

Artificial Intelligence - Caso Prático 2

Caso Prático 2

Nuno Vianna Capão

Instituto Europeo de Posgrado

La equidad en los modelos de detección de fraude

El caso presentado presenta un escenario en el que un modelo de clasificación de fraude basado en el uso de energía predice una menor probabilidad de ser mujer que de ser hombre. A pesar de que este refleja un patrón observado en la realidad social, es fundamental examinar si este modelo es justo de acuerdo con el concepto de “*fairness*” en la Inteligencia Artificial.

Además, la justicia debe proporcionar la base para que los resultados previstos sean precisos y no perpetúen la discriminación implícita. Parece que el modelo considera el género como una forma de hacer predicciones, lo cual no es ético. A pesar de que hay una posibilidad de que las estadísticas muestren que las mujeres no cometen delitos de manera frecuente, el género puede llevar a sesgos implícitos que estereotipan a los hombres al exagerar sus posibilidades de cometer otros tipos de delitos.

Considerando la escuela filosófica de “*fairness*”, este modelo aún no sería del todo justo. La equidad algorítmica no se limita a la precisión estadística, sino más bien a si todos los individuos son evaluados utilizando los mismos criterios, independientemente de las características especiales o protegidas, como el género. Por lo tanto, el modelo ideal se centraría en variables que claramente estén vinculadas al comportamiento de consumo en sí y sean independientes de factores demográficos.

Finalmente, el modelo podría ser considerado injusto ya que recurre a un particular criterio de género que, aunque esté preocupado sobre la base de patrones reales, lleva consigo un sesgo estructural. Este tipo de decisión algorítmicamente puede resultar en una amplificación de desigualdades y es antitético a la equidad: el concepto es el mismo de infraestructura semántica que garantiza igualdad de oportunidades.