Big Data in Finance

Tackling Discrimination in Machine Learning Algorithms

Group: DL2
Lupo Benatti
Xavier Gueniot
Mark Kutemeyer
Jakob Merz
Andrea Perl
Rolf Stirnimann

esade
RAMON LLULL UNIVERSITY

Overview

- Problem Definition
- Data Preparation
- Model Creation
- Model Interpretation

A Project in 4 Sprints

ALL | Team PINK | Team ORANGE | Team GREEN

Kick-Off 28.05 | Review 1 04.06 | Review 2 08.06 | Review 3 18.06 | Final Rev. 25.06

	Sprint 1 10.02. - 24.02.	Sprint 2 25.02 - 09.03.	Sprint 3 10.03. - 30.03.	Sprint 4 31.03. - 12.04.
Problem Definition	Datasets/ Resources; Decide on project			
Data Preparation	Data Exploration; Preprocess ing; Data Viz	Write Report; Review Report		
Model Creation		First models in Python; Create Logistic Regression Base Model	DL Models; Review Models & Metrics	Down Sampling; Disparate Impact Remover
Model Interpretation			Presentation: First draft	Presentation: Final version; Review Final Presentation

Discrimination in Algorithms

Historically,

vulnerable groups to discrimination have been treated differently, tearing our society apart causing social injustice & economic inequality

Women & Minorities

Bias and prejudices,

are rooted in our decisions as well the data originated from them. Nevertheless, we keep training our Machine Learning models with this biased data

In an automated world,

we can't afford to let these discriminatory biases spill over into our digital future, especially if we rely on machines to carry out decisions

Objective: fight data-driven inequality & discrimination by eliminating or minimizing the effect of bias on human decisions

1. **Identify barriers to fair datamining**
2. **Detect and minimize wrongful discrimination**
3. **Investigate possible solutions to this rising problem**

Preprocessing

Dataset consists of 15 personal attributes that are usually evaluated to provide loans

We tested whether there are any outliers: the column “payment_timing” has ~8,800 outliers, which we removed from the dataset

The minority feature is clearly unevenly distributed. Almost all defaults are by minority clients

We applied a correlation matrix using seaborn and we found out that some features are highly positively or negatively correlated

Result: The data is potentially biased; the dataset is not imbalanced; defaults for gender are not unevenly distributed whereas for minorities they clearly are

Identification of Biased Variables

We applied 3 different methods to get an initial idea on which features are important for predicting our outcome variable :

1. Univariate chi
2. Recursive elimination
3. Tree based approaches

Takeaways:

- Minority is always the most important feature: not surprising as it is one of the features perfectly correlated with default
- Moreover, the ZIP codes, rent and job stability seem to be important for determining whether a client defaults or not

We believe 4 variables: ZIP, rent, job stability and occupation cause the bias
Does removing these four variables reduce bias?

