
# Treating Unfairness in Machine Learning

DS Group 01

## Intro & Team

A quick overview of our topic and our team.

### Data

Dataset of credit granting for different applicants:

Gender, Marital Status, Dependents, Self Employed, Education, Applicant & Co Appliant Income, Property Area, Loan Amount, Loan Amount Term, Credit History, **Loan Status** (target)

### Task

Treating gender fairness in a credit granting model.

### Team

Ruslan – Test Engineer

Isabella – Finance Consultant

### Mentor

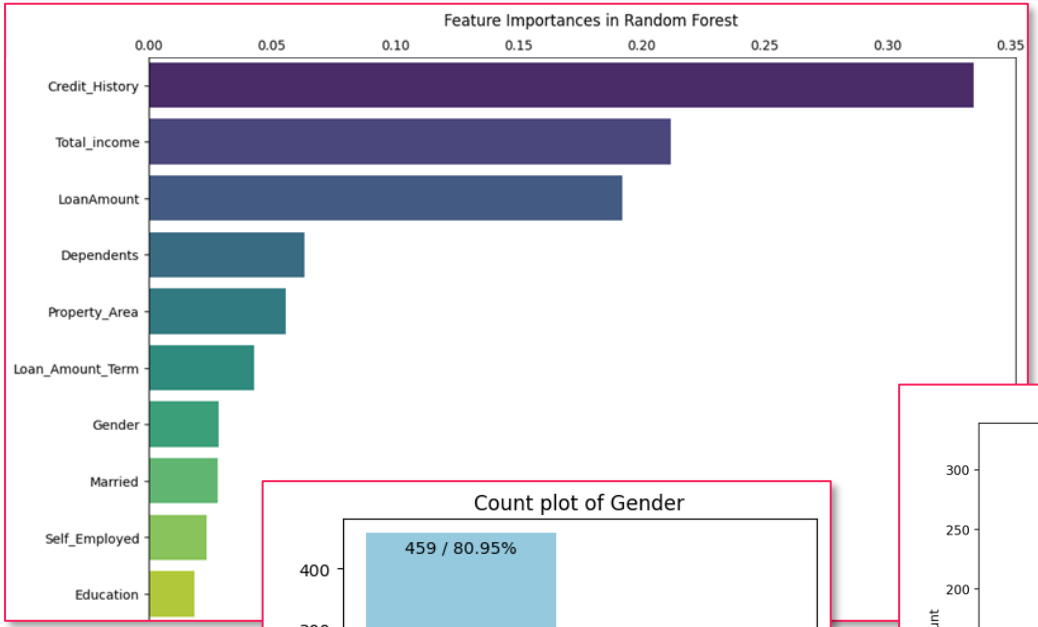
Patrick – Data Scientist

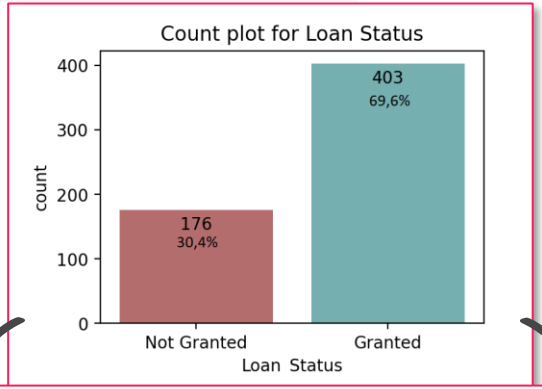
## Goal

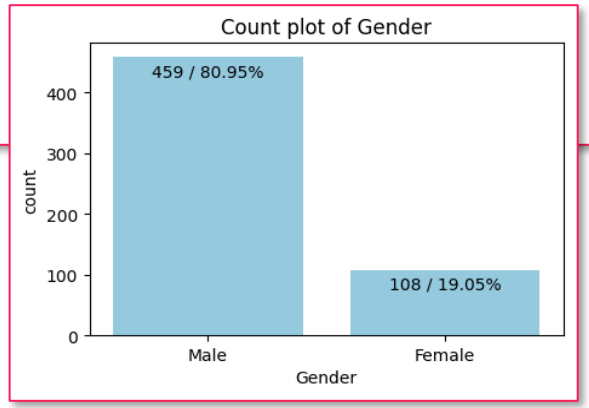
No gender biases in terms of fairness when granting credit.

