Investigating fairness in data-driven allocation of public resources

Eva Achterhold (LMU München)
MZES Social Science Data Lab
26.10.2022

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Agenda

Part I

- Fairness in ML
- Measuring fairness
- Use case

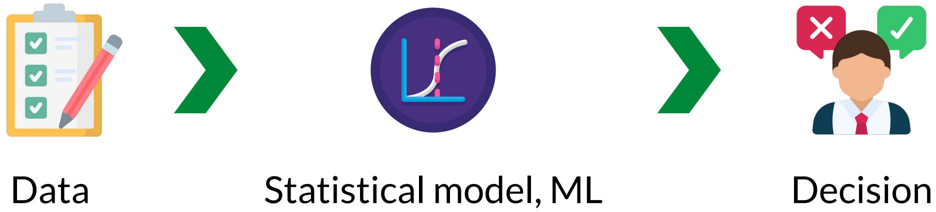
Part II

- Live coding session
- Interpretation

Part I

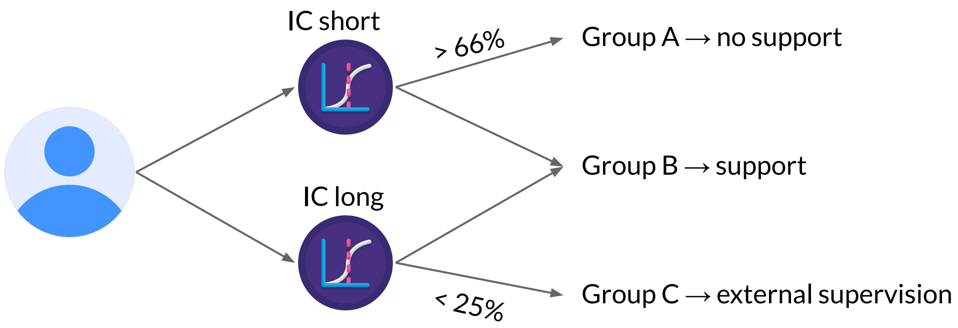
Data-driven decision making

- Using large amounts of data from the past to find patterns under the assumption that things will turn out the same in the future
- Varies in degree of human intervention and complexity
- Is being used for initial risk assessment, e.g. risk of recidivism or LTU



AMAS Arbeitsmarkt-Chancen-Assistenzsystem

- Predicted Integration Chances (IC) of job seekers
- 2 criteria: short-term and long-term re-employment
- IC score was shown to case workers who decide on the allocation of measures



Holl et al. (2018)

AMAS criticism

- Negative coefficients for women and people with non-EU citizenship
- Decision to not support people with high risk for LTU
- No agency for the job seeker to influence score
- Lack of transparency in communication about the implementation of the system
- Studies show that case workers hardly adapted algorithmic suggestions

Allhutter et al. (2020), Niklas et al. (2015)

Socio-technical sources of bias

Data generation	Data preparation and analysis	Implementation
- Sampling bias	- Variable construction	- Decision rules
- Historical discrimination	- Modelling choices	- Human behavior
- Measurement bias	- Error metrics	- Communication
- Data collection		- Feedback loops

Gerdon et al. (2022)

Measuring fairness in ADM

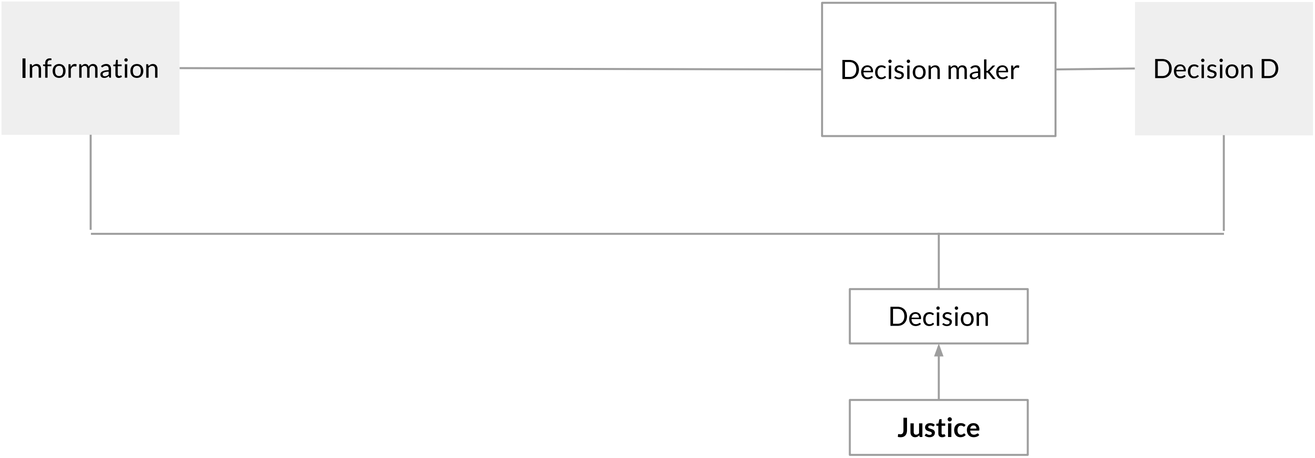
Fairness: Towards whom?

