



UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

Bias in Machine Learning

Group 6

www.units.it

October 23, 2023

Members: Castagnotto Alessandro
Coceani Elisa
Majer William
Mingrone Tommaso
Secci Marco



TOC

1. Introduction
2. Causes
3. Strategies to mitigate
4. Conclusion



Introduction

Introduction

Causes

Strategies to mitigate

Conclusion

Many researchers have pointed to the possibility that machine learning algorithms will produce classifiers that display racial, gender, or other forms of bias.

- How does this bias arise?
- Is it possible to constrain machine learning algorithms to produce rigorously fair predictions?



Causes of Bias in ML

Introduction

Causes

Strategies to
mitigate

Conclusion

- Bad training data
- Data selection
- Discriminating features



Bad training data

Introduction

Causes

Strategies to
mitigate

Conclusion

If the dataset used to train a machine learning algorithm contains racial, gender, or other biases, the algorithm can learn and reflect those biases. For example, if historical data contains gender or race discrimination in the past, the algorithm can learn to reproduce such discrimination.



Data selection

Introduction

Causes

Strategies to
mitigate

Conclusion

The selection of training data can influence the presence of bias. If the data used to train a model is collected in ways that reflect existing biases or is incomplete, the model may reflect those biases.



Discriminating features

Introduction

Causes

Strategies to
mitigate

Conclusion

The features selected for model training can lead to discrimination. For example, if you use characteristics that are related to racial or gender factors, the model can learn to make predictions based on those characteristics, even if it is not appropriate.



Strategies to mitigate bias

Introduction

Causes

**Strategies to
mitigate**

Conclusion

- Data Cleaning
- Balanced representation
- Fairness-aware learning
- Continuous auditing and evaluation
- Regulation and legislation



Data Cleaning

Introduction

Causes

**Strategies to
mitigate**

Conclusion

It is important to carefully examine your training data to identify and remove existing biases. This may require removing problematic data or reducing the importance of certain features.



Balanced representation

Introduction

Causes

**Strategies to
mitigate**

Conclusion

Make sure that the training dataset equally represents the different categories in play, so as to avoid overfitting or underfitting of some categories.



Fairness-aware learning

Introduction

Causes

**Strategies to
mitigate**

Conclusion

Use fairness-aware machine learning approaches that explicitly seek to reduce bias in models and produce fairer and more balanced predictions.



Introduction

Strategies to mitigate

Continuously monitor machine learning models in production to identify and correct any emerging biases. This may include the use of fairness metrics to evaluate model performance.



Regulation and legislation

Introduction

Causes

**Strategies to
mitigate**

Conclusion

Authorities and organizations can introduce regulations and guidelines to ensure justice and fairness in the use of machine learning algorithms.



Conclusion

Introduction

Causes

Strategies to
mitigate

Conclusion

While it is not possible to completely eliminate all biases, it is possible to reduce them and manage them to produce more rigorously fair and balanced forecasts. Awareness and commitment to addressing this issue are key to developing and using AI responsibly.

Thank You.

Group 6

