Bias-Aware Machine Learning for Equitable Credit Lending Decisions

Sophia Mena, Mentor: Navid Bahadoran

Applied and Computational Mathematics, Department of Mathematics, Florida State University

Background

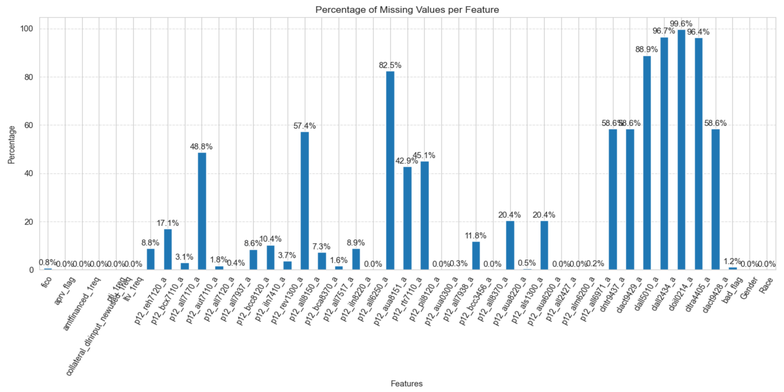
Nearly every industry uses machine learning models. It is common practice in banking, healthcare, and cybersecurity, playing a critical role decision-making processes. However, it is easy for these models to view people as numbers without realizing the real-world impact of their decisions on people's lives. Identifying and solving biases should be at the forefront of making these models to delivery precision and serve those who need it, particularly marginalized communities. The focus of this research is on auto loans. In the United States, a car is essential for daily life and auto loans are a common way that millions of Americans can afford it. The purpose of this research is to develop a model that is accurate and unbiased to ensure equitable credit lending decisions. This research contributes to building more inclusive financial systems that empower individuals and reduce disparities.

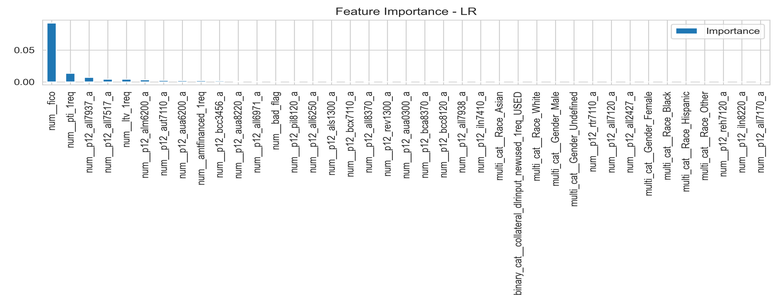
Methodology

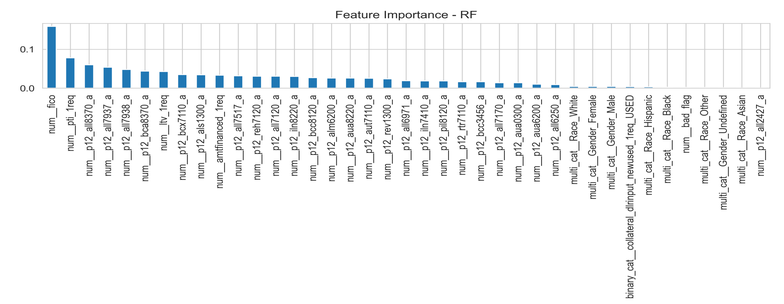
- **Preprocessing:**
  - Utilized a comprehensive dataset with 43 qualitative and quantitative features, including FICO score, payment-to-income ratio, loan-to-value ratio, gender, and race.
  - Removed columns with >80% missing data to ensure data quality.
  - Applied KNN Imputer to fill in remaining missing values, preserving dataset integrity.
- **Feature Analysis:**
  - Identified key features influencing approval odds, such as FICO score, total number of never delinquent trades, and loan-to-value ratio.
- **Model Development:**
  - Developed two predictive models. A logistic regression and a random forest classifier. We were also curious on which model would produce the most precise results
  - Both models were trained and evaluated to ensure accuracy and fairness in predicting credit approval.
- **Bias Mitigation:**
  - Analyzed data outputs to identify and address potential biases, through looking at each features p value and VIF number.
  - Also analyzed the demographics of the dataset to ensure that there was diversity among the models being trained.

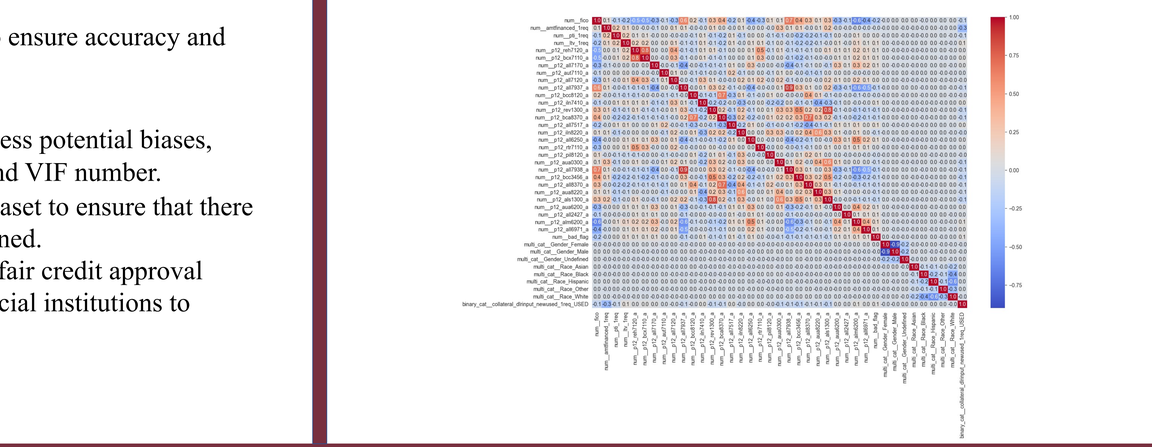
This structured approach ensures a robust and fair credit approval system, providing actionable insights for financial institutions to improve equity in lending.

Data



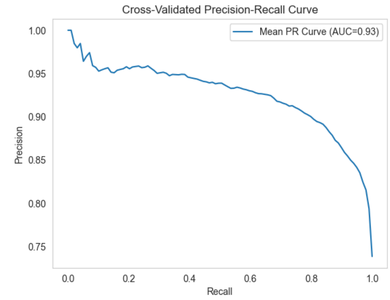




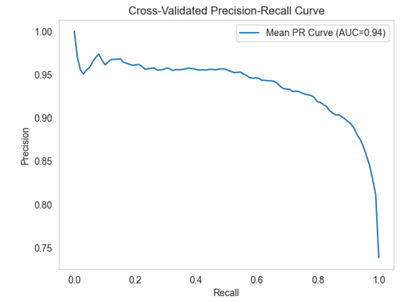


Results

1. A summary of the data revealed that the p values of gender and race is nan, indicating that those features are irrelevant in estimating an applicant's approval odds.
2. A VIF number was also calculated for each feature and revealed an infinite value. The infinite value indicates multicollinearity and further research into its effects in the model.
3. Precision of logistic regression model



4. Precision of random forest classifier



Future Studies

Since the model that we made was unbiased, the next step would be to find an existing biased model and correct that error while also making it more precise. Further research would also need to be done in experimenting with more advance models to improve the precision.

Acknowledgements

I want to especially thank my mentor, Navid Bahadoran. He has guided me throughout this whole process and let me make this project my own by allowing me to explore. It has been great to talk about the project weekly as well as figuring out different ways to tackle different issues and limitations that came along the way.