Assessing Bias in Mortgage Lending and Payment Trends During the COVID-19 Pandemic Using Machine Learning

Erika Atoma

Department of Mathematics and Statistics (Data Science Program), American University, Washington DC

INTRODUCTION AND MOTIVATION

The COVID-19 pandemic reshaped the financial landscape, disproportionately impacting first-time homebuyers and exacerbating economic disparities across demographic groups. The chosen machine learning models analyze key predictors such as *Total Monthly Income*, *Loan-to-Value Ratio (LTV)*, *Debt-to-Income Ratio*, *Credit Score*, *Property Type*, and *Mortgage Type* to uncover systemic patterns in mortgage delinquency risk. Additionally, ML models assess whether borrower demographics such as *Gender*, *Ethnicity* and *Race* is associated with mortgage delinquency outcomes, which led to crippling effect post-pandemic times from the year 2023. By leveraging fairness-aware machine learning techniques and interpretability methods, the models aim to quantify potential biases in mortgage lending and identify important features when mortgage borrower groups faces disproportionate risk exposure during this economic downturn.

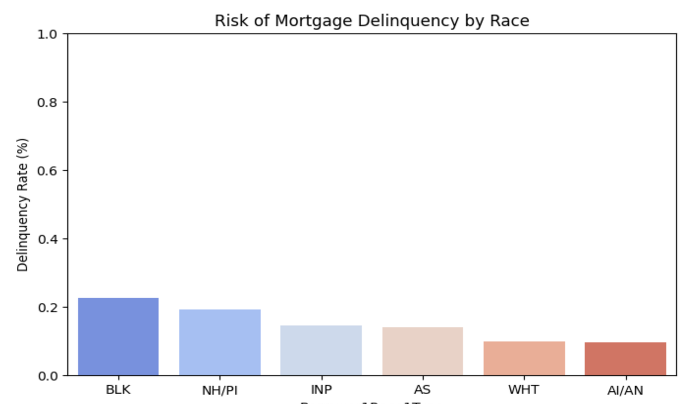
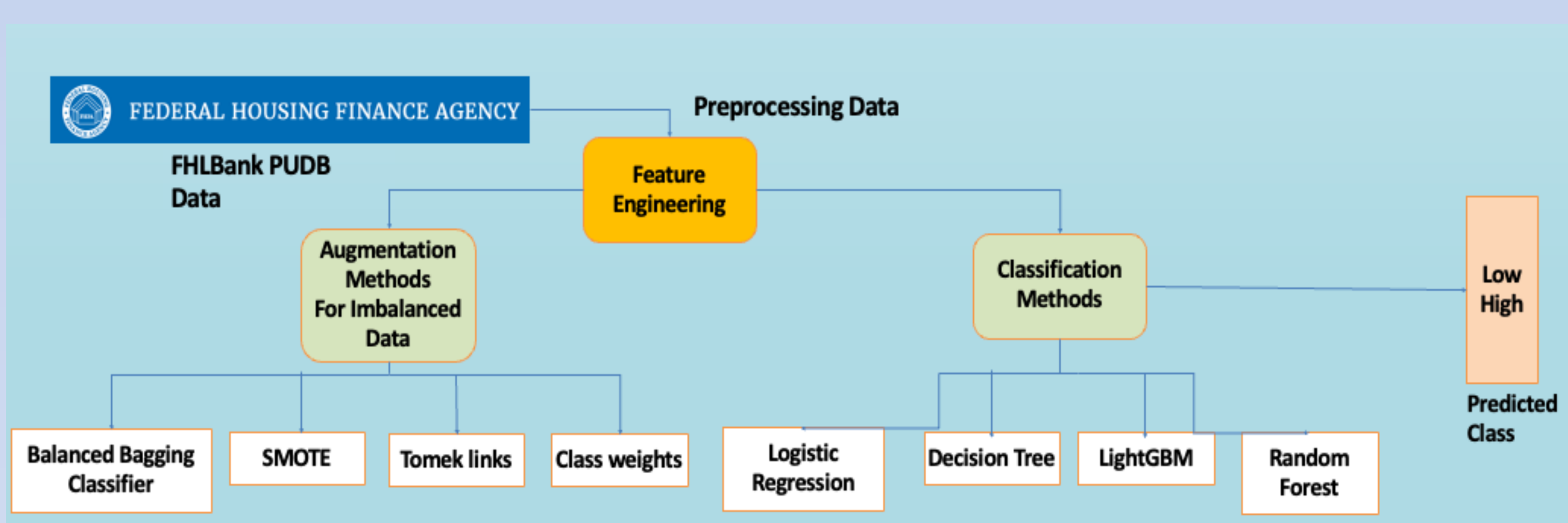