- Group vs individual fairness
  - Comparing averages over social groups?
  - Treating similar individuals similarly?

Fairness: Towards whom?

- Group vs individual fairness
  - Comparing averages over social groups?
  - Treating similar individuals similarly?
- Protected attributes
  - What are protected attributes?
  - Protected attributes and responsibility?

Fairness and Justice

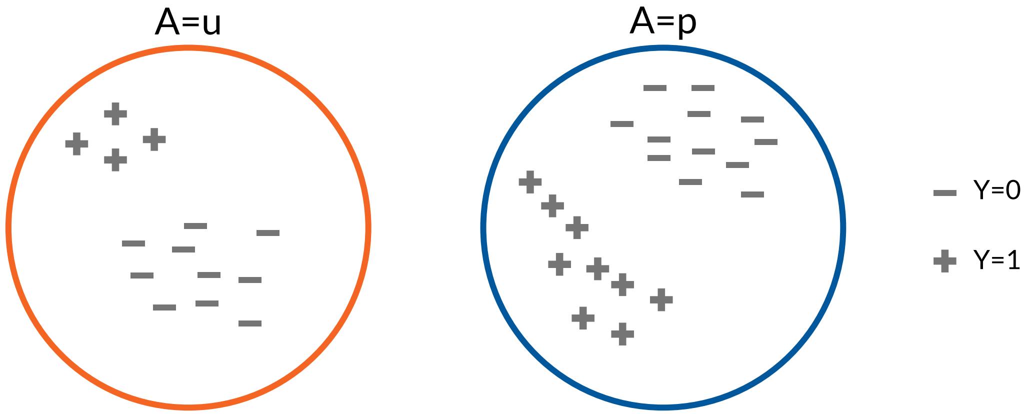


Just decisions

- Which distributive justice principle?
- How to allocate (public) goods in a just manner?
- How to implement the decisions in the system?
- Metrics:
    - (Conditional) Statistical Parity Difference
    - Disparate Impact
    - …

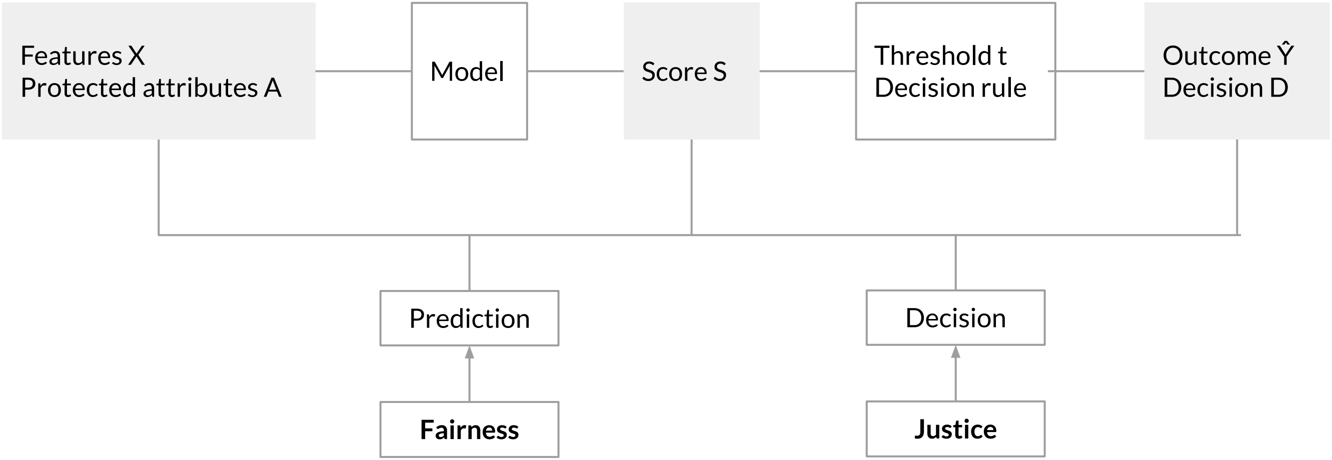
Statistical Parity Difference

$$SPD(z) = P(z=1 \mid A=u) - P(z=1 \mid A=p) \quad z \in \{Y, \hat{Y}\}$$



SPD(Y) = 4⁺/14 - 9⁺/23 = -0.11

Fairness and Justice



Kuppler et al. (2022)

