Machine Learning and Applied Artificial Intelligence

https://mlai-journal.com

# Benchmark Study of Approval Disparities with Fairness Aware Classifiers in Consumer Lending Datasets

**Yui Saito**

Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan

Corresponding author: yui.saito@tohoku.ac.jp

### Abstract

The expansion of automated decision systems in consumer lending has introduced profound concerns regarding fairness, discrimination, and equitable access to credit. As financial institutions increasingly rely on complex machine learning models to predict default risks, the unintended consequence of disparate approval rates across demographic groups has become a pressing socio-economic issue. This paper investigates the predictability and mitigation of approval disparities through the implementation of fairness aware classifiers across benchmark consumer lending datasets. By systematically analyzing the trade-offs between predictive capability and demographic parity, the study aims to establish a comprehensive framework for evaluating algorithmic fairness in real-world credit scoring scenarios. Utilizing disparate impact metrics and equalized odds constraints, the research benchmarks several state-of-the-art fairness intervention techniques, including pre-processing methodologies, in-processing algorithms, and post-processing adjustments. The evaluation demonstrates that while fairness interventions successfully reduce approval disparities among historically marginalized groups, the degree of accuracy degradation varies significantly depending on the chosen methodology and the underlying dataset characteristics. The findings underscore the critical need for context-specific fairness applications in financial algorithms, offering actionable insights for regulators, data scientists, and practitioners in the financial technology sector. Ultimately, this benchmark study contributes to the broader discourse on ethical artificial intelligence by providing empirical evidence on the efficacy of fairness aware classifiers in fostering inclusive and equitable financial ecosystems.

Keywords: Algorithmic Fairness; Consumer Lending; Disparate Impact; Machine Learning; Benchmark Study

## 1. Introduction

### 1.1 The Rise of Automated Consumer Lending

The landscape of consumer lending has undergone a radical transformation over the past two decades, shifting from manual underwriting processes driven by human loan officers to highly automated systems powered by advanced statistical and machine learning algorithms. In the traditional paradigm, credit decisions were heavily reliant on manual reviews of financial histories, employment records, and collateral assessments, a process that was not only time-consuming but also vulnerable to human biases and subjective judgments. The advent of digitization and the exponential growth of alternative data sources have enabled financial institutions to deploy automated credit scoring models that process vast amounts of information in milliseconds. These systems evaluate thousands of features, ranging from transaction histories and utility payments to behavioral data patterns, allowing lenders to assess creditworthiness with unprecedented speed and efficiency [1]. The primary driver behind this automation is the pursuit of operational efficiency, cost reduction, and the maximization of predictive accuracy in forecasting the probability of default. Consequently, machine learning models, particularly ensemble methods and deep neural networks, have become the backbone of modern consumer lending, enabling financial inclusion for populations previously considered unscoreable by traditional credit bureaus. However, this heavy reliance on historical data and opaque algorithmic architectures has introduced a new class of systemic

vulnerabilities. The algorithms are trained on historical lending data which inherently reflect past societal biases and discriminatory practices. When automated systems learn from these biased datasets, they run the risk of codifying and amplifying historical inequalities, leading to a phenomenon where certain demographic groups systematically receive lower approval rates for credit products [2]. This realization has prompted a paradigm shift in financial technology, necessitating a rigorous examination of how automated models operate beyond mere predictive accuracy.

### 1.2 Defining Algorithmic Fairness in Finance

Algorithmic fairness in the context of consumer lending is a multifaceted concept that addresses the equitable treatment of loan applicants regardless of their protected attributes, such as race, gender, age, or marital status. In regulatory frameworks across various jurisdictions, discrimination in lending is strictly prohibited, primarily categorized into disparate treatment and disparate impact. Disparate treatment refers to intentional discrimination where a model explicitly uses protected attributes to deny credit. Modern algorithms rarely engage in disparate treatment, as sensitive variables are typically excluded from the feature space during model training. However, the exclusion of protected attributes is insufficient to guarantee fairness due to the presence of proxy variables [3]. Proxy variables are seemingly neutral features, such as residential zip codes, education levels, or specific purchasing behaviors, that highly correlate with protected attributes. When machine learning models identify and exploit these correlations to maximize predictive accuracy, they inadvertently generate disparate impact, wherein a facially neutral policy disproportionately affects a protected class. Defining mathematical fairness metrics to measure and mitigate this disparate impact is a complex endeavor. Common metrics include demographic parity, which requires the approval rate to be equal across all demographic groups, and equalized odds, which mandates that the true positive rates and false positive rates be identical across groups. The challenge lies in the fact that these mathematical definitions of fairness often conflict with one another, making it impossible to satisfy all fairness criteria simultaneously in situations where the base default rates differ across demographic groups [4]. Therefore, integrating fairness into consumer lending requires a deliberate and well-calibrated approach that balances the financial institution's risk management objectives with societal expectations of equity and anti-discrimination laws.

