German Credit Datasets, A Biased Mess?

Bennett, Cody, Ian

GERMAN CREDIT DATA

Agenda

- What Is German Credit Data? [01]
- Data Bias [02]
- Reading the Data [03]
- Interpreting the Data [04]
- Our Findings [05]
- Wrap Up [06]

status	duration	credit_history	purpose	amount	savings	employment_duration	installment_rate	personal_status_sex	other_debtors
... < 100 DM	6	critical account/other credits existing	domestic appliances	1169	unknown/no savings account	... >= 7 years	4	male : single	none
0 <= ... < 200 DM	48	existing credits paid back duly till now	domestic appliances	5951	... < 100 DM	1 <= ... < 4 years	2	female : divorced/separated/married	none
no checking account	12	critical account/other credits existing	retraining	2096	... < 100 DM	4 <= ... < 7 years	2	male : single	none
... < 100 DM	42	existing credits paid back duly till now	radio/television	7882	... < 100 DM	4 <= ... < 7 years	2	male : single	guarantor
... < 100 DM	24	delay in paying off in the past	car (new)	4870	... < 100 DM	1 <= ... < 4 years	3	male : single	none
no checking account	36	existing credits paid back duly till now	retraining	9055	unknown/no savings account	1 <= ... < 4 years	2	male : single	none
no checking account	24	existing credits paid back duly till now	radio/television	2835	500 <= ... < 1000 DM	... >= 7 years	3	male : single	none
0 <= ... < 200 DM	36	existing credits paid back duly till now	car (used)	6948	... < 100 DM	1 <= ... < 4 years	2	male : single	none

present_residence	property	age	other_installment_plans	housing	number_credits	job	people_liable	telephone	foreign_worker	credit_risk
4	real estate	67	none	own	2	skilled employee/official	1	yes	yes	1
2	real estate	22	none	own	1	skilled employee/official	1	no	yes	0
3	real estate	49	none	own	1	unskilled - resident	2	no	yes	1
4	building society savings agreement/life insurance	45	none	for free	1	skilled employee/official	2	no	yes	1
4	unknown/no property	53	none	for free	2	skilled employee/official	2	no	yes	0
4	unknown/no property	35	none	for free	1	unskilled - resident	2	yes	yes	1
4	building society savings agreement/life insurance	53	none	own	1	skilled employee/official	1	no	yes	1
2	car or other	35	none	rent	1	management/self-employed/highly qualified employee/officer	1	yes	yes	1

What is German Credit Data?

In general terms, German Credit Datasets are those containing information on German bank customers' credit risk.

Each row represents an individual's financial status and history, with the factors contributing being make or break for some looking for a loan.

It begs the question, are the factors used bias towards certain groups? And if they are, can AI be trusted to make equitable decisions based on the data?

Data Bias:

Data Bias is when data or information is limited in a certain way, this then paints a false representation of the population, or does not tell us the whole story.

Regarding AI usage, biased data can sway decision making in favor of certain groups.

We grouped some of the factors we found, some of which we could see being problematic.

1. Gender, Age, Marital Status

2. Employment Duration, Job Type, Foreign Worker Status

3. Housing, Savings, Number of Credits, Personal Possessions

4. Credit History, Purpose, Ownership

Reading the Data

We incorporated a few different methods for reading the data, stemming from the ideals of Lindsay Poirier.

Connotative Reading: Based on the notion of looking how the data's meaning could have shifted over time, possibly making what was once an unbiased factor into one that marginalizes certain groups.

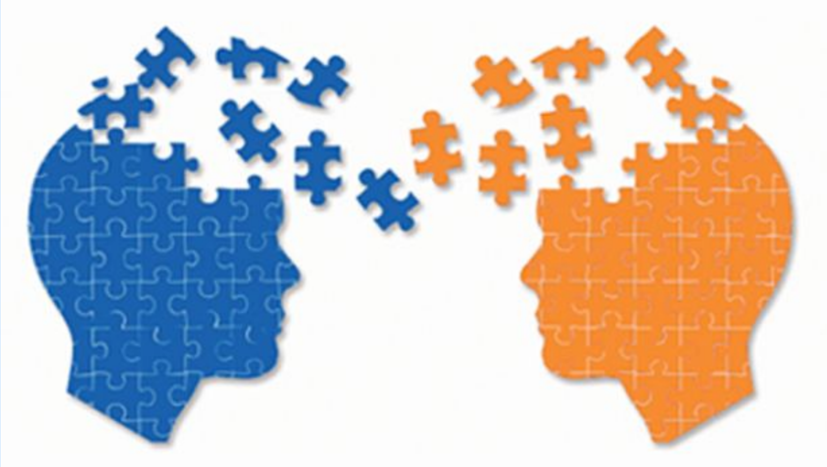
Deconstructive Reading: Seeing what was left out of incorporated factors that could lead to bias. Essentially asking the question; Who is being excluded? What factors are being overlooked or counted out that shouldn't be?

Interpreting the Data

While reviewing the data, we related it back to Professor Koopmans writing about interpretive formats.

Social Construct of Data: Data reflects and continues to reinforce power dynamics within systematic biases. This also brings up the idea of over surveillance and the panopticon.

Institutional Incentives: Data can be manipulated to benefit the existing data already to favor outcomes. This bias continues inequality and outcasts marginalized groups.



Our Findings

Based on our reading of the data, as well as considering interpretive format anatomies of context, we have determined a few things about the data:

- The inputs into the data are out of date and thus bias towards certain groups, specifically, single females, those within LGBTQIA+ communities, and foreign workers. The dataset lacks societal context and cannot properly show the hardships that are associated within these communities.
- Data overtime has and continues to reflect social power dynamics. This context is lost within the dataset, which can lead to further entrenchment within said power dynamics. Thus, factors like personal possessions, housing, and savings, mixed with those such as foreign worker status, can be biased towards certain people, as they do not consider the opportunities that are necessary to build upon these inputs.

Wrap Up

Based on our findings, we have come to the conclusion that there is too much data bias within German credit datasets, or at the very least the one we analysed, for AI to be able to make equitable decisions when determining credit risk. It would marginalize certain communities, as the data lacks the necessary context, and thus should not be a part of these crucial decisions.

Thank you