Fair predictions

- Which error metric(s) are minimized?
- Whose errors are minimized?
- Metrics:
    - Performance metrics, e.g. cross-entropy loss, F1 score, accuracy…
    - Fairness metrics, e.g. Equal Opportunity Difference, Average Odds Difference…

Confusion matrix

		Predicted Outcome	
		LTU	Non LTU
Actual	LTU +	True Positive +	False Negative +
	Non LTU −	False Positive −	True Negative −

Confusion matrix

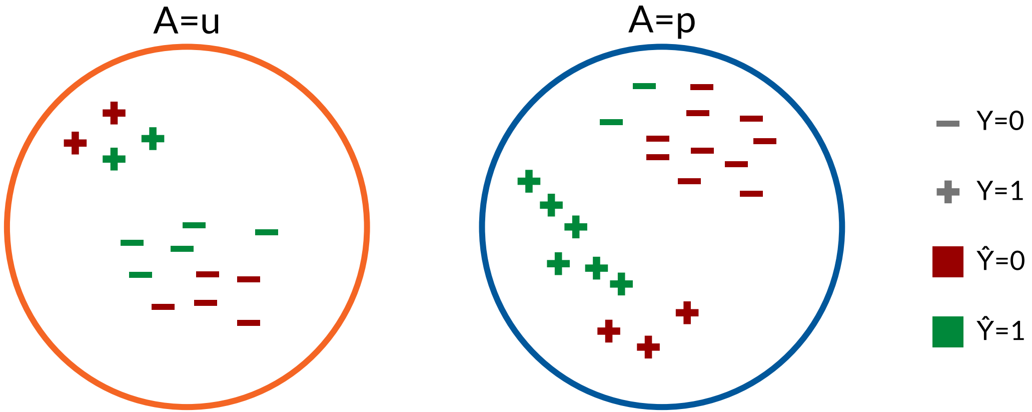
		Predicted Outcome	
		LTU ■	Non LTU ■
Actual	LTU ✚	True Positive ✚	False Negative ✚
	Non LTU ▬	False Positive ▬	True Negative ▬

What proportion of actual positives was identified correctly?

Recall = TPR = TP / (TP + FN) = ✚/ (✚+✚)

Equal Opportunity Difference

$$EOD = P(\hat{Y}=1 \mid A=u, Y=1) - P(\hat{Y}=1 \mid A=p, Y=1)$$



+	2	+	2
−	5	−	5

+	6	+	3
−	2	−	10

EOD = $(2^{+}/4^{+})-(6^{+}/9^{+})$ = -0.17

From theory to practice

The data at hand

- Data of 5000 job seekers
- Data collection was done by a third party
- Assume no sampling bias
- Possible fairness considerations:
    - Historical discrimination
    - Measurement bias
    - Data collection

Notation

Features:

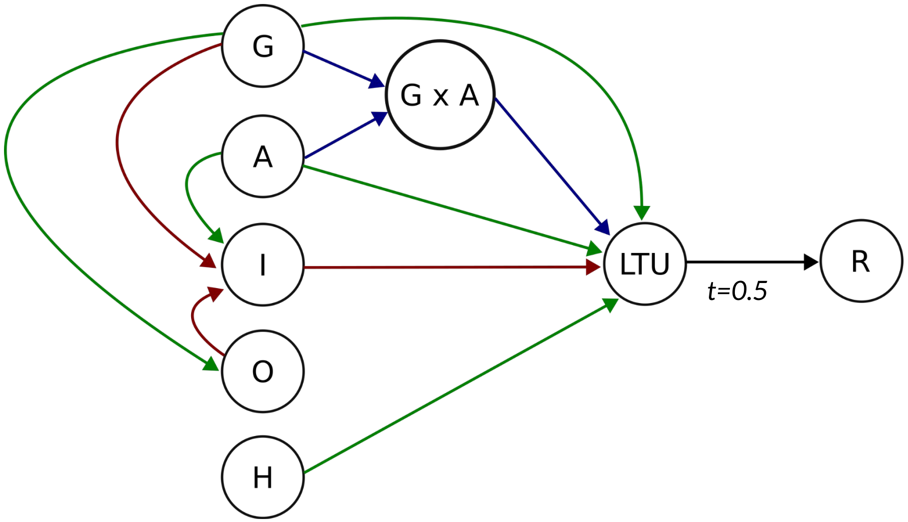
- Gender G ∈ {0, 1}; 1=female
- Age: A ∈ [16, 67]
- Last income: I ∈ R+
- Occupation sector: O ∈ {0, 1}; 0=Production, 1=Service
- Unemployment history: H ∈ {0, 1}; 1=was unemployed

