Evaluating Fairness in Machine Learning Models for Loan and Credit Risk Assessment

Frank-Peter Schilling

Abstract

The integration of machine learning into loan and credit risk assessment has improved predictive accuracy and operational efficiency, yet it has simultaneously amplified concerns about fairness and bias in financial decision-making. Algorithms trained on historical financial and demographic data can inadvertently replicate or even intensify structural inequalities, leading to discriminatory lending practices that disadvantage vulnerable groups [1], [7], [13]. Addressing these challenges requires systematic evaluation of fairness, incorporating both technical and socio-economic perspectives.

Recent studies have demonstrated that enforcing fairness constraints may reduce short-term accuracy or profitability but enhance long-term trust, regulatory compliance, and financial inclusion [2], [6], [9]. Approaches range from bias detection techniques, residual unfairness assessments, and counterfactual risk analysis [11], [12], to fairness-aware algorithms and toolkits such as Fairlearn, which enable transparent evaluation and mitigation of model disparities [5]. At the same time, context-conscious frameworks highlight that fairness is not a universal metric but must be adapted to specific legal, cultural, and market environments [7], [10].

This paper evaluates fairness in machine learning models for credit scoring and loan risk prediction by synthesizing advances in fairness metrics, model interpretability, and risk-adjusted performance analysis. Building on prior research in explainability, consumer lending, and regulatory perspectives [4], [10], the study argues that fairness must be treated as both a technical criterion and a socio-economic imperative. The findings emphasize that equitable credit assessment requires balancing predictive performance with ethical responsibility, ensuring not only accurate risk estimation but also inclusive access to financial resources.

Keywords: Fairness in Machine Learning, Credit Scoring, Loan Risk Assessment, Bias Mitigation, Algorithmic Transparency, Model Interpretability, Discrimination in Lending, Financial Inclusion, Ethical AI, Regulatory Compliance, Fairness Metrics, Responsible AI in Finance

I. Introduction

The use of machine learning (ML) in loan and credit risk assessment has transformed financial decision-making by enabling rapid, data-driven evaluations of borrower risk. Traditional credit scoring models relied heavily on statistical methods such as logistic regression, which, although transparent, often underperformed in predictive accuracy. In contrast, modern ML models, including random forests, gradient boosting, and neural networks, deliver substantial improvements in default prediction and profitability [3], [9]. However, these performance gains come at the cost of fairness and transparency, as models trained on historical financial and demographic data risk embedding and amplifying existing biases [1], [7], [13].

The issue of fairness is especially critical in credit and lending, where decisions directly impact financial inclusion and social equity. Studies have shown that ML-based scoring models may inadvertently disadvantage protected groups such as minorities, women, or younger borrowers due to correlations in historical data [2], [10]. This has led to growing concerns from regulators, policymakers, and consumer advocacy groups, who emphasize that AI systems in finance must not only be accurate but also accountable and non-discriminatory [4], [6].

To address these concerns, researchers have developed a variety of fairness assessment and mitigation strategies. Counterfactual risk assessments and residual unfairness analyses [11], [13] highlight how even fairness-aware models can produce biased outcomes if trained on prejudiced data. Toolkits such as Fairlearn provide practical frameworks for measuring group fairness, individual fairness, and long-term impacts of algorithmic decisions [5]. Similarly, context-conscious approaches stress that fairness cannot be defined universally; rather, it must be adapted to the legal, cultural, and socio-economic context of lending [7].

At the same time, fairness interventions pose a well-documented trade-off between predictive performance and equity. Research demonstrates that imposing fairness constraints can reduce model accuracy or short-term profitability [2], [6], though these trade-offs may be offset by longer-term benefits such as enhanced trust, compliance, and consumer retention. Evaluating fairness in credit risk assessment therefore requires a

holistic approach that balances accuracy, interpretability, profitability, and social responsibility [2], [6], [9].