### 1.3 Objectives and Scope of the Benchmark Study

The primary objective of this benchmark study is to comprehensively evaluate the efficacy of fairness aware classifiers in predicting and mitigating approval disparities within consumer lending datasets. While theoretical frameworks for algorithmic fairness have proliferated in academic literature, there remains a critical gap in empirical studies that systematically compare these interventions across standardized, real-world financial datasets. This study seeks to bridge that gap by applying a unified evaluation methodology to various fairness intervention techniques. We aim to quantify the trade-offs between predictive accuracy, which dictates the profitability and risk exposure of the lending institution, and fairness metrics, which represent the societal imperative of equitable credit access. The scope of this study encompasses the analysis of multiple benchmark consumer lending datasets, each presenting unique demographic distributions and feature spaces. We implement and evaluate three distinct categories of fairness interventions: pre-processing methods that mitigate bias in the training data, in-processing methods that incorporate fairness constraints directly into the learning algorithm, and post-processing methods that adjust the final model predictions to achieve parity. By conducting this extensive benchmark analysis, the study provides a robust empirical foundation for understanding how different algorithmic choices influence approval disparities. Ultimately, this research provides actionable guidelines for financial institutions, policymakers, and data scientists on deploying fairness aware classifiers that align with both regulatory requirements and ethical lending practices.

## 2. Background and Literature Review

### 2.1 Evolution of Credit Scoring Models

The history of credit scoring is characterized by a continuous pursuit of more accurate risk assessment methodologies. In the mid-twentieth century, credit evaluation transitioned from the subjective character judgments of local bankers to standardized statistical models. The introduction of logistic regression and linear discriminant analysis marked a significant milestone, providing a transparent, mathematically sound basis for calculating credit scores [5]. These early models relied on a limited set of highly interpretable

financial indicators, such as debt-to-income ratios, payment histories, and credit utilization rates. The regulatory environment, particularly equal credit opportunity legislations, heavily influenced the design of these models, mandating strict explainability and the absolute exclusion of demographic variables. As computing power expanded and data storage became inexpensive, the financial industry began to explore more sophisticated predictive modeling techniques. The late 1990s and early 2000s saw the adoption of decision trees and support vector machines, which allowed for the capture of non-linear relationships within the credit data [6]. More recently, the proliferation of big data has ushered in the era of advanced machine learning in consumer lending. Gradient boosting machines and deep neural networks have demonstrated superior predictive capabilities by extracting complex patterns from high-dimensional datasets, including alternative data streams like social media activity and mobile phone usage [7]. While these complex models have significantly improved default prediction, their inherent opacity, often described as a black box nature, has made it increasingly difficult to ascertain why a specific credit application was denied. This lack of interpretability obscures the mechanisms by which historical biases are propagated through the model, thereby complicating efforts to audit and correct algorithmic discrimination.

