research proposal-updated-version-2.pdf

by Niteesh Dubey

Credit Risk Prediction Using Machine Learning Models with Bias Mitigation

Niteesh Kumar Dubey

Research Proposal

FEBRUARY 2025

Abstract

Financial institutions rely on accurate credit risk prediction to make informed lending decisions and maintain stability. However, machine learning models used for this purpose often inherit biases from historical data, resulting in discriminatory outcomes for individuals based on protected characteristics such as race, gender, and socioeconomic status. This research aims to develop a machine learning framework for credit risk prediction that integrates bias detection and mitigation techniques to ensure both predictive accuracy and fairness. The study will implement and compare multiple bias mitigation approaches, including Fair-SMOTE and re-weighting, using publicly available datasets (German Credit and Lending Club). Model evaluation will incorporate both traditional performance metrics (F1-score, ROC-AUC) and fairness measures (Disparate Impact, Equal Opportunity Difference). This research contributes to the growing field of responsible AI by establishing a methodological framework for developing equitable credit risk models that balance predictive performance with fairness considerations, thereby promoting ethical compliance and addressing the technical and regulatory challenges faced by financial institutions in implementing fair lending practices.

Keywords: Credit risk prediction, machine learning, bias mitigation, fairness metrics, responsible AI, Fair-SMOTE, re-weighting, disparate impact, equal opportunity difference, ethical finance

Table of Contents

1. Introduction

1.1 Background and Problem Statement

Credit risk assessment has become an essential function for financial institutions, enabling them to evaluate loan applicants' likelihood of default and make informed lending decisions. Traditional statistical methods for credit risk assessment have increasingly been supplanted by machine learning (ML) algorithms, which can analyze vast datasets and detect complex patterns indicative of default risk (Byanjankar et al., 2015; Butaru et al., 2016). Despite their superior predictive capabilities, these models present a significant ethical challenge: they can inherit and amplify biases present in historical lending data, resulting in discriminatory outcomes that disproportionately affect marginalized groups.

Recent studies have revealed alarming disparities in loan approval rates across demographic groups. For instance, (Bartlett et al., 2022) found that Latinx and African-American borrowers paid, on average, 7.9 basis points more for home purchase mortgages compared to similarly qualified white borrowers. This disparity translates to an additional $765 million in aggregate interest annually. Notably, while FinTech algorithms also exhibited discriminatory pricing, the degree of discrimination was approximately 40% less than that observed with face-to-face lenders. Similarly, (Kelley and Ovchinnikov, 2020) examined how anti-discrimination laws impact gender bias in non-mortgage fintech lending, revealing that prohibiting the use of gender data can inadvertently increase discrimination against female applicants. These statistics underscore the severity and persistence of bias in credit evaluation systems.

1.2 Ethical and Legal Implications

The perpetuation of bias in credit risk prediction extends beyond technical concerns to encompass significant ethical and legal implications. Discriminatory lending practices violate fundamental principles of fairness and equity, contributing to systemic economic disparities and financial exclusion (Fuster et al., 2022). From a legal perspective, financial institutions utilizing biased models risk violating regulations such as the Equal Credit Opportunity Act (ECOA) in the United States and the European Union's General Data Protection Regulation (GDPR), which prohibit discrimination in lending practices and algorithmic decision-making (Goodman and Flaxman, 2017).

The financial industry has faced increased regulatory scrutiny regarding algorithmic fairness. In 2023, the Consumer Financial Protection Bureau (CFPB) issued guidelines explicitly addressing the responsibility of financial institutions to ensure fairness in AI-driven lending decisions (CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence, 2023).

1.3 Sources of Bias in Machine Learning Models

Bias in machine learning models for credit risk prediction stems from multiple sources. Historical lending data often reflects past discriminatory practices, where certain demographic groups were systematically denied access to credit or offered unfavorable terms (Blattner and

Nelson, 2021). When models are trained on such data, they learn to replicate these patterns, creating a feedback loop that perpetuates discrimination (Hardt et al., 2016).

