



# Tecnológico de Monterrey

Inteligencia artificial avanzada para la ciencia de datos II (Gpo 101)

Profesor: Félix Ricardo Botello Urrutia

## **Momento de Retroalimentación**

Sofía Cantú Talamantes	A01571120
Ozner Leyva	A01742377
Nallely Serna Rivera	A00833111
Fernanda Perez	A01742102

Noviembre 2024

# ***Reto Entendimiento del Problema***

## ***1. Introducción***

Arca Continental es una de las principales embotelladoras de Coca-Cola en el mundo, y en su constante búsqueda por mejorar la eficiencia y efectividad de sus operaciones, ha identificado la necesidad de desarrollar un indicador que calcule la probabilidad de éxito de nuevos productos en el mercado. La empresa cuenta con grandes volúmenes de datos sobre ventas, comportamiento de clientes, y preferencias de consumo, pero enfrenta el reto de transformar estos datos en información accionable que le permita tomar decisiones estratégicas más informadas.

El problema radica en que, aunque la empresa posee una vasta cantidad de datos históricos, no cuenta con un mecanismo claro que le permita predecir con precisión qué tan exitoso será un producto en función de las características del cliente y sus hábitos de consumo. La ausencia de este indicador dificulta la personalización de las estrategias comerciales, resultando en esfuerzos de marketing menos efectivos y una asignación subóptima de recursos.

El proyecto propone el uso de modelos de machine learning, específicamente el CatBoostClassifier, para calcular la probabilidad de éxito de productos en función de características del cliente, patrones de consumo y datos externos. Esto permite identificar clientes con mayor potencial y mejorar la efectividad de las estrategias de marketing.

## ***2. Objetivos***

Objetivo general: Desarrollar un modelo predictivo basado en IA para identificar productos exitosos, definidos como aquellos consumidos durante 5 meses consecutivos (2 de prueba + 3 regulares).

Objetivos específicos:

- Analizar patrones de ventas y comportamiento de clientes.
- Crear indicadores clave que relacionan tendencias de ventas con datos de consumo.
- Desarrollar un modelo de clasificación que permita identificar clientes con alto potencial para nuevos lanzamientos, en orden de dicho potencial.
- Validar el indicador de éxito desarrollado, evaluando su precisión y eficacia.

## ***3. Antecedentes***

La Inteligencia Artificial (IA) está transformando el escenario corporativo contemporáneo, transformando la forma en que las empresas procesan datos y toman decisiones estratégicas. Su habilidad para examinar grandes cantidades de información en tiempo real no solo mejora las

operaciones, sino que también ofrece una personalización nunca antes vista en la interacción con los clientes.

En el ámbito empresarial, la Inteligencia Artificial funciona como un impulsor de la innovación, favoreciendo la automatización de procesos complejos y la identificación de patrones que antes eran inadvertidos. Se destacan particularmente sus usos en la segmentación de mercado y los sistemas de recomendación, en los que algoritmos sofisticados anticipan conductas de consumo y mejoran la experiencia del usuario, revolucionando áreas como el comercio al por menor y el ocio digital. La Inteligencia Artificial (IA) está revolucionando el escenario empresarial.

No obstante, este avance tecnológico plantea importantes consideraciones éticas. La gestión responsable de datos personales y la transparencia en su utilización se han convertido en pilares fundamentales para mantener la confianza del consumidor. En respuesta, surge la necesidad de establecer marcos regulatorios robustos que equilibren la innovación tecnológica con la protección de la privacidad individual, asegurando un desarrollo sostenible y ético de estas tecnologías.

Para asegurar hacer un uso ético y responsable de la IA en el marketing, están establecidos marcos normativos como la Recomendación sobre la ética de la inteligencia artificial de la UNESCO y el Reglamento General de Protección de Datos (GDPR) en Europa, que buscan respaldar la privacidad de los clientes y promueven la transparencia en el procesamiento de datos personales y sensibles. En tanto a la aplicación de IA en la segmentación de clientes, se emplean técnicas como el clustering, ayudando a las empresas a agrupar a sus clientes de acuerdo a patrones de comportamiento similares. Estudios e investigaciones como el de The Machine Learners (2024), muestran que esta segmentación puede impulsar y mejorar las estrategias de lanzamiento en sectores minoristas, incrementando la precisión y efectividad en campañas de marketing."

