



www.aileaders-project.eu

EJERCICIO DE ESCENARIO

Equidad y sesgo

- Sistema de justicia penal



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.

EJERCICIO DE ESCENARIO

Equidad y sesgo

Sistema de justicia penal

01



Resumen 3

02



Introducción 3

03



Presentación de herramientas 4

04



Actividades prácticas 6

05



Conclusiones 8

06



Referencias 8

• 01 Resumen



Tipo de REA

Ejercicio de escenario

Objetivo/Finalidad

Análisis del sesgo en un algoritmo de aprendizaje automático para el sistema de justicia penal

Resultados de aprendizaje esperados

- Concienciación sobre el sesgo en las predicciones de los algoritmos de aprendizaje automático
- Detección de sesgos en las predicciones

Palabras clave

- Sesgo
- Equidad
- Aprendizaje automático
- Sistema de justicia penal

Enfoque Metodológico Sugerido:

Aprendizaje basado en casos

• 02 Introducción



- 01 Más concretamente, aborda el sesgo del algoritmo de aprendizaje automático COMPAS utilizado en el sistema de justicia penal de los Estados Unidos.
- 02 El algoritmo de aprendizaje automático COMPAS calcula una puntuación de riesgo que se utiliza para ayudar al juez a decidir sobre la prisión preventiva y la fijación de la fianza.
- 03 El juez decide basándose en su evaluación del riesgo de que un acusado en libertad no se presente al juicio o cause daño al público (reincidencia).
- 04 El algoritmo COMPAS calcula una puntuación de riesgo, que mide el riesgo de que un acusado reincida (predicción de reincidencia).
- 05 Esa puntuación de riesgo se utiliza para ayudar al juez a tomar decisiones sobre la prisión preventiva y la fianza.

ProPublica

- 01 ProPublica publicó un artículo sobre el sesgo en el algoritmo COMPAS.
- 02 ProPublica también obtuvo datos de COMPAS y los publicó.

• 03 Presentación de herramientas



Este ejercicio de simulación requiere conocimientos sobre:

01 El lenguaje de programación Python.

02 El software Jupyter Notebook.

- Para este ejercicio también se necesita un ordenador con Python y Jupyter.
- Además de los módulos integrados de Python, también se requieren las bibliotecas pandas, NumPy, Matplotlib y scikit-learn.

- ▼
- 01** Este ejercicio utiliza contenidos desarrollados por NOS (un operador de telecomunicaciones portugués), pero que están disponibles públicamente en la plataforma para desarrolladores GitHub.
 - 02** Más concretamente, en este ejercicio utilizaremos el contenido del módulo SLU17: Ética y equidad.
 - 03** Este módulo forma parte del curso Introducción a la ciencia de datos de NOS.
 - 04** A su vez, este curso forma parte de la ruta de aprendizaje más amplia FFAST Advance Data Science.
 - 05** NOS utiliza todos estos contenidos para formar a los nuevos empleados, pero están disponibles públicamente en GitHub.
 - 06** Además, puede ver todos los repositorios públicos de NOS en su página principal de GitHub.
 - 07** Todos los archivos necesarios para este ejercicio de escenario se proporcionan en la carpeta «oer_files».
 - 08** Por lo tanto, no es necesario descargar ningún archivo de la página de GitHub para el módulo SLU17: Ética y equidad.
 - 09** De hecho, todos esos archivos ya se proporcionan en la carpeta «oer_files».
 - 10** Naturalmente, todo el mérito de esos archivos corresponde a NOS.
 - 11** Observamos que hay algunas diferencias bastante menores entre un par de archivos de la carpeta «oer_files» y los archivos correspondientes en GitHub.
 - 12** En el archivo «README.md» proporcionado, hemos eliminado un enlace a un archivo de Google Docs que solo está disponible para los empleados de NOS. Sin embargo, este archivo no es necesario para el ejercicio del escenario.

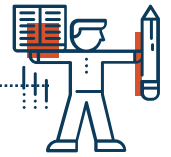


En el archivo «Exercise notebook.ipynb» Hemos realizado los siguientes cambios

- 01** Se hizo editable la primera celda de código y se eliminó una llamada a un estilo Matplotlib obsoleto.
- 02** Añadimos una celda de código justo debajo del cálculo del FPR para los acusados negros, con el resultado «fpr_b», para que fuera idéntico al realizado anteriormente en el caso de los acusados blancos.
- 03** La última celda de código estaba vacía y, por lo tanto, se eliminó.

La carpeta «oer_files» también incluye un archivo que no está presente en la página de GitHub. Se trata del archivo «**Exercise notebook solved.ipynb**». Por comodidad y como referencia, este archivo contiene el código y las soluciones de todos los ejercicios del archivo «Exercise notebook.ipynb».