Figure 1: Risk of Mortgage Delinquency Rate by Race (year 2020-2022)

DATA

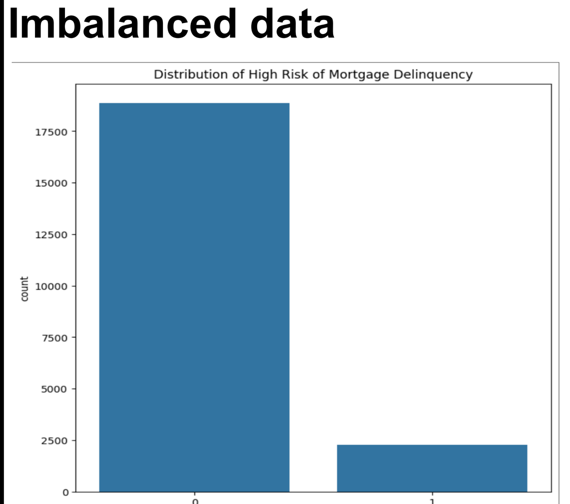
Public USE Database- Federal Home Loan Bank System, where the Federal Home Loan Bank (FHLBank) PUDB Datasets are collected from 2020 to 2022 to investigate the impact of financial and socio-economic characteristics of first-time home buyers to predict the likelihood of mortgage delinquency.

METHODS



Creation of a binary target variable $y = HighRiskofMortgageDelinquency$ based on the variable $TotalDebtExpense$ where it is classified 1 if $y > 45$ and 0 otherwise.

Imbalanced data



The number of low-risk borrowers: 2268
The number of high-risk borrowers: 18716
This was an imbalanced dataset, which was in need for augmentation methods.

Preprocessing

The dataset was into 80% of training set and 20% of testing set.

The log transformation method $y = \log(x)$ provide the tools to normalize the dataset when performing a principal component analysis (PCA) for the classification methods. The PCA dimensionality reduction transformation involved to explain 95% variance of the dataset. This was reduced from 13 to 10 principal components.

Analysis:

The underlying goal of comparing the PCA results. I compute the machine learning models with and without PCA and involved using augmented methods for imbalanced methods.

For Logistic Regression: $\log\left(\frac{X}{1-X}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \epsilon$

For Decision tree: $\hat{y} = f(X) = \sum_{j=1}^{J} c_j \cdot I(X \in R_j)$, and then it was pruned using $\alpha = 0.001$ to limit the growth of the trees which could lead to overfitting.

For Random Forest: $\hat{y} = 1/T \sum_{t=1}^{T} f_t(X)$ using the random state = 42.

For LightGBM: $\hat{y_t} = \widehat{y_{t-1}} + \eta \sum_{j=1}^{J} c_j I(X \in R_j)$

RESULTS

Experimental Results:

The standard metrics in machine learning-based classification, such as Precision, Recall, F-Measure, ROC-AUC were adopted, which gives a better assessment of the classification models when considering an imbalanced data classes.