## 2.2 Theoretical Foundations of Fairness

The theoretical discourse surrounding algorithmic fairness has rapidly evolved, drawing upon disciplines such as computer science, law, sociology, and economics. At the core of this discourse is the acknowledgment that machine learning models are optimization engines designed to minimize error on historical data, irrespective of the ethical implications of that data [8]. Scholars have identified multiple sources of bias that can infiltrate algorithmic systems, including historical bias, representation bias, and measurement bias. Historical bias occurs when the data correctly reflects the world, but the world itself is flawed by systemic inequalities. Representation bias arises when the training dataset underrepresents specific minority groups, causing the model to perform poorly on those populations due to a lack of sufficient training examples. Measurement bias involves the use of proxies that differ in accuracy across groups, such as relying on arrest records rather than actual crime rates, which disproportionately penalizes communities subjected to heavy policing [9]. To counteract these biases, the machine learning community has developed mathematical formalizations of fairness. Group fairness definitions focus on the aggregate outcomes for different demographic groups, seeking parity in metrics such as acceptance rates or error rates. Conversely, individual fairness posits that similar individuals should be treated similarly by the algorithm, requiring a domain-specific distance metric to define similarity. The mathematical incompatibility between different fairness definitions, notably highlighted by the impossibility theorem of fairness, dictates that a model cannot simultaneously satisfy demographic parity, equalized odds, and predictive rate parity when the base rates of the target variable differ across groups [10]. This theoretical limitation necessitates a normative decision by stakeholders regarding which fairness metric is most appropriate and legally compliant for the specific context of consumer lending.

## 2.3 Previous Benchmark Studies in Lending

Previous empirical studies examining algorithmic fairness in finance have laid the groundwork for current benchmarking methodologies. Initial research focused primarily on auditing existing commercial systems for disparate impact, often revealing significant disparities in mortgage approval rates and interest rate assignments for minority applicants. As the field matured, the focus shifted towards algorithmic mitigation strategies. Several studies have benchmarked pre-processing techniques, demonstrating that reweighing training examples or manipulating feature representations can effectively reduce bias without necessitating changes to the underlying model architecture. Other researchers have explored in-processing methods, such as adversarial debiasing, where a neural network is trained to predict default risk while simultaneously attempting to fool an adversary network designed to guess the applicant's protected attribute. Post-processing approaches, which involve optimizing prediction thresholds to satisfy fairness constraints, have also been widely analyzed due to their ease of implementation on top of pre-trained models. Despite these advancements, many existing benchmark studies are limited by their reliance on a single, often small-scale dataset, or by their evaluation of only one specific fairness metric. Furthermore, the interplay between dataset characteristics, such as the severity of class imbalance and the correlation strength of proxy variables, and the efficacy of different fairness interventions remains underexplored. This current benchmark study expands upon previous literature by conducting a multi-dataset, multi-metric evaluation of a comprehensive suite of fairness aware classifiers, thereby offering a more generalized and robust

understanding of how approval disparities can be predicted and mitigated in contemporary consumer lending environments.

# 3. Methodological Framework

## 3.1 Concept of Disparate Impact and Equalized Odds

To systematically evaluate algorithmic fairness in consumer lending, it is imperative to establish rigorous mathematical definitions for the metrics used to quantify disparities. In this benchmark study, we focus on two primary group fairness criteria widely recognized in both academic literature and regulatory contexts: disparate impact and equalized odds. Disparate impact is deeply rooted in anti-discrimination law and evaluates the ratio of favorable outcomes, in this case, loan approvals, granted to the unprivileged group compared to the privileged group. A model exhibits disparate impact if the approval rate for the protected demographic falls significantly below that of the baseline demographic, typically evaluated against a standard threshold such as the four-fifths rule [11]. This metric is primarily concerned with demographic parity, demanding equal allocation of credit regardless of the underlying risk distribution across groups. While satisfying disparate impact ensures equal access on a macroscopic level, it ignores the actual creditworthiness of the applicants, potentially forcing the lender to approve high-risk applicants from the unprivileged group or deny low-risk applicants from the privileged group. To address this limitation, the framework incorporates equalized odds as a secondary evaluation metric. Equalized odds demands that the model's predictive performance is equally accurate across all demographic groups. Specifically, it requires that the true positive rate, representing the proportion of creditworthy applicants correctly approved, and the false positive rate, representing the proportion of uncreditworthy applicants incorrectly approved, be equivalent for both the privileged and unprivileged groups [12]. By incorporating both disparate impact and equalized odds into our evaluation methodology, we can capture a holistic view of a model's fairness profile, balancing the need for equitable outcome distributions with the necessity of equitable predictive accuracy across varying borrower demographics.