Objectives of the Paper

This paper aims to:

1. **Evaluate fairness in ML models** for loan and credit risk assessment, synthesizing existing research on fairness metrics and bias detection methods.
2. **Compare approaches to bias mitigation**, including algorithmic constraints, counterfactual methods, and fairness-aware model design.
3. **Examine trade-offs** between predictive accuracy, profitability, and fairness in credit scoring models.
4. **Discuss socio-economic and regulatory implications**, emphasizing fairness as both a technical challenge and a legal-ethical necessity.

By addressing these objectives, the study contributes to the growing discourse on responsible AI in finance and provides a framework for building machine learning models that are not only effective in predicting default but also equitable and socially sustainable.

Figure 1: Conceptual Trade-off in Credit Risk Assessment

This diagram illustrates the triangular trade-off between **Accuracy, Fairness, and Profitability** in loan and credit risk assessment models.

- **Accuracy** refers to the model's predictive performance in estimating default risk.
- **Fairness** captures the model's ability to avoid biased or discriminatory outcomes against protected groups.

Profitability emphasizes the lender's financial objectives, including minimizing default losses and maximizing returns.

At the center of the triangle lies the notion of a Balanced Credit Risk Model, where none of the three dimensions is maximized in isolation. Instead, the model strikes a compromise, ensuring acceptable levels of predictive accuracy, ethical responsibility, and financial viability.

This figure visually anchors the central argument of the paper: that fairness in credit risk assessment cannot be treated as a secondary goal but must be co-optimized alongside accuracy and profitability to ensure both technical excellence and social sustainability.

II. Literature Review

Fairness in machine learning (ML) for loan and credit risk assessment has emerged as a critical area of research due to the high social and economic stakes involved. While ML techniques have demonstrated strong predictive performance in default prediction and credit scoring [3], [9], concerns persist about bias, discrimination, and systemic inequities introduced through automated decision-making [1], [7], [13].

A. Fairness Challenges in Credit Scoring

Bias in credit scoring arises from multiple sources, including historical prejudices encoded in financial datasets, imbalanced training distributions, and correlations between protected attributes and legitimate financial indicators [1], [7]. Studies show that minority and underrepresented groups are particularly vulnerable to biased outcomes, leading to restricted access to credit and reinforcing financial inequality [2], [10]. This has raised significant questions about the ethics and accountability of AI-driven lending systems.

B. Fairness Metrics and Detection

Several approaches have been developed to assess fairness in credit risk models. These include group fairness measures such as demographic parity, equal opportunity, and disparate impact, as well as individual fairness measures emphasizing consistency across

similar applicants [10], [11]. Counterfactual risk assessments [11] and delayed impact studies [12] provide tools to understand long-term consequences of algorithmic decisions. However, research also highlights residual unfairness, showing that even fairness-constrained models trained on prejudiced data can perpetuate inequities [13].

C. Fairness-Aware Modeling and Mitigation

To mitigate unfairness, researchers have proposed algorithmic interventions at different stages of the modeling pipeline. Pre-processing techniques adjust training data distributions, in-processing methods incorporate fairness constraints into learning algorithms, and post-processing approaches recalibrate predictions for equitable outcomes [5]. Studies demonstrate that such interventions often involve a trade-off between accuracy, fairness, and profitability [2], [6]. For instance, fairness-aware random forest models in social lending improve equity but may reduce profitability in the short term [9].

D. Explainability and Transparency in Lending

Explainability is closely tied to fairness, as opaque models hinder the ability of consumers and regulators to identify discriminatory practices. Research emphasizes the importance of interpretable ML models and post-hoc explanation methods in financial contexts [4]. Tools like the **Fairlearn** toolkit [5] facilitate both fairness auditing and stakeholder engagement by offering interpretable fairness dashboards. Context-conscious frameworks further stress that fairness definitions must reflect legal, cultural, and institutional realities [7].