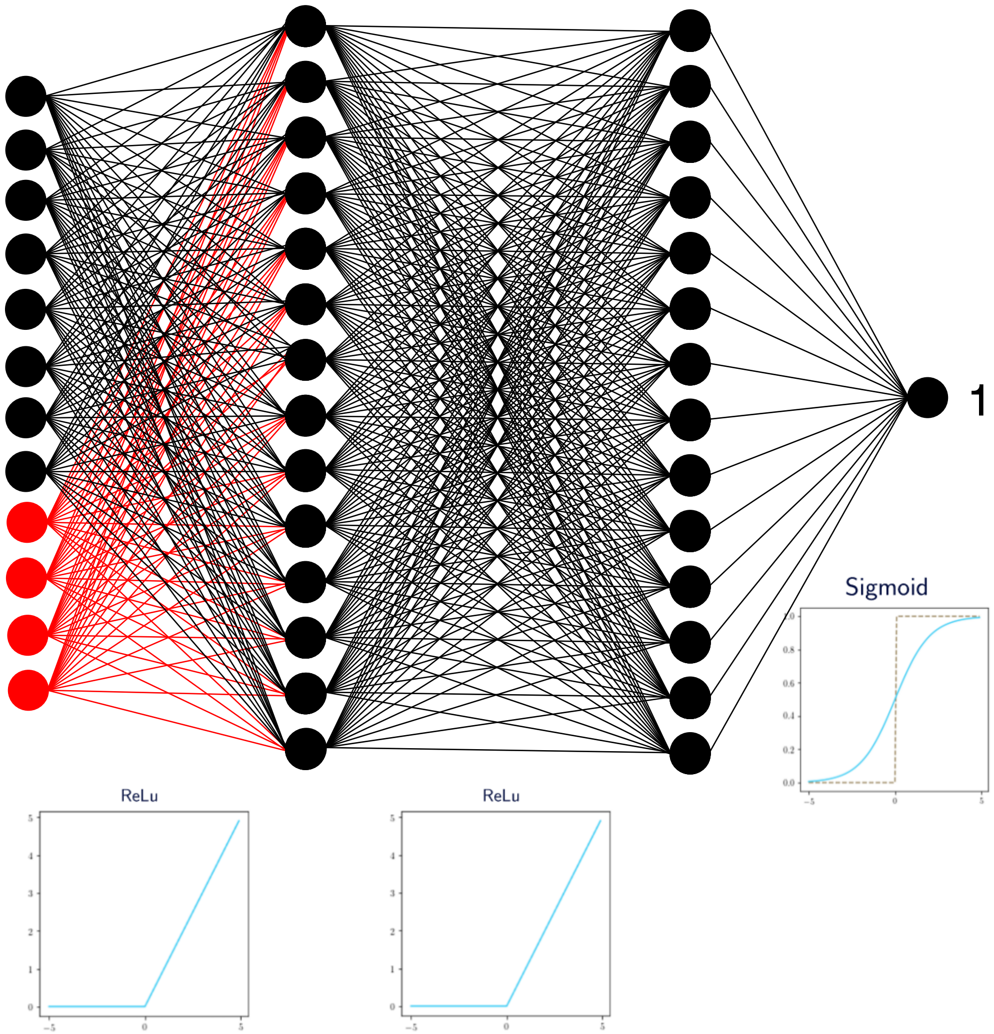
DL Models: Loan Default Prediction

64 neurons

36 neurons

8 Non-Biased Features

4 Biased Features



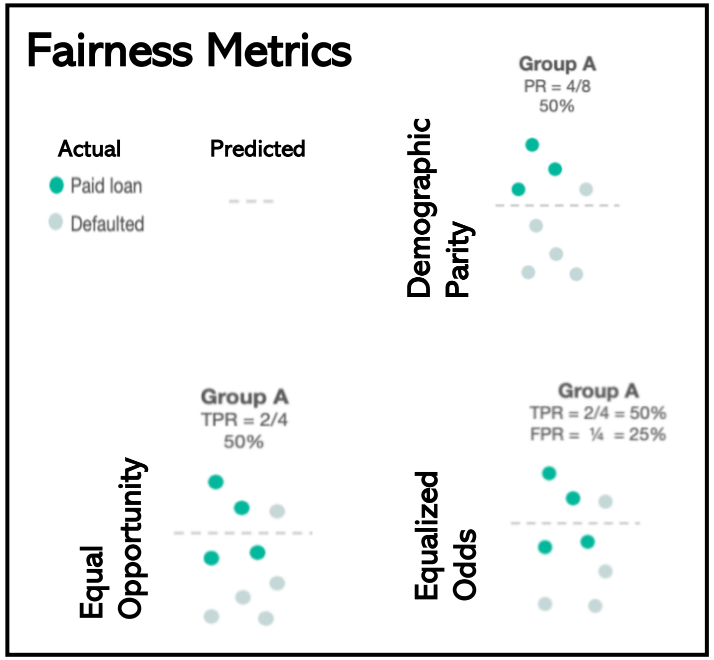
1D output

Model 1:
"Biased Model"
12 features

Model 2:
"Unbiased Model"
8 features

Precision > 85%
Recall > 47%

Fairness Evaluation



	Biased Model		Unbiased Model		Down Sampling	
	Majority	Minority	Majority	Minority	Majority	Minority
Demographic Parity	96%	9%	4%	4%	58%	45%
Equal Opportunity	94%	9%	4%	4%	67%	73%
Equalized Odds	94%	9%	4%	4%	67%	73%
	3%	91%	95%	95%	51%	46%

Alternative Solutions

Disparate Impact Remover is an algorithm that tries to reduce the bias in the dataset by ensuring that the features in the un-privileged case (in our case minority) are similar to the features in the privileged case (non-minority).

