

# Priorización de Clientela con Aprendizaje Automático

## Inteligencia Artificial Avanzada para la Ciencia de Datos II (Gpo 101)

### Pixel Duck

#### Instituto Tecnológico y de Estudios Superiores de Monterrey

Cleber Gerardo Pérez Galicia, Juan Pablo Bernal Lafarga, Jacobo Hirsch Rodríguez, Eryk Elizondo González  
Noviembre 2024



### Introducción

Arca Continental, la segunda embotelladora de Coca-Cola más grande de América Latina, se destaca por su diversificada oferta de bebidas y su compromiso con la sostenibilidad y la responsabilidad social. En el competitivo mercado de bebidas, la empresa utiliza el análisis predictivo y la ciencia de datos para anticipar el comportamiento del consumidor y mejorar el lanzamiento de nuevos productos.



### Problemática

Arca Continental busca desarrollar un algoritmo de inteligencia artificial capaz de predecir el éxito de nuevos productos en el mercado, basándose en el comportamiento de venta de sus clientes y sus características demográficas.

### Metodología

#### Extracción de Datos

Para la extracción, se prepararon dos datasets principales. El primero contiene un registro histórico de transacciones, donde cada entrada corresponde a la venta de un producto específico a un cliente en un mes dado. El segundo proporciona información detallada sobre las características de cada producto disponible para la venta. Cada entrada corresponde a un producto específico y contiene campos relevantes.