• 04 Actividades prácticas



El primer paso es leer el contenido del cuaderno «Learning notebook.ipynb».

Este cuaderno cubre conceptos básicos relacionados con

01 Los componentes de un sistema de aprendizaje

02 privacidad por defecto, y;

03 sesgo y equidad.

- A continuación, el ejercicio de escenario consiste en realizar los ejercicios del cuaderno «Cuaderno de ejercicios.ipynb».
- Este cuaderno comienza con una descripción básica del contexto: el sesgo en las puntuaciones de riesgo calculadas por el algoritmo COMPAS utilizado en el sistema de justicia penal de los Estados Unidos.
- Para más información, el cuaderno incluye enlaces a un [libro](#) sobre equidad y aprendizaje automático, y al [artículo](#) de ProPublica sobre el sesgo de las máquinas en las sentencias penales.

EJERCICIO 01

- El ejercicio 1 consiste en trazar distribuciones para la puntuación de riesgo calculada por el algoritmo COMPAS.
- El objetivo es trazar la distribución general, así como las distribuciones por raza.
- En concreto, se deben trazar las distribuciones de la puntuación de riesgo tanto para los acusados blancos como para los negros.

EJERCICIO 02

- El ejercicio 2 consiste en trazar distribuciones para las puntuaciones de riesgo recibidas por la clase positiva (reincidentes).
- El objetivo es trazar la distribución general, así como las distribuciones por raza.
- Una vez más, se trazarán las distribuciones de la puntuación de riesgo tanto para los acusados blancos como para los negros (reincidentes).

EJERCICIO 03

- El ejercicio 3 considera a los acusados que fueron clasificados como de alto riesgo de reincidencia (es decir, que tenían un valor alto en la puntuación de riesgo).
- El objetivo es calcular la tasa de falsos positivos (FPR) para estos acusados de alto riesgo.
- La FPR se calculará tanto para los acusados blancos como para los negros (de alto riesgo).

La FPR también se conoce como **tasa de falsas alarmas** o **probabilidad de falsas alarmas**. La FPR se describe en el cuaderno. Sin embargo, dado que se trata de una medida crucial y la clave para comprender el sesgo en este ejercicio de simulación, aquí también se ofrece una explicación más detallada de la FPR.

El FPR se calcula como

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{FP}}{N}$$

- 01** FP es el número de falsos positivos, es decir, en nuestro caso, el número de personas que realmente no son reincidentes y que han sido clasificadas erróneamente como reincidentes.
- 02** TN es el número de negativos verdaderos, es decir, en nuestro caso, el número de personas que realmente no son reincidentes y que fueron clasificadas correctamente como no reincidentes;
- 03** N es el número total de negativos de la verdad fundamental, es decir, en nuestro caso, el número total de personas que realmente no son reincidentes.

- 01** Por lo tanto, la FPR es la tasa de negativos de la verdad fundamental que se clasificaron incorrectamente como positivos.
- 02** En nuestro contexto, la FPR es la tasa de personas que no son reincidentes y que fueron clasificadas incorrectamente como reincidentes.
- 03** Los valores más altos son, por lo tanto, más indeseables, ya que una FPR más alta significa que una mayor proporción de personas que no son reincidentes fue clasificada incorrectamente como reincidente.
- 04** Como se ha mencionado, el objetivo es escribir el código para realizar todos los ejercicios del cuaderno «**Exercise notebook.ipynb**».
- 05** Como se ha mencionado anteriormente, por comodidad y como referencia, también proporcionamos un cuaderno que ya está relleno con el código adecuado («**Cuaderno de ejercicios resuelto.ipynb**»).

• 05 Conclusión



Este ejercicio de simulación muestra que las predicciones de los algoritmos de aprendizaje automático pueden estar sesgadas. Esto quedó claramente demostrado por la marcada diferencia en las tasas de falsos positivos entre los acusados blancos y negros. Por lo tanto, el sesgo en los algoritmos puede tener un impacto importante en las personas o grupos afectados por las decisiones influenciadas por dichos algoritmos.

• 06 Referencias



- [Equidad y aprendizaje automático – Libro](#)
- [Artículo de ProPublica sobre el sesgo de las máquinas en las sentencias penales](#)
- [Unidad de aprendizaje sobre ética y equidad de NOS en GitHub](#)
- [Curso de introducción a la ciencia de datos de NOS en GitHub](#)
- [Ruta de aprendizaje FAAST de NOS en GitHub](#)
- [Página principal de NOS en GitHub](#)



Sigue nuestro viaje



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.