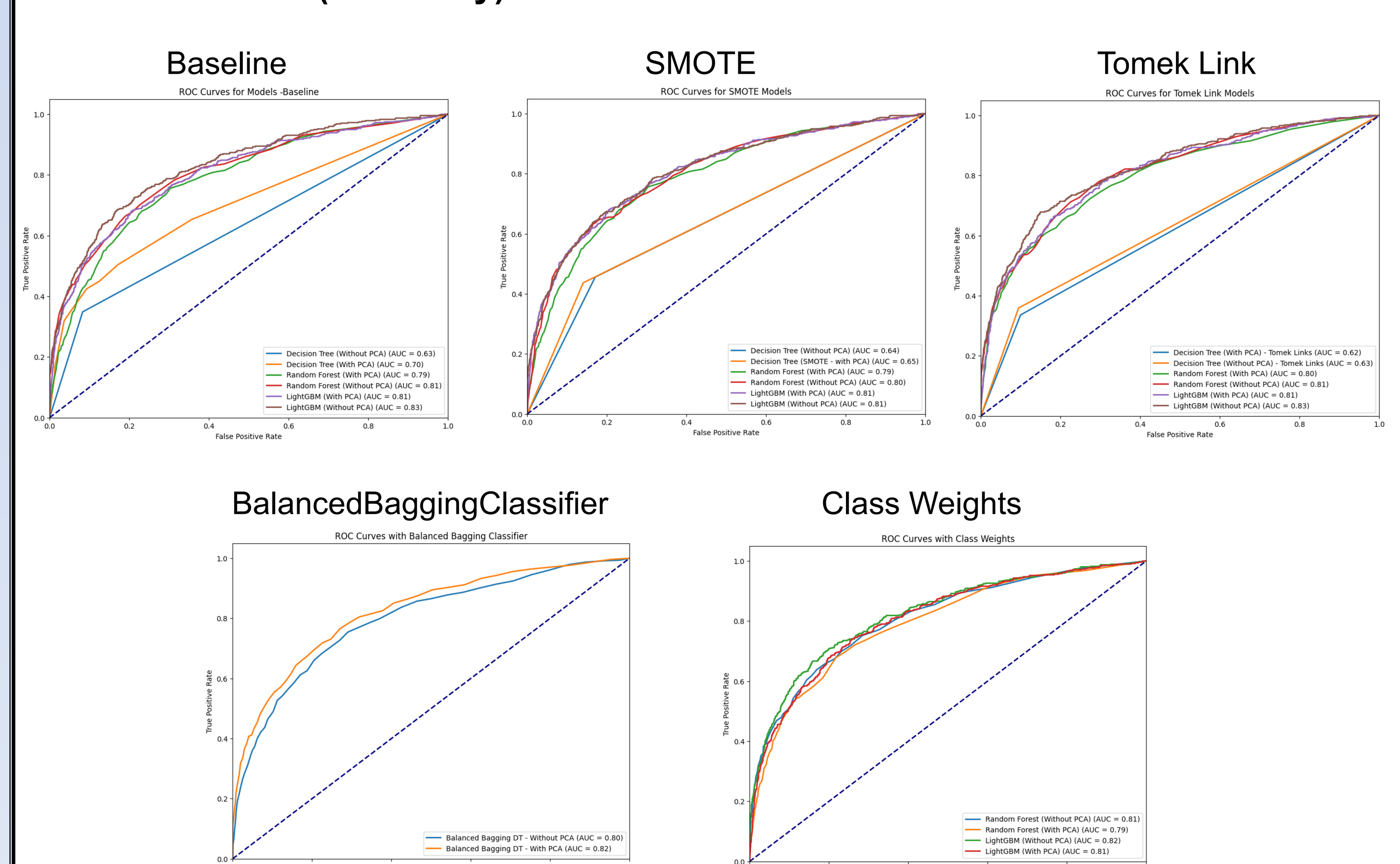
Table 1: Without PCA

Classifier	Data Augmentation	Label	Precision	Recall	F1-Score
Decision Tree	Baseline	0	0.92	0.92	0.92
		1	0.35	0.35	0.35
	Pruning	0	0.90	1.00	0.95
		1	0.89	0.16	0.28
	SMOTE	0	0.92	0.86	0.89
		1	0.28	0.44	0.34
	Tomek -Links	0	0.92	0.91	0.91
		1	0.33	0.36	0.34
	BalancedBaggingClassifier	0	0.95	0.85	0.90
		1	0.35	0.61	0.44
Random Forest	Baseline	0	0.91	0.99	0.95
		1	0.73	0.27	0.39
	SMOTE	0	0.93	0.93	0.93
		1	0.47	0.46	0.47
	Tomek-Links	0	0.91	0.98	0.95
		1	0.68	0.27	0.38
	Class Weights	0	0.91	0.99	0.95
		1	0.72	0.22	0.34
Logistic Regression	Baseline	0	0.91	0.99	0.95
		1	0.72	0.26	0.39
LightGBM	Baseline	0	0.91	0.99	0.95
		1	0.72	0.26	0.39
	SMOTE	0	0.93	0.93	0.93
		1	0.47	0.45	0.46
	Tomek-Links	0	0.92	0.98	0.95
		1	0.68	0.28	0.40
	Class weights	0	0.95	0.84	0.89
		1	0.35	0.67	0.46

Table 2: With PCA

Classifier	Methods	Label	Precision	Recall	F1-Score