Additionally, bias can emerge through feature selection and data preprocessing. Seemingly neutral variables such as zip code or education level may serve as proxies for protected attributes like race or socioeconomic status (Corbett-Davies et al., 2023). The algorithmic design itself can also introduce bias through optimization objectives that prioritize profit maximization over fairness considerations (Kleinberg et al., 2016).

1.4 Bias Mitigation Approaches

Recent advances in fairness-aware machine learning have produced various techniques to address bias in predictive models. Fair-SMOTE (Synthetic Minority Over-sampling Technique), introduced by (Chakraborty et al., 2021), addresses data imbalances by oversampling underrepresented groups to create more balanced training datasets. This technique has shown promise in reducing disparities in model predictions across demographic groups without significantly compromising predictive accuracy.

Re-weighting methods assign different weights to samples during model training to counteract bias. Additionally, fairness metrics such as Disparate Impact (DI) and Equal Opportunity Difference (EOD) provide quantitative measures to detect and evaluate bias in model predictions (Hardt et al., 2016).

1.5 Research Significance and Contribution

This research addresses the critical intersection of machine learning, financial inclusion, and ethical AI. By developing a framework for bias-mitigated credit risk prediction, the study contributes to both theoretical understanding of algorithmic fairness and practical implementation of responsible AI in the financial sector. The findings will provide financial institutions with actionable insights and methodologies to develop more equitable credit evaluation systems while maintaining predictive performance.

The study's unique contribution lies in its comprehensive comparison of multiple bias mitigation techniques within the specific context of credit risk prediction, evaluating their effectiveness across diverse datasets and demographic groups. By establishing best practices for balancing accuracy and fairness, this research advances the emerging field of responsible AI in finance and supports the industry's transition toward more ethical lending practices.

2. Background and Related Research

2.1 Evolution of Credit Risk Prediction Methods

The field of credit risk prediction has evolved substantially from simple heuristic approaches to sophisticated machine learning algorithms. Traditional methods relied primarily on expert judgment and linear statistical models such as logistic regression (Altman, 1968). While these approaches offered interpretability, they struggled to capture complex, non-linear relationships in financial data (Lessmann et al., 2015).

Modern machine learning techniques have demonstrated superior predictive performance in credit risk assessment. Random Forest algorithms have shown particular effectiveness in handling the high-dimensional feature spaces typical of credit data (Barboza et al., 2017). Similarly, Gradient Boosting Machines (GBMs) have consistently outperformed traditional methods in benchmark studies (Lessmann et al., 2015; Fitzpatrick and Mues, 2016).

Despite their predictive advantages, these advanced models present challenges related to interpretability and fairness. This opacity becomes particularly problematic in the context of credit decisions, where regulatory frameworks often mandate explainability (Goodman and Flaxman, 2017).

2.2 Bias Detection in Machine Learning Models

Identifying bias in machine learning models requires specialized metrics that evaluate fairness across different demographic groups. Disparate Impact (DI), inspired by legal standards in anti-discrimination law, measures the ratio of favorable outcomes between protected and unprotected groups (Feldman et al., 2015). A DI value significantly below 1.0 indicates potential discrimination against the protected group. Similarly, Equal Opportunity Difference (EOD) assesses whether a model maintains equal true positive rates across demographic categories (Hardt et al., 2016).

Several studies have applied these metrics to reveal bias in credit scoring models. (Kozodoi et al., 2022) analyzed commercial credit scoring systems and found DI values as low as 0.67 for racial minorities, indicating substantial disparities in approval rates. (Zhang and Ntoutsi, 2019) demonstrated that even models with high overall accuracy can exhibit significant EOD values, suggesting discriminatory patterns in their predictions. Beyond statistical measures, researchers have developed visualization techniques to identify bias.

2.3 Bias Mitigation Techniques

Approaches to mitigating bias in machine learning models fall into three main categories: pre-processing methods (addressing bias in training data), in-processing methods (modifying the learning algorithm), and post-processing methods (adjusting model outputs) (d'Alessandro et al., 2017).