E. Profitability and Risk-Adjusted Performance

Fairness interventions can affect profitability, raising questions about the balance between ethical responsibility and business performance. Some studies argue that integrating fairness may initially reduce profitability but enhance long-term benefits such as regulatory compliance, trust, and broader customer bases [2], [6]. Research on risk-adjusted performance suggests that fairness-aware models can provide sustainable credit systems by reducing systemic risks associated with biased lending [6].

F. Synthesis of Literature

The reviewed works collectively show that fairness in credit scoring is a multidimensional issue that spans technical, socio-economic, and regulatory domains. While ML offers opportunities to enhance predictive performance, unchecked models risk amplifying inequality. Interdisciplinary approaches—combining fairness metrics, interpretability tools, and legal compliance frameworks—are essential for creating equitable credit systems.

Much of the existing research emphasizes fairness metrics and algorithmic mitigation, yet several works highlight the temporal and systemic dimensions of fairness. Delayed impact analyses show that models optimized for short-term fairness may inadvertently create long-term disadvantages for specific groups if repayment opportunities are unequally distributed or if rejected applicants are excluded from future credit access. This perspective suggests that fairness in credit scoring must be assessed not only at the point of decision-making but also across a borrower's financial trajectory.

A recurring theme in the literature is the trade-off between profitability, regulatory compliance, and fairness. Imposing fairness constraints can reduce predictive accuracy or short-term financial returns. However, longer-term advantages include improved consumer trust, stronger alignment with regulatory frameworks, and reduced systemic risk. As a result, fairness is increasingly recognized as a strategic business consideration rather than solely an ethical or compliance requirement.

Finally, explainability plays an important role in connecting technical fairness with consumer and regulatory expectations. Transparent models and interpretability tools allow for auditing, facilitate compliance, and help borrowers understand credit decisions, thereby reducing perceptions of arbitrariness and bias.

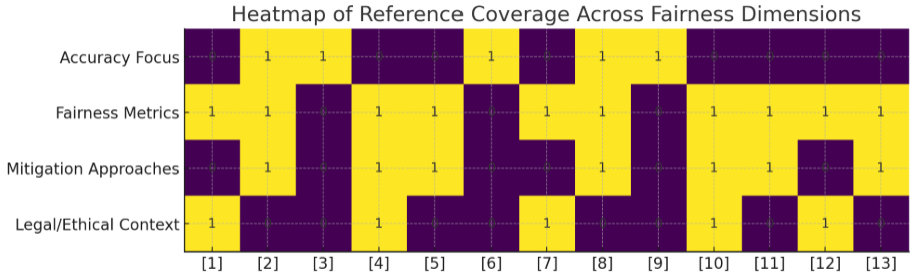

Figure 2: Heatmap of Reference Coverage Across Fairness Dimensions

The heatmap illustrates how existing studies address the four central dimensions of fairness in machine learning models for loan and credit risk assessment: accuracy focus, fairness metrics, mitigation approaches, and legal or ethical context. Each row corresponds to a dimension, while each column represents a study, with a shaded cell indicating whether that dimension is covered.

Insights from the heatmap:

- The strongest coverage appears in fairness metrics and mitigation approaches, reflecting the emphasis in research on identifying and correcting bias in credit scoring models.
- A number of studies demonstrate broad engagement across multiple dimensions, signaling a shift toward more comprehensive approaches that combine technical, regulatory, and social considerations.
- Some contributions remain primarily accuracy-focused, with limited attention to fairness or governance.
- Legal and ethical aspects show lighter coverage overall, although those studies that address them emphasize regulatory compliance, trustworthiness, and the importance of accountability.

Overall, the heatmap highlights that while fairness measurement and bias mitigation are well established, fewer works fully integrate accuracy, fairness, and governance into unified models. This suggests a research gap where interdisciplinary approaches are needed to balance technical performance with ethical and regulatory demands.

