Hito 2 - Grupo 8

Fairness & Bias
(Data Science for Social Good)

Vicente González

Rodrigo Iturrieta

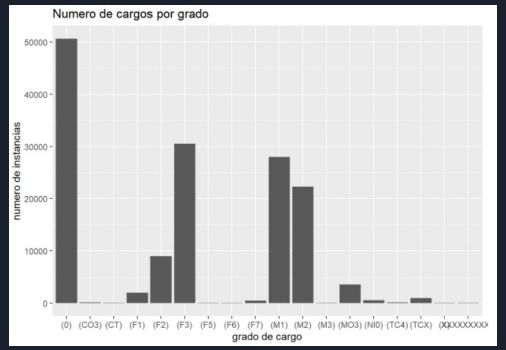
Pablo Jaramillo

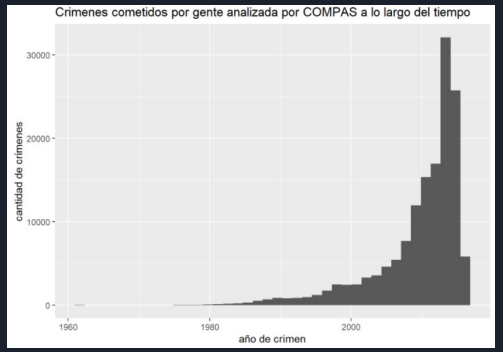
Cristian Lillo

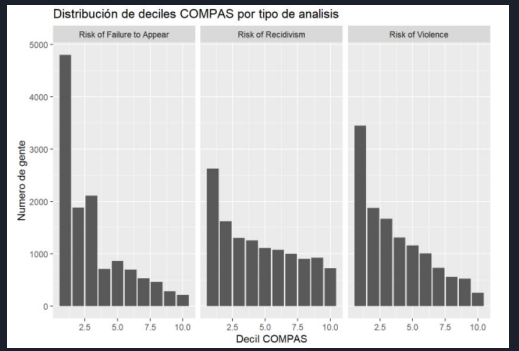
Benjamín Valenzuela

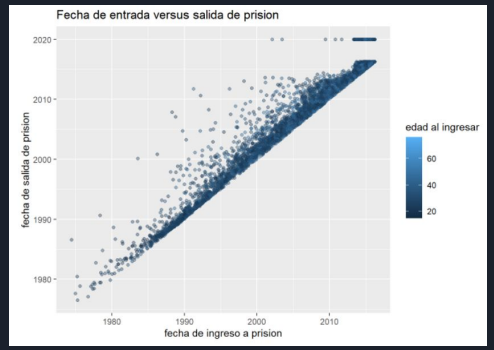
Mejoras al Hito 1

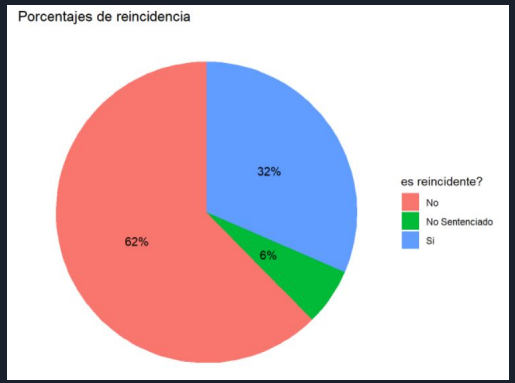
- Se crearon gráficos para visualizar los datos de las 6 tablas utilizadas en el proyecto





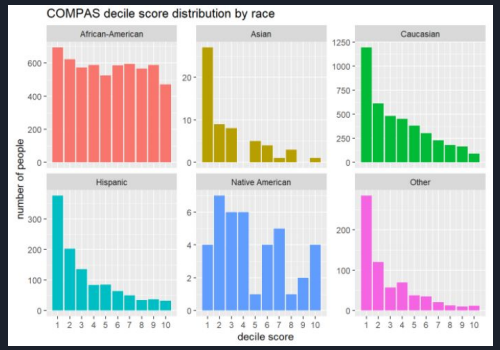




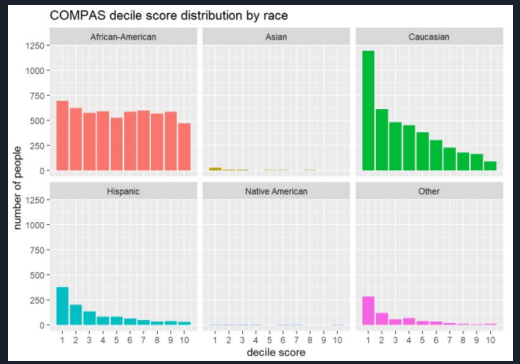


Mejoras al Hito 1

- Se unificaron los ejes de gráficos conjuntos para realizar una mejor comparación de sus datos en la visualización



Antes



Ahora

- Cambio de preguntas a unas más adecuadas para la minería de datos, donde sean los datos los que nos muestren las relaciones entre variables y no exista un condicionamiento a buscar una respuesta positiva para las preguntas, como ocurrió en el Hito 1

Propuesta experimental

Introducción

Objetivo principal:

- Generar un modelo capaz de predecir la reincidencia de una persona, evitando crear un modelo sesgado por la etnia.
- Analizar el grado de sesgo que contiene COMPAS.