Down sampling as a methodology to reduce bias without deleting important features.

Implications for Businesses

Types of Fairness

A — ***Individual fairness:*** *where the idea is to ensure that similar individuals are treated similarly*

B — ***Group fairness:*** *Here, the idea is to ensure that the error rates for different groups of people are similar*

Biased AI in our financial lives

AI systems are biased in ways that may harm consumers and employees

What can businesses do to prevent it?

1. *Maintain Transparency*
2. *Substantiate Assumptions*
3. *Ensure Data Quality and Security*
4. *Perform Regular Bias Assessments*

GitHub Link

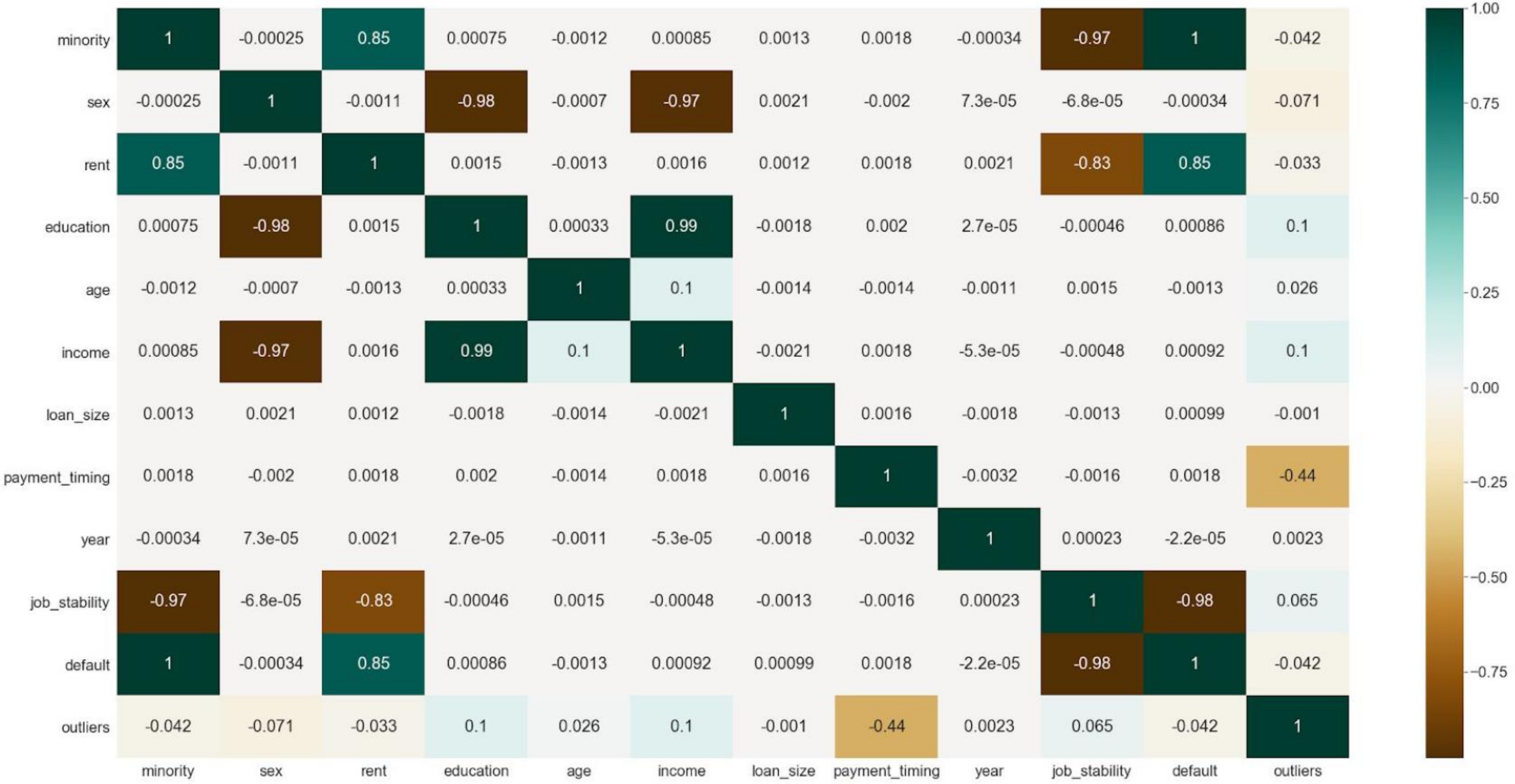
[GitHub with the whole project]

