



Inteligencia artificial avanzada para la ciencia de datos

Reto

Etapas 5. Reporte de evaluación

Diego Arturo Padilla Domínguez - A01552594

Keyuan Zhao - A01366831

Carolina Herrera Martínez - A01411547

Cutberto Arizabalo Nava - A01411431

Jose Pablo Cobos Austria - A01274631

Campus Querétaro

18 de noviembre de 2022

Evaluación de resultados

Esta evaluación se hará al modelo con mejores resultados, el modelo 5 XGBoost.

Se tiene como objetivo determinar cuáles son los aspectos estructurales que determinan la cantidad de viajes que se realizan hacia una comuna. Nuestro modelo cumple con ese objetivo, ya que con él podemos obtener que tanta importancia tiene cada uno de nuestros atractores de viajes analizados.

Importancia de cada variable determinada por el modelo XGBoost

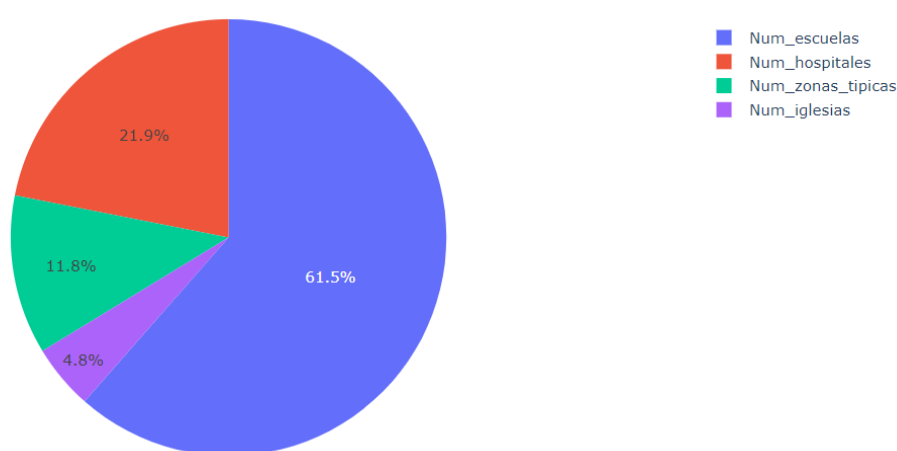


Figura 1. Importancia de cada variable en porcentaje.

Nuestro criterio de éxito es obtener hallazgos significativos sobre la causalidad de los datos y que el socio formador determine si los hallazgos son significativos.

Los hallazgos que obtuvimos son los siguientes:

- En la Fig.1 se observa que **el principal atractor de viajes son los hospitales**. El contexto es que el año 2021 fue uno de los años con mayor cantidad de casos de COVID-19 a nivel mundial, provocando que hubiera una mayor movilidad hacia los lugares que tuvieran una mayor cantidad de hospitales.

- El segundo factor con mayor impacto son las escuelas. Esto está relacionado en que en 2021 Chile comenzó el regreso remoto a las aulas (T13, 2020), causando una mayor movilidad hacia comunas que contasen con ellas.
- Los otros dos factores (iglesias y zonas típicas) tienen un impacto bajo, volviendo al punto en que fue un año con pandemia, se puede determinar que esos lugares normalmente estaban clausurados, causando así que estos no sean de gran relevancia para la movilidad.

El socio formador ha determinado que **los hallazgos obtenidos son significativos** y de alta calidad, **recomendando el incremento de variables analizadas**.

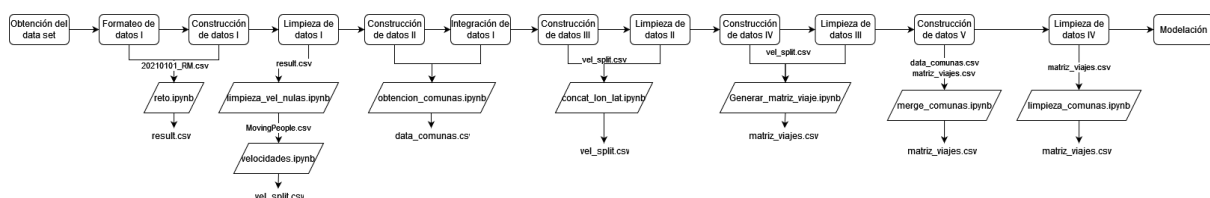
Identificación de fallas

Además de lo visto anteriormente, durante esta misma evaluación de nuestro proceso nos encontramos con una serie de fallas en este mismo, que influyeron de forma negativa en su progreso. A continuación se presentarán algunos de los que tuvieron mayor impacto:

- **Falta de claridad** : esto fue debido a que en algunas etapas del proceso, las tareas que teníamos que realizar eran ambiguas o no eran entendidas al 100% por el equipo, de tal modo que no cumplimos con la tarea correctamente o la incluyamos en otra etapa.

Revisión de procesos

Diagramas de proceso usando para minería de datos:



En cuanto a las fases de minería de datos todas fueron necesarias ya que cada etapa fue utilizando los resultados obtenidos en la etapa anterior para mejorarlos y así llegar a la versión final de nuestro dataset de entrenamiento.

La generación de la matriz de viajes al inicio no se hizo de manera óptima, ya que fue necesario hacer correcciones de formato al documento de matriz de viaje, generando múltiples versiones del mismo. Esto puede resultar contraproducente ya que se puede llegar a perder la trazabilidad de cual es la versión correcta a utilizar.

El proceso podría ser mejorado si al momento de generar la matriz de viajes hubiéramos agregado una actividad de verificación de errores antes de generar el archivo final de matriz de viajes, de esa manera hubiéramos evitado aplicar correcciones sobre la matriz ya generada.

Implementar la búsqueda automatizada de hiperparámetros en modelos previos nos pudo haber ayudado a encontrar más rápido un modelo que tenga la precisión deseada.

Asimismo es importante intentar reducir los ciclos de retroalimentación con el socio formador para tener certeza de que el trabajo realizado es el correcto y no tener retrocesos en caso de alguna discordancia. Esto se hará mediante la comunicación por medio de la herramienta **Slack**, quedando dependiente de la disponibilidad del socio formador.

Revisión de los objetivos de negocio con respecto a sus criterios de éxito.

Los insights generados por nuestro modelo nos permite obtener información clara con respecto a cuáles son los principales atractivos de viajes dentro del contexto de nuestro dataset. Asimismo, el socio formador considera que los datos obtenidos son de valor, por los que confirmamos que se cumplen los objetivos de negocio.

Siguientes pasos

Opciones a seguir:

Llegados a este punto, contamos con las siguientes posibilidades:

- Definir el modelo como validado y lanzarlo a producción
- Hacer una nueva iteración con un modelo que contenga más variables

Decisión:

Debido a la fecha de aceptación del modelo y las recomendaciones del socio formador, se llevará a cabo una nueva iteración en donde se analizarán una mayor cantidad de variables.

Las variables a agregar son:

- Densidad de la población
- Áreas verdes en metros cuadrados
- Empresas grandes, medianas y pequeñas
- Número de conexiones fijas de internet

Bibliografía

T13. (2020, 16 noviembre). *Modalidad híbrida: La estrategia de algunos colegios de cara al regreso a clases en*: <https://www.facebook.com/teletrece>.
<https://www.t13.cl/noticia/nacional/colegios-preparan-regreso-clases-2021-modalidad-hibrida-16-11-2020>