Among pre-processing techniques, Fair-SMOTE has shown particular promise for credit risk prediction. (Chakraborty et al., 2021) demonstrated that Fair-SMOTE reduced discrimination in loan approval predictions, while maintaining comparable accuracy to standard models. The technique works by generating synthetic samples for underrepresented groups, creating a more balanced dataset that leads to fairer model predictions.

Post-processing methods adjust model outputs to ensure fairness across groups. (Hardt et al., 2016) introduced a technique that modifies decision thresholds for different demographic groups to equalize error rates. (Pleiss et al., 2017) extended this approach with calibrated equal opportunity, which maintains both calibration and equal opportunity across protected groups.

2.4 Fairness-Accuracy Trade-offs

A central challenge in developing fair ML models is balancing fairness with predictive performance. Multiple studies have explored this trade-off in the context of credit risk prediction. (Kleinberg et al., 2016) proved mathematically that different fairness criteria can be inherently incompatible, suggesting that perfect fairness across all metrics may be unattainable. However, empirical research indicates that moderate improvements in fairness can often be achieved with minimal accuracy reduction.

2.5 Regulatory and Industry Perspectives

Financial regulators worldwide have increasingly focused on algorithmic fairness in lending. The European Banking Authority (EBA) has been actively monitoring the adoption of General Purpose Artificial Intelligence (GPAI) in the banking sector, emphasizing the need to mitigate discrimination and bias that could lead to financial exclusion. In April 2024, the EBA organized a workshop with EU stakeholders to discuss the risks and opportunities associated with GPAI, highlighting the importance of robust data governance frameworks to ensure accuracy and reliability in AI models (Special topic – Artificial intelligence | European Banking Authority, 2025).

Similarly, the U.S. Office of the Comptroller of the Currency (OCC) has established an office dedicated to responsible innovation. In March 2016, the OCC published a white paper titled "Supporting Responsible Innovation in the Federal Banking System: An OCC Perspective," outlining principles to guide the development of a framework that supports responsible innovation in financial services (Greenfield, n.d.).
These initiatives reflect a growing regulatory emphasis on ensuring fairness and accountability in AI-driven credit decision processes.

Fairness in algorithmic decision-making is widely recognized, yet adoption of fairness-aware ML in finance remains uneven. Many institutions acknowledge the need to mitigate bias, but few have integrated comprehensive fairness frameworks. Key barriers include the complexity of embedding fairness measures into existing systems. Regulatory uncertainty and concerns about trade-offs between fairness and accuracy further hinder progress. Studies (Kleinberg et al., 2016; Goodman and Flaxman, 2017) highlight these challenges. This research aims to develop practical, evidence-based frameworks to mitigate bias in credit risk models and promote equitable lending.

3. Research Questions

This research will address the following key questions:

1. How can biases in machine learning models for credit risk prediction be effectively detected and mitigated?
2. What impact do bias mitigation techniques have on the overall predictive performance of credit risk models?
3. Is it possible to achieve fairness in credit risk prediction without significantly reducing model accuracy?

4. Aim and Objectives

The primary aim of this research is to develop a streamlined framework for bias-mitigated credit risk prediction using two credit datasets. This framework will integrate bias detection and mitigation techniques to create a proof-of-concept model that optimally balances predictive accuracy with fairness. Ultimately, the model will offer actionable insights to support the development of ethical lending practices in the financial industry.The objectives of the study are:

Data Analysis and Bias Detection:

1. Analyze the two available credit datasets to identify patterns of bias across key sensitive attributes (e.g., gender and race).
2. Quantify bias using a focused set of fairness metrics, such as Equal Opportunity Difference (EOD) and Disparate Impact (DI)

Bias Mitigation Implementation:

1. Implement one or two bias mitigation techniques (e.g., re-weighting and Fair-SMOTE) within a baseline machine learning model for credit risk prediction.
2. Apply these techniques across both datasets to assess consistency and generalizability.