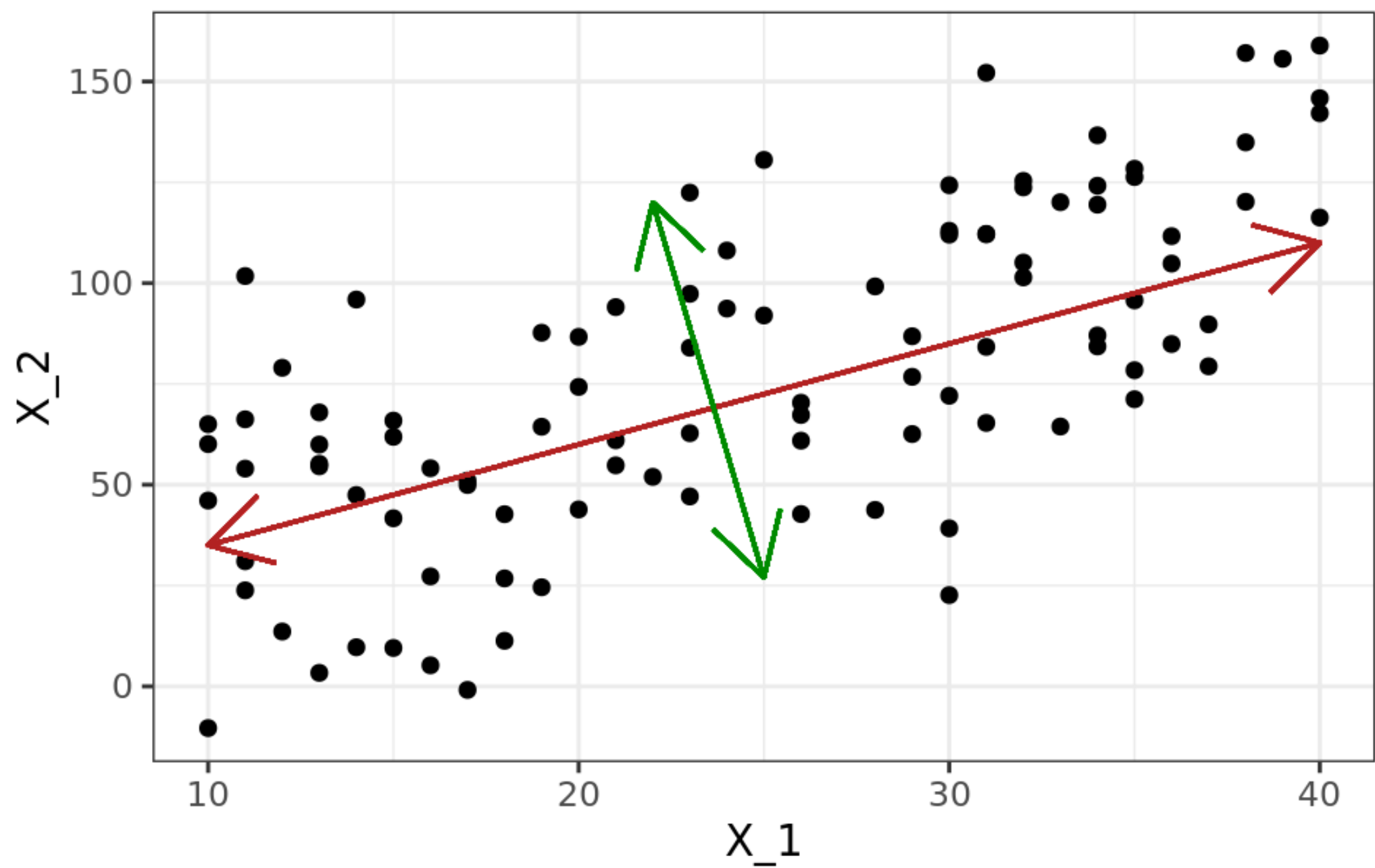
#### Limpeza de Datos

Para la limpieza, se identificó redundancia entre las variables categóricas que identifican a los productos y con el objetivo de reducir la dimensionalidad de las tablas, se utilizaron 2 técnicas de procesamiento de lenguaje natural con el fin de identificar las similitudes semánticas entre las categorías. Las 2 técnicas son:

- Vector Embeddings
- Similitud de Jaccard

#### Transformación de Datos

Para la transformación de los productos, se aplicó la función de One-Hot Encoding de Scikit Learn en las columnas categóricas para convertirlas a binarias. Luego, se normalizaron los valores de las columnas utilizando la función de Min-Max Scaling. Finalmente, se redujo la dimensionalidad con la técnica de Análisis de Componentes Principales.



Para la transformación de las ventas, se reformateó la variable de la fecha, la cual se encontraba como tipo entero, a un datetime. Luego, se calculó la frecuencia de ventas de los productos por cliente al obtener una proporción entre los meses donde

hubieron ventas entre el total de meses a evaluar de todos los productos. Finalmente, se obtuvo la magnitud de ventas relativas de los productos al obtener la proporción de todos los galones vendidos por producto por cliente entre la venta de galones totales de todos los productos del mismo cliente.

