AI Risk & Fairness Audit Report

Bias Detection and Fairness Auditing in Mortgage Loan Approvals

Ishan Juneja

1. Problem Overview

Task:
Detect and explain unusual patterns in AI decision-making for mortgage loan approvals using a provided dataset.Build a working model that is both accurate and fair backend by comprehensive bias reporting and fairness-aware techniques.

Real-World Importance:
This challenge matters deeply in real-world and ethical contexts because biased financial algorithms can worsen inequality by unfairly denying loans to individuals based on race, gender, age, or socioeconomic status.

Dataset:
Custom dataset (loan_access_dataset.csv) containing the following sensitive attributes:

- Gender
- Race
- Disability Status
- Criminal Record
- Income
- Age
- Zip code

2. Model Summary

I used a **Random Forest Classifier** for this task.

Why Random Forest?

- Handles categorical and numerical data well
- Resistant to overfitting
- Provides variable importance

Preprocessing Steps:

- One-hot encoding of categorical variables
- Standard scaling of numeric columns (Income, Credit Score Loan Amount)
- Handling missing values with default values (e.g., 0)

Performance:
Accuracy: 61.85%

Precision (Class 0 – Denied): 64%

Recall (Class 0 – Denied): 74%

Precision (Class 1 – Approved): 57%

Recall (Class 1 – Approved): 45%

F1-Score (Macro Avg): 60%

3. Bias Detection Process

I performed **group-level audits** using pandas SQL queries and visualizations.

Audited Aspects:

- Raw data (EDA)
- Model output (predictions)

Bias Detection Techniques Used:

- Approval rate by group comparisons
- Proportion calculations with pandas SQL
- Fairness visuals (stacked bar plots)
- Intersectional analysis by combining Gender + Race or Age + Disability

The analysis focused on clearly interpretable bias patterns using group statistics and visualization.

4. 📉 Identified Bias Patterns

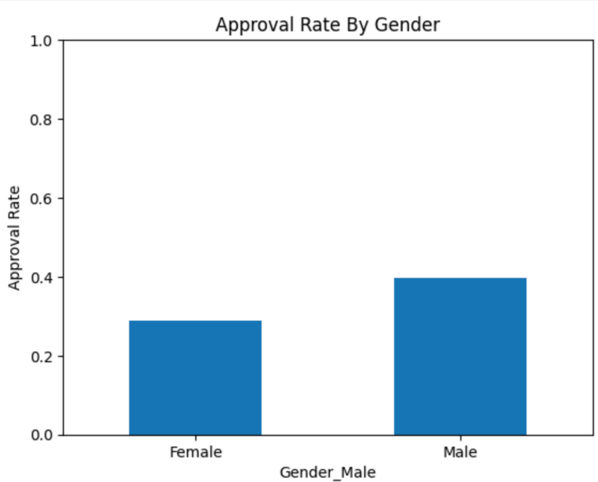
Group	Observed Bias
Gender	Female applicants had lower approval rates than males
Race	White applicants were approved at higher rates than Non-White applicants
Employment Type	Full-time workers were approved more often than gig or self-employed individuals
Criminal Record	Applicants with a criminal record had significantly lower approval chances

Other insights:

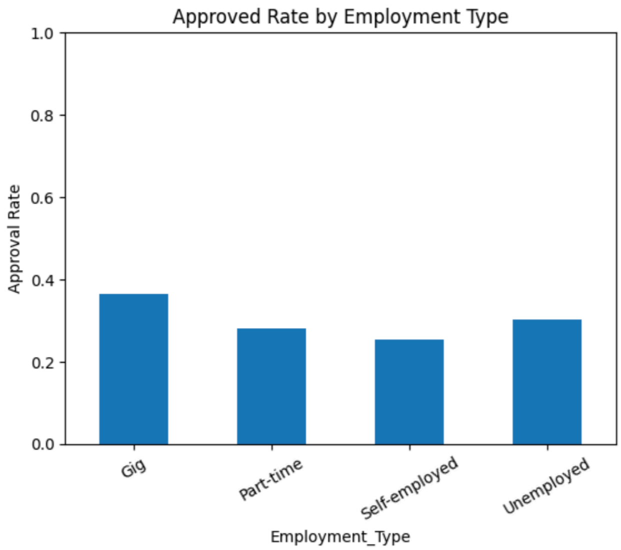
- Citizens were favoured over visa holders or permanent residents
- Older applicants (60+) showed different approval behaviour compared to those aged 25–60

5. Visual Evidence

```
gender_results.plot(kind='bar', title="Approval Rate By Gender", ylabel="Approval Rate")
plt.xticks(ticks=[0, 1], labels=['Female', 'Male'], rotation=0)
plt.ylim(0,1)
plt.show()
```



```python
employment_results.plot(kind='bar', title="Approved Rate by Employment Type", ylabel="Approval Rate", ylim=(0, 1))
plt.xticks(rotation=30)
plt.show()
```



6. Real-World Implications

If this model were deployed without fairness safeguards:

- **Female**, **Non-White**, and **Gig economy workers** could face systemic exclusion from homeownership opportunities.
- Individuals with criminal records, regardless of reformation, would likely be denied access to housing loans.

In a regulated environment, such a model would likely fail a fairness audit and could face legal or ethical repercussions.

7. Limitations & Reflections

- Some fairness analysis tools like SHAP or Fairlearn weren't fully implemented due to time constraints.

What I'd try next time:

- Use adversarial de-biasing or reweighing methods
- Implement fairness constraints during model training
- Conduct intersectional fairness audits at deeper granularity

Lessons Learned:

- Fairness in AI is not automatic — you must test for it
- Bias can exist even if you don’t use sensitive features directly
- Visualization is a powerful tool for exposing unfair treatment