Performance Comparison:

1. Evaluate and compare the performance of the bias-mitigated models against traditional models in terms of both predictive accuracy and fairness metrics.
2. Conduct cross-dataset validation if feasible to ensure the framework's robustness.

5. Research Methodology

5.1 Research Approach

This study employs a quantitative research approach, focusing on empirical evaluation of machine learning models andbias mitigation techniques. The methodology combines experimental design with statistical analysis to systematically assess fairness-accuracy trade-offs in credit risk prediction.

5.2 Data Selection and Justification

Two publicly available datasets will be utilized for this research:

1. **German Credit Dataset:** This dataset contains 1,000 credit applications with 20 attributes including age, gender, and employment status. It has been selected due to its wide usage in fairness research and the presence of sensitive attributes that allow for bias analysis (Grömping, 2019).

2. **Lending Club Dataset:** This dataset contains 2.2 million loan records with detailed borrower information and loan performance metrics. It has been selected to evaluate the scalability of the proposed methods and their effectiveness in the context of marketplace lending, which represents a growing segment of the credit market (Jagtiani and Lemieux, 2019).

These datasets have been selected to represent different lending contexts, feature spaces, and demographic compositions, ensuring that the findings have broad applicability across various financial institutions and lending scenarios.

5.3 Data Preprocessing

Data preprocessing will follow a structured approach:

1. **Missing Data Handling:** Multiple imputation techniques will be employed based on the nature of the missing data (Missing Completely at Random, Missing at Random, or Missing Not at Random) as recommended by (Buuren, 2018).
2. **Feature Engineering:** Domain-specific features will be created to enhance predictive performance while carefully evaluating their potential to introduce or amplify bias (Dwork et al., 2012).
3. **Outlier Detection and Treatment:** Robust statistical methods including Isolation Forest and Local Outlier Factor will be used to identify and handle outliers while preserving legitimate data patterns (Liu et al., 2008).
4. **Feature Scaling:** Standardization and normalization techniques will be applied to ensure algorithm convergence and prevent features with large magnitudes from dominating the model.
5. **Sensitive Attribute Identification:** Protected characteristics (race, gender, age) and potential proxy variables will be identified through correlation analysis and domain knowledge to inform bias detection and mitigation strategies (Corbett-Davies et al., 2023).

5.4 Model Development

Four machine learning algorithms will be implemented and compared:

1. **Logistic Regression:** Selected for its interpretability and baseline performance. Despite its simplicity, logistic regression remains widely used in credit scoring due to regulatory preferences for explainable models (Thomas, 2000).
2. **Random Forest:** Chosen for its ability to handle non-linear relationships and high-dimensional feature spaces. Random Forest has consistently performed well in credit risk prediction benchmarks (Lessmann et al., 2015).
3. **XGBoost:** Selected for its state-of-the-art performance in tabular data problems. XGBoost has demonstrated superior predictive capabilities in numerous credit scoring competitions (Chen and Guestrin, 2016).

Each algorithm will be optimized using grid search with cross-validation to determine optimal hyperparameters that balance performance and fairness considerations.

5.5 Bias Detection
Fairness metrics will be implemented to quantify bias across protected groups:

1. **Disparate Impact (DI):** Will measure the ratio of favorable outcomes between protected and unprotected groups, with values significantly below 1.0 indicating potential discrimination (Feldman et al., 2015).
2. **Equal Opportunity Difference (EOD):** Will assess differences in true positive rates across demographic groups to identify disparities in the ability of qualified applicants to obtain credit (Hardt et al., 2016).
3. **Statistical Parity Difference (SPD):** Will evaluate differences in approval rates across groups regardless of qualification, highlighting systemic disparities in access to credit (Dwork et al., 2012).
4. **Theil Index:** Will provide an aggregate measure of inequality in model predictions, capturing overall fairness beyond binary group comparisons (Speicher et al., 2018).

Visualization techniques including fairness confusion matrices and slice analysis will complement these metrics to provide intuitive understanding of bias patterns across intersectional demographics.

