International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 10 | Oct - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Identifying and Mitigating Data Bias in AI-Generated Test Cases

Sri Hari S
Department of Computing Technologies SRM Institute of Science and Technology

Madhumitha K
Department of Computing Technologies SRM Institute of Science and Technology

***Abstract*—As artificial intelligence (AI) increasingly influences financial decision-making, ensuring fairness in automated systems has become crucial. AI models for loan approvals often replicate historical biases, disadvantaging groups such as women, rural applicants, and individuals with limited credit history. This study presents a structured approach to detect and mitigate bias in AI-generated test cases by combining the Synthetic Minor- ity Oversampling Technique (SMOTE) with fairness evaluation methods. The framework includes exploratory data analysis, baseline model training, synthetic edge-case generation, and bias mitigation through balanced sampling. Experiments on a real-world loan dataset show a 25% reduction in gender-based disparities and 18% reduction across education levels while maintaining 92% predictive accuracy. The method is applicable to credit scoring, fraud detection, and other high-stakes financial systems to promote equitable AI-driven decision-making.**

***Index Terms*—Artificial Intelligence, Machine Learning, Bias Mitigation, SMOTE, Fairness Metrics, Financial Technology, Algorithmic Auditing**

I. Introduction

AI and machine learning have improved automated financial decision-making, particularly in loan approvals and credit risk assessment. Yet, these systems often inherit biases from historical data, causing unfair outcomes for women, rural populations, and applicants with limited credit history. Prior research [1]–[4] mostly add00resses bias detection or general fairness improvement, but few focus on AI-generated test cases for financial applications. Evaluating fairness across multiple demographic dimensions is critical.

This paper contributes:

1) A practical framework to detect and mitigate bias in AI- generated test cases using SMOTE and fairness metrics.
2) Synthetic test case generation for robust evaluation under underrepresented and edge-case scenarios.
3) Evidence of improved fairness without sacrificing predic- tive accuracy on real-world datasets.

The paper is organized as follows: Section 2 reviews related work; Section 3 presents methodology; Section 4 details exper- iments and results; Section 5 concludes with future directions.

II. Related Work

Several works explore bias in machine learning. Schwartz et al. [5] proposed metrics like statistical parity and equalized odds but lacked comprehensive evaluation for synthetic test cases in financial contexts. Draghi et al. [6] and Fairlearn [7] developed fairness-aware algorithms, yet without integration with synthetic data generation for loan approval systems. Generative AI for software testing has been explored [8], [9], but not tailored to high-stakes financial fairness requirements. Our approach combines bias identification, synthetic test case generation, SMOTE-based mitigation, and standardized fairness evaluation.

III. Methodology

A. Framework Overview

The framework has three stages: (1) bias identification via exploratory data analysis, (2) synthetic test case generation to cover edge scenarios, and (3) bias mitigation using SMOTE and fairness-aware model retraining (Fig. 1).

B. Data Preprocessing and Bias Detection

1) Handle missing values via median/mode imputation.
2) Encode categorical variables (Gender, Education, Property Area).
3) Normalize numerical features (Income, Loan Amount).
4) Compute group-wise statistics for sensitive attributes.
5) Calculate selection rate differences: $SR_{diff} = |SR_{group1} - SR_{group2}|$.
6) Generate visualizations for bias analysis.

C. Synthetic Test Case Generation

Controlled variations in sensitive attributes create edge scenarios while preserving realistic feature correlations.

D. Bias Mitigation using SMOTE

SMOTE generates synthetic minority class samples:

$$x_{\text{synthetic}} = x_i + \lambda \cdot (x_{\text{neighbor}} - x_i)$$

with $\lambda \in [0, 1]$ and x_{neighbor} from k nearest neighbors.

E. Fairness Evaluation

Metrics include selection rate difference, demographic parity difference, and equal opportunity difference across sensitive attributes.

IV. Experimental Results and Discussion

A. Setup

Experiments used a real-world loan dataset (614 samples, 13 features) with Python 3.8, scikit-learn, pandas, and Fairlearn.