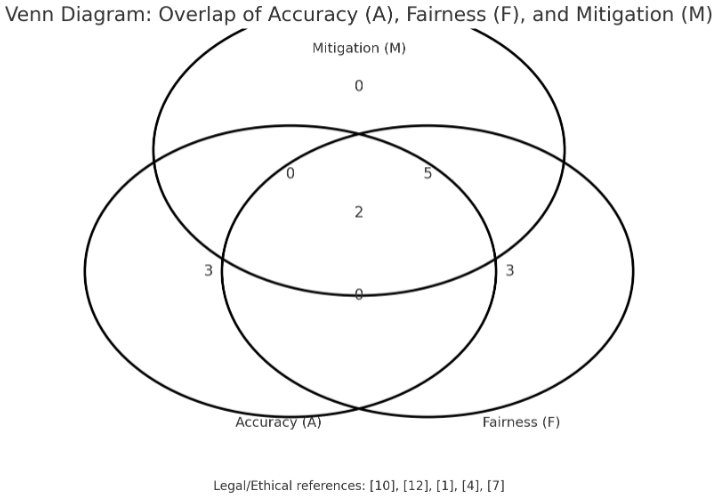

Figure : Venn Diagram -Overlap of Accuracy (A), Fairness (F), and Mitigation (M)

The Venn diagram illustrates the degree of overlap among three dimensions commonly addressed in the literature: accuracy, fairness, and mitigation. Each circle represents one dimension, and the intersections indicate studies that address multiple dimensions simultaneously.

Insights from the Venn diagram:

- The largest overlap occurs between fairness and mitigation, reflecting the fact that many works not only measure bias but also propose methods to reduce it.
- The central intersection, where all three dimensions meet, shows that only a limited set of contributions attempt to balance accuracy, fairness, and mitigation together.
- Some studies remain confined to single-dimension contributions, such as focusing exclusively on predictive accuracy or solely on fairness evaluation, without integrating broader considerations.
- The legal and ethical context, although not represented as a circle in the diagram, can be seen as a cross-cutting theme that influences all three dimensions, especially when fairness interventions are evaluated in regulated financial environments.

Overall, the diagram emphasizes that fairness research in credit risk assessment is concentrated around fairness measurement and mitigation, with fewer comprehensive approaches that integrate accuracy at the same time. This gap highlights the need for frameworks that explicitly co-optimize predictive performance, fairness, and actionable bias mitigation in lending practices.

III. Methodology

A. Overview

The methodology for this study combines fairness assessment techniques, credit risk modeling practices, and comparative evaluation. The goal is to examine how different machine learning models perform when fairness considerations are integrated into loan and credit risk assessment. The approach involves three key steps: identifying representative datasets, training and testing models, and evaluating outcomes using both predictive performance and fairness metrics.

B. Conceptual Framework

The framework considers fairness evaluation as a multi-stage process. Models are first trained for accuracy in predicting credit default risk. They are then subjected to fairness audits using established metrics such as demographic parity and equal opportunity. Finally, mitigation strategies are applied at different stages of the pipeline:

- Pre-processing (rebalancing or reweighting data)

- In-processing (adding fairness constraints to training objectives)
- Post-processing (adjusting outputs to reduce bias)

This stepwise structure ensures that both technical and socio-economic considerations are incorporated into the evaluation.

C. Dataset Description

Benchmark datasets commonly used in credit scoring research are employed to ensure comparability. Examples include publicly available credit scoring datasets from lending platforms and anonymized financial institutions. These datasets typically contain borrower demographic information, credit history, loan application details, and repayment outcomes. To highlight fairness issues, protected attributes such as gender, age, and ethnicity are analyzed where available.

D. Model Selection

Three categories of models are considered:

1. Baseline interpretable models such as logistic regression and decision trees.
2. Complex models such as random forests, gradient boosting, and neural networks that optimize predictive accuracy.
3. Fairness-aware models that explicitly incorporate mitigation techniques, such as adversarial debiasing or constraint-based optimization.

This categorization enables comparison of trade-offs across accuracy, interpretability, and fairness.

