

# BIG DATA: EL MODELO K-Means Y SU APLICACIÓN EMPRESARIAL

AXEL DOUGLAS HERRERA CARCAMO

Centro Universitario Tecnológico, CEUTEC, Universidad Tecnológica Centroamericana, UNITEC Tegucigalpa, Honduras

## INTRODUCCIÓN

Actualmente, el mercado empresarial de Honduras enfrenta diversos desafíos en la segmentación de clientes, debido a brechas tecnológicas, sociales, académicas, económicas y culturales. Estas dificultades sin duda complican la identificación de segmentos claros y significativos, especialmente en un mercado saturado por pequeñas y medianas empresas distribuidas en toda la economía del país.

## DESARROLLO

A continuación, se presenta un análisis preciso y detallado que permitirá una segmentación efectiva del mercado, facilitando así la toma de decisiones estratégicas en materia de marketing y ventas en el sector retail.

- Método:** Para llevar a cabo el estudio, se emplearon el IDE RStudio y el algoritmo K-MEANS para calcular y asignar los centroides de cada clúster basados en las siguientes variables: volumen de compra, frecuencia de compra, valor monetario o ingreso por transacción, antigüedad del cliente y categoría del producto. Además, se definieron el número de 4 clúster y se aplicó la minimización de la suma de cuadrados de las distancias entre cada punto de datos y su centroide asignado, representado por la fórmula  $\text{Min } \sum_{i=1}^K \sum_{j \in S_i} \|X_j - \mu_i\|^2$ . Posteriormente se contrató un estudio de mercado en la empresa Euromonitor con acceso a una base de datos SQL Server y se construyó un ETL para obtener información de los últimos 3 años para las diferentes categorías del tabaco. Esto se utilizó con la intención de entender el mercado en Honduras y la mentalidad de los clientes con respecto al consumo por categoría de cigarrillos para categorizar cada clúster de los 4 que se definieron para el modelo y de esta forma seleccionar el clúster con mayor oportunidad para ingresar al mercado con una nueva marca.

La intención fue crear segmentos de mercado más efectivos y relevantes para las estrategias de marketing y ventas.

- Recomendaciones:** Dado que el modelo asume linealidad se recomienda utilizar una base de datos robusta. Se debe evitar utilizar datos de texto o imágenes para que el rendimiento del modelo no se vea afectado.
- Financiamiento:** El pago para el uso de la base de datos fue patrocinado por los directores ejecutivos de Grupo Kortabaco de Honduras.

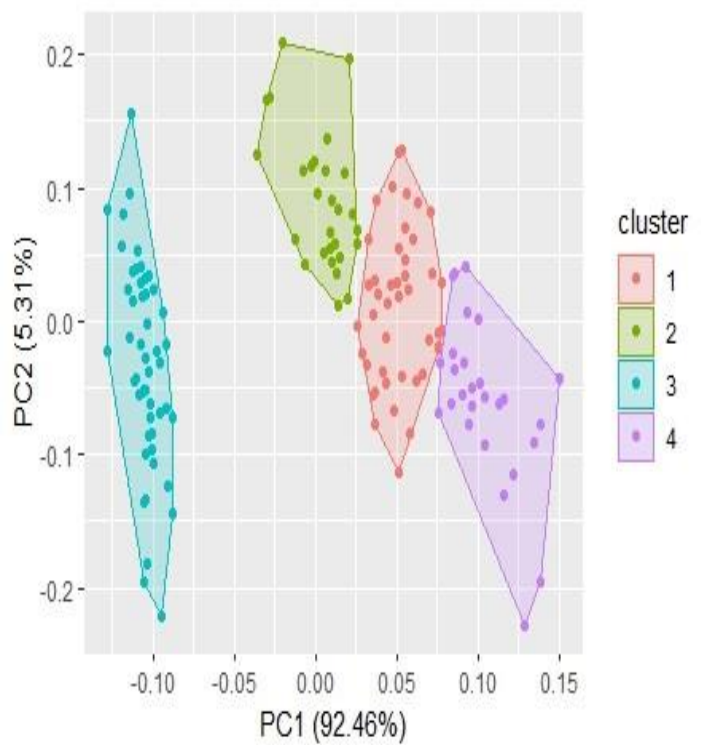


Figura 1. Modelo K-Means Clústeres

## EUROMONITOR INTERNATIONAL

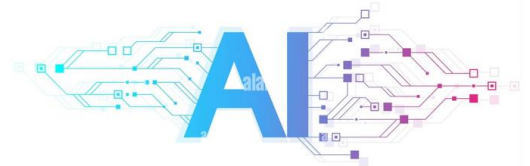
Investigaciones de mercado y análisis de la industria completos.



Corporación Coreana de Tabaco y Ginseng



(IDE) para el lenguaje de programación R



ML Algoritmos no supervisados

## CONCLUSIONES

El algoritmo K-Means permitió identificar de manera efectiva diferentes segmentos de mercado basados en el comportamiento de compra de los clientes (ver figura 1). Por ejemplo, se distinguió entre clientes que realizaban compras frecuentes, pero de bajo valor (Clúster 3), aquellos que hacían compras esporádicas, pero de alto valor (Clúster 4, y otros patrones de comportamiento de compra (Clúster 1 y 2). Además, se encontró que el segmento de la base de la pirámide fue ideal para introducir un nuevo producto en el mercado, ya que se identificó que la lealtad y la recompra son más altas en las regiones rurales de Honduras.

Contacto: [axeldhcode@unitec.edu](mailto:axeldhcode@unitec.edu)

Conflicto de interés: Existe un acuerdo de confidencialidad con Grupo Kortabaco de Honduras.

CONICIETI 2024

I Conferencia Internacional de Ciencia, Tecnología e Innovación