

Análisis del contexto y la normatividad

A00831314- Sofia Reyes

2023-08-29

Normativa de datos

Tipo de datos

La base de datos California Housing Prices, es una versión modificada de su primera aparición en 1997 en una investigación de regresión lineal dispersa. La información que contiene fue recuperada de un censo en el estado de California en 1990, por lo que las variables se encuentran en uno de dos escenarios:

1- Son la mediana de dicha variable entre las casas de un bloque o manzana 2- La sumatoria de los valores de dicha variable en las casas de un bloque o manzana.

1. Análisis del Dataset

Según la información sobre el tipo de datos en este contexto, las normativas y leyes que aplican en estos, son las que rigen al Buró de Censos en Estados Unidos. Esta organización gubernamental está regida principalmente por el título 13 del código de Estados Unidos, el cual básicamente es la política de privacidad y confidencialidad para personas naturales y jurídicas. Este título estipula que la información privada que permita identificar a una persona o empresa (información personal), no puede ser compartida o publicada.[1] Así también, en el título trece se especifica que el buró de censos tiene la responsabilidad de publicar la información de los censos en forma de datos estadísticos, que no revelen información personal de ningún ciudadanx. Este principio, incluso se puede encontrar en la Monografía de Confidencialidad, donde se cuenta la historia de cómo se ha ido adaptando la política de privacidad: “the agency has the responsibility of releasing data for the purpose of statistical analysis” [2]

Los datos California Housing Prices, cumplen con estas normativas del Buró de Censos. Esto gracias a cómo están generadas las distintas variables, donde las mismas son estadísticos que representan de manera general las diferentes manzanas dentro de un distrito. Al trabajar con el valor de la mediana, o la suma total, es imposible identificar una casa específica, pero es posible tener una representación de los datos de casas en una manzana, o barrio.

2. Análisis de la herramienta de ML

Con respecto al modelo de regresión lineal creado, este está fundamentado dentro de los principios mostrados en la Monografía de Confidencialidad, e incluso en la cita textual comentada anteriormente. Como tal el objetivo de este modelo es de objetivo académico y estadístico, por lo que se encuentra dentro de la normatividad.

2.1. Posibles usos indebidos de la herramienta de ML

En el caso de que el algoritmo creado llegue a manos de una persona que no lo vaya a utilizar con fines estadísticos, sino, por ejemplo, con fines lucrativos, la Monografía de Confidencialidad de los datos del Buró de Censos Estadounidense estaría siendo violado. Independientemente de que es imposible descubrir la información personal de habitantes de una casa en específico por medio del algoritmo, el hecho de utilizarlo para otro fin que no sea el académico sin un permiso previo, ya representa una falta de ética.

Referencias

- [1] Census Bureau. (s.f.). *Title 13, U.S. Code* . Recuperado de: https://www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html#:~:text=The%20Census%20Bureau%20is%20bound,collect%20from%20in
- [2] Gatewood, G. (2001). *Confidentiality Monography*. Recuperado de: <https://www.census.gov/history/pdf/ConfidentialityMonograph.pdf>