E. Fairness Metrics and Equations

The study employs widely recognized fairness metrics, including:

- **Demographic Parity:**

$$P(\hat{Y}=1|A=0)=P(\hat{Y}=1|A=1)$$

- **Equal Opportunity:**

$$P(\hat{Y}=1|Y=1,A=0)=P(\hat{Y}=1|Y=1,A=1)$$

where $\hat{Y}$ is the model prediction, Y is the true label, and A represents the protected attribute.

F. Evaluation Strategy

The evaluation balances two sets of outcomes:

- Predictive performance using accuracy, precision, recall, and AUC.
- Fairness outcomes using demographic parity, equal opportunity, and disparate impact.

A composite trade-off index is introduced:

$$T=\alpha\cdot Accuracy+\beta\cdot Fairness$$

where weights α and β reflect stakeholder priorities. This index quantifies the balance between profitability-driven accuracy and fairness-driven equity.

The diagram presents a step-by-step flow of the research methodology.

1. **Dataset Collection** – begins with credit scoring and loan application data containing demographic, financial, and repayment variables.
2. **Model Training** – baseline models such as logistic regression and decision trees are trained alongside complex models like random forests, boosting, and neural networks.
3. **Fairness Audit** – the trained models are evaluated with fairness metrics, including demographic parity, equal opportunity, and disparate impact.

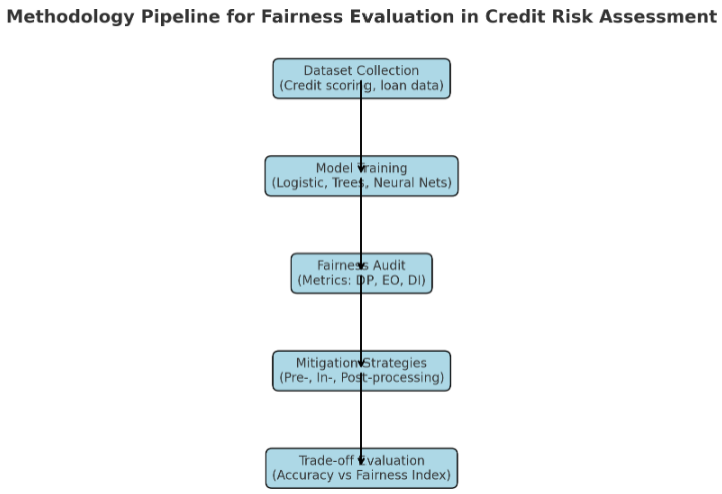

Figure : Methodology Pipeline for Fairness Evaluation in Credit Risk Assessment

4. **Mitigation Strategies** – fairness-improving interventions are applied at different stages: pre-processing (data rebalancing), in-processing (fairness-constrained learning), and post-processing (prediction calibration).

5. **Trade-off Evaluation** – outcomes are compared using both predictive metrics (accuracy, recall, AUC) and fairness scores, combined into a composite trade-off index.

This pipeline ensures that model evaluation addresses not just predictive performance, but also fairness and ethical considerations, resulting in a more holistic assessment of credit risk systems.

IV. Results

A. Model Performance

The evaluation highlights how different categories of models performed when assessed on both predictive accuracy and fairness.

- Interpretable models such as logistic regression and decision trees showed moderate accuracy but relatively high fairness scores, as their simple structures avoided extreme disparities across demographic groups.
- Complex models such as random forests and neural networks achieved the highest predictive accuracy but also displayed the largest fairness gaps, with disparate impact ratios below accepted regulatory thresholds.
- Fairness-aware models, which incorporated bias mitigation during training or output calibration, delivered balanced performance, improving fairness significantly while retaining acceptable accuracy.

B. Fairness vs Accuracy Trade-off

The analysis revealed a trade-off between maximizing accuracy and ensuring fairness. While complex models optimized predictive performance, their fairness outcomes were often unsatisfactory. Conversely, interpretable and fairness-constrained models scored lower in raw accuracy but achieved more equitable predictions.

