



Curso: Aprendizaje de Máquina
Professor: Andrés Gregori Altamirano

Estudio sobre las Encarcelaciones en EEUU

Estudiantes - ID:
Ramón Andrés Galindo Gerardo - 0248040
Fernando Voticky Osorio - 0222993

Deadline: 17 / Julio / 2025
Periodo: 1254

Introducción e Hipótesis

El encarcelamiento en Estados Unidos es un tema que genera gran discusión social. El escrutinio bajo el que vive este tema proviene de acusaciones de discriminación contra los sentenciados, particularmente por raza. Aprovechando que estos procesos judiciales están muy bien documentados, decidimos indagar en el tema y revisar qué factores son los determinantes en los juicios estadounidenses. De aquí proviene nuestra hipótesis - Existe una desigualdad sistémica en la asignación de condenas.

Descripción del Dataset

Para trabajar este caso, utilizamos el dataset del historial judicial de sentencias Justfair, que contiene 557 variables y cerca de 600,000 registros. Este historial comprende los registros entre 1970 y 2018. Sin embargo, por la escala del proyecto, limitaciones computacionales y la cantidad abrumadora de información del dataset, decidimos reducir sus dimensiones, hasta obtener un dataset de menos de 500,000 filas y 25 columnas. Para dar perspectiva a esta reducción, el archivo original pesa 1.3 GB, mientras que el que utilizamos para el proyecto pesó tan solo 47 MB.

Métodos y Desarrollo

Para poner a prueba nuestra hipótesis, designamos al tiempo de sentencia total como nuestra variable objetivo, y elegimos tantas variables independientes relevantes como pudimos, entre las cuales se encontraban las demográficas. Hecho esto, decidimos realizar modelos basados en CART (random forest, XGBoost y GBM) para entender el dataset. Por limitaciones computacionales, no pudimos realizar grid search para encontrar a los modelos más óptimos. Sin embargo, con los parámetros dados, los tres modelos mostraron eficiencia similar, sin ningún modelo que se coronara frente a los demás.

Resultados

Los modelos, aunque en el test set hayan salido con resultados similares, mostraron resultados distintos en los pesos que asignaron a sus variables. Por ejemplo, mientras que XGBoost le dio un gran peso a la raza negra (mientras que la raza blanca no aparece), los otros modelos consideraron a esta variable como significativa, pero no le asignaron la misma relevancia. Las variables que sí permanecieron consistentes entre modelos, que también fueron las más altas fueron aquellas relacionadas a los cargos y el pasado criminal del acusado. En el análisis de correlaciones encontramos correlación significativa entre la longitud de la sentencia y las razas blanca y negra, la longitud de la sentencia y ciertas cortes específicas, y la longitud de la sentencia y el ser proveniente de uno de dos países específicos, muy probablemente México y EU. Se requiere una mayor investigación para afirmar que el que las sentencias sean dadas en ciertas cortes las hace más o menos largas, pero la evidencia y el que sea congruente con otros sutiles hallazgos relacionados con la inclinación política del juez apunta a que la relación no es insignificante. Aunque no se encontró una relación significativa entre la ideología política y la longitud de la sentencia, todos los datos analizados de distritos con política más liberal como California y Arizona tienen tasas de

sentencia menores, mientras que otros con políticas conservadoras como Texas tienen tasas de sentencia mayores. No se encontraron diferencias significativas con base en el género del juez. Es importante recalcar que los modelos predictivos demostraron que los principales factores son los legales, pero hemos hallado una mayor concentración de cadenas largas en partes específicas de la constitución demográfica de los datos.

Conclusión

Los resultados apuntan a darnos la razón. Si bien no dan absoluta certeza, existe la constante lucha entre correlación y causalidad para estos temas, la inclinación es evidente. Mostrar estos patrones con respaldo numérico es importante para contribuir a la discusión sobre la justicia y prudencia dentro del proceso de encarcelamiento, y probar la utilidad del aprendizaje de máquina fuera del ámbito empresarial. Estos resultados denotan un claro hoyo en el sistema judicial dentro de un país que se presenta como autoridad moral - el problema debe ser atendido

Bibliografía

Justfair. (2018). *JUSTFAIR: Judicial System Transparency for Fairness through Archived/Inferred Records*. Consultado el 15 de julio del 2025, de:
<https://qsideinstitute.org/research/criminal-justice/justfair/>