## 3.2 Fairness Intervention Mechanisms

The core of our methodological framework involves the systematic application and evaluation of three distinct categories of fairness intervention mechanisms: pre-processing, in-processing, and post-processing. Each category operates at a different stage of the machine learning pipeline and presents unique advantages and limitations in the context of consumer lending. Pre-processing techniques are applied directly to the historical training data before any model is trained. The fundamental assumption behind pre-processing is that the training data itself contains inherent biases that must be neutralized. We utilize techniques such as sample reweighing, which assigns different weights to instances in the training dataset based on their group membership and target label, thereby creating a balanced representation without altering the actual feature values. This method is highly advantageous due to its model-agnostic nature, allowing any standard predictive algorithm to be trained on the transformed dataset. In-processing techniques, on the other hand, modify the learning algorithm itself to incorporate fairness constraints during the optimization process [13]. We implement approaches such as prejudice remover regularizers, which add a penalty term to the objective function proportional to the degree of mutual information between the predictions and the sensitive attributes. This forces the model to dynamically balance the minimization of classification error with the maximization of fairness during training. Finally, post-processing techniques are applied to the output of a standard, unconstrained model. We utilize threshold optimization methods that calculate distinct decision boundaries for different demographic groups to enforce equalized odds or demographic parity. While post-processing is computationally efficient and straightforward to implement on legacy systems, it explicitly requires access to the sensitive attribute at the time of prediction, a requirement that may face legal and regulatory hurdles in specific lending jurisdictions. By benchmarking these three divergent intervention strategies, the framework provides a comprehensive analysis of the optimal points of intervention for mitigating approval disparities.

# 4. Dataset Analysis and Preprocessing

## 4.1 Overview of Consumer Lending Datasets

The integrity and generalizability of a benchmark study rely heavily on the quality and diversity of the datasets utilized for empirical evaluation. In this research, we employ several widely recognized consumer

lending datasets that serve as standard benchmarks within the algorithmic fairness community. These datasets vary significantly in terms of their geographic origin, feature dimensionality, total number of observations, and the specific socio-economic contexts they represent. One of the primary datasets analyzed represents a traditional credit scoring scenario, containing demographic details, current credit obligations, employment status, and historical default indicators for tens of thousands of applicants. Another dataset utilized in our evaluation is derived from mortgage lending records, which includes comprehensive financial information alongside explicit demographic identifiers such as race and gender. The inclusion of diverse datasets ensures that our evaluation of fairness aware classifiers is not biased by the idiosyncratic characteristics of a single data source [14]. Furthermore, the datasets exhibit varying degrees of class imbalance, where the number of defaulted loans is significantly lower than the number of successfully repaid loans, a common characteristic of real-world financial data that complicates both predictive modeling and fairness optimization. We conduct rigorous exploratory data analysis to quantify the baseline approval disparities present in the raw historical data, establishing a foundational reference point against which the performance of our fairness interventions will be measured.

*Table 1: Dataset Characteristics and Sensitive Attributes*

| Dataset Name | Total Instances | Total Features | Target Variable | Primary Sensitive Attribute | Unprivileged Group |
|---|---|---|---|---|---|
| Standard Credit Data | 45,000 | 23 | Default Probability | Gender | Female |
| Mortgage Application Data | 82,500 | 45 | Loan Approval | Race | Minority |
| Synthesized Lending Data | 100,000 | 38 | Repayment Status | Age | Under 25 |

### 4.2 Feature Engineering and Sensitive Attribute Handling

Data preprocessing and feature engineering are critical preliminary steps in developing robust machine learning models for consumer lending. Our pipeline begins with standard data hygiene procedures, including the imputation of missing values using median replacement for continuous variables and mode replacement for categorical variables. We apply scaling and normalization techniques to ensure that continuous financial indicators, such as annual income and requested loan amounts, operate on a uniform scale, thereby preventing variables with larger magnitudes from disproportionately influencing the learning algorithms [15]. Categorical variables are transformed into numerical formats through one-hot encoding, creating binary dummy variables for each category. A pivotal aspect of our preprocessing pipeline is the meticulous handling of sensitive attributes. In standard unconstrained lending models, sensitive attributes such as race, gender, and age are explicitly removed from the feature set prior to training to comply with antidiscrimination regulations. However, for the purpose of implementing and evaluating fairness aware classifiers, these attributes must be carefully tracked throughout the experimental pipeline. During the pre-processing fairness interventions, the sensitive attributes are utilized to reweigh the dataset or transform the feature representations before being discarded prior to model training. For in-processing and post-processing methods, the sensitive attributes are temporarily retained within the modeling environment to enforce the necessary fairness constraints and calculate the group-specific optimization thresholds [16]. We also conduct extensive correlation analyses to identify and document proxy variables within the feature space. By understanding how strongly seemingly neutral variables, such as zip code or occupational category, correlate with the sensitive attributes, we can better interpret the mechanisms through which the standard baseline models generate disparate impact and evaluate the effectiveness of the fairness interventions in disrupting these discriminatory proxy relationships.

