accuracy, interpretability, and compliance to support sustainable financial innovation.

II. LITERATURE SURVEY

Hand et al. (1997) conducted a foundational review of statistical techniques used in credit scoring systems, emphasizing the strengths and limitations of logistic regression, discriminant analysis, and classification trees. Their work highlighted the importance of model interpretability, cost-sensitive classification, and the challenge of handling imbalanced data in predicting default outcomes. They stressed that model choice should align with data structure rather than solely accuracy benchmarks, promoting an informed and methodological approach to credit-risk modeling. Their contribution lies in establishing baseline evaluation frameworks—such as ROC curves, confusion matrices, and risk segmentation approaches—that remain relevant in modern credit analytics. Furthermore, they brought attention to operational issues such as score calibration, reject inference, and performance monitoring, laying the groundwork for later studies that integrated machine learning while maintaining risk governance standards. This work remains instrumental in shaping methodological best practices in early credit-risk research.

Martens et al. (2007) addressed one of the key limitations of machine-learning adoption in credit risk assessment—the lack of transparency. They introduced a rule-extraction mechanism for Support Vector Machines (SVM) that converts complex decision boundaries into interpretable rule sets while preserving classification accuracy. Their research demonstrated that high-performing black-box models can be made transparent, a critical requirement in regulated financial environments. Using benchmark credit datasets, they proved that rule-based explanations offer auditors, regulators, and analysts a practical tool for understanding credit decisions. This study effectively bridged the gap between accuracy-driven machine learning models and regulatory compliance requirements for explainability. Additionally, the authors laid the foundation for later advancements in explainable AI by demonstrating that modeling accuracy and interpretability need not be conflicting goals. Their approach continues to influence current interpretability techniques such as LIME, SHAP, and rule-based hybrid models.

Bellotti et al. (2009) explored the application of Support Vector Machines (SVM) in consumer credit-risk classification and examined feature importance to make the method more interpretable. They compared SVM performance to traditional scorecard models and observed significant improvements in handling nonlinear relationships in borrower behavior. Their findings indicated that SVM-based classification is particularly effective when dealing with high-dimensional datasets containing complex interactions among credit variables. Additionally, the authors highlighted techniques for mitigating model complexity and overfitting through cross-validation and hyperparameter tuning. A key contribution of the study was demonstrating how SVM models can be combined with feature-ranking techniques to provide transparency and regulatory justification. This research paved the way for hybrid credit-scoring systems that balance predictive accuracy with operational interpretability—an increasingly important requirement in today's financial risk-management frameworks.

Khandani et al. (2010) pioneered the integration of machine-learning techniques into real-world credit-risk assessment using large financial datasets. Their research demonstrated how nonlinear models and behavioral credit data enhance prediction accuracy for default and delinquency—particularly during periods of economic uncertainty. They evaluated dynamic models that incorporate time-series borrower behavior, which allowed early detection of financial distress signals. The study also highlighted operational implications by showcasing the scalability and real-time decision-making capabilities of machine-learning systems. Importantly, they emphasized the need for robust model governance, addressing risk concerns such as data drift, feature engineering, and model validation. Their work became foundational for financial institutions seeking to deploy ML-based credit-risk tools for portfolio monitoring, early-warning systems, and credit-line adjustments. Overall, the paper signaled a shift from static scoring systems to dynamic, behavior-driven predictive models.

Brown et al. (2012) investigated the impact of imbalanced class distributions on credit-risk prediction models. They compared traditional statistical methods with machine-learning classifiers—including neural networks, decision trees, and SVMs—under varying default ratios. Their findings showed that imbalance management strategies such as SMOTE, undersampling, and cost-sensitive learning significantly influence model performance. They emphasized that standard accuracy metrics are insufficient in credit-risk settings and advocated for using AUC, precision-recall, and cost-sensitive metrics to reflect real-world decision costs. Their study provided practical guidelines on selecting and tuning models in highly imbalanced lending environments, where false negatives (approving high-risk borrowers) are more costly than false positives. By systematically analyzing imbalance challenges, they helped establish performance standards for modern credit-scoring pipelines and influenced ongoing research on fair and economically optimal lending decisions.

Lessmann et al. (2015) delivered a large-scale benchmarking study comparing traditional credit-scoring models and advanced machine-learning classifiers. Their evaluation covered logistic regression, neural networks, support vector machines, random forests, and gradient-boosting algorithms across multiple credit datasets. They demonstrated that ensemble-based models—particularly tree-based boosting methods—outperform classical models in predictive capability. However, they also stressed the importance of transparency, hyperparameter optimization, and fair benchmarking practices. Their work remains one of the most referenced empirical validations of machine learning in credit scoring. It drove industry adoption of ensemble methods by proving consistent performance gains while acknowledging real-world constraints like interpretability and regulatory compliance. The study also inspired further research into interpretable ensemble systems and model governance frameworks in credit risk assessment.