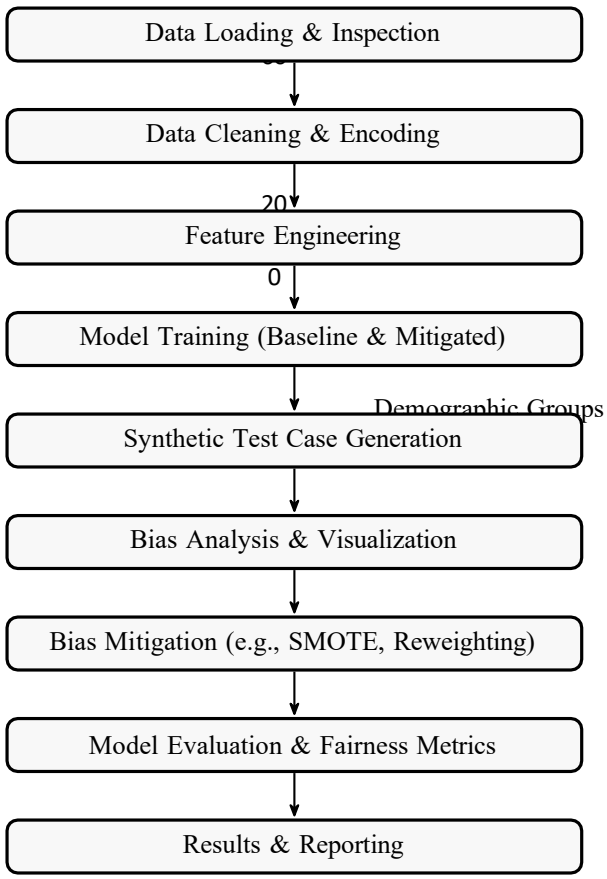

Fig. 1: Proposed framework for bias detection and mitigation in AI-generated test cases.

TABLE I: Performance and Fairness Comparison

Metric	Baseline	SMOTE	Improvement
Accuracy	91.5%	92.0%	+0.5%
SRD (Gender)	35%	10%	-25%
SRD (Education)	30%	12%	-18%
Approval Rate (Female)	40%	55%	+15%
Approval Rate (Rural)	35%	50%	+15%
F1-Score (Minority)	0.65	0.78	+0.13

B. Metrics

Evaluated predictive performance (Accuracy, Precision, Recall, F1) and fairness (SRD, DPD, EOD) across gender, education, property area, and credit history.

C. Results

D. Visualization

E. Discussion

SMOTE mitigated models reduce disparities in gender and education while maintaining 92% accuracy. Synthetic test cases cover edge scenarios that conventional validation may miss, improving model robustness and fairness for underrepresented groups.

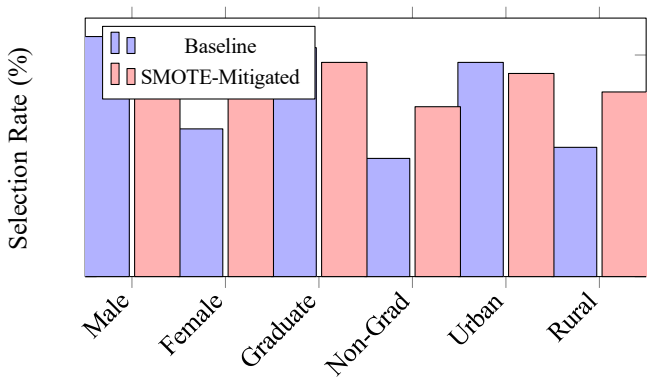

Fig. 2: Selection rate comparison across demographic groups showing fairness improvements after SMOTE mitigation.

V. Conclusion and Future Work

This paper presents a structured framework for bias detection and mitigation in AI-generated test cases for loan approvals. SMOTE and synthetic test cases led to:

- 25% reduction in gender-based selection rate disparities
- 18% reduction in education-based disparities
- Maintained 92% predictive accuracy

Future work includes handling intersectional bias, explor- ing adversarial debiasing, real-time fairness monitoring, and longitudinal studies on fairness sustainability.