Preprocesamiento:

- Se realiza un sondeo de datos para encontrar datos no deseados como NA y no numéricos y se adaptan acordemente o eliminan de no poder ser rescatables.

Propuesta experimental

Pregunta 4

Se partirá con la pregunta 4, ya que permite identificar los datos claves:

- Se añade la columna `decile_score` a todas las tablas de la base de datos.
- Se crean modelos de clasificación (K-Means y Decision Trees) con respecto a este valor.
- Se busca ir reduciendo la dimensionalidad de las tablas.
- Iterativamente, se crean modelos para todos los posibles subconjuntos de combinaciones de las columnas, y se conserva la que obtiene mejores resultados con respecto al parámetro F1.
- Se reducen las tablas a estas combinaciones con mejores resultados.
- Se concluye con la respuesta a la pregunta basado en las columnas que “sobrevivan”.

Propuesta experimental

Pregunta 2

Luego se continúa con la pregunta 2, ya que esta permitirá tener un sub-sampling más avanzado:

- Usando las tablas filtradas, se juntan para tener una tabla única con todos los datos.
- Se eliminan las columnas relacionadas con la reincidencia.
- Se usa una reducción de dimensionalidad a 2D, usando PCA para hacer un análisis visual de los clúster.
- Se aplican modelos de clustering (KNN, Jerárquico y DBSCAN) sobre los datos, utilizando los métodos del codo y rodilla según correspondan.
- La idea es buscar clústers con muy bajo porcentaje de personas reincidentes o con un porcentaje muy alto.
- Usando una exploración de datos sobre los clúster encontrados, se analizan patrones que permitan responder la pregunta.

Propuesta experimental

Preguntas 1 y 3

Las preguntas 1 y 3 permitirán concluir el análisis y dar a conocer un modelo no sesgado, si es que es posible. Para ello utilizaremos una misma metodología para realizar los modelos de clasificación:

- Se identifican los atributos importantes usando una matriz de correlación.
- Se entrenan los modelos de KNN y Decision Trees usando GridSearch con respecto al parámetro F1.
- Se aplicará oversampling y subsampling a los datos para evitar desequilibrios.
- Dependiendo si el modelo resultante es lo suficientemente satisfactorio, se puede concluir afirmativamente o no.

En particular, para la pregunta 1 se aplicará esto con respecto a la reincidencia de las personas y para la pregunta 3 se hará con respecto a la etnia, usando sólo los datos que COMPAS entrega como output.

Resultado preliminar

Con fin de responder a la pregunta “¿Qué características describen mejor un factor de reincidencia?” se realizó lo siguiente:

- Utilizar matrices de correlación para identificar atributos relevantes.
- En base a estos, entrenar clasificadores utilizando distintas combinaciones de atributos.
- Utilizar K-Neighbours y Decision Tree.

El objetivo es obtener el mejor clasificador dentro de todas las configuraciones, y en base a la ‘calidad’ de este evaluar si efectivamente se pueden distinguir atributos característicos de la reincidencia.

Para esto se usó la tabla people_cl, la cual es una versión procesada de la tabla people, tal que todos sus campos de fechas fueron pasadas a números decimales normalizados, campos tipo string catalogados numéricamente de forma binaria o incremental cuando existe una relación de tal tipo entre los strings y atributos identificadores eliminados.

Resultado preliminar

Luego de formar la matriz de correlación de `people_cl`, se distinguieron los distintos coeficientes de los atributos con respecto a 'is_recid'. En base a esto se realizaron los siguientes entrenamientos:

- Utilizando todos los atributos.
- Omitiendo los tres atributos menos significativos (correlación baja).
- Omitiendo todos los atributos con coeficiente de correlación entre -0.1 y 0.1.

Además de realizar subsampling y oversampling para todos los entrenamientos. Todo esto se realizó dos veces por clasificador, considerando o descartando el atributo 'race' para así abarcar aún más posibilidades.

Utilizando **K-Neighbours** se obtuvo una precisión máxima de 0.68 al utilizar todos los atributos y no haber realizado subsampling ni oversampling.

Resultado preliminar

Al repetir lo anterior, pero utilizando **Decision Tree**, la precisión más alta fue de 0.71. Esta se obtuvo al omitir los tres atributos menos significativos, sin realizar subsampling ni oversampling.

Si bien con estos resultados se ***podrían*** describir ciertos atributos más característicos de la reincidencia, esta no sería la conclusión más correcta.

Tras 42 entrenamientos y el uso de dos clasificadores distintos, solo se obtuvo una precisión de 0.71, la cual no consideramos como un valor lo suficientemente alto como para afirmar que ciertos atributos describen de buena manera un factor de reincidencia.

Estas características que describen mejor un factor de reincidencia existen, pero no con una influencia que nos permita asegurar que, conociendo estas, se pueda estimar con seguridad la reincidencia de un individuo.