Thank You

Back-Up

esade
RAMON LLULL UNIVERSITY

Seaborn Matrix



Minority Discrimination

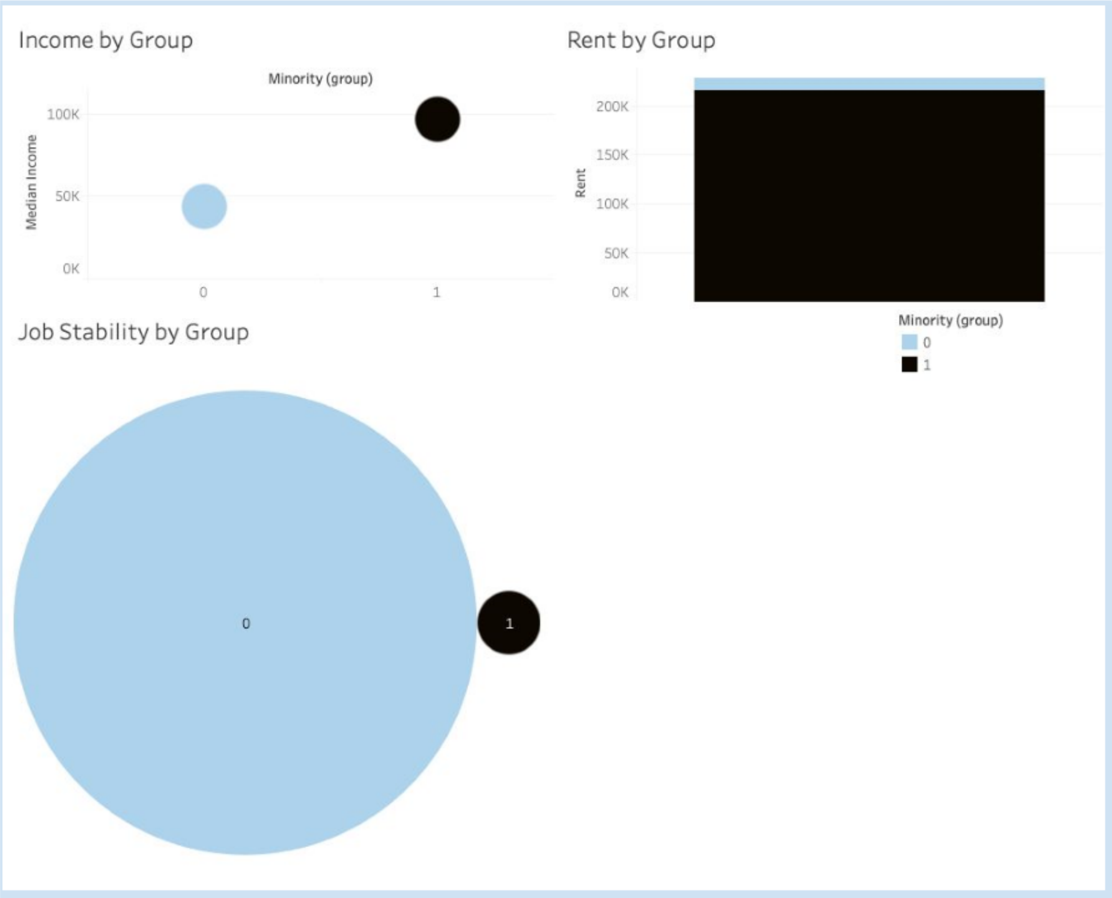


Figure 1: Exploratory Analysis by Group. We notice that minority groups tend to have less job stability than their counterparts.

