

# Sesgo Racial en Sistemas de Salud de Estados Unidos

DAVID MONTERO LOAIZA

## Seguridad, Privacidad y Aspectos Legales

El sesgo racial hace parte de los sesgos implícitos, también conocidos como estereotipos implícitos, mediante los cuales se le atribuye a algún grupo social ciertas cualidades. Este tipo de sesgo suele ser inconsciente, y aunque alguna persona particular no mantenga algún estereotipo conscientemente (estereotipo explícito), es probable que involuntariamente mantenga este tipo de sesgo.

El sesgo racial, como sesgo implícito, se mantiene en diversos ámbitos, no de manera explícita, pero sí llevado involuntariamente a ser parte de sesgos en diferentes áreas como la medicina, educación y justicia. En la ciencia de datos, el sesgo racial también aparece, como el caso en que Google etiquetó por error a dos afro-americanos como «Gorilas», o el caso de la aplicación FaceApp, que aclaraba el color de piel de clientes negros ya que sus datos de entrenamiento eran mayormente gente blanca [1]. El aumento del uso de herramientas digitales, en las cuales se ve incluida la ciencia de datos, está generando en algunos sectores la «*automatización de discriminación racial*» [2].

En el caso de los sistemas de salud, se han desarrollado múltiples herramientas para la detección de riesgos de salud, en cuyo estado, el riesgo de salud es predicho según una cantidad  $n$  de variables predictoras. El objetivo de predecir el riesgo de salud se basa en obtener de manera temprana un indicador del riesgo que puede tener una determinada persona de padecer una enfermedad grave, que si no se trata a tiempo, puede causar un gasto mayor en un futuro a menos que logre detectarse con anterioridad, lo que causaría un gasto menor, y una atención temprana a pacientes que pueden sufrir una enfermedad grave.

En el sistema de salud de los Estados Unidos, un algoritmo que se usa para predecir niveles de riesgo de salud para aproximadamente 200 millones de pacientes al año, se encuentra sesgado racialmente sin especificar explícitamente la «raza» como una variable predictora del nivel de riesgo [3]. En este caso, la variable predictora que produjo el sesgo racial fue el «costo total de atención de salud», la cual se introdujo en el modelo asumiendo que a un mayor costo de atención, una patología mucho más grave se atendía, y por ende, un mayor riesgo de salud.

El diseño de este algoritmo provocó que pacientes negros, los cuales aparentemente tenían un mismo nivel de riesgo que pacientes blancos —e incluso niveles inferiores—, tuvieran en realidad enfermedades mucho más graves debido a que los proveedores de salud gastan mucho menos dinero en su atención, haciendo que la variable de costo de atención disminuyera su nivel de riesgo [2].

Las consecuencias de este sesgo son graves, ya que, aunque un paciente negro esté mucho más enfermo que un paciente blanco, el paciente negro no será atendido por un programa de alto riesgo —programas dedicados a mejorar el cuidado para pacientes con condiciones clínicas complejas— ya que ambos tienen el mismo nivel de riesgo según la predicción, o incluso un riesgo menor. Si las predicciones del riesgo se hubieran hecho teniendo en cuenta como predictores base la cantidad y la severidad de enfermedades crónicas activas, quitando gran parte del sesgo del modelo, de un 17.7 % de pacientes negros detectados por el algoritmo original, se aumentaría a un 46.5 % de pacientes negros que habría sido detectado según el nivel de riesgo como acreedores de programas de alto riesgo [4].

Teniendo en cuenta el avance en herramientas digitales, el sesgo racial se puede volver peligrosamente sistemático mientras no se diseñe con cuidado sistemas y modelos que tengan en cuenta variables que pueden traer consigo errores con este tipo de sesgo. Los datos de entrenamiento para este tipo de predicciones suele ser histórico y trae consigo ya un sesgo racial explícito —evidenciado en el menor gasto aplicado a pacientes negros— que viene de épocas pasadas. Esto lo vuelve sumamente peligroso para predicciones a futuro, ya que el sesgo se vería aplicado a todos los pacientes que entren en el algoritmo, lo que aumenta la cantidad de víctimas de este en comparación a que si se tratara de un sesgo racial por una persona particular que evalúa el riesgo «manualmente».

## Referencias

- [1] Nicol Turner Lee. Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, 16(3):252–260, 2018.
- [2] Ruha Benjamin. Assessing risk, automating racism. *Science*, 366(6464):421–422, 2019.
- [3] Heidi Ledford. Millions affected by racial bias in health-care algorithm. *Nature*, 574:608–608, 2019.
- [4] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.