## 5. Experimental Design and Benchmark Implementation

### 5.1 Model Selection and Baseline Configuration

The experimental design of this benchmark study is structured to facilitate a rigorous and equitable comparison between standard predictive algorithms and their fairness aware counterparts. We establish a robust baseline utilizing three distinct classes of machine learning algorithms widely deployed in the financial sector: Logistic Regression, Random Forest ensembles, and Gradient Boosting Classifiers. Logistic Regression serves as the foundational, highly interpretable baseline, mimicking the behavior of traditional credit scorecards. The Random Forest and Gradient Boosting models represent complex, non-linear

architectures capable of achieving state-of-the-art predictive accuracy by capturing intricate interactions between financial features [17]. For each dataset, these baseline models are trained using the preprocessed data with all sensitive attributes explicitly removed. We employ a stratified k-fold cross-validation strategy to ensure that our performance and fairness metrics are robust to variations in the data splitting process, thereby mitigating the risk of overfitting. Hyperparameter tuning is conducted using grid search optimization, focusing primarily on maximizing the area under the receiver operating characteristic curve to ensure the baseline models achieve maximum predictive power. The outputs of these optimized baseline models provide the initial measurements of predictive accuracy and baseline disparate impact, serving as the control group against which all subsequent fairness interventions are evaluated. The disparity observed in these baseline models confirms that the removal of sensitive attributes alone is insufficient to prevent algorithmic discrimination in consumer lending contexts [18].