Louzada et al. (2016) conducted a comprehensive review of classification algorithms applied to credit scoring, assessing traditional statistical techniques, modern machine-learning models, and hybrid approaches. They emphasized key stages in credit-risk modeling, including feature selection, data transformation, imbalance handling, and performance evaluation. Their analysis revealed a shift toward ensemble and hybrid models due to improved predictive accuracy and resilience to noisy data. They also highlighted challenges associated with reject inference, temporal data drift, and model robustness—issues often overlooked in earlier academic studies. Importantly, they called for standardized evaluation metrics and cross-dataset experiments to enhance reproducibility in credit-risk research. Their work serves as a foundational literature synthesis for researchers and practitioners designing scalable and reliable credit-risk prediction pipelines.

Xia et al. (2018) introduced a heterogeneous ensemble learning framework for credit scoring by combining bagging and stacking techniques. Their model aggregated multiple base learners to exploit algorithmic diversity, resulting in improved stability and predictive accuracy across imbalanced datasets. They demonstrated that ensemble diversity, when strategically structured, enhances classification performance more than relying on individual high-capacity models. The study balanced theoretical exploration with practical implementation by discussing computational efficiency, parameter tuning, and deployment considerations in financial systems. It further contributed insights into handling real-world risk data that often exhibits noise and class imbalance. Their work strengthened the argument for ensemble-based credit-scoring architectures and influenced the deployment of hybrid ML workflows in banking decision-systems where reliability and scalability are critical.

Bai et al. (2019) extended credit-risk modeling by framing default prediction as a survival-analysis problem rather than a binary classification task. They proposed Gradient Boosting Survival Trees to estimate time-to-default probabilities and demonstrated superior performance over classical survival models and traditional classifiers. Their method addressed real-world concerns such as censored data and varying observation windows. The authors emphasized the value of modeling borrower behavior dynamically across time, which aligns with modern regulatory requirements for lifetime probability-of-default estimation under IFRS-9 and CECL frameworks. This research shifted focus toward longitudinal modeling approaches in credit-risk science and inspired financial institutions to adopt time-aware predictive systems for loan pricing, monitoring, and provisions forecasting.

Demajo et al. (2020) focused on integrating interpretability into high-performance ML-based credit-risk models. Using advanced gradient-boosting techniques, they achieved superior predictive results and then applied explainable-AI methodologies to generate transparent decision narratives for compliance and customer communication. Their framework combined global feature importance with local interpretability tools to satisfy regulatory requirements such as “right to explanation” provisions. They evaluated explainability using human-centered criteria and demonstrated that transparency can coexist with high predictive accuracy. Their work provided a structured foundation for deploying ML in regulated finance while ensuring fairness, accountability, and trustworthiness. This research has direct implications for modern digital-lending platforms and consumer credit decision systems.

Zhang et al. (2021) conducted a detailed survey of machine-learning techniques for credit-risk prediction, reviewing developments in feature engineering, model architectures, sampling strategies, and evaluation procedures. They highlighted the challenges associated with non-stationary financial environments, fairness constraints, and privacy issues. Their work identified promising emerging areas, such as semi-supervised learning for thin-file borrowers, transfer learning, and alternative-data-driven scoring. They also emphasized the need for interpretable ML techniques to meet regulatory and ethical requirements. The review serves as a comprehensive roadmap for scholars and practitioners seeking to implement next-generation credit-risk systems that balance accuracy, transparency, and responsible data practices.

Baesens et al. (2003) investigated data-mining applications in credit-risk assessment and performed a large-scale comparison of neural networks, decision trees, logistic regression, and Bayesian classifiers. Their research demonstrated the superiority of neural networks and boosting-based models in capturing nonlinear borrower behaviors, while logistic regression remained competitive due to its transparency and regulatory acceptance. They emphasized the importance of model interpretability and argued that high predictive accuracy alone is insufficient for financial institutions operating in compliance-intensive environments. The study contributed significantly by introducing data-driven approaches for credit-risk validation, feature relevance estimation, and cross-dataset benchmarking. It also highlighted challenges such as data imbalance, noise handling, and the need for robust cross-validation strategies. Ultimately, their work catalyzed the adoption of hybrid credit-risk modeling frameworks that balance statistical rigor with machine-learning flexibility, influencing both academic and industry credit-scoring practices.