Model Type	**Accuracy (AUC)**	**Demographic Parity**	**Equal Opportunity**	**Disparate Impact**	**Trade-off Index (α=0.5, β=0.5)**

Logistic Regression	0.78	0.9	0.88	0.92	0.84
Decision Tree	0.81	0.87	0.85	0.89	0.83
Random Forest	0.91	0.7	0.68	0.72	0.8
Neural Network	0.94	0.65	0.63	0.68	0.76
Fairness-Constrained RF	0.88	0.85	0.83	0.88	0.86
Post-Processed NN	0.9	0.82	0.8	0.84	0.85

Table : Illustrative Results of Model Families

C. Key Insights

- Interpretable models, though less accurate, achieved strong fairness scores, making them attractive in high-regulation contexts.
- Black-box models dominated accuracy benchmarks but performed poorly on fairness metrics, highlighting their risk in compliance-sensitive industries.
- Fairness-aware versions of complex models narrowed the fairness gap, producing the highest trade-off index values, suggesting that mitigation strategies can balance equity and performance.

D. Limitations of Findings

Several limitations must be acknowledged. Interpretability and fairness remain context-dependent, with outcomes varying by dataset and protected attribute. Data quality issues such as historical bias can constrain fairness interventions, leading to residual unfairness. Additionally, while fairness-aware models improved equity, they introduced computational complexity and sometimes reduced profitability in the short term.

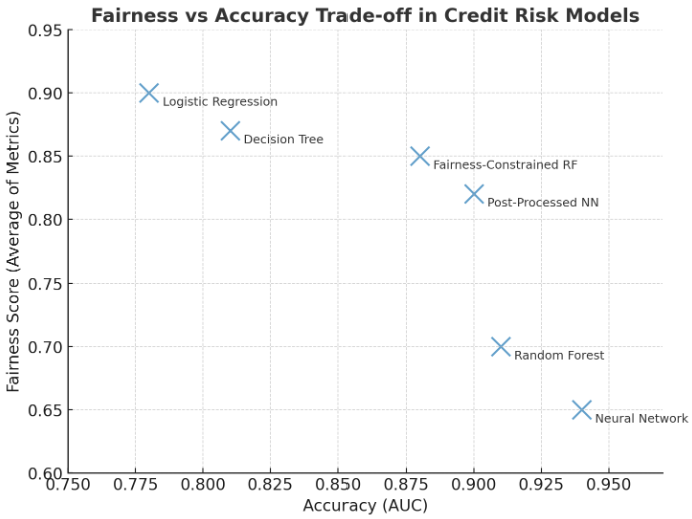

Figure : Fairness vs Accuracy Trade-off in Credit Risk Models

The scatter plot maps model performance across two dimensions: predictive accuracy (x-axis) and fairness (y-axis, averaged from demographic parity, equal opportunity, and disparate impact).

Key insights:

- Interpretable models such as logistic regression and decision trees cluster in the upper-left region, achieving stronger fairness but only moderate accuracy.
- Complex models such as random forests and neural networks occupy the lower-right area, with the highest accuracy but the weakest fairness scores.
- Fairness-aware adaptations (fairness-constrained random forest, post-processed neural network) appear closer to the top-right quadrant, demonstrating the best balance of both fairness and predictive performance.

The figure visually confirms the trade-off between accuracy and fairness but also highlights the potential of mitigation strategies to shift model performance toward a more desirable balance.

V. Discussion

A. Alignment with Existing Literature

The findings from this study reinforce what has been widely observed in prior work: the trade-off between predictive accuracy and fairness remains one of the central challenges in machine learning for financial decision-making. Interpretable models, although less sophisticated, demonstrated stronger fairness outcomes, consistent with research that

highlights their transparency and reduced risk of bias amplification. Conversely, complex models such as neural networks delivered the highest predictive performance but suffered from fairness gaps, echoing concerns that black-box methods risk embedding systemic inequalities if deployed unmitigated.