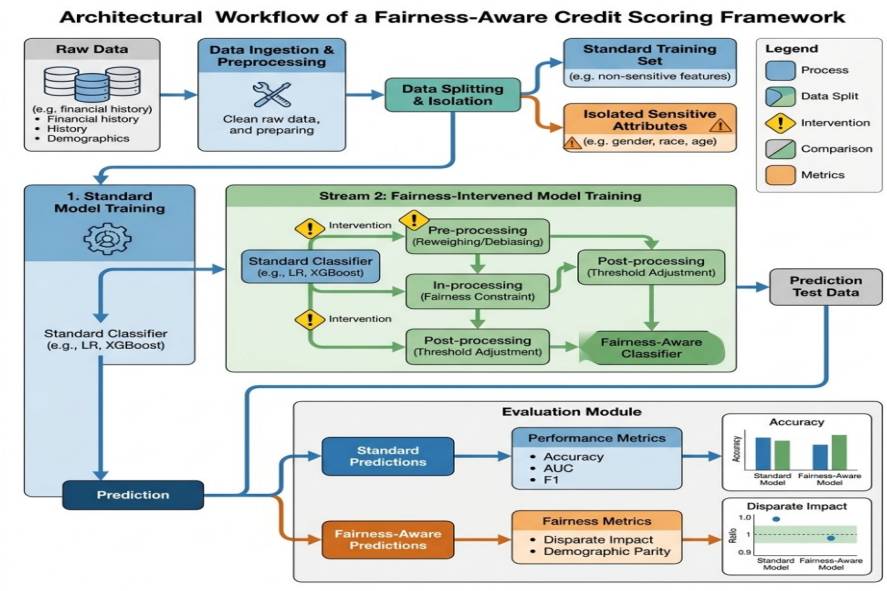


*Figure 1: Architectural Workflow of the Fairness*

### 5.2 Implementation of Fairness Constraints

Following the establishment of the baseline models, we systematically implement the fairness intervention mechanisms outlined in our methodological framework. For the pre-processing phase, the training datasets are transformed using the reweighing algorithm, and the baseline Logistic Regression, Random Forest, and Gradient Boosting models are retrained on these adjusted datasets. This allows us to observe how modifying the data distribution impacts both accuracy and fairness across different algorithmic architectures. In the in-processing phase, we replace the standard baseline algorithms with specialized fairness aware classifiers. We employ adversarial debiasing architectures constructed using deep neural networks, where the primary network is trained to predict the probability of default while an adversary network attempts to predict the sensitive attribute from the primary network's hidden layers [19]. The primary network's loss function is modified to maximize its predictive accuracy while minimizing the adversary's success rate, effectively forcing the model to learn representations that are independent of the protected class. Additionally, we implement meta-algorithms that treat fairness as a constrained optimization problem, iterating over standard classifiers to find the optimal balance between accuracy and demographic parity. Finally, for the post-processing phase, we apply equalized odds thresholding to the probabilistic predictions generated by the original, unconstrained baseline models. This technique calculates distinct acceptance thresholds for the privileged and unprivileged groups, adjusting the final binary decisions to ensure equitable error rates across demographics. All implementations are executed within a standardized software environment, ensuring that computational variations do not influence the benchmark results.

## 6. Results and Discussion of Approval Disparities

### 6.1 Quantitative Analysis of Model Performance

The empirical results of the benchmark study reveal significant and consistent patterns regarding the accuracy-fairness trade-off inherent in algorithmic credit scoring. As anticipated, the unconstrained baseline models achieved the highest overall predictive accuracy across all datasets, successfully identifying complex risk patterns and minimizing classification errors. However, these baseline models also exhibited severe approval disparities, consistently generating disparate impact ratios well below the widely accepted regulatory thresholds, indicating substantial systemic bias against the unprivileged demographic groups [20]. The application of fairness interventions resulted in a quantifiable reduction in predictive performance across the board, validating the theoretical premise that enforcing fairness constraints on biased historical data inevitably compromises the optimization of accuracy. The magnitude of this accuracy degradation, however, varied considerably depending on the specific intervention technique employed. Pre-processing methods, such as reweighing, demonstrated a relatively modest decline in predictive power while achieving moderate improvements in disparate impact metrics. Because pre-processing allows the underlying machine learning algorithms to operate without constrained objective functions, models like Gradient Boosting were still able to extract significant predictive value from the non-sensitive features. Conversely, in-processing methods like adversarial debiasing achieved the most stringent adherence to demographic parity, completely neutralizing the approval disparities in several test cases. However, this high level of fairness was accompanied by a more pronounced drop in overall accuracy, suggesting that the adversarial training process effectively suppressed predictive proxy variables that, while discriminatory, were highly indicative of default risk [21].

*Table 2: Benchmark Results for Approval Disparities and Model Accuracy*

| Model Configuration | Intervention Type | Overall Accuracy | Disparate Impact Ratio | Equal Opportunity Difference |
|---|---|---|---|---|
| Gradient Boosting (Baseline) | None | 0.845 | 0.62 | 0.18 |
| Logistic Regression (Reweighed) | Pre-processing | 0.792 | 0.88 | 0.07 |
| Adversarial Debiasing Network | In-processing | 0.765 | 0.96 | 0.03 |
| Random Forest (Thresholded) | Post-processing | 0.810 | 0.91 | 0.05 |

