

Entrega 1: Proyecto final

Integrantes:

Andrés Parra Garzón

Carlos Cañas Díaz

Shirley Sánchez Sedano

Primera propuesta: Análisis no supervisado sobre series de tiempos

- **Motivación:** Haciendo uso de la información histórica de las ventas de los últimos 5 años de un sistema DSD (Direct Store Distribution) del sector de exportación de flores, que cuenta con 200 puntos de venta en Estados Unidos, y con 4 categorías principales de productos, se quiere poder clasificar las series de tiempo de dichas ventas con el fin de poder agruparlas y hacer pronósticos de ventas sobre grupos de series y no sobre las 800 series iniciales, suponiendo que el comportamiento de las ventas de algunas tiendas pueden agruparse con el comportamiento de las ventas de otras tiendas.
- **Explicar porque pertenecen al área de aprendizaje no supervisado:** Este problema pertenece al área de aprendizaje no supervisado ya que no tenemos a priori una clasificación del tipo de serie de tiempo a la cual queremos llegar, no tenemos ningún objetivo, se quiere que los mismos datos permitan agrupar las diferentes series y generan un nuevo grupo.
- **Identificar el potencial cliente interesado:** El cliente interesado es toda la cadena de abastecimiento del sistema DSD, ya que con la posibilidad de generar grupos de series de tiempo hace la tarea de pronóstico de ventas sea menos intensiva en tiempo y puede generar información más ágil para la toma de decisiones

Factibilidad:

- **Describir la fuente de donde los obtendría:** Los datos están disponibles en el data warehouse de la empresa al nivel de detalle requerido, normalizadas y ya procesada. Tenemos la información en unidades vendidas por semana del año, por tienda, por categoría y por año, desde el 2018 hasta el año actual.
- **Evaluar si existen barreras o no para obtenerlos:** Se debe asegurar la confidencialidad de la data, que no sea posible identificar la ubicación de las tiendas ni el tipo de categoría de las series de tiempo que se están utilizando, se tendría que solicitar permiso a la empresa para poder usarla de manera académica. En general no habría barreras para obtener la información

Segunda propuesta: Creación de la calificación interna para una Cooperativa Financiera

- **Motivación:** Una Cooperativa Financiera ubicada en la ciudad de Medellín tiene como propósito crear una calificación interna donde se pueda identificar las características tanto de sus mejores como peores clientes. Esta calificación debe tener en cuenta tanto la información de las centrales de riesgo, de cada uno de los servicios adquiridos con la entidad y las variables sociodemográficas de cada uno de los consumidores.
La anterior segmentación servirá como insumo para crear/actualizar los modelos de riesgo de crédito.
- **Explicar porque pertenecen al área de aprendizaje no supervisado:** Pertenecen al área de no supervisado porque aún no se cuenta con esta segmentación, por lo tanto, la Cooperativa no cuenta con esta etiqueta en sus datos.
- **El potencial cliente interesado:** El principal interesado es la Cooperativa Financiera ya que al realizar diferentes análisis se concluyó la necesidad de una segmentación de los clientes con el fin de mitigar el riesgo crediticio al momento de otorgar los créditos.

Factibilidad:

- **Describir la fuente de donde los obtendría:** Se obtiene desde las centrales de riesgo y desde las bases de datos de la Cooperativa.
- **Evaluar si existen barreras o no para obtenerlos:** No existe barreras para obtener los datos ya que en el momento se cuenta con la información proveniente de las centrales de riesgo y sociodemográfica de los clientes. Sin embargo, aún faltaría la información sobre los créditos recientes obtenidos por los usuarios, esta información es suministrada por el área de sistemas y que podría demorar aproximadamente una semana en ser extraída.

Tercera propuesta: Análisis de sentimientos y agrupación de quejas y reclamos

- **Motivación:** Las quejas y reclamos de los clientes son un problema complejo que si no se maneja con cuidado puede tener consecuencias serias para la imagen de una compañía.
- **Explicar porque pertenecen al área de aprendizaje no supervisado:** Este problema pertenece al área de lenguaje no supervisado ya que la información proporcionada por los clientes no está clasificada per se, es decir, para poder responder a la pregunta planteada debemos encontrar patrones en cada caso a

través de la data, que nos permitan saber lo que origina el reclamo, complejidad y severidad.

- **Identificar el potencial cliente interesado:** Cualquier compañía que ofrezca servicios que haga manejo de las quejas y reclamos de sus clientes.

Factibilidad:

- **Describir la fuente de donde los obtendría:** Data histórica del sistema de quejas y reclamos de una aerolínea (data de 2018 y 2019), la cual se encuentra en el lago de datos de dicha compañía.
- **Evaluar si existen barreras o no para obtenerlos:** Para tener acceso a la data es necesario hacer la solicitud y obtener la aprobación del custodio del activo analítico, quien debe garantizar que no sea posible identificar a los clientes, ni sus transacciones, y en general, que se haga un uso adecuado de los datos desde un contexto académico. Esto no debería ser un problema ya que uno de los miembros del equipo es el custodio de este activo.

Roles del equipo en el desarrollo del proyecto:

Rol	Descripción	Responsable
Ingeniero de datos	Responsable de habilitar los datos que alimentarán el modelo analítico (extraer, integrar, transformar, limpiar, etc.)	Andrés Parra
Científico de datos (Exploración y preparación)	Responsable de analizar exhaustivamente cada una de las variables, buscar patrones y relaciones entre ellas, que permitan tener un entendimiento general de la data que habiliten la selección de las variables relevantes para el modelo a través de técnicas para dicho propósito	Shirley Sánchez
Científico de datos (Modelamiento)	Responsable de aplicar técnicas de machine learning y procesamiento de lenguaje natural (NLP) para construir el modelo de aprendizaje no supervisado que solventa la problemática planteada	Carlos Cañas