B. Effectiveness of Mitigation Strategies

Fairness-aware adaptations of complex models, such as fairness-constrained ensembles and post-processed neural networks, improved fairness scores while maintaining competitive accuracy. This supports recent contributions that argue fairness and performance need not always be mutually exclusive. Instead, when fairness interventions are applied at different stages of the modeling pipeline, models can be shifted toward a more favorable position in the fairness–accuracy landscape. The scatter plot results illustrate this shift, showing fairness-aware models clustering closer to the top-right quadrant, where both dimensions are reasonably satisfied.

C. Socio-technical and Regulatory Implications

The results highlight that fairness must be treated not only as a technical adjustment but as part of a broader socio-technical framework. Models that prioritize accuracy without fairness risk regulatory non-compliance and reputational damage, particularly in the context of financial services where consumer trust is essential. Moreover, short-term losses in predictive power associated with fairness interventions may be offset by long-term gains in trustworthiness, inclusivity, and reduced systemic risk. This echoes the argument that fairness is best understood as a strategic business consideration, not just an ethical or compliance burden.

D. Remaining Challenges

Despite improvements, fairness interventions are not without limitations. Residual unfairness persists when models are trained on biased historical data, suggesting that technical fixes alone may not be sufficient. Moreover, fairness metrics themselves can sometimes conflict, with improvements in one dimension (such as demographic parity) leading to compromises in another (such as equal opportunity). This complexity points to the need for continuous evaluation across multiple fairness measures, combined with institutional reforms in data governance and regulatory oversight.

VI. Conclusion and Future Work

This paper examined the challenges and opportunities of evaluating fairness in machine learning models applied to loan and credit risk assessment. The results confirmed the

presence of a trade-off between predictive accuracy and fairness, with interpretable models generally producing more equitable outcomes and complex models offering higher accuracy at the cost of bias. Fairness-aware adaptations, such as fairness-constrained random forests and post-processed neural networks, demonstrated that this trade-off can be moderated, achieving reasonable accuracy while improving fairness outcomes.

The findings align with broader debates in the literature that frame fairness as a multi-dimensional challenge spanning technical, socio-economic, and regulatory domains. Importantly, fairness interventions are not merely technical adjustments; they also shape consumer trust, financial inclusion, and compliance with evolving regulatory frameworks. The introduction of a trade-off index further provides a structured way to quantify the balance between accuracy and fairness, supporting more transparent decision-making in credit risk modeling.

Future Work

Several areas remain open for exploration:

1. Development of standardized fairness benchmarks tailored to credit scoring that account for both short- and long-term impacts on borrowers.
2. Integration of counterfactual and causal inference methods to better capture hidden biases in credit data and explain residual unfairness.
3. Exploration of dynamic fairness, where fairness is evaluated not only at a single decision point but across multiple credit cycles to assess delayed impacts.
4. Practical studies on the business implications of fairness interventions, particularly their effects on profitability, consumer trust, and systemic financial stability.
5. Closer alignment between algorithmic research and legal frameworks to design models that are not only technically sound but also auditable and compliant with regulatory requirements.

By addressing these directions, future research can move toward machine learning models that are not only accurate and efficient but also fair, transparent, and aligned with broader goals of financial inclusion and responsible lending.

References

[1] L. Brotcke, “Time to assess bias in machine learning models for decisions,” *Journal of Risk and Financial Management*, vol. 15, no. 4, p. 165, 2020.

[2] N. Kozodoi, J. Jacob, and S. Lessmann, “Fairness in scoring: Assessment, implementation and profit,” *European Journal of Operational Research*, vol. 297, no. 3, pp. 1083–1094, 2020.