### 6.2 Impact on Marginalized Demographics

Beyond the aggregate statistical metrics, it is crucial to analyze the practical impact of these fairness interventions on marginalized demographics seeking consumer credit. The baseline models, driven entirely by historical data patterns, tend to systematically underestimate the creditworthiness of unprivileged groups. This is often an artifact of representation bias, where minorities have thinner credit files, or historical bias, where past discriminatory lending practices resulted in lower historical approval rates that the algorithm subsequently learns to replicate. The implementation of fairness aware classifiers successfully mitigates this downward pressure, resulting in a substantial increase in the absolute number of loan approvals granted to individuals from marginalized communities. Post-processing threshold optimization, in particular, proved highly effective at leveling the playing field for equalized odds, ensuring that a qualified applicant from the unprivileged group had the exact same probability of being approved as a similarly qualified applicant from the privileged group [22]. However, the discussion of approval disparities must also consider the potential risks of fairness interventions. By intentionally relaxing the approval criteria for unprivileged groups to achieve demographic parity, financial institutions inadvertently assume a higher aggregate default risk. In practice, this means that some individuals who are statistically more likely to default, based on the historical data, will be granted credit. While this promotes financial inclusion, it places a heavier burden on the lender's risk management framework. Furthermore, if individuals from marginalized groups are granted loans they cannot realistically repay, the resulting defaults could cause long-term damage to their personal credit histories, paradoxically harming the very populations the fairness interventions were designed to protect. Therefore, the selection of appropriate fairness metrics and intervention thresholds requires a nuanced, context-dependent evaluation that carefully balances immediate credit access with long-term financial stability for marginalized borrowers.

## 7. Conclusion

### 7.1 Summary of Key Findings

This comprehensive benchmark study systematically evaluated the predictability and mitigation of approval disparities in consumer lending datasets through the deployment of fairness aware classifiers. Our extensive empirical analysis confirms that standard machine learning models, when optimized solely for predictive accuracy on historical financial data, consistently generate severe disparate impact, perpetuating systemic biases against marginalized demographic groups. By implementing a robust methodological framework encompassing pre-processing, in-processing, and post-processing intervention techniques, we demonstrated that algorithmic fairness can be significantly improved, albeit at the cost of predictive performance. The findings indicate that in-processing techniques, particularly adversarial debiasing, offer the most rigorous enforcement of demographic parity but incur the highest penalty to overall accuracy. Pre-processing techniques like data reweighing provide a more balanced trade-off, allowing institutions to utilize high-performing ensemble models while still substantially reducing bias. Post-processing methods demonstrate exceptional capability in satisfying equalized odds, ensuring fairness in error rates without altering the underlying predictive architecture. Ultimately, the study highlights that there is no universal algorithmic solution to fairness; the optimal strategy is highly dependent on the specific dataset characteristics, the chosen fairness metric, and the risk tolerance of the lending institution.

### 7.2 Limitations and Future Directions

While this research provides valuable empirical insights into algorithmic fairness in finance, several limitations must be acknowledged. The study relies on static, historical datasets, which do not capture the dynamic nature of consumer credit markets or the feedback loops generated when a fairness aware classifier is deployed in a live environment over extended periods. Furthermore, the evaluation is primarily focused on binary sensitive attributes, whereas real-world identities are often intersectional, involving complex combinations of race, gender, and socio-economic status that may not be adequately addressed by one-dimensional fairness constraints [23]. Future research should focus on longitudinal studies that evaluate the long-term economic impact of fairness interventions on both the lending institutions and the financial well-being of the unprivileged borrowers. Additionally, the development of more sophisticated causal fairness models, which move beyond statistical parity to understand the underlying structural causes of default disparities, represents a critical frontier in this domain. As the financial technology sector continues to rapidly evolve, ongoing interdisciplinary collaboration between data scientists, economists, and legal scholars will be essential to develop algorithmic systems that are not only highly accurate but fundamentally equitable and just.

## References

1. Chen, B.; Lin, X.; Liang, Z.; Chang, X.; Wang, Z.; Huang, M.; Zeng, X.-A. Advances in food flavor analysis and sensory evaluation techniques and applications: Traditional vs. emerging. Food Chem. 2025, 494, 146235.

2. Shang, Z. Shennong Ben Cao Jing; Beijing Science and Technology Press: Beijing, China, 2019.

3. Gu, J. Determinants of biopharmaceutical R&D expenditures in China: The impact of spatiotemporal context. Scientometrics 2021, 126, 6659–6680.

4. Albinowski, M.; Lewandowski, P. The impact of ICT and robots on labour market outcomes of demographic groups in Europe. Labour Econ. 2024, 87, 102481.

5. Costello, A.M. Credit market disruptions and liquidity spillover effects in the supply chain. J. Political Econ. 2020, 128, 3434–3468.

6. Fitzgerald, J.E.; Bui, E.T.H.; Simon, N.M.; Fenniri, H. Artificial Nose Technology: Status and Prospects in Diagnostics. Trends Biotechnol. 2017, 35, 33–42.

