**Treating Unfairness in Machine Learning**

**Abstract**

In this project, we tackled gender bias in a credit granting machine learning model using a dataset of loan applicants. The primary goal was to ensure fairness between male and female applicants by adjusting the threshold of the model. We used a Random Forest classifier to predict loan status and evaluated fairness through the Positive Prediction Ratio (PPR) for both genders. By fine-tuning the threshold, we minimized gender disparity.

**Introduction**

Machine learning models can unintentionally propagate biases present in the data, leading to unfair outcomes. In this case, we focused on reducing gender bias in a credit granting model. The dataset includes a variety of features about applicants, such as Gender, Marital Status, Income, Education, Loan Amount, and Credit History, with the target variable being whether a loan was granted or not.

Since fairness is crucial in ensuring equal access to credit, we focused on minimizing gender bias while maintaining a well-performing model. Our primary metric for fairness was the Positive Prediction Ratio (PPR), which measures the rate of positive predictions (i.e., loan approvals) for each gender. The challenge was to reduce the difference between male and female applicants, ensuring that both genders have a comparable likelihood of receiving a loan.

**Method**

* **Data Preparation**: We cleaned the dataset by handling missing values, removing outliers, and creating a total income feature by summing the Applicant and Co-applicant incomes. The data was then split into training and test sets.
* **Modeling**: We utilized the Random Forest algorithm due to its robustness and efficiency. Since accuracy was not a suitable metric due to class imbalance, we aimed to optimize the F1 Score.
* **Hyperparameter Optimization**: Cross-validation and hyperparameter optimization were performed to fine-tune the model for best performance.
* **Fairness Metric**: To evaluate fairness, we calculated the Positive Prediction Ratio (PPR) for male and female applicants. We adjusted the decision threshold to reduce the difference in PPR between genders.
* **Fairness Adjustment**: Post-modeling, we adjusted the default prediction threshold from 0.5 to 0.44 to minimize the difference in the PPR of male and female female applicants.

**Result**

The project gave us insightful findings about the trade-offs between fairness and performance. Below are the key results before and after adjusting the threshold:

|  |  |
| --- | --- |
| **Before Threshold Adjustment** | **After Threshold Adjustment** |
| * Threshold: 0.5 * F1 Score: 0.852 * PPR Male: 0.804 * PPR Female: 0.767 * PPR Difference: 0.037 | * Threshold: 0.44 * F1 Score: 0.846 * PPR Male: 0.855 * PPR Female: 0.867 * PPR Difference: 0.012 |

By adjusting the threshold, we reduced the difference in PPR between male and female applicants. The PPR difference shrank from 0.037 to 0.012, an improvement in fairness. However, this came with a slight decrease in the F1 score from 0.852 to 0.846.

**Conclusion:**

This project highlights the importance of addressing fairness in machine learning models, particularly in sensitive domains like credit granting. Our analysis confirms that while Random Forest can produce a balanced and accurate model, this does not guarantee fairness between different groups—in this case, male and female applicants.

By adjusting the model’s prediction threshold to 0.44, we were able to improve fairness, as evidenced by the reduced PPR disparity between genders. However, this adjustment slightly impacted the model’s overall performance. This underscores the delicate balance between optimizing a model for fairness and maintaining high predictive accuracy. Finally, while we succeeded in improving gender fairness, the initial gender imbalance in the data suggests deeper structural biases that need to be addressed at the data collection and representation levels for long-term fairness.

***GitHub Link:*** [***RuslanZar/TechLab\_DSProjectTeam1 (github.com)***](https://github.com/RuslanZar/TechLab_DSProjectTeam1/tree/main)

*Ruslan Zarafutdinov –Data Science*

*Isabella Bismanns – Data Science*

*Patrick Bormann –Mentor*