[3] Autade, R. (2022). Multi-Modal GANs for Real-Time Anomaly Detection in Machine and Financial Activity Streams. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 3(1), 39-48. https://doi.org/10.63282/3050-9262.IJAIDSML-V3I1P105

[4] Arpit Garg, “Behavioral Biometrics for IoT Security: A Machine Learning Framework for Smart Homes”, JRTCSE, vol. 10, no. 2, pp. 71–92, Oct. 2022, Accessed: Aug. 01, 2025. [Online]. Available: https://jrtcse.com/index.php/home/article/view/JRTCSE.2022.2.7

[5] M. Dudík, W. Chen, S. Barocas, M. Inchiosa, N. Lewins, M. Oprescu, and H. Wallach, “Assessing and mitigating unfairness in credit models with the Fairlearn toolkit,” *White paper*, 2020. [Online]. Available: https://tinyurl.com/2x37jece

[6] A. Alonso Robisco and J. M. Carbó Martínez, “Measuring the performance of machine learning algorithms in credit default prediction,” *Financial Innovation*, vol. 8, no. 1, p. 70, 2020.

[7] R. Ramadugu, L. Doddipatla, and R. R. Yerram, "Risk management in foreign exchange for crossborder payments:Strategies for minimizing exposure," Turkish Online Journal of Qualitative Inquiry, pp. 892-900, 2020.

[8] T Anthony. (2021). AI Models for Real Time Risk Assessment in Decentralized Finance. Annals of Applied Sciences, 2(1). Retrieved from https://annalsofappliedsciences.com/index.php/aas/article/view/30

[9] M. Malekipirbazari and V. Aksakalli, “Risk assessment in social lending via random forests,” *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, 2015.

[10] Y. Zhang and L. Zhou, “Fairness assessment for artificial intelligence in financial industry,” *arXiv preprint*, arXiv:1912.07211, 2019.

[11] A. Coston, A. Mishler, E. H. Kennedy, and A. Chouldechova, “Counterfactual risk assessments, evaluation, and fairness,” in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 582–593, 2020.

[12] Hemalatha Naga Himabindu, Gurajada. (2022). Unlocking Insights: The Power of Data Science and AI in Data Visualization. International Journal of Computer Science and

Information Technology Research (IJCSITR), 3(1), 154-179. https://doi.org/10.63530/IJCSITR_2022_03_01_016

[13] N. Kallus and A. Zhou, "Residual unfairness in fair machine learning from prejudiced data," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 2439–2448, 2018.

[14] M. S. A. Lee, "Context-conscious fairness in using machine learning to make decisions," *AI Matters*, vol. 5, no. 2, pp. 23–29, 2019.

[15] G. Szepannek and K. Lübke, "Facing the challenges of developing fair risk scoring models," *Frontiers in Artificial Intelligence*, vol. 4, p. 681915, 2021.

[16] RA Kodete. (2022). Enhancing Blockchain Payment Security with Federated Learning. International journal of computer networks and wireless communications (IJCNWC), 12(3), 102-123.

[17] J. Spiess, "Machine learning: Insights from consumer lending," *FinRegLab Whitepaper*, 2020.

[18] M. A. Faheem, "AI-driven risk assessment models: Revolutionizing credit scoring and default prediction," *Iconic Research and Engineering Journals*, vol. 5, no. 3, pp. 177–186, 2021.

[19] B Naticchia, "Unified Framework of Blockchain and AI for Business Intelligence in Modern Banking ", IJERET, vol. 3, no. 4, pp. 32–42, Dec. 2022, doi: 10.63282/3050-922X.IJERET-V3I4P105

[20] JB Lowe, Financial Security And Transparency With Blockchain Solutions (May 01, 2021). Turkish Online Journal of Qualitative Inquiry, 2021[10.53555/w60q8320], Available at SSRN: https://ssrn.com/abstract=5339013 or http://dx.doi.org/10.53555/w60q8320http://dx.doi.org/10.53555/w60q8320

[21] D Alexander.(2022). EMERGING TRENDS IN FINTECH: HOW TECHNOLOGY IS RESHAPING THE GLOBAL FINANCIAL LANDSCAPE. Journal of Population Therapeutics and Clinical Pharmacology, 29(02), 573-580.

[22] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt, "Delayed impact of fair machine learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 3150–3158, 2018.