Minority Discrimination

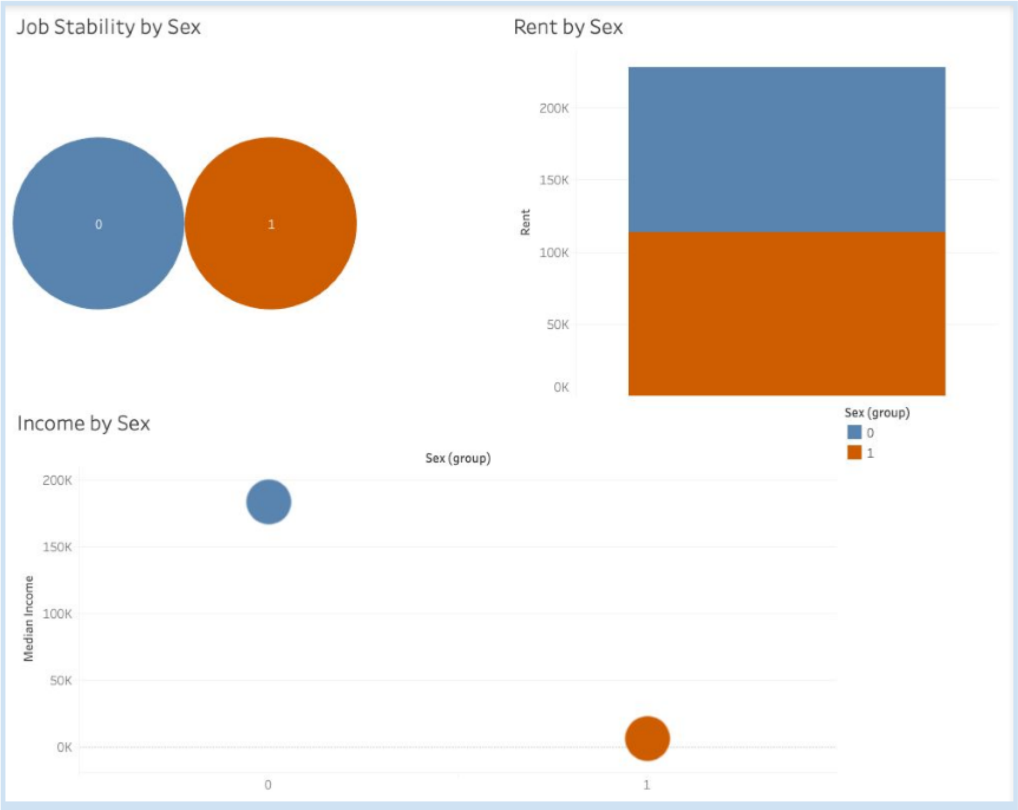


Figure 2: Exploratory Analysis by Gender. We notice that women tend to have much lower income than men.

Result Overview

Model	Process	Accuracy on Test	Default	Precision	Recall	F-1 Score	Minority	Accuracy on test	Demographic Parity	Equal Opportunity	Equalised Odds
		0.37	0	0.82	0.33	0.47					
LR	Bias		1	0.14	0.58	0.22	Not calculated				
		0.41	0	0.84	0.38	0.53					
LR	No Bias		1	0.14	0.59	0.23	Not calculated				
		0.37	0	0.84	0.32	0.46					
LR	Down-S		1	0.14	0.65	0.24	Not calculated				
		0.47	0	0.85	0.47	0.60	0	0.79	0.96	0.93	0.03
DL	Bias		1	0.15	0.52	0.23	1	0.17	0.09	0.09	0.91
		0.85	0	0.85	0.95	0.90	0	0.82	0.05	0.05	0.95
DL	No Bias		1	0.15	0.04	0.07	1	0.82	0.05	0.05	0.95
		0.52	0	0.91	0.49	0.64	0	0.51	0.58	0.67	0.51
DL	Down-S		1	0.20	0.71	0.31	1	0.56	0.45	0.73	0.46

Source: Results from 4-Deep Learning.ipynb notebook

The Journey