



Inteligencia artificial avanzada para la ciencia de datos

Reto

Etapa 1. Entendimiento del negocio

Diego Arturo Padilla Domínguez - A01552594

Keyuan Zhao - A01366831

Carolina Herrera Martínez - A01411547

Cutberto Arizabalo Nava - A01411431

Jose Pablo Cobos Austria - A01274631

Campus Querétaro

27 de septiembre de 2022

Misión

Como equipo de desarrollo buscamos la satisfacción de nuestros clientes en costo, tiempo y alcance.

Objetivo

Se busca modelar los viajes del 25% de la población de Santiago de Chile, que representan la cantidad de usuarios de la compañía Movistar en un día normal. Para antes del día 28 de noviembre del 2022.

El presente proyecto tiene por objetivo obtener y aplicar una metodología que permita saber cuáles son los principales atractores de viajes en la ciudad de Santiago entre distintas comunas.

Background

El presente proyecto busca estimar los **Viajes Origen-Destino** del 30% de la población de la Región Metropolitana de Chile (dato brindado por Loreto Bravo), mediante el uso de información de movilidad, almacenada por Movistar Chile, posibles de procesar para convertirlos en información de movilidad de personas. Todo lo anterior bajo los estándares de completo anonimato y sin entrega de datos personales de los usuarios de los terminales celulares, y de aquella forma, en consistencia con la normativa vigente sobre uso de datos personales y protección de la vida privada de las personas. Los aspectos éticos y legales se pueden ver más a fondo en el documento de [‘Privacidad y Seguridad de datos’](#).

Información del socio formador que sea relevante para el contexto

Nuestro socio formador es el Instituto de Data Science UDD bajo la dirección de la Doctora Loreto Bravo. Los datos se nos entregaron previamente anonimizados, sin que tengamos acceso a la información personal de los usuarios.

Objetivo del negocio

Generar un tablero que permita visualizar la cantidad de viajes que se realizan en un día entre las distintas comunas de Santiago Chile.

Determinar cuáles son los aspectos estructurales que determinan la cantidad de viajes que se realizan hacia una comuna.

Criterio de éxito

Obtener hallazgos significativos sobre la causalidad de los datos. El socio formador determinará si los hallazgos fueron significativos.

Inventario de recursos

- Expertos:
 - MTI Eduardo Daniel Juárez Pineda.
 - Dr. Benjamin Valdés Aguirre.
 - Dr. Ismael Solís Moreno.
 - Dr. José Antonio Cantoral Ceballos.
 - Dr. Carlos Alberto Dorantes Dosamantes.
 - Dra. Loreto Bravo
- Datos:
 - Dataset compartido por el socio formador
- Recursos de cómputo:
 - Nuestras computadoras
 - infraestructura de AWS
- Software:
 - Canvas
 - Slack
 - AWS
 - Visual Studio
 - Drive
 - Git
 - Spark
 - Python
 - TensorFlow
 - Tableau

Requerimientos

- Calendario de entrega
- Documentación del código
- Compresibilidad de los datos

Restricciones


Todos los datos fueron por un proceso de [seudonimización](#) debido a las normas internas de tratamientos de datos.

Supuestos

Sobre los datos:

- Tenemos al menos una torre por cada Comuna
- Cada torre le da servicio a una sola Comuna
- Las fechas de cada instancia de los datos se encuentran anonimizadas, por lo que solo la hora nos dará información veraz.
- La antena a la que se conecta el usuario es la más cercana.

Riesgos y contingencias

 Matriz de riesgos

Terminología

Seudoanonimización. Es un procedimiento de gestión de datos donde se reemplazan campos de información personal dentro de un registro de datos por uno o más identificadores artificiales o pseudónimos.

Minería de datos. Es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados.

Costo y beneficios

El proyecto tiene un costo estimado de 246 horas de esfuerzo con un personal de 5 ingenieros.

El beneficio de este proyecto es la mejora de la toma de decisiones basadas en analítica de datos.

Objetivo de minería de datos

Obtener datos representativos sobre las rutas Origen-Destino mediante la limpieza, transformación y análisis del set de datos inicial. Utilizar dichos datos para entrenar un modelo que sea capaz de predecir la cantidad de viajes a una comuna con base en las características de dicha comuna.

Criterio de éxito de minería de datos

Los datos obtenidos deben permitirnos generar una matriz de movilidad, generar una predicción de cantidad de viajes entre distintas comunas con una variación menor al 15% y con ello obtener la importancia de cada atractor de viaje analizado.

Plan de proyecto

 Plan de trabajo.xlsx