5.6 Bias Mitigation
Three bias mitigation techniques will be implemented and compared:

1. **Fair-SMOTE:** Will adjust the training data distribution by oversampling underrepresented protected groups and generating synthetic examples to balance class distributions (Chakraborty et al., 2021). Implementation will use the imblearn library with custom modifications to address credit-specific imbalances.
2. **Re-weighting:** Will assign different importance weights to training samples based on their protected attributes and labels to counteract historical bias (Kamiran and Calders, 2012). The weighting function will be optimized through validation to maximize fairness improvement while minimizing accuracy loss.

Each technique will be applied to all four machine learning algorithms to identify the most effective combinations across different contexts.

5.7 Model Evaluation

Models will be evaluated using a comprehensive framework that considers both traditional performance metrics and fairness measures:

3

1. **Performance Metrics:** Accuracy, precision, recall, F1-score, ROC-AUC, and profit-based metrics (Expected Maximum Profit) will assess predictive performance.
2. **Fairness Metrics:** DI, EOD, SPD, and Theil Index will quantify fairness across protected groups. A fairness dashboard will visualize these metrics to facilitate comparison.
3. **Statistical Significance:** Bootstrap resampling and hypothesis testing will determine whether observed differences in fairness and performance are statistically significant (Efron and Tibshirani, 1998).
4. **Cross-Validation:** Stratified k-fold cross-validation will ensure robust evaluation across different data partitions while maintaining demographic representation.

6. Expected Outcomes

The anticipated outcome of this research is a machine learning model designed to predict credit risk with enhanced fairness, particularly concerning sensitive factors like race and gender. This study will illustrate how implementing bias mitigation techniques can substantially improve the equity of credit risk assessments without compromising predictive accuracy. By addressing biases inherent in historical data, the model aims to deliver more equitable outcomes for diverse demographic groups.

This approach will enable financial institutions to utilize a trustworthy and ethical tool for making well-informed lending decisions while ensuring compliance with regulatory frameworks. In doing so, it seeks to foster greater trust among consumers and promote responsible lending practices. By integrating fairness into credit risk predictions, this research not only addresses ethical concerns but also contributes to the development of responsible artificial intelligence (AI) systems in the financial sector. Ultimately, the findings of this study will serve as a valuable resource for organizations striving to create a fairer and more transparent credit evaluation process.

7. Required Resources

7.1 Data Resources

1. **Public Datasets:** German Credit Dataset and Lending Club Dataset, all freely available through repositories such as Kaggle and UCI Machine Learning Repository.
2. **Synthetic Data Generation Tools:** For validating approaches in controlled environments with known bias patterns.

7.2 Technical Resources

1. **Programming Environment:** Python programming language with the following libraries:
   - scikit-learn, XGBoost for machine learning implementation
   - AIF360, Fairlearn for bias detection and mitigation
   - imblearn for Fair-SMOTE implementation
   - Pandas, NumPy for data manipulation
   - Matplotlib, Seaborn, Plotly for visualization
2. **Computing Resources:** High-performance computing environment with:
   - Minimum 16GB RAM for processing larger datasets
   - Cloud computing resources (e.g., Google Colab Pro) for scalability testing
3. **Development Tools:**
   - Jupyter Notebooks for exploratory analysis and documentation
   - Git for version control
   - Docker for creating reproducible environments

7.3 Literature Resources

1. **Academic Databases:** Access to IEEE Xplore, ACM Digital Library, ScienceDirect, and arXiv for literature review and staying current with research developments.
2. **Regulatory Guidelines:** Documentation from financial regulatory bodies regarding algorithmic fairness requirements.

8. Plan of Work/Research Plan

Activities	1st to 2 nd week	3 rd to 4th week	5 th to 6th	7 th to 8 th week	9 th to 10 th week	11th to 13th Week	14th to 16th week
Background Study							
Literature Review							
Proposal Submission							
Questionnaire Preparation							
Secondary Data Collection							
Secondary Data Analysis							
Interim Report Writing							
Interim Report Submission							
Final Report Writing							
Final Report Submission							

Conclusion

This study seeks to develop a framework for bias-mitigated credit risk prediction, integrating bias detection and mitigation techniques within machine learning models. By evaluating Fair-SMOTE and re-weighting across two credit datasets, this research provides empirical insights into improving fairness while maintaining predictive accuracy.