Decision Tree	Baseline	0	0.92	0.91	0.91
		1	0.32	0.34	0.33
	Pruning	0	0.90	0.99	0.94
		1	0.62	0.17	0.26
	SMOTE	0	0.95	0.80	0.87
		1	0.30	0.66	0.41
	Tomek -Links	0	0.91	0.90	0.91
		1	0.30	0.34	0.32
	BalancedBaggingClassifier	0	0.94	0.85	0.89
		1	0.33	0.58	0.42
Random Forest	Baseline	0	0.91	0.98	0.94
		1	0.63	0.21	0.31
	SMOTE	0	0.93	0.89	0.91
		1	0.35	0.47	0.41
	Tomek-Links	0	0.91	0.98	0.95
		1	0.65	0.26	0.37
	Class Weights	0	0.90	0.99	0.94
		1	0.62	0.16	0.25
Logistic Regression	Baseline	0	0.90	0.99	0.94
		1	0.73	0.11	0.19
LightGBM	Baseline	0	0.91	0.99	0.95
		1	0.66	0.21	0.32
	SMOTE	0	0.95	0.80	0.87
		1	0.30	0.66	0.41
	Tomel-Links	0	0.91	0.98	0.95
		1	0.64	0.28	0.38
	Class weights	0	0.94	0.83	0.88
		1	0.31	0.62	0.42

ROC-AUC Curve (Accuracy) Models



Important Determinants for the Prediction of High-Risk Mortgage Borrowers

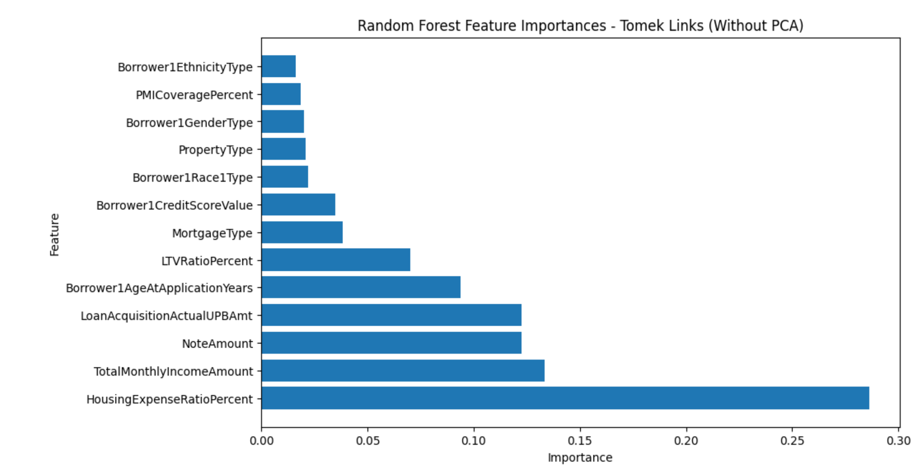


Figure 2: Random Forest Feature Importance (Tomek Link without PCA)

