ai leaders

SCENARIO EXERCISE

Fairness and Bias – Criminal Justice System

www.aileaders-project.eu

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

CONTENTS

01 Abstract

02 Introduction

03 Tools presentation

04 Hands-on activities

05 Conclusion

06 References

Co-funded by the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.

Abstract

Type of OER:	• Scenario Exercise
Goal/Purpose:	• Analysis of bias in a machine learning algorithm for the criminal justice system
Expected Learning Outcomes:	• Awareness of bias in machine learning algorithms predictions • Detecting bias in predictions
Suggested Methodological Approach (Case-Based Learning, Problem-Based Learning...):	• Case-Based Learning
Keywords:	• bias, fairness, machine learning, criminal justice system

Introduction

Introduction

This scenario exercise analyses fairness and bias in the context of the criminal justice system.

More specifically, it addresses bias in the COMPAS machine learning algorithm used in the United States criminal justice system.

Introduction

The COMPAS machine learning algorithm calculates a risk score that is used to help a judge decide on pre-trial detention and setting bail.

The judge decides based on their assessment of the risk a released defendant would fail to appear at trial or cause harm to the public (recidivism).

The COMPAS algorithm calculates a risk score, which measures the risk that a defendant will re-offend (recidivism prediction).

That risk score is used to help the judge make the decisions on pre-trial detention and bail.

Introduction

ProPublica published an article on the bias in the COMPAS algorithm.

Data from COMPAS was also obtained and publicly released by ProPublica.

Tools presentation

Tools presentation

This scenario exercise requires knowledge of:

- The Python programming language;
- The Jupyter notebook software.

A computer with both Python and Jupyter is also needed for this scenario exercise.

In addition to built-in Python modules, the pandas, NumPy, Matplotlib and scikit-learn libraries are also required.

Tools presentation

This scenario exercise uses contents that were developed by NOS (a Portuguese telecom operator), but are publicly available on the GitHub developer platform.

More specifically, in this scenario exercise we will use the contents of the module SLU17 – Ethics and Fairness.

Tools presentation

This module is part of the NOS’ course Intro to Data Science.

In turn, this course is part of the larger FAAST Advance Data Science learning path.

All of these are used by NOS to onboard new employees, but are publicly available on GitHub.

Additionally, you can see all of NOS’ public repositories at their main GitHub page.

Tools presentation

All files required for this scenario exercise are provided in the folder “oer_files”.

So, there is no need to download any files from the GitHub page for the module SLU17 – Ethics and Fairness.

Indeed, all those files are already provided in the “oer_files” folder.

Naturally, all credit for those files goes to NOS.

Tools presentation

We remark that there are some quite minor differences between a couple of files in the “oer_files” folder and the corresponding files in GitHub.

In the provided “README.md” file, we deleted a link to a Google Docs file that is only available to NOS’ employees.

However, this file is not required for the scenario exercise.

Tools presentation

In the “Exercise notebook.ipynb” file we made the following changes:

- The first code cell was made editable, and a call to a deprecated Matplotlib style was deleted;
- We added a code cell just below the calculation of the FPR for black defendants, outputting “fpr_b”, to make it identical to what is done above in the case of white defendants;
- The last code cell was empty and as such was deleted.

Tools presentation

The “oer_files” folder also includes a file that is not present in the GitHub page.

This is file “Exercise notebook solved.ipynb”.

For convenience and reference, this file contains the code / solutions for the all the exercises in the “Exercise notebook.ipynb” file.

Hands-on activities

Hands-on activities

The first step is to read the contents of the notebook “Learning notebook.ipynb”.

This notebook covers basic concepts regarding:

1. the components of a learning system;
2. privacy by default, and;
3. bias and fairness.

Hands-on activities

Then, the scenario exercise itself consists in doing the exercises in the notebook "Exercise notebook.ipynb".

This notebook starts with a basic description of the setting: bias in the risk scores calculated by the COMPAS algorithm used in the US criminal justice system.

For further reference, the notebook has links to a book on fairness and machine learning, and to the ProPublica article on machine bias in criminal sentencing.

Hands-on activities

Exercise 1 involves plotting distributions for the risk score calculated by the COMPAS algorithm.

The goal is to plot the overall distribution, as well as the distributions by race.

In particular, the distributions of the risk score for both white and black defendants are to be plotted.

Hands-on activities

Exercise 2 involves plotting distributions for the risk scores received by the positive class (recidivists).

The goal is to plot the overall distribution, as well as the distributions by race.

Again, the distributions of the risk score for both white and black (recidivist) defendants are to be plotted.

Hands-on activities

Exercise 3 considers the defendants who were classified as high-risk of recidivism (that is, had a high value of the risk score).

The goal is to calculate the False Positive Rate (FPR) for these high-risk defendants.

The FPR is to be calculated for both white and black (high-risk) defendants.

