ENTREPRENEURIAL FINANCING AND MACHINE LEARNING

CAN WE MITIGATE GENDER BIAS?

Marlene Dahle
June 10th, 2022

Research Question

Which gender bias mitigation strategies are most effective for machine learning models created to assist in early-stage investment decision-making?

THE THESIS IN SHORT

- Theoretical background
  - Machine learning in early-stage investment decision-making
  - Gender bias in early-stage investment decision-making
  - Gender bias in machine learning models
- Methods:
  - Industry interviews
  - Creating a machine learning model
  - Experiments with gender bias mitigation

THEORETICAL BACKGROUND - FINDINGS

MACHINE LEARNING IN EARLY-STAGE INVESTMENTS

- Early stages but increasing with big potential
- Investors interested in using machine learning
- Much of the literature use Crunchbase data
- None address gender bias or fairness

THEORETICAL BACKGROUND - FINDINGS

GENDER BIAS IN EARLY-STAGE INVESTMENT DECISION MAKING

- The low funding rate of female founded ventures on its own is not proof of gender bias
- The literature presented shows gender bias in many stages of the decision-making process
- Well-researched field but still no agreement on why the funding is lower with female founders, or how to mitigate even in real life decision-making processes

THEORETICAL BACKGROUND - FINDINGS

GENDER BIAS IN MACHINE LEARNING MODELS

- Machine learning models use real life data, prone to human bias
- Focus on gender bias in ML models is a new field with increasing focus

THEORETICAL BACKGROUND - ASSUMPTIONS

MACHINE LEARNING MODELS ARE INCREASINGLY USED IN EARLY-STAGE INVESTMENT DECISION-MAKING

THERE IS GENDER BIAS FOUND IN THE INVESTMENT DECISION-MAKING PROCESS

MACHINE LEARNING MODELS EXHIBIT AND AMPLIFY GENDER BIAS FROM DATA

INDUSTRY INTERVIEWS

HOW ARE INVESTMENT DECISIONS MADE?

- Much focus on the founders themselves
- Gut feel and likability is an important part for investors in evaluating early-stage startups
- Not just facts and numbers
- Some important ‘features’: work experience, founding experience, team size, personal traits

INDUSTRY INTERVIEWS

ROLE OF GENDER IN ENTREPRENEURIAL FINANCING

- Increasing focus on attracting and investing in more female founders
- No agreement on why there are few female founders - not even agreement that there are few female founders
- Male founder stereotype

INDUSTRY INTERVIEWS

USE OF MACHINE LEARNING IN THE DECISION-MAKING PROCESS

- All were interested in how machine learning can assist in the decision-making process
- Only one investor firm were actively creating their own models
- Techniques and tools such as checklists were sometimes used

CREATING A MACHINE LEARNING MODEL

- Crunchbase data, 18.467 companies with gender information
- Target variable: ‘acquired’ or ‘ipo’ as successful, ’closed’ as non-successful
- Logistic Regression as the classifier
- No features with ‘look-ahead bias’

CREATING A MACHINE LEARNING MODEL

Feature number	Feature name	Short description
1	city_success_ranking	The ranking of the city in terms of number of successes divided with total number of companies
2	region_success_rank	The ranking of the region in terms of number of successes divided with total number of companies
3	multiple_degrees_sum	The sum of founders with multiple degrees per company
4	multiple_degrees_average	The sum of founders with multiple degrees per company
5	Is_completed_sum	The average of founders with completed degrees per company
6	Is_completed_avg	The average of founders with completed degrees per company
7	work_experience_avg	The average number of days of work experience of the founders
8	work_experience_sum	The total number of days of work experience of the founders
9	education_time_avg	The average education time of the founders
10	education_time_sum	The sum of education time of the founders in number of days
11	male_founders	Number of male founders per company
12	female_founders	Number of female founders per company
13	unknown_founders	Number of founders of unknown gender
14	total_num_founders	Total number of founders per company
15	category_list1	First category word describing company
16	category_list2	Second category word describing company
17	category_list3	Third category word describing company
18	category_group_list1	First category word describing company group/sector
19	category_group_list2	Second category word describing company group/sector
20	category_group_list3	Third category word describing company group/sector
21	mostly_male_founders	True if there are strictly more male founders. Used as protected attribute, not in model.
Target	status	Target variable. The status of the company, 0 if it is closed, 1 if acquired/ipo

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 1: FAIRNESS THROUGH UNAWARENESS NOT ENOUGH

- No difference in accuracy (77%)
- No difference in balanced accuracy (70%)
- No difference in any fairness metrics
- Difference in SPD DI from prior data and model

	Dataset metrics	
	Train	Test
SPD	-0.07	0.08
DI	0.90	0.90

	Accuracy	Balanced accuracy	AOD	DI	SPD	EOD
With gender attributes	0.77	0.70	-0.11	0.80	-0.15	-0.13
Without gender attributes (baseline)	0.77	0.70	-0.11	0.81	-0.15	-0.13

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 2: REWEIGHING IS AN EFFECTIVE TOOL

- No effect of reweighing on accuracy or balanced accuracy (77% / 70%)
- Improvement in all fairness metrics

	Accuracy	Balanced accuracy	AOD	DI	SPD	EOD
Baseline	0.77	0.70	-0.11	0.81	-0.15	-0.13
Reweighing	0.77	0.70	0.01	0.94	-0.05	-0.04

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 3: OVER- AND UNDER-SAMPLING WILL AFFECT FAIRNESS METRICS

- If gender is known, one possibly likely scenario is that intuitively one tries to balance the gender representation to make it more equal
- If one does not at the same time reviews fairness metrics and use mitigation tools such as reweighing, model becomes more unfair