### Modelo de Solución

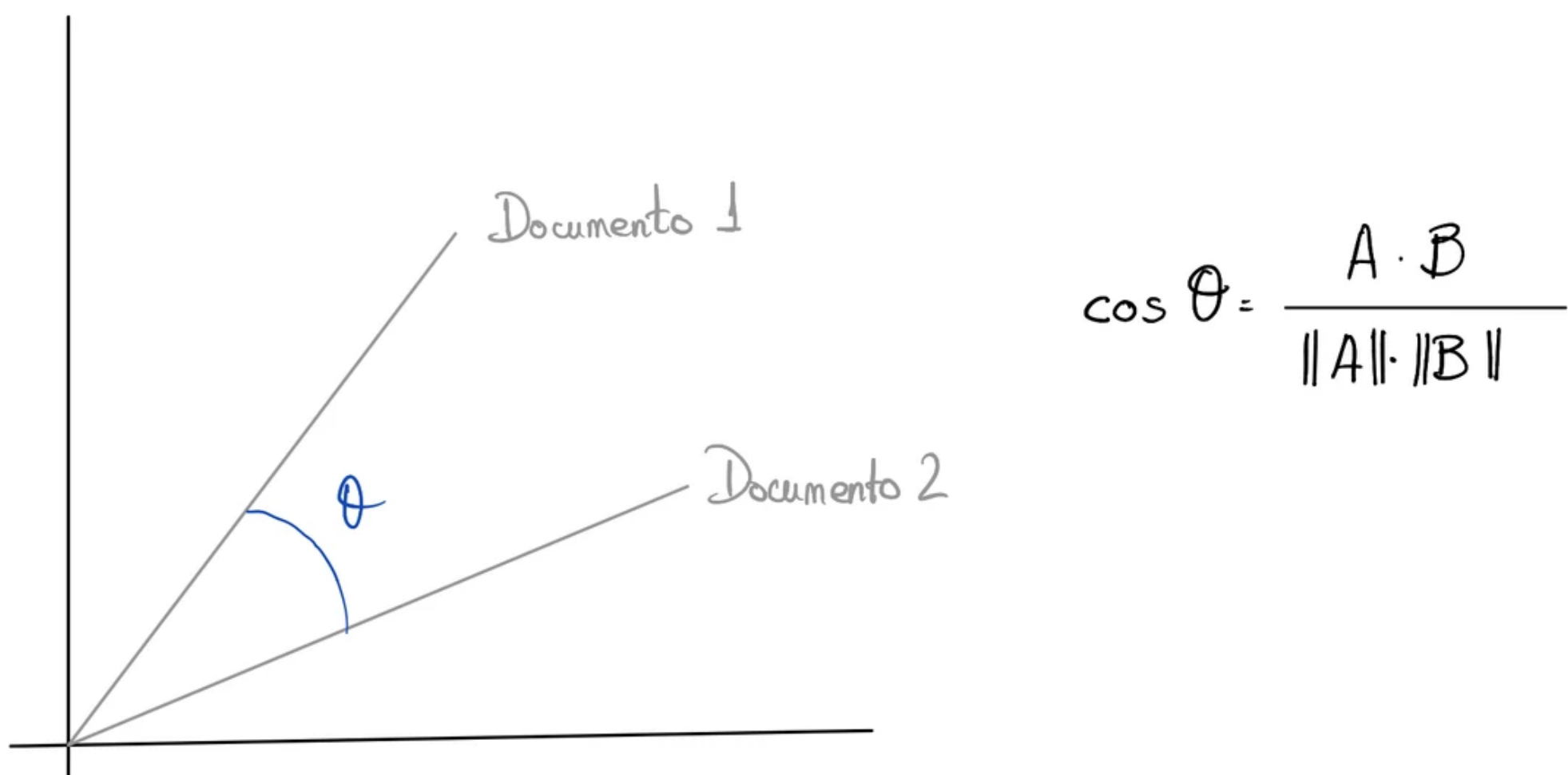
#### Definición de Nuevo Producto

Para la solución de la problemática primero se debe definir un nuevo producto a comparar el cual posee un listado de características que describen al producto. Este producto es la entrada para el modelo, el cual usará los datos previos como entrenamiento para poder priorizar la clientela con base a su compatibilidad con el nuevo producto.

Una vez que este producto se define, se transforman las columnas del producto de forma igual a los datos previos, permitiendo que se puedan comparar entre sí.

#### Similitud Coseno

Una vez que se obtiene el nuevo producto, se obtiene la similitud coseno entre el nuevo producto y todos los productos en la base de datos, con el objetivo de definir numéricamente la similitud entre los productos.



### Compatibilidad

Una vez que se obtiene la similitud coseno, se une con la frecuencia y magnitud de ventas calculadas previamente de forma que se obtienen 3 indicadores, estos se les multiplica un peso personalizado para poder determinar la influencia de cada indicador en la determinación de la compatibilidad del producto.

$$compatibility = \alpha \times SF + \beta \times SP + \gamma \times PS$$

Donde:

- $\alpha = 0,2$  (peso de la frecuencia de ventas),
- $\beta = 0,2$  (peso de la proporción de ventas),
- $\gamma = 0,6$  (peso de la similitud de productos).

### Evaluación

Una vez que se obtiene el listado ordenado de clientes y su compatibilidad con el nuevo producto, es posible extraer la información del cliente y su producto comparado y visualmente determinar la congruencia de los resultados.

### Resultados

Lo que resulta en una lista ordenada de todos los clientes con su mejor producto y el indicador calculado de compatibilidad. Este listado ayuda a asignar una prioridad a los clientes de forma que los clientes más arriba en la lista poseen una mejor compatibilidad con el producto introducido y tendrán mejores resultados de adoptar el producto nuevo en un acercamiento, esto en contraste con los que se encuentran más debajo en el listado. Esto nos ayuda a optimizar la asignación de recursos de marketing enfocado a clientes mejor receptivos.