Hands-on activities

As mentioned, the goal is to write the code to do all the exercises in the notebook “Exercise notebook.ipynb”.

As previously mentioned, for convenience and reference, we also provide a notebook that is already filled in with appropriate code (“Exercise notebook solved.ipynb”).

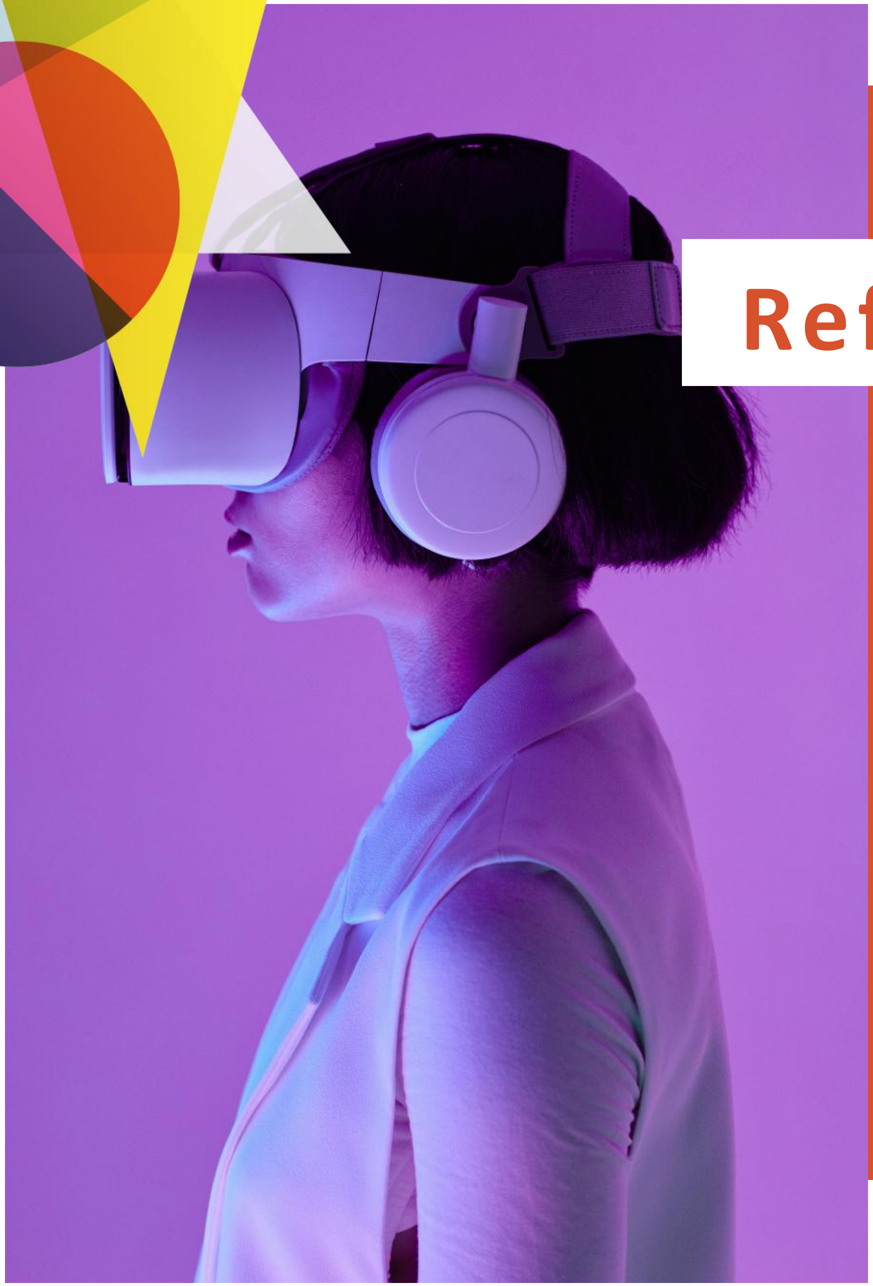
Conclusion

Conclusion

This scenario exercise shows that the predictions of machine learning algorithms may be biased.

This was clearly shown by the distinct difference in false positive rates between white and black defendants.

So, bias in algorithms may have a major impact on people or groups affected by decisions influenced by those algorithms.



References

References

Fairness and Machine Learning – Book

ProPublica Article on Machine Bias in Criminal Sentencing

NOS Ethics and Fairness Learning Unit on GitHub

NOS Intro to Data Science Course on GitHub

NOS FAAST Learning Path on GitHub

Main NOS Page on GitHub

Follow our journey

www.aileaders-project.eu

Co-funded by
the European Union

Co-funded by the European Union. Views and opinions expressed are however those of the author or authors only and do not necessarily reflect those of the European Union or the Foundation for the Development of the Education System. Neither the European Union nor the entity providing the grant can be held responsible for them.