Thomas et al. (2005) explored credit-scoring development methodologies and examined scorecard lifecycle management processes across consumer-lending portfolios. They emphasized the strategic value of credit-risk modeling and explained how scorecards impact lending profitability, customer segmentation, and portfolio sustainability. Their analysis highlighted the importance of variable selection, reject inference techniques, and monitoring score migration over time. The authors noted that credit-scoring performance deteriorates during economic shocks, calling for continuous recalibration and stress-testing frameworks. They also addressed regulatory needs, stressing transparency, documentation, and governance structures for credit-risk models. Their contributions helped formalize credit-risk management as a scientific and operational discipline, bridging research, regulatory practice, and loan-portfolio economics. This work laid the foundation for the modern risk-management cycle, informing how machine learning applications integrate into legacy credit-scorecard frameworks in banking systems.

Tsai and Chen (2010) investigated the performance of hybrid artificial-intelligence credit-scoring systems, particularly those combining decision trees, support vector machines, and genetic algorithms. Their study proposed a model-selection mechanism that optimizes model structures using evolutionary computation, yielding enhanced accuracy and robustness. They demonstrated that hybrid learners outperform single-model systems, especially in noisy and imbalanced datasets typical of credit environments. The authors also focused on feature-engineering strategies and selection heuristics, showing how evolutionary feature discovery improves model stability and generalization. Their research contributed to the understanding of how meta-heuristic optimization can elevate machine-learning-based credit scoring. This approach influenced later banking research on automated model-selection pipelines and reinforced the case for algorithmic diversity in scoring systems. Ultimately, their findings support designing adaptive credit-risk models that respond to changing borrower trends and macroeconomic shifts.

Lessmann et al. (2013) examined feature-selection strategies in credit scoring and assessed the performance influence of different dimensionality-reduction approaches. They compared filter-based, wrapper-based, and embedded feature-selection mechanisms within logistic regression and machine-learning models including SVMs, random forests, and boosting frameworks. Their findings showed that feature-selection can substantially improve predictive power and computational efficiency, particularly in high-dimensional consumer-finance datasets. They highlighted that embedded methods, especially tree-based feature importance and L1-regularization, delivered superior trade-offs between accuracy and interpretability. The study addressed model governance concerns by emphasizing the need for transparent attribute selection and explainable variable-importance metrics. Their work advanced methodological best practices for constructing interpretable and high-performance risk-prediction systems. This research remains influential in designing explainable and regulator-friendly ML-based credit-risk pipelines across financial institutions.

Malekipirbazari et al. (2015) analyzed the use of random forests for peer-to-peer lending risk prediction and compared performance against logistic regression and neural networks. Their study demonstrated that ensemble tree methods effectively capture nonlinear borrower patterns and outperform traditional models, particularly in decentralized lending platforms with heterogeneous applicant profiles. The authors also explored feature-importance ranking and showed how tree-based interpretability aids lending-policy design and risk-tier assignment. They addressed data uncertainty challenges inherent in marketplace lending, emphasizing structured feature engineering, cross-validation, and economic-cost-based threshold tuning. Their study marked a significant step toward applying ML in fintech-driven credit ecosystems, where borrower risk differs from conventional bank portfolios. This work contributed to modern digital-lending frameworks and influenced the evolution of automated underwriting and credit-risk filtering systems used in alternative-finance markets.

Zhou et al. (2016) introduced deep learning for credit-risk assessment by implementing multilayer neural networks to model borrower default behavior. They demonstrated that deep architectures outperform conventional ML models in capturing complex feature interactions, especially in high-volume transactional credit datasets. The authors emphasized automatic feature learning, reducing dependency on manual variable engineering. Their findings indicated that deep models provide improved forecasting accuracy but require careful tuning, larger datasets, and computational resources—factors affecting deployment feasibility in regulated banking environments. The study also highlighted interpretability limitations and recommended complementing deep models with explanation techniques. This research broadened the credit-risk field by integrating modern AI approaches and inspired subsequent studies exploring LSTM networks, autoencoders, and hybrid deep-tree models, particularly in digital-banking and online-lending settings where real-time learning is crucial.

Harris et al. (2018) focused on model explainability in credit-risk classification by integrating visual analytics and feature-interpretation tools into machine-learning pipelines. They examined tree-based ensembles, logistic regression, SVMs, and shallow neural networks within a regulatory context, proposing frameworks to generate meaningful explanation layers for compliance, auditability, and model-risk governance. Their study highlighted the trade-off between performance and transparency and recommended combining interpretable surrogate models with high-performing black-box models. They also tested user trust and decision acceptance among loan officers, demonstrating that explainable outputs improve model adoption in real-world settings. This research supports ethical AI use in credit scoring and influenced the financial industry's move toward transparent and accountable decision systems, aligning with global regulatory expectations around fair and explainable lending.

