“Momento de retroalimentación: Análisis del contexto y la normatividad”

Rodolfo Jesús Cruz Rebollar

A01368326

Grupo 101

En primera instancia, uno de los desafíos éticos más significativos de la inteligencia artificial en la actualidad es la toma de decisiones realizada de forma automatizada por algoritmos basados en la IA, mismas que antes eran tomadas por personas [1], especialmente aquellas cuyo efecto tiene un impacto importante y directo en la vida de las personas (por ejemplo, un algoritmo para determinar si contratar a alguien o no para un puesto laboral), lo cual tiene implicaciones éticas bastante significativas en el sentido de que es posible que el algoritmo automatizado no se encuentre adecuadamente entrenado para determinar si otorgarle el puesto laboral a un candidato o no, al haber sido entrenado inicialmente con datos cuya calidad sea deficiente (tipos de datos erróneos, datos ausentes, entre otras inconsistencias en ellos), por lo cual, en ese caso, el modelo no podrá aprender correctamente los patrones, o correlaciones detrás de los datos y como consecuencia arrojará predicciones erradas y con un bajo nivel de confiabilidad, provocando que el algoritmo decida que no se le otorgue el puesto laboral a un candidato cuando éste mismo en realidad sí posee las habilidades y aptitudes necesarias para ejercer adecuadamente el puesto, lo cual sería injusto, además de una posible discriminación, dado que no se tomarían en cuenta por otro lado todas las habilidades y cualidades positivas de la persona al momento de decidir si contratarla o no para el puesto, lo cual no permitiría evaluar con la profundidad requerida todo el potencial de la persona para tener un buen desempeño en el entorno laboral, motivo por el cual, para garantizar que los algoritmos de aprendizaje automático e IA operen de una manera ética en beneficio de la población, primeramente es necesario asegurar que éstos sean entrenados en un principio con datos de buena calidad y libres de errores o inconsistencias que puedan entorpecer el aprendizaje de los mismos, para lo cual antes de elaborar cualquier modelo predictivo o clasificatorio, es necesario realizar el preprocesamiento de los datos a utilizar para corregir o eliminar inconsistencias en la estructura de los datos que puedan afectar el rendimiento de los modelos, además de asegurar que al entrenar los modelos, los datos usados para dicho fin tengan clases balanceadas, es decir que exista la misma proporción de datos de cada clase posible para entrenar a los modelos, ya que en caso de que dicha proporción sea desigual, se corre el riesgo de que éstos mismos aprendan que los datos de una clase en específico, tienen obligatoriamente una cualidad en común, lo cual puede introducir sesgo a las predicciones del modelo [2], por ejemplo, si se

tiene un modelo que determine si contratar a un candidato o no para el puesto de gerente de una empresa, y los datos con los que se entrena dicho modelo tienen más registros de gerentes anteriores que fueron hombres que de las que fueron mujeres, se podría dar la situación en la que una candidata mujer que aplica para el puesto de gerente de la empresa resultara no ser contratada en base a la decisión del modelo clasificatorio mientras que un candidato hombre sí sería contratado, lo cual constituye una violación a los principios éticos de nuestra sociedad, ya que sería una situación de discriminación hacia una persona debido a alguna de sus características físicas o de personalidad, por lo que en ese tipo de casos, una de las mejores alternativas consiste en agregar parámetros de configuración adicionales al proceso de selección aleatorio de los datos para entrenar el modelo, esto para garantizar que al hacer la selección, se elijan cantidades equilibradas de datos de cada clase posible y con ello, garantizar a su vez que el modelo no se incline por una tendencia específica de valores o datos al momento de clasificar los datos, por lo cual una posible solución para el ejemplo visto anteriormente radica en que al generar el conjunto de datos de entrenamiento para el modelo, se especifique como criterio de selección el hecho de que al elegir aleatoriamente los datos, la cantidad recopilada de datos de mujeres sea la misma que la de hombres, para que con ello, tanto un hombre como una mujer tendrán ambos la misma probabilidad de ser seleccionados como dato de entrenamiento para el modelo y como consecuencia eliminar el sesgo de desigualdad de género que tenía el modelo al momento de decidir si contratar o no a una persona para el puesto de trabajo.

En resumen, en base a todo lo descrito con anterioridad, se concluye que todas las etapas del trabajo cotidiano de un científico de datos son muy importantes, ya que de todas ellas dependerá el hecho de que los modelos creados y entrenados posteriormente sean altamente efectivos, precisos y confiables, entre lo cual destaca el que dichos modelos no incurran en sesgos de ningún tipo al momento de realizar las predicciones de alguna variable de interés y con ello, también evitar que esos mismos modelos propicien alguna situación de desigualdad o discriminación hacia ciertos grupos de personas desde el punto de vista ético, no obstante, también es importante recalcar que en situaciones en las que se requiera tomar decisiones que a su vez vayan a tener un impacto mayormente fuerte en la vida de población general, es necesario que no deleguemos al modelo predictivo toda la responsabilidad de elegir si otorgar o no un cierto beneficio a un usuario, sino que nosotros como programadores del modelo, examinemos a conciencia los resultados que el modelo nos proponga y a su vez también conozcamos a profundidad cómo es aquella persona candidata para el puesto laboral y si realmente cumple o no con las habilidades y conocimientos que se requieren para desempeñar adecuadamente el puesto, por lo cual el resultado arrojado por el modelo deber ser solamente un elemento auxiliar para llegar a nuestra conclusión final, pero no debe sustituir a nuestra capacidad de pensamiento crítico que tenemos como personas para determinar el mejor curso

de acción a seguir al momento de tomar una decisión, por lo cual, la decisión final deberá ser nuestra.

Referencias

[1] C. F. Breidbach and P. Maglio, “Accountable algorithms? The ethical implications of data-driven business models,” *J. Serv. Manag.*, vol. 31, no. 2, pp. 163–185, 2020, doi: 10.1108/JOSM-03-2019-0073.

[2] Khan Academy, “Sesgo en algoritmos predictivos,” *Khan Academy*. https://es.khanacademy.org/computing/ap-computer-science-principles/data-analysis-101/x2d2f703b37b450a3:machine-learning-and-bias/a/bias-in-predictive-algorithms (accessed Aug. 26, 2024).