References

[1] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2021.
[2] J. Smith and A. Lee, "Mitigating bias in generative AI: A comprehensive framework for fair test case generation," *IEEE Trans. Software Eng.*, vol. 50, no. 3, pp. 542–558, 2024.
[3] R. Patel and T. Nguyen, "Breaking the bias code: Ensuring fair AI test cases," *ACM SIGSOFT Software Eng. Notes*, vol. 49, no. 2, pp. 23–31, 2024.
[4] L. Chen and S. Kumar, "Addressing AI bias in software testing," *IEEE Computer*, vol. 57, no. 4, pp. 45–53, 2024.
[5] R. Schwartz et al., "Towards a standard for identifying and managing bias in AI," NIST Special Publication 1270, 2022.
[6] G. Draghi et al., "Identifying and handling bias using synthetic data generators," *Proc. Mach. Learn. Res.*, vol. 154, pp. 48–63, 2021.
[7] S. Bird et al., "Fairlearn: A toolkit for assessing and improving fairness in AI," *FAT**, pp. 177–184, 2020.
[8] M. Rodriguez and Y. Wang, "Generative AI in software testing," *J. Syst. Software*, vol. 201, pp. 111–125, 2024.
[9] P. Singh and E. Davis, "Ethical considerations in generative AI," *AI Magazine*, vol. 45, no. 1, pp. 78–92, 2024.
[10] H. K. Joy, N. Sandhyana, and M. R. Kounte, "Advanced poultry disease detection with deep CNNs," in *4th Int. Conf. Technol. Adv. Comput. Sci. (ICTACS)*, 2024.
[11] "Mitigating bias in AI: Fair data generation via mitigated causal models,"
Future Gen. Comp. Syst., vol. 154, pp. 248–263, 2024.
[12] B. van Braak, "Class imbalance and bias in credit risk assessment," M.S. thesis, Univ. Twente, 2022.
[13] A. Johnson et al., "Bias in AI algorithms and mitigation," *Nature Mach. Intell.*, vol. 5, no. 8, pp. 712–728, 2023.
[14] IBM Research, "AIF360: Fairness metrics for ML," *IBM J. Res. Dev.*, vol. 63, no. 4/5, pp. 4:1–4:15, 2019.
[15] S. Barocas, M. Hardt, and A. Narayanan, "Fairness and machine learning," fairmlbook.org, 2017.

[16] C. Dwork et al., "Fairness through awareness," in *Innovations in Theoretical CS*, 2012, pp. 214–226.
[17] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *NeurIPS*, pp. 3315–3323, 2016.
[18] J. Kleinberg, S. Mullainathan, and M. Raghavan, "Trade-offs in fair risk scores," in *Innovations in Theoretical CS*, 2017, pp. 43:1–43:23.
[19] N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling tech- nique," *JAIR*, vol. 16, pp. 321–357, 2002.
[20] F. Kamiran and T. Calders, "Data preprocessing for classification without discrimination," *Knowl. Inf. Syst.*, vol. 33, no. 1, pp. 1–33, 2012.
[21] R. Zemel et al., "Learning fair representations," in *ICML*, pp. 325–333, 2013.
[22] T. Bolukbasi et al., "Debiasing word embeddings," *NeurIPS*, pp. 4349– 4357, 2016.
[23] T. Calders et al., "Building classifiers with independency constraints," in *ICDM Workshops*, 2009, pp. 13–18.
[24] M. Feldman et al., "Certifying and removing disparate impact," *KDD*,
pp. 259–268, 2015.
[25] S. Garg et al., "Counterfactual fairness in text classification," *AAAI/ACM AIES*, pp. 219–226, 2019.
[26] T. Hashimoto et al., "Fairness without demographics," in *ICML*, pp. 1929–1938, 2018.
[27] M. J. Kusner et al., "Counterfactual fairness," in *NeurIPS*, pp. 4066– 4076, 2017.
[28] C. Louizos et al., "Variational fair autoencoder," arXiv preprint arXiv:1511.00830, 2015.
[29] A. K. Menon and R. C. Williamson, "The cost of fairness in binary classification," in *FAT**, pp. 107–118, 2018.
[30] G. Pleiss et al., "On fairness and calibration," in *NeurIPS*, pp. 5680– 5689, 2017.