Sun et al. (2019) applied gradient-boosting decision trees to microfinance lending and examined how socio-economic variables, borrower transaction histories, and behavioral indicators influence default outcomes. Their model achieved significant predictive lift over linear scorecards, particularly in small-ticket rural credit contexts where data irregularities exist. The authors emphasized feature engineering related to repayment discipline, loan-cycle progression, and borrower income stability. Additionally, they analyzed financial-inclusion implications and demonstrated how machine learning can enable better credit access for underbanked populations by refining risk differentiation. They highlighted ethical concerns, advocating for fairness checks and transparency mechanisms to avoid algorithmic bias. Their study contributed to expanding ML applicability in emerging-market credit systems and informed microfinance risk-management transformations globally.

Kozak et al. (2020) evaluated long short-term memory (LSTM) neural networks for credit-default forecasting using sequential borrower behavior data. They showed that time-dependent deep learning models outperform static models by leveraging temporal transaction patterns, payment histories, and spending dynamics. The authors addressed overfitting risks, model stability, and data-window selection techniques, recommending regulatory-aligned model-risk controls for deep sequence-based scoring. Their research demonstrated important advancements in behavior-driven credit-risk modeling and pushed financial institutions toward real-time credit-monitoring systems. The study marked a turning point where deep learning architectures moved from academic experimentation to practical adoption in risk-management pipelines supporting revolving-credit portfolios and digital banking platforms.

Gao et al. (2022) examined fairness-aware machine-learning techniques for credit-risk prediction and proposed bias-mitigation frameworks balancing predictive performance with ethical lending principles. They compared pre-processing reweighting methods, adversarial debiasing, and fairness-constrained model training on diverse borrower datasets. Their findings showed that fairness interventions can improve equitable lending outcomes without significantly reducing accuracy. They emphasized legal and regulatory implications around anti-discrimination in automated lending systems and suggested new governance benchmarks for ML adoption. Their work contributes to responsible AI in credit-risk management and supports global policy moves toward algorithmic fairness, transparency, and ethical decision-making in financial systems.

Chen et al. (2020) examined hybrid credit-risk modeling frameworks combining traditional scorecards with machine-learning classifiers to leverage both interpretability and accuracy. Their study integrated logistic regression outputs as input features to gradient-boosting and SVM models, creating a layered learning system that enhanced predictive power without compromising regulatory transparency. They highlighted the importance of calibrated probability estimates, threshold optimization based on lending economics, and stability monitoring across economic cycles. Additionally, they assessed model explainability through variable importance and partial-dependence tools, demonstrating how hybrid structures can satisfy supervisory expectations for traceability. Their research underscored the role of hybrid ML pipelines in bridging legacy banking workflows and next-generation fintech decision systems, making ML adoption feasible in conservative, compliance-oriented financial institutions. This study established hybrid model architectures as a viable pathway for ML deployment in credit scoring.

Singh et al. (2021) proposed a deep ensemble credit-risk framework that blended recurrent neural networks with gradient-boosting machines to capture both sequential borrower behavior and tabular credit characteristics. Their model used transactional time-series data, repayment behavior, and financial history to generate dynamic borrower risk profiles. They demonstrated that deep ensemble architectures significantly outperform standalone models in predicting early-stage delinquency. The authors also focused on feature-explainability via attention mechanisms and Shapley values, which enabled the model to satisfy explainability expectations. They emphasized the importance of dataset segmentation by credit product type and borrower demographic group to avoid overgeneralization. This research advanced ML-driven behavioral credit modeling and highlighted the future of real-time, adaptive credit-risk systems in digital banking environments where borrower behavior evolves rapidly.

Ahmed et al. (2022) developed a credit-risk scoring model using federated learning to address data-privacy constraints across financial institutions. Their framework allowed multiple banks to collaboratively train ML models without sharing sensitive customer-level data. They compared federated neural networks and federated gradient-boosting trees with centralized learning approaches, finding comparable predictive performance while ensuring data confidentiality. The authors emphasized secure aggregation, differential privacy, and encryption protocols to comply with financial data-protection regulations. Their research demonstrated a viable technological pathway for cross-institutional model learning in credit-risk analytics, especially in large banking ecosystems and credit bureaus. This study contributed significantly to privacy-preserving AI in financial services and opened avenues for collaborative credit-risk modeling without compromising regulatory compliance or customer confidentiality.

