

## 1. Contexto de aplicación

La predicción de la demanda de productos es una tarea habitual para las empresas y recientemente quienes están a su cargo son los profesionales del área de la ciencia de datos. Las previsiones son esencialmente importantes para tiendas de comestibles, pues pronosticar por encima implica un exceso de inventario de productos perecederos y pronosticar por debajo finalmente se traduce en pérdida de ventas, reputación y clientes. En general, las empresas quieren conocer cuál es la demanda de sus productos en cada uno de sus puntos de venta para hacer una distribución correcta de los mismos y además llegar a los puntos de venta en el tiempo correcto.

Se construirá un modelo que prediga las ventas de cada familia producto por punto de venta con mayor precisión que los métodos de previsión subjetiva.

## 2. Objetivo de machine learning (queremos predecir X, dada tal información)

El objetivo es construir un modelo que permita predecir las ventas (demanda) de cada familia de productos en cada punto de venta de la empresa Favorita de Ecuador dada información histórica de ventas, información geoespacial de cada punto de venta, información histórica del precio del petróleo durante el tiempo de entrenamiento y test y finalmente información de los días festivos ocurridos durante este tiempo.

## 3. Dataset: tipo de datos, tamaño (número de datos y tamaño en disco), distribución de las clases

El dataset contiene un total de 33 familias de productos para los cuáles se debe predecir la demanda por punto de venta.

Para realizar la predicción, se cuenta con las siguientes tablas:

### **Histórico de ventas:**

Tamaño en disco: 121.8 MB

Número de registros: 3.000.888

Id (int): Identificador (llave primaria)

Store\_nbr (int): Identifica la tienda en la cual fueron vendidos los productos

Date (date): La fecha de venta de los productos

Family (String): Identifica el tipo de producto vendido

Sales (int): Número total de productos pertenecientes a la familia que fueron vendidos en el punto de venta para la fecha

Onpromotion (int): El número de productos de la familia que estaban en promoción en la fecha

**Tiendas:**

Tamaño en disco: 1.39 kB

Número de registros: 54

Store\_nbr (int): Identificador de la tienda (llave primaria)

City (string): Ciudad donde se encuentra la tienda

State (string): Estado donde se encuentra la tienda

Type (string): Tipo de tienda

Cluster (int): Grupo de tiendas similares (homologación)

**Precio petróleo:**

Tamaño en disco: 20.58 kB

Número de registros: 1218

Date (date): Fecha de la medición

Dcoilwtico (int): Precio del petróleo

**Festivos y eventos:**

Tamaño en disco: 22.31 kB

Número de registros: 350

Date (date): Fecha

Type (string): Evento o festivo

Locale (string): El evento fue local o regional

Locale\_name (string): Nombre de la localidad

Descripcion (string): A qué se debió el evento o el festivo

Transferred (bool): El evento o festivo fue transferido a otra fecha

**Transacciones:**

Tamaño en disco: 1.55 MB

Número de registros: 83488

Date (date): Fecha

Store\_nbr (int): Identificación de la tienda

Transactions (int): Número de transacciones realizadas en la tienda durante la fecha

4. Métricas de desempeño (de machine learning y negocio)

**Negocio:**

Se espera predecir la demanda de tal forma que el nivel de stock para cada producto en por cada punto de venta no decaiga más del 5% pues a partir

de este punto se va a considerar que se están perdiendo ventas ya sea con respecto a productos de la misma familia o de la competencia.

### **Machine learning:**

Root Mean Squared Logarithmic Error

#### **5. Referencias y resultados previos**

[1] Y. Kaneko and K. Yada, "A Deep Learning Approach for the Prediction of Retail Store Sales," 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 2016, pp. 531-537, doi: 10.1109/ICDMW.2016.0082.

[2] Pavlyshenko, B. M. (2022). Forecasting of Non-Stationary Sales Time Series Using Deep Learning. arXiv.  
<https://doi.org/10.48550/ARXIV.2205.11636>

[3] A.L.D. Loureiro, V.L. Miguéis, Lucas F.M. da Silva, Exploring the use of deep neural networks for sales forecasting in fashion retail, Decision Support Systems, Volume 114, 2018, Pages 81-93, ISSN 0167-9236, <https://doi.org/10.1016/j.dss.2018.08.010>.