

**Universidad Autonoma de Nuevo León**

FACULTAD DE CIENCIAS FÍSICO MATEMATICAS

**PRIMER AVANCE PROYECTO  
INTEGRADOR**

*Minería de datos*

Autor:  
Luis Rodolfo Torres Contreras

Octubre 2020

1. Título de la base de datos.
  - Nombre: Groceries dataset
  - URL: <https://www.kaggle.com/heeraldedhia/groceries-dataset>
2. Descripción de los datos.
  - Recopilación de mercancía en tienda minorista de abarrotes.
  - Columnas:
    - (a) Member\_number: id del usuario, type Int64.
    - (b) Date: Fecha de compra type Object o String.
    - (c) itemDescription: descripción producto comprado, type Object o String.
3. Justificación del uso de los datos.
  - Son datos de una tienda minorista, las tiendas minoristas no tienen implementado sistemas de recomendación como Amazon, Netflix, etc. Mi principal objetivo es realizar este sistema de recomendación de manera genérica es decir que cualquier tienda minorista pueda tener este sistema de recomendación.
  - Porque esta BD, la economía de México se basa en pequeñas empresas, la mayoría de ellas no cuenta con un sistema de recomendación que las ayude a vender más, mi objetivo es vender una solución para que se pueda implementar en sus sistemas y estas puedan generar más ingresos.
4. Planteamiento del problema.
  - Las pequeñas y medianas empresas (pymes) generan 72% del empleo y 52% del Producto Interno Bruto (PIB) del país, según estimaciones de Banca Empresarial Banamex.
  - Debido al Covid\_19 la economía en México se ha visto afectada por esta razón sería conveniente implementar algoritmos (apriori, clustering, prediccion, etc) en tiendas minoristas para maximizar sus ventas.
5. Objetivo final.
  - Entregar una lista de reglas de asociación para que los clientes puedan implementar recomendaciones de los productos en sus tiendas.
  - Si la tienda cuenta con listado de clientes, segmentar las preferencias de los clientes, de esta manera se tendrá un indicador sobre los productos que selecciona cada cliente. Con esta información podremos adelantarnos para comprar los productos con ayuda de una prediccion.

6. Planeación de la herramienta a utilizar.

- Reglas de Asociación: algoritmo a priori, la base de datos se presta para poder realizar este tipo de algoritmo debido a los productos que compra cada usuario.
- Clustering debido a que la base de datos cuenta con registro de usuarios podemos implementar este algoritmo para poder agruparlos por tipo de producto que consumen y prestar atención en los grupos que se forman y no en toda la población de usuarios.
- Sistemas de recomendación: ítems en gustos del usuario, de acuerdo con esto podemos destacar un top 10 de productos y recomendarlos a otros usuarios, podría ser algo que consuman todos los grupos que clasificamos antes.