Targets:

- Long term unemployment: LTU ∈ {0, 1}
- Receives ressources: R ∈ {0, 1}

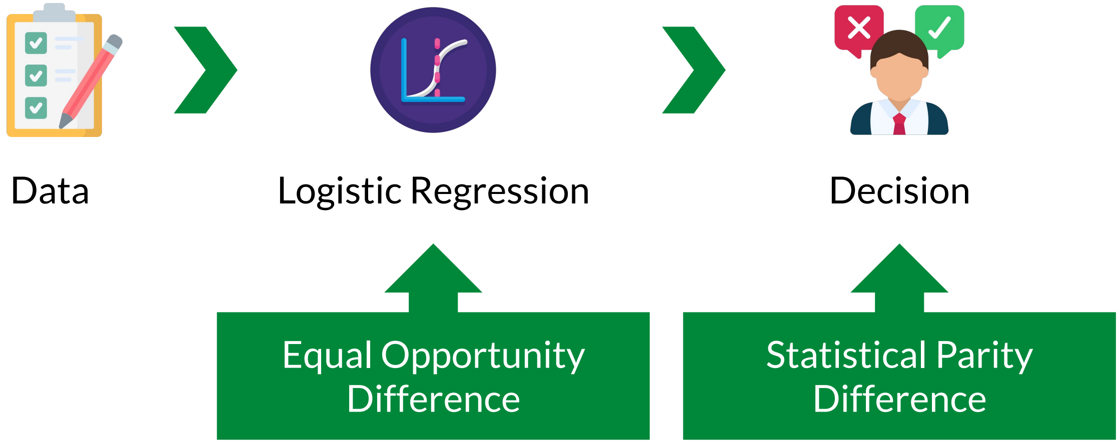
Relations in the data

→ Positive relation
→ Negative relation
→ Interaction



G	Gender
A	Age
I	Last Income
O	Occupation Sector
H	Unemployment History
LTU	Long Term Unemployment
R	Receives Resources

Coding part



We use IBM's AIF360 toolkit in Python, which is also available in R.

Fairness packages in R

- aif360: https://github.com/Trusted-AI/AIF360/tree/master/aif360/aif360-r
- mlr3fairness: https://github.com/mlr-org/mlr3fairness/
- fairmodels: https://github.com/ModelOriented/FairModels
- fairness: https://github.com/kozodoi/fairness
- fairlearn: https://github.com/fairlearn/fairlearn
- ...

Part II

Please open the following link:

https://tinyurl.com/fair-ml-ssdl

Questions:

1. What do the fairness metrics tell us about the predicted LTU risk for each social group?

2. Does the ADM process discriminate against women?

Fairness metrics

Statistical Parity Difference

- Aim: $\hat{Y}$ is independent of attribute G
- Positive value → women are more likely to be labeled LTU → higher chances for support

Equal Opportunity Difference

- Aim: The TPR are equal for social groups
- Around 0 → the model can equally well detect actual positives for men and for women

Does the process discriminate?

SPD: Women are more likely to be labeled LTU ❯ They have a higher chance of receiving resources

EOD: Almost equal TPR across social groups ❯ Given someone needs resources, gender is not a factor for actually receiving them

What next?

- Consider long-term effects
  - Feedback loops: Supporting women can result in change of data generating process → achieving SPD in the long run?
- Include model uncertainty
  - Individuals with high uncertainty could be forwarded to case workers
- Involve interdisciplinary research
  - We can draw on insights from social sciences, philosophy, law…
- Engage in public debates
  - Involving those who are affected in the discourse

References

Allhutter, D., Cech, F., Fischer, F., Grill, G., and Mager, A. (2020). Algorithmic profiling of job seekers in Austria: How austerity politics are made effective. *Frontiers in Big Data*, 3.

Gerdon, F., Bach, R. L., Kern, C., & Kreuter, F. (2022). Social impacts of algorithmic decision-making: A research agenda for the social sciences. *Big Data & Society, 9*(1).

Holl, J., Kernbeiß, G., Wagner-Pinter, M. (2018): Das AMS-Arbeitsmarktchancen-Modell. Dokumentation zur Methode. URL: https://www.ams-forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf (Zugriff am 14.09.2022)

Kuppler, M., Kern, C., Bach, R., & Kreuter, F. (2022). From fair predictions to just decisions? Conceptualizing algorithmic fairness and distributive justice in the context of automated decision-making. *Frontiers in Sociology*, 155.

Niklas, J., Sztandar-Sztanderskal, K., and Szymielewicz, K. (2015). Profiling the unemployed in Poland: Social and political implications of algorithmic decision making. Technical report, Fundacja Panoptykon.

Icons are taken from https://www.flaticon.com/