The findings will contribute to **responsible AI development** in financial institutions, ensuring compliance with **regulatory frameworks** and promoting **fairer lending practices**. The study will also highlight the trade-offs between fairness and accuracy, providing actionable recommendations for model development in the financial sector.

Future research could explore **scalability across larger datasets** and **applicability in different financial domains**, such as **insurance underwriting or mortgage lending**. These extensions could further strengthen ethical AI implementation in finance, fostering **trust and transparency** in credit risk assessments.

References

d'Alessandro, B., O'Neil, C. and LaGatta, T., (2017) Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big data*, 52, pp.120–134.

Altman, E.I., (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 234, pp.589–609.

Anon (2023) *CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence*. [online] Consumer Financial Protection Bureau. Available at: https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/ [Accessed 26 Feb. 2025].

Anon (2025) *Special topic – Artificial intelligence | European Banking Authority*. [online] Available at: https://www.eba.europa.eu/publications-and-media/publications/special-topic-artificial-intelligence?utm_source=chatgpt.com [Accessed 26 Feb. 2025].

Barboza, F., Kimura, H. and Altman, E., (2017) Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, pp.405–417.

Bartlett, R., Morse, A., Stanton, R. and Wallace, N., (2022) Consumer-lending discrimination in the FinTech Era. *Journal of Financial Economics*, 1431, pp.30–56.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A.W. and Siddique, A., (2016) Risk and risk management in the credit card industry. *Journal of Banking & Finance*, 72, pp.218–239.

Byanjankar, A., Heikkilä, M. and Mezei, J., (2015) Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. In: *2015 IEEE Symposium Series on Computational Intelligence*. [online] 2015 IEEE Symposium Series on Computational Intelligence. pp.719–725. Available at: https://ieeexplore.ieee.org/document/7376683 [Accessed 26 Feb. 2025].

Chakraborty, J., Majumder, S. and Menzies, T., (2021) Bias in machine learning software: why? how? what to do? In: *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ESEC/FSE 2021. [online] New York, NY, USA: Association for Computing Machinery, pp.429–440. Available at: https://dl.acm.org/doi/10.1145/3468264.3468537 [Accessed 25 Feb. 2025].

Chen, T. and Guestrin, C., (2016) XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. [online] New York, NY, USA: Association for Computing Machinery, pp.785–794. Available at: https://dl.acm.org/doi/10.1145/2939672.2939785 [Accessed 25 Feb. 2025].

Corbett-Davies, S., Gaebler, J.D., Nilforoshan, H., Shroff, R. and Goel, S., (2023) *The Measure and Mismeasure of Fairness*. Available at: http://arxiv.org/abs/1808.00023 [Accessed 26 Feb. 2025].

Efron, B. and Tibshirani, R., (1998) *An introduction to the bootstrap*. Nachdr. ed. Monographs on statistics and applied probability. Boca Raton, Fla.: Chapman & Hall.

Goodman, B. and Flaxman, S., (2017) European Union Regulations on Algorithmic Decision Making and a 'Right to Explanation'. *AI Magazine*, 383, pp.50–57.

Greenfield, K., (n.d.) Supporting Responsible Innovation in the Federal Banking System: An OCC Perspective.

Grömping, U., (2019) South German Credit Data: Correcting a Widely Used Data Set. *Report Series in Production Management, 33.*

Kamiran, F. and Calders, T., (2012) Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 331, pp.1–33.

Kelley, S. and Ovchinnikov, A., (2020) Anti-discrimination Law, AI, and Gender Bias in Non-mortgage Fintech Lending.

Kozodoi, N., Jacob, J. and Lessmann, S., (2022) Fairness in credit scoring: Assessment, implementation and profit implications. *European journal of operational research*, 2973, pp.1083–1094.

