project canvas

Project name: Fairness in classification

Project owner:

Purpose
What is the intent of this project?
Why are we doing this project?

This project will investigate what it means to have bias in data and how it may affect the algorithms we use to help with decision making, likewise to distinguish between good and bad bias. Additionally, we want to implement a classifier to beable to understand what exactly is going wrong when we use an algorithm to assess and generate risk scores, as well as validating. Lastly we want to have an in depth discussion and ethical considerations of using such technology and what it inherently means that we as a society sentence people to years in prison

Scope
What does this project contain?
What does this project not contain?

Project contains:
- Description of data
- Statistical analysis of data
- Implementation of a classification algorithm on the COMPAS dataset
- Manipulation of attribute values to document bias
- Possible bias adjustments of the classifier
- Ethical discussion on the use of COMPAS

Success Criteria
What do we need to achieve in order for the project to be successful?
How can the Success Criteria be measured?

Milestones
When will we start the project and when is the final deadline ?
What are the key milestones and when will they occur?
How can the milestones be measured?

Project start: 26th of february
Final deadline: Midle of june

Milestones
06.03: Found and read a decent amount of relevant literature.
11.03: Understand the data (feature, structure etc.).
18.03: Hand-in method, introduction, data description.
06.04: Started finding possible bias in own data using statistics.
13.04: Started implementing adjustments in model to remove bias.
06.06: Implementation work mostly done.
18.06: Discussion, method and analysis done.

Actions
Which activities need to be executed in order to reach a certain milestone?

Find articles: Search on google, Findit, google scolar
Decide which statistics to compute
Write python code in order to:
- Import data
- Write a neural network
- Statistics analysis: Confusion matrix etc.

1. To train a classifer on the COMPAS data, and complete an investigation on whether or not the data contains bias, and how the bias influences our classifier.
2. Adjust classifier to reduce bias.
3. A discussion which contains multiple opinions and view points.

Outcome
What is the end result?

- A book
- A website
- An event

Repport
Presentation
Pitch
Obtain knowledge

Team
Who are the team members?
What are their roles in the project?

Sunniva Olsrud Punsvik: Invastigator
Rasmus Stokholm Bryld: Investigator
Matilde Maria de Place: Investigator

Stakeholders
Who has an interest in the success of the project?
In what way are they involved in the project?

Aasa Feragen
Sune Hannibal Holm
Melanie Ganz
Morten Mørup

Users
Who will benefit from the outcome of the project?

Society
Data analysts
Aasa Feragen
Sune Hannibal Holm
Melanie Ganz

Resources
What resources do we need in the project?
- Physical (office, building, server)
- Financial (money)
- Human (time, knowledge)

DTU's supercomputer
Aasa Feragen (superviser)
Morten Mørup
Overleaf
Python
13 weeks + 3 weeks

Constraints
What are the known limitations of the project?
- Physical (office, building, server)
- Financial (money)
- Human (time, knowledge, politics)

Time constraint: We only have limited time, and we have other projects.
Knowledge of some of the needed libraries in python
Programming and philosophy experience

Risks
Which risks may occur during the project?
How do we treat these risks?

Subjective discussion
What bias in data means, could be different from person to person
Some might mean, that the bias we remove, is not bias.