	Number of companies with mostly male founders	Number of companies with mostly female founders
Original	16.519	1.948
Oversampling	16.519	16.519
Undersampling	1.948	1.948
Combination	9.910	4.955

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 3: OVER- AND UNDER-SAMPLING WILL AFFECT FAIRNESS METRICS

	Accuracy	Balanced accuracy	AOD	DI	SPD	EOD
Baseline	0.77	0.70	-0.11	0.81	-0.15	-0.13
SMOTE Oversampling						
No reweighing	0.72	0.71	-0.37	0.44	-0.46	-0.41
Reweighing	0.72	0.69	0.02	0.89	-0.07	-0.12
Random Oversampling						
No reweighing	0.74	0.71	-0.14	0.77	-0.17	-0.14
Reweighing	0.74	0.71	0.02	0.98	-0.01	0.01
Random Undersampling						
No reweighing	0.69	0.69	-0.07	0.83	-0.10	-0.08
Reweighing	0.67	0.68	0.02	0.95	-0.03	-0.02
Random Combination Sampling						
No reweighing	0.74	0.72	-0.1	0.78	-0.15	-0.12
Reweighing	0.73	0.72	0.04	0.99	-0.01	0.02

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 4: COMBINATION SAMPLING OF TARGET VARIABLE

- If gender variables are not known, or fairness/gender bias is not a focus, a possible/relevant strategy could be to sample the target variables to create better balanced accuracy.
- Without reweighing, this will significantly affect fairness by all fairness metrics
- It improves balanced accuracy (70% / 74%)
- With reweighing the balanced accuracy and accuracy are slightly worse by 1 percentage point, but fairness improves significantly

RESULTS FROM MITIGATION EXPERIMENTS

EXPERIMENT 4: COMBINATION SAMPLING OF TARGET VARIABLE

	Accuracy	Balanced accuracy	AOD	DI	SPD	EOD
Baseline	0.77	0.70	-0.11	0.81	-0.15	-0.13
Combination sampling of target variable						
No reweighing	0.74	0.74	-0.26	0.4	-0.32	-0.32
Reweighing	0.73	0.73	0.01	0.87	-0.07	0.1

RESULTS FROM MITIGATION EXPERIMENTS

ALL EXPERIMENTS

	Accuracy	Balanced	AOD	DI	SPD	EOD
Experiment 1						
Without gender attributes (baseline)	**0.77**	0.70	-0.11	0.80	-0.15	-0.13
With gender attributes	**0.77**	0.70	-0.11	0.81	-0.15	-0.13
Experiment 2						
Reweighing	**0.77**	0.70	**0.01**	0.94	-0.05	-0.04
Experiment 3						
SMOTE Oversampling						
No reweighing	0.72	0.71	-0.37	0.44	-0.46	-0.41
Reweighing	0.72	0.69	0.02	0.89	-0.07	-0.12
Random Oversampling						
No reweighing	0.74	0.71	-0.14	0.77	-0.17	-0.14
Reweighing	0.74	0.71	0.02	0.98	-0.01	**0.01**
Random Undersampling						
No reweighing	0.69	0.69	-0.07	0.83	-0.10	-0.08
Reweighing	0.67	0.68	0.02	0.95	-0.03	-0.02
Random Combination Sampling						
No reweighing	0.74	0.72	-0.10	0.78	-0.15	-0.12
Reweighing	0.73	0.72	0.04	**0.99**	**-0.01**	0.02
Experiment 4						
Combination sampling of target variable						
No reweighing	0.74	**0.74**	-0.26	0.40	-0.32	-0.32
Reweighing	0.73	0.73	**0.01**	0.87	-0.07	0.10

DISCUSSION

LIMITATIONS OF DATA

- Real life ‘features’ (investment criteria) vs Crunchbase features
- Choice of target variable
- How to capture those who would be successful if it was not for gender bias in investments
    - Funding is important to succeed - data cannot capture those who would have succeeded if getting an investment

DISCUSSION

PROTECTED ATTRIBUTES

- Fairness through unawareness is not enough
  - How do we work with gender bias mitigation techniques without knowing the founder gender?
- Not known whether only including the companies where founder gender is known is represents overall data

DISCUSSION

FAIRNESS METRICS

- No current agreement on which fairness metrics to use in this context (or in general)
- Disparate Impact and Statistical Parity::
  - Addresses unintentional bias in a selection process
  - Forces to accept the same number of males and females
- Equal Opportunity Difference:
  - True positives for both genders should be equal
  - Addresses the qualification problem
- Average Odds: includes both false positive rates and true positive rates for both groups
- Based on historical success
- Open question: how do we measure fairness in this context?

DISCUSSION

APPLICATIONS AND RELEVANCE

- Unrealistic data features might make it harder to use models on their own because of interpretability
- Need to incorporate more of an understanding between the models and human interaction
- Investors rely on a ‘gut feel’ - unlikely to change a lot with a model that is not interpretable or use similar features
  - Especially when considering how even checklists and similar tools are often put aside in the process
- Transparency in models combined with lack of fairness metrics and tools might pose a large problem in terms of gender bias in future and current machine learning models
- A possible lack of understanding and agreement on gender bias in the field

FUTURE WORK

- Creating a framework for working with ML models in this context for industry use
- More realistic datasets mimicking investor decision-making
  - Different data modalities, NLP on evaluations
- More models, more tools/techniques, different target variables
- Fairness metrics use, agreement on goals
- How to create interpretable models that are more suited to investors' work process
- Gender bias mitigation when gender variable is unknown