Liu, F.T., Ting, K.M. and Zhou, Z.-H., (2008) Isolation Forest. In: *2008 Eighth IEEE International Conference on Data Mining*. [online] 2008 Eighth IEEE International Conference on Data Mining. pp.413–422. Available at: https://ieeexplore.ieee.org/document/4781136/?arnumber=4781136 [Accessed 26 Feb. 2025].

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J. and Weinberger, K.Q., (2017) *On Fairness and Calibration*. Available at: http://arxiv.org/abs/1709.02012 [Accessed 26 Feb. 2025].

Zhang, W. and Ntoutsi, E., (2019) *FAHT: An Adaptive Fairness-aware Decision Tree Classifier*. Available at: http://arxiv.org/abs/1907.07237 [Accessed 26 Feb. 2025].

Blattner, L. and Nelson, S., (2021) *How Costly is Noise? Data and Disparities in Consumer Credit*. Available at: http://arxiv.org/abs/2105.07554 [Accessed 26 Feb. 2025].

Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R., (2012) Fairness through awareness. In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12. [online] New York, NY, USA: Association for Computing Machinery, pp.214–226. Available at: https://doi.org/10.1145/2090236.2090255 [Accessed 25 Feb. 2025].

Feldman, M., Friedler, S.A., Moeller, J., Scheidegger, C. and Venkatasubramanian, S., (2015) Certifying and Removing Disparate Impact. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15. [online] New York, NY, USA: Association for Computing Machinery, pp.259–268. Available at: https://doi.org/10.1145/2783258.2783311 [Accessed 25 Feb. 2025].

Fitzpatrick, T. and Mues, C., (2016) An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *European Journal of Operational Research*, 2492, pp.427–439.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T. and Walther, A., (2022) Predictably Unequal? The Effects of Machine Learning on Credit Markets. *The Journal of Finance*, 771, pp.5–47.

Hardt, M., Price, E., Price, E. and Srebro, N., (2016) Equality of Opportunity in Supervised Learning. In: *Advances in Neural Information Processing Systems*. [online] Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html [Accessed 26 Feb. 2025].

Jagtiani, J. and Lemieux, C., (2019) The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 484, pp.1009–1029.

Kleinberg, J., Mullainathan, S. and Raghavan, M., (2016) *Inherent Trade-Offs in the Fair Determination of Risk Scores*. Available at: http://arxiv.org/abs/1609.05807 [Accessed 26 Feb. 2025].

Lessmann, S., Baesens, B., Seow, H.-V. and Thomas, L.C., (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European journal of operational research*, 2471, pp.124–136.

Thomas, L.C., (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 162, pp.149–172.

Buuren, S. van, (2018) *Flexible Imputation of Missing Data, Second Edition.* 2nd ed. New York: Chapman and Hall/CRC.

Speicher, T., Heidari, H., Grgic-Hlaca, N., Gummadi, K.P., Singla, A., Weller, A. and Zafar, M.B., (2018) A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual &Group Unfairness via Inequality Indices. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* [online] KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. London United Kingdom: ACM, pp.2239–2248. Available at: https://dl.acm.org/doi/10.1145/3219819.3220046 [Accessed 27 Feb. 2025].

research proposal-updated-version-2.pdf

ORIGINALITY REPORT

6%	5%	2%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Liverpool John Moores University Student Paper	1%
2	ijrpr.com Internet Source	1%
3	Tarid Wongvorachan, Okan Bulut, Joyce Xinle Liu, Elisabetta Mazzullo. "A Comparison of Bias Mitigation Techniques for Educational Classification Tasks Using Supervised Machine Learning", Information, 2024 Publication	1%
4	Submitted to Nims University Rajasthan, Jaipur Student Paper	1%
5	uwe-repository.worktribe.com Internet Source	1%
6	www.ibm.com Internet Source	1%
7	akhilkumarreddymth522.sites.umassd.edu Internet Source	1%
8	export.arxiv.org Internet Source	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	On		