

CASO AMANDA¹

La tuberculosis es una de las metas del objetivo 3 de los Objetivos de Desarrollo Sostenible (ODS) de la Organización Mundial de la Salud (OMS). Según un estudio² del Ministerio de Salud (MINSAL, 2018), la morbilidad total de tuberculosis en todas sus formas (TBTF) -la suma de casos nuevos y las recaídas- fue de 15,6 por 100.000 habitantes para el año 2017 (una incidencia menor a 10 por 100.000 habitantes define a un país de baja incidencia). El mismo estudio señala que respecto a los grupos de riesgo, se tiene que adultos mayores y extranjeros (principalmente latinoamericanos) presentan la mayor proporción de casos de TBTF (21,4% y 15,5% respectivamente) en Chile.

Amanda es ingeniera en computación y cumple la función de cientista de datos en IA-Salud, una consultora que presta asesoría a distintas organizaciones privadas y públicas ligadas al mundo de la salud. En esta ocasión está realizando un trabajo de asesoría para un hospital perteneciente a uno de los servicios de salud del país. Para realizar una detección y vigilancia temprana de pacientes con tuberculosis (una infección bacteriana contagiosa que afecta principalmente a los pulmones), le piden que implemente un algoritmo de inteligencia artificial que ayude a determinar si un paciente tiene o no tuberculosis, utilizando distintos datos y mediciones del paciente. Para esto, Amanda entrenó una variedad de algoritmos con datos de la población de salud que atiende el hospital, tomando una proporción similar de pacientes con tuberculosis que pacientes sin tuberculosis.

Actualmente ella está en la etapa de probar y seleccionar el algoritmo a utilizar para la predicción de tuberculosis. Debe entregar los resultados en un mes al hospital y hay cuatro algoritmos susceptibles de ser utilizados para la labor encomendada: el algoritmo X, el algoritmo Y, el algoritmo W y el algoritmo Z. Al examinar sus desempeños, advierte que el algoritmo X es más exacto³ en promedio del total de la muestra, pero tiene una tasa de error⁴ más elevada en un subconjunto de la muestra correspondiente a la población migrante. Por otra parte, el algoritmo Y tiene tasas de error similares en la población de distintas nacionalidades y orígenes, pero tiende a fallar más en promedio. En cuanto a los algoritmos restantes, encuentra que el algoritmo W tiene mayor exactitud, pero menor explicabilidad⁵ mientras que el algoritmo Z es menos exacto pero los resultados que entrega son explicables.

En relación a la predictibilidad de un algoritmo, hay dos consecuencias de que este genere predicciones equivocadas. La primera es que si una persona no tiene tuberculosis pero el algoritmo predice que la tiene, se realizaría un mayor gasto en tratamientos innecesarios. La segunda es que si una persona tiene tuberculosis pero el algoritmo predice que no la tiene, entonces su enfermedad no sería detectada y por lo tanto no se trataría, causando potencialmente su muerte. Si se utiliza un

¹ Caso elaborado colaborativamente entre la Asociación de Ética en Datos e Inteligencia Artificial (AEDIA) y la Unidad de Ética de ETHICS.

² https://diprece.MINSAL.cl/wp-content/uploads/2018/10/2018.10.05_TUBERCULOSIS-2017.pdf

³ La exactitud se refiere a la proporción de datos bien predichos con respecto al total de datos.

⁴ La tasa de error se refiere a la proporción de datos *mal* predichos con respecto al total de datos. En este caso a detectar la enfermedad cuando no la tiene o viceversa.

⁵ La explicabilidad se refiere al grado en el que el comportamiento del modelo se puede explicar en términos humanos, considerando tanto el resultado como todo el proceso de la toma de decisión.

(<https://blogs.upm.es/catedra-idanae/wp-content/uploads/sites/698/2019/09/3T19-Quarterly-iDANAE.pdf>)

algoritmo menos exacto en la población entera o en una muestra de ella, las consecuencias mencionadas se presentan en mayor cantidad.

En cuanto a la explicabilidad de un algoritmo, si este no es explicable entonces ni el equipo médico ni el equipo de ingenieras/os podrían estudiar sus resultados para entender potenciales errores, afectando así su confiabilidad. Además, se estaría infringiendo el *derecho a una explicación* del paciente, reconocido legalmente en otras legislaciones⁶ como el derecho que tiene un/a usuario/a a pedir una explicación de la salida de un algoritmo cuya decisión pueda tener un impacto importante en su vida, como ocurre en este caso por las consecuencias graves de un posible error.

⁶ En la Unión Europea, aplica el *General Data Protection Regulation* (<https://www.privacy-regulation.eu/en/r71.htm>). También existe una regulación similar en Estados Unidos para puntuación crediticia automatizada, por el *Equal Credit Opportunity Act* (<https://www.ecfr.gov/current/title-12/chapter-X/part-1002/subpart-A/section-1002.9>)