

# Sistema ReBaF



## **Nuestro Equipo:**

- Álvarez Mateo, Software Developer, Especialista en Machine Learning
- Mazzucco Uriel, Data Engineer, Especialista en Big Data
- Pilla María del Pilar, Data Analyst, Especialista en Business Intelligence
- Rojas Martín, Software Developer, Especialista en Natural Language Processing

## **¿Qué es ReBaF?:**

ReBaF (Review Based Features) es un sistema de procesamiento de datos para empresas de E-Commerce, que obtiene información más certera sobre los productos en base a las reseñas de los usuarios. Esta información es utilizada para mejorar los análisis de inteligencia empresarial e implementar un modelo de recomendación que conecta de manera más efectiva a los consumidores con los productos de la plataforma.

## ¿Quiénes somos?

Colombia Argentina Data Solutions es una empresa internacional de tecnología que brinda soluciones de procesamiento de datos para mantener a tu empresa en la cima del mercado.

*“Scientia potentia est”, ‘El conocimiento es poder’, El Leviatán, Thomas Hobbs, 1668.*

## ¿Por qué debería aplicar nuestro sistema en su empresa?

El negocio del comercio electrónico nace de conectar compradores con vendedores. Conexiones de calidad significan más ventas y por ende más ingresos. Para mejorar estas conexiones se debe conocer lo mejor posible a los vendedores, los productos y los compradores.

ReBaF analiza las reseñas de los usuarios para obtener beneficios por partida doble. Por un lado nos permite conocer mejor a los compradores y sus patrones de compra, y por otro lado brinda información más certera sobre los productos. Con esto, podemos perfeccionar los sistemas de recomendación.

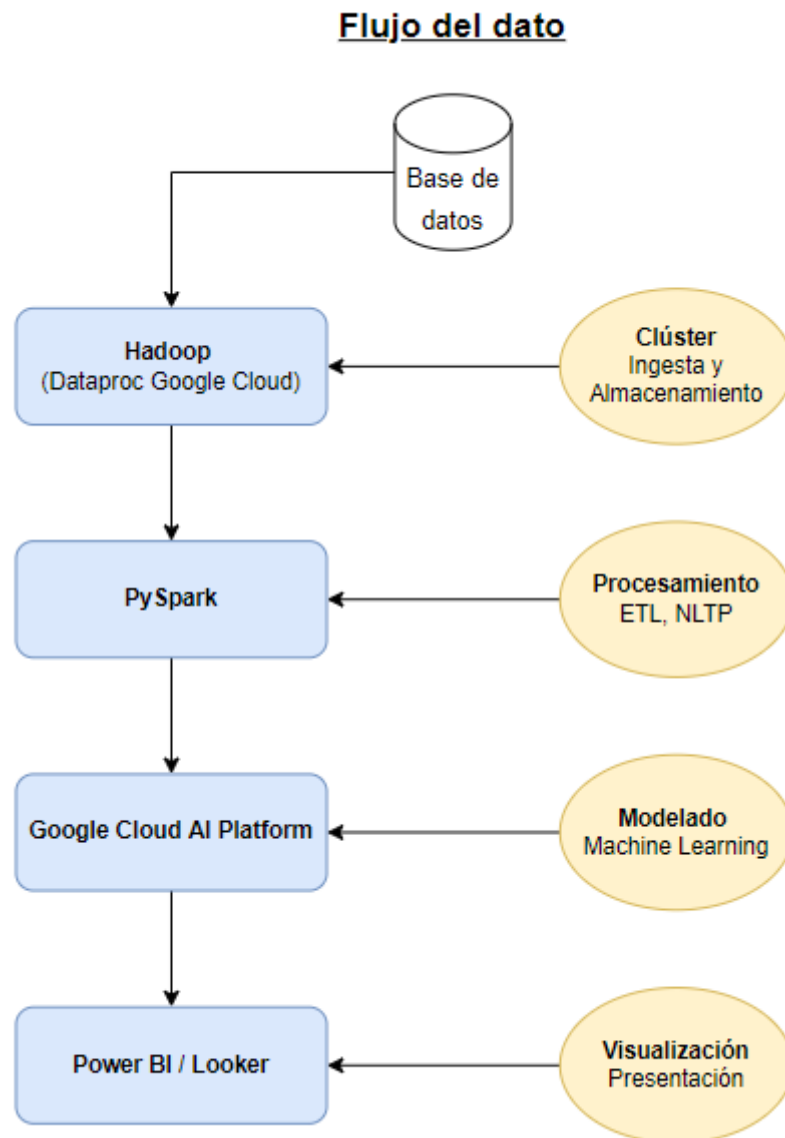
## ¿Qué haremos?

Para demostrar la utilidad de nuestro sistema, lo implementaremos sobre reseñas de Amazon desde 1996 hasta 2014. La selección utilizada abarca aquellos productos con al menos 5 reseñas, de los cuales también tendremos metadatos.

Procesaremos los datos para armar un modelo de recomendación, y realizar un análisis con base en la nueva información: NPS (Net Promoter Score), CSAT (Índice de Satisfacción del Cliente), Productos más populares e Impacto de las características obtenidas.



## Procedimiento:



### **Ingesta y Almacenamiento: Hadoop.**

Realizaremos el almacenamiento y procesamiento de los datos a través de servicios de Cloud Computing, logrando un método que se adapte a cualquier limitación de hardware que su empresa pudiera tener.

Para ello crearemos un clúster con Dataproc (Google Cloud).

Vía comandos bash obtendremos los datasets, para luego ingestarlos y procesarlos a través de Spark.

## Procesamiento : PySpark

Spark nos posibilita implementar soluciones de procesamiento de datos integradas con Python, lo cual será de mucha ayuda en el procesamiento de texto.

## I. ETL:

Aunque los datos recibidos ya tienen realizado un proceso de limpieza y selección, realizaremos algunos ajustes de normalización de columnas y nos desharemos de algunas columnas que no aportan utilidad.

## II. Procesamiento de Texto: NLP

Buscaremos palabras claves que aporten información extra sobre el producto desde la percepción del comprador, como la calidad, el precio, la facilidad de uso o el acabado; y la satisfacción del cliente.

### III. Nuevos indicadores:

Con los datos obtenidos del texto crearemos nueva información, en forma de columnas extras.

## Modelado: Google Cloud AI, Sklearn

Implementaremos un modelo de Machine Learning que recomiende productos en base a la información tanto previa como recién generada.

## Visualización: PowerBI , Looker, BigQuery

Crearemos una presentación analizando los patrones y KPIs obtenidos, seleccionando la herramienta de acuerdo a las necesidades.

**Gantt:**

[illegible]

### **¿Se puede trabajar con otras herramientas?**

- Se puede utilizar SQL si la empresa tiene una base de datos y la capacidad de cómputo.
- Como alternativa de Cloud Computing se puede utilizar Azure o AWS.

### **¿Cómo se podría continuar?**

De acuerdo a las necesidades del negocio, se podrían construir las siguientes características en base a ReBaF:

- Implementando un sistema de etiquetas review-based para guiar al usuario.
- Trazando perfiles para cada usuario de la plataforma.
- Aplicando aquellas métricas que su empresa considere más adecuadas.
- Añadiendo otra información que aporte al modelo de recomendación.

Además, el ReBaF se puede adaptar a las estructuras de datos de cada empresa, puesto que su núcleo es el texto de las reseñas.

Ante cualquier consulta, envíe un correo a [uriel.mazzucco@mi.unc.edu.ar](mailto:uriel.mazzucco@mi.unc.edu.ar)