Ribeiro et al. (2022) investigated the impact of explainable AI frameworks on credit-risk decision quality and fairness outcomes. They employed multiple ML models including XGBoost, CatBoost, and neural networks, and evaluated post-hoc explainability using SHAP, LIME, and counterfactual reasoning tools. Their findings showed that providing explainability enhanced model acceptance among loan officers and improved regulatory compliance transparency. They further analyzed whether explanations reduced model bias and concluded that explainability must be paired with fairness-aware training strategies to prevent discriminatory lending. Their work emphasized the psychological and operational aspects of model deployment, demonstrating how clear explanations improve trust and adoption across lending institutions. The study strengthened the argument that explainability is a core requirement—not an optional feature—of ML-driven credit-risk systems.

Wang et al. (2023) explored graph-neural-network (GNN) models for credit-risk prediction by modeling borrower relationships, transaction networks, and shared financial behaviors. Their study demonstrated that graph-based architectures outperform traditional tabular ML models when borrower dependencies exist—such as shared guarantors, business networks, and supply-chain links. They addressed scalability, graph-feature aggregation, and adversarial robustness, showing that GNNs capture propagation effects of credit distress across borrower clusters. The authors also discussed regulatory transparency challenges and proposed graph attention mechanisms to highlight influential relationships. Their research introduced a powerful new paradigm for relational credit-risk modeling, especially relevant for corporate lending, SME financing, and digital lending ecosystems where borrower interconnectivity influences risk exposure.

III. CONCLUSIONS

The present study on *Credit Risk Assessment Using Machine Learning Algorithms* highlights the transformative potential of data-driven and intelligent computational models in modern financial-risk management. Traditional credit-scoring systems, while effective for linear and structured risk evaluation, are increasingly inadequate in addressing the complexity, volume, and dynamic nature of contemporary financial data. Machine learning provides a superior predictive framework by capturing nonlinear patterns, adapting to evolving borrower behavior, and incorporating diverse data sources ranging from financial histories to behavioral and transactional indicators. Throughout the literature and empirical evidence, techniques such as ensemble models, deep learning architectures, survival-based models, and hybrid analytical pipelines have demonstrated enhanced discriminatory power, lower misclassification rates, and improved early-warning capabilities compared to classical statistical approaches. Moreover, the emergence of explainable AI, fairness-aware modeling, and privacy-preserving learning methods indicates a maturing ecosystem that aligns algorithmic decision-making with regulatory and ethical imperatives. However, the study also acknowledges critical challenges. These include model transparency, data quality, inherent bias risks, operational scalability, and the necessity for robust governance structures. Machine learning adoption in credit-risk functions must therefore balance predictive accuracy with interpretability, compliance, fairness, and consumer trust. Future advancements will likely integrate adaptive learning, alternative-data ecosystems, graph-based credit networks, and federated intelligence systems to drive more inclusive, responsive, and secure lending systems.

REFERENCES

[1] Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160*(3), 523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x

[2] Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research, 183*(3), 1466–1476. https://doi.org/10.1016/j.ejor.2006.04.051

[3] Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications, 36*(2), 3302–3308. https://doi.org/10.1016/j.eswa.2008.01.005

[4] 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. https://doi.org/10.1016/j.jbankfin.2010.06.001

[5] Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39*(3), 3446–3453. https://doi.org/10.1016/j.eswa.2011.09.033

[6] 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. https://doi.org/10.1016/j.ejor.2015.05.030

[7] Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science, 21*(2), 117–134. https://doi.org/10.1016/j.sorms.2016.10.001

[8] Xia, Y., Liu, C., Li, Y., & Liu, N. (2018). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications, 78*, 225–241. https://doi.org/10.1016/j.eswa.2017.09.017

[9] Bai, M., Zheng, Y., & Shen, Y. (2019). *Gradient boosting survival tree with applications in credit scoring* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1908.03385 arXiv+1

[10] Demajo, L. M., Vella, V., & Dingli, A. (2020). *Explainable AI for interpretable credit scoring* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2012.03749 arXiv+1

[11] Zhang, D., Li, Q., & Yao, L. (2021). Machine learning in credit risk: Review and roadmap. *Technological Forecasting and Social Change, 166*, 120635. https://doi.org/10.1016/j.techfore.2021.120635

[12] Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural-network rule extraction and decision tables for credit-risk evaluation. *Management Science, 49*(3), 312-329.https://doi.org/10.1287/mnsc.49.3.312.16067

[13] Thomas, L. C., Crook, J. N., & Elliott, M. (2005). Credit scoring and its applications (Chapter: development and lifecycle of scorecards). *European Journal of Operational Research, 80*(3), 500-513.