



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Estado de México

Inteligencia Artificial Avanzada para la Ciencia de Datos I

Normatividad en el contexto del Reto

TC3006C

Grupo: 101

Profesores

Víctor Adrián Sosa Hernández

Elisabetta Crescio

Julio Guillermo Arriaga Blumenkron

Jorge Adolfo Ramirez Uresti

David Higuera Rosales

Presenta

David Damián Galán - A01752785

Fecha de entrega: 16 de septiembre de 2023

Introducción

El aprendizaje máquina es una herramienta que ha revolucionado de manera increíble los avances que hemos tenido en resolución de problemas en los últimos años, desde predicción de valores hasta clasificación y generación de texto, imágenes y videos. Sin embargo, también existe una problemática en cuanto al uso de esta tecnología. Si los datos de entrenamiento no son adecuados (contienen sesgo), entonces el modelo generado también podría tener problemas de este estilo, resultando en injusticias y afectaciones para ciertos sectores de la población. Por ejemplo, en el pasado se han evidenciado proyectos con esta clase de sesgos, como una herramienta de reclutamiento que utiliza machine learning que elegía a hombres sobre mujeres, herramientas de reconocimiento facial que no funcionan con personas de color, y una cuenta de Twitter basada en IA que publicaba comentarios racistas (Saturia, 2022).

Principios éticos aplicados al reto

En el proyecto, se utiliza el “Ames Housing Dataset”, el cual fue recopilado por Dean De Cock en la Truman State University. Este dataset contiene 3970 ventas realizadas en el área en los años de 2006 a 2010. El autor menciona en su artículo que la motivación para dicha compilación es que el dataset se estaba volviendo muy desactualizado y ya no podían reflejarse los precios del mercado actual (De Cock, 2011). Más importante, posteriormente se descubrió que este dataset tenía un problema ético: los autores del dataset incluyeron una variable referida a la auto-segregación racial, asumiendo que esto tendría un efecto en el precio de la vivienda. Esto es un ejemplo de cómo problemas en el dataset pueden llevar ocasionar problemas por sesgos en los datos de entrenamiento, ya que, en principio, si el modelo que proviene de esos datos fuera llevado a producción y a un servicio en la vida real, ocasionaría un efecto negativo por racismo.

Afortunadamente, el dataset de Ames Housing, el cual es en el que se basa el modelo de machine learning del reto no contiene dichas variables controversiales, lo que es sumamente beneficioso para evitar que nuestro modelo aprenda de los sesgos del dataset. Además, otra ventaja es que los datos están anonimizados, lo que significa que no se puede identificar información personal de la gente entre la que se realizó la compra-venta, sino que solo hay datos de la venta en sí misma. No queremos que nuestro modelo aprendiera de esos datos si es que existieran, ya que los datos no anonimizados también podrían llevar a problemas éticos.

Ya son varias entidades las que han tratado de regular el uso y mitigación de los sesgos perjudiciales en la inteligencia artificial y aprendizaje máquina, como lo es la Unión Europea con el “Artificial Intelligence Act”, el cual sirve como guía para identificar y tratar los problemas éticos en sistemas de alto riesgo que utilizan inteligencia artificial. Sin embargo, estos deben de ser diferentes para cada dataset. En el caso que nos confiere, podemos verificar que no existen sesgos hacia las personas porque las variables que se utilizaron finalmente para entrenar al modelo solamente se refieren a la casa como tal. De hecho, ni siquiera se utiliza información acerca del vecindario o de su localización, por lo que es complicado que los problemas éticos aparezcan.

Referencias

De Cock, D. (2011). Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project. Journal of Statistics Education, 19(3).

European Parliament. (2021). Artificial Intelligence Act.

[https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792)

Saturia, N. (2022). Bias and Ethical Concerns in Machine Learning. ISACA Journal. 2022(4).