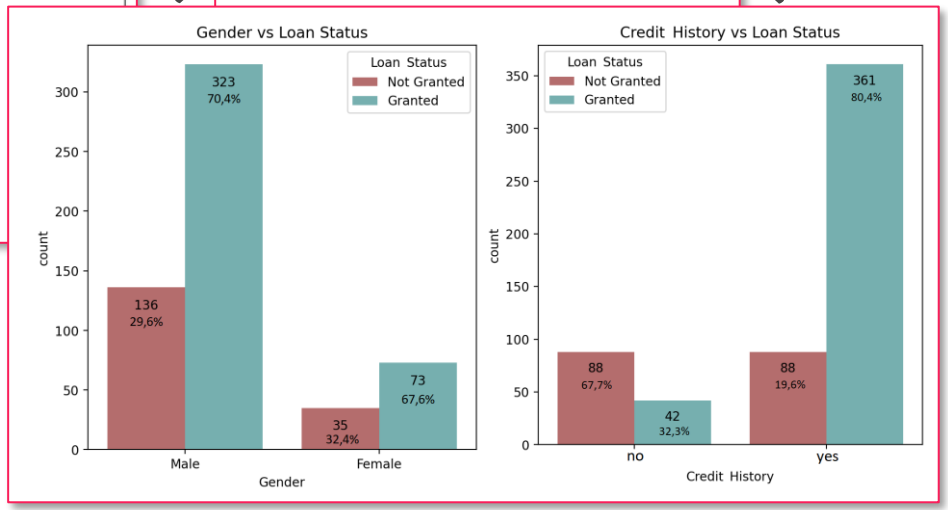
## Data

A deeper look into our dataset.



## Methodology

Our approach to solve the problem.

### Preparation

- Analysis of the dataset and dividing it into training and test data
- Using different methods and comparing the results
- Utilizing **Random Forest** as our Method
- **Selecting Positive Prediction Ratio (PPR)** as fairness measure (proportion of positive predictions for each group)

### Modeling

- Hyperparameter optimization and cross-validation
- Accuracy is not a suitable metric because the data is not balanced
- Training of the model to get the best **F1 Score** (balance of precision and recall)
- Determine the **PPR scores** for men and women

$$F1\ Score = \frac{TP}{TP + 0{,}5(FP + FN)}$$

$$Precision = \frac{TP}{TP + FP}$$

$$PPR = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

### Fairness Analyzation

- Treating the unfairness of the model ex-post by **changing** the models default **threshold**
- Adjusting the threshold can minimize the difference between ppr_male and ppr_female and **improve fairness** in the results

## Conclusion

Our findings and learnings.

- The data is unbalanced, and we assume there is bias in the data, **as 81%** of applicants in the dataset are **male**
- A balanced model in terms of **performance** does **not** mean **fairness** across different groups in the results
- We have the smallest difference in the ppr score for female and male applicants by using a **threshold** of **0.44**
- **Increasing** the **fairness** between gender **affects** the **performance** of the model

| | Optimized Performance | Optimized Fairness |
|---|---|---|
| **Threshold** | 0.5 | 0.44 |
| **F1 Score** | 0.852 | 0.846 |
| **PPR Male** | 0.804 | 0.855 |
| **PPR Female** | 0.767 | 0.867 |
| **Difference in PPR** | **0.037** | **0.012** |

Thank you for your attention.

# Happy to answer your questions!