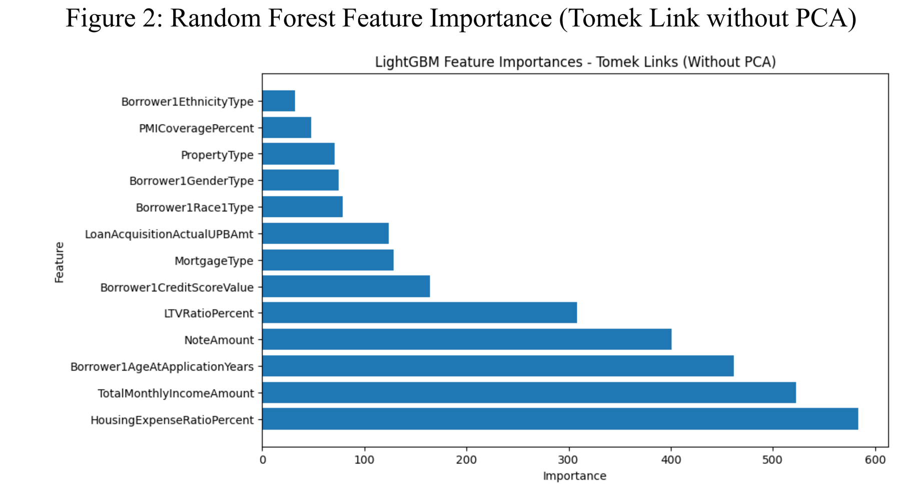


Figure 2: LightGBM Feature Importance (Tomek Link without PCA)

DISCUSSION

- PCA negatively impacts minority class recall, which reduces the model's ability to detect underrepresented cases.
- Pruning (for Decision Tree) and Tomek Link (for Random Forest & LightGBM) perform best without PCA.
- Random Forest and LightGBM achieved the highest ROC-AUC scores (~0.83), indicating strong predictive power for mortgage delinquency risk.
- Financial standing (Housing Expense Ratio%) is the most critical predictor, highlighting the role of affordability in mortgage delinquency risk.
- Borrower attributes such as Ethnicity, Gender, and Race have relatively low importance in both models, suggesting that the direct financial indicators are more decisive in predicting delinquency risk. However, this does not rule out systemic biases in loan approval or lending policies.

CONCLUSIONS/LIMITATIONS/FUTURE DIRECTIONS

The study demonstrates that LightGBM and Random Forest are the most effective models for predicting mortgage delinquency, but addressing class imbalance is crucial to accurately detect high-risk borrowers. The key financial predictors such as housing expense ratio, credit score, and loan-to-value ratio play a significant role, yet demographic disparities suggest that financial metrics alone cannot fully explain inequities.

It is limited by the scope of available data, which may not be generalizable to all lending institutions. Therefore, future research should incorporate alternative fairness-aware machine learning techniques and real-time prediction frameworks to better understand first-time home buyers' experiences and explore how local housing policies contribute to systemic disparities.

REFERENCES

Adam Zewe | MIT News Office (2022). Fighting discrimination in mortgage lending. (2022, March 30). Retrieved from https://news.mit.edu/2022/machine-learning-model-discrimination-lending-0330#:~:text=Caption%3A%20MIT%20researchers%20have%20developed,borrowers%20receive%20a%20mortgage%20loan

Office of Research (2021). Consumer Financial Protection Bureau. Retrieved from Characteristics of Mortgage Borrowers During the COVID-19 Pandemic. Retrieved from Consumer Financial Protection Bureau: https://www.consumerfinance.gov/data-research/research-reports/characteristics-mortgage-borrowers-during-covid-19-pandemic/

Public Use Database | FHFA. (2025, March 7). Retrieved from https://www.fhfa.gov/data/pudb

ACKNOWLEDGEMENTS

Special thanks to my advisor Professor Elizabeth Malloy, the Department of Mathematics and Statistics, and the Department of Computer Science.