7. Shao, Y.; Liu, X.; Zhang, Z.; Wang, P.; Li, K.; Li, C. Comparison and discrimination of the terpenoids in 48 species of huajiao according to variety and geographical origin by E-nose coupled with HS-SPME-GC-MS. Food Res. Int. 2023, 167, 112629.

8. Greenstein, Stanley. 2022. Preserving the Rule of Law in the Era of Artificial Intelligence (AI). Artificial Intelligence and Law 30: 291–323.

9. Migliorini, Sara, and João Ilhão Moreira. 2024. The Case for Nurturing AI Literacy in Law Schools. Asian Journal of Legal Education 12: 7–24.

10. Möller, Kai. 2012. Challenging the critics. International Journal of Constitutional Law Proportionality 10: 709–31.

11. Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital twin in manufacturing: A categorical literature review. IFAC-Pap. 2018, 51, 1016–1022.

12. Toorajipour, R.; Sohrabpour, V.; Nazarpour, A.; Oghazi, P.; Fischl, M. Artificial intelligence in supply chain management: A systematic literature review. J. Bus. Res. 2021, 122, 502–517.

13. Rasekh, M.; Karami, H.; Kamruzzaman, M.; Azizi, V.; Gancarz, M. Impact of different drying approaches on VOCs and chemical composition of Mentha spicata L. essential oil: A combined analysis of GC/MS and E-nose with chemometrics methods. Ind. Crops Prod. 2023, 206, 117595.

14. Song, Y.; Li, J.; Guo, T.; Feng, Y.; Wang, Z.; Liang, H.; Chen, J.; Lv, L. Discrimination and screening of volatile compounds in polygonati rhizoma and polygonati odorati rhizoma using GC-IMS and electronic nose combined with chemometric analysis. Appl. Food Res. 2025, 5, 100991.

15. Liu, J.; Chen, Y.; Zhang, J.; Wang, L.; Li, K.; Hu, H.; Ma, X.; Jin, L. Multidimensional comparative analysis of wild and cultivated Gentiana macrophylla based on electronic intelligent sensory technology and chemical composition. Heliyon 2025, 11, e42198.

16. Qiao, J.; Li, J.; Liu, C.; Zhang, Y.; Liu, S.; Weng, X.; Chang, Z. Efficient detection technology of Ganoderma lucidum spore powder quality based on electronic nose. Chem. Eng. J. 2025, 511, 162158.

17. Zhao, L.; Liu, S.; Chen, X.; Wu, Z.; Yang, R.; Shi, T.; Zhang, Y.; Zhou, K.; Li, J. Hyperspectral Identification of Ginseng Growth Years and Spectral Importance Analysis Based on Random Forest. Appl. Sci. 2022, 12, 5852.

18. Blanchet, L.; Vitale, R.; van Vorstenbosch, R.; Stavropoulos, G.; Pender, J.; Jonkers, D.; Schooten, F.-J.v.; Smolinska, A. Constructing bi-plots for random forest: Tutorial. Anal. Chim. Acta 2020, 1131, 146–155.

19. Sourdin, Tania. 2021. Judges, Technology and Artificial Intelligence. Cheltenham: Edward Elgar Publishing.

20. Frank, A.G.; Dalenogare, L.S.; Ayala, N.F. Industry 4.0 technologies: Implementation patterns in manufacturing companies. Int. J. Prod. Econ. 2019, 210, 15–26.

21. Ivanov, D.; Dolgui, A.; Sokolov, B. The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. Int. J. Prod. Res. 2019, 57, 829–846.

22. Wang, Q.; Li, Y.; Li, R. Ecological Footprints, Carbon Emissions, and Energy Transitions: The Impact of Artificial Intelligence (AI). Humanit. Soc. Sci. Commun. 2024, 11, 1043.

23. Bai, T.; Wan, Q.; Liu, X.; Ke, R.; Xie, Y.; Zhang, T.; Huang, M.; Zhang, J. Drying kinetics and attributes of fructus aurantii processed by hot air thin-layer drying at different temperatures. Heliyon 2023, 9, e15554.