#### **4. *Herramientas***

Los recursos utilizados y a utilizar son los siguientes:

- Software y frameworks: Python, bibliotecas como Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, CatBoostClassifier y SciPy.
- Hardware y servicios: Computadora con capacidad suficiente para procesar datos de gran volumen y servicios de almacenamiento local.
- Fuentes de información: Archivos CSV con datos de ventas, productos y clientes. Estos datos deben estar bien estructurados, en formato tabular y con atributos clave como ID de cliente, descripción del producto, ingresos y gastos promedio.

#### **5. *Metodología***

El ciclo de vida de la ciencia de datos, aplicada a este proyecto, fue la siguiente:

- Extracción: Los datos se obtuvieron de archivos CSV estructurados, con variables clave como ventas por cliente, categorías de productos, ingresos promedio y gasto promedio.
- Limpieza: Manejo de valores nulos en atributos críticos y estandarización de datos para garantizar consistencia en las escalas.
- Indicadores: Se calculó un indicador clave basado en la relación entre el porcentaje de ventas de un cliente sobre el total de ventas y la tendencia de crecimiento (pendiente de ventas)
- Transformación: Normalización y estandarización de datos para asegurar una escala consistente.
- Generación de modelos: Implementación del modelo CatBoostClassifier para manejo eficiente de datos categóricos, regularización L1 y L2 para evitar sobreajuste, y balanceo de clases ajustando pesos.
- Evaluación: Se trabajó con métricas para poder medir el éxito del modelo.
- Análisis de resultados: Se interpretaron los resultados para llegar a conclusiones sólidas.

## 6. *Resultados*

### **Segmentación de Clientes**

Se obtuvieron tres grupos de clientes con características distintas:

- Cluster 0: Clientes con ingresos medios y gastos relativamente bajos.
- Cluster 1: Clientes de alto ingreso con gastos elevados.
- Cluster 2: Clientes de ingreso medio-alto y gasto moderado.

Los porcentajes de gasto sobre ingreso fueron del 69.7%, 65.8% y 67.8% para los clusters 0, 1 y 2, respectivamente. A pesar de que esto indica variaciones notables en los patrones de consumo de cada grupo, también resulta fascinante que ninguno de los tres clusters se distancie más del 4% entre sí.

### **Desempeño del Modelo**

El modelo CatBoostClassifier mostró resultados prometedores:

- Recall: 70% para la clase de éxito, indicando una buena capacidad para identificar casos exitosos
- F1-score del Cluster 0 del 0.77 (alto rendimiento) y Cluster 1 del 0.09 (necesita mejora)
- Accuracy general: 63%

Adicionalmente, se hizo un caso de prueba con el lanzamiento de Monster Energy. Se le dió al modelo la siguiente información:

- Marca: Monster Energy

- Tamaño: Individual
- Tipo de Producto: Bebidas Energéticas
- Categoría Global: Bebidas Emergentes
- Sabor Global: Mixto
- Categoría de Producto: Categorías en Expansión

En la siguiente tabla, se puede apreciar como hay una clara diferencia entre los primeros puestos de la lista y los últimos que tenemos en el top 50, demostrando la eficiencia del modelo.

Número en el Top	ID de la tienda	Probabilidad de éxito
1	499930861	0.471995
2	499968994	0.443180
3	500147703	0.439536
4	500009869	0.436760
5	500118167	0.429275
46	500010092	0.296036
47	500120816	0.292944
48	500196332	0.292756
49	500270761	0.283161
50	500038823	0.273766

**Tabla 1.** Resultados del caso de prueba con el lanzamiento de Monster Energy.

De ahí se calculó la precisión de la predicción, en donde se obtuvo:

- Se predijo correctamente el cliente top 1
- 4 de 5 clientes en el top 5
- 9 de 10 clientes en el top 10
- 16 de 20 clientes en el top 20
- 41 de 50 clientes evaluados

Cabe considerar que el 94% de los clientes predichos mantuvieron su actividad hasta diciembre de 2022.

## **7. Conclusiones y trabajo a futuro**

El proyecto de desarrollo de un modelo predictivo para Arca Continental representa un avance significativo en la aplicación de inteligencia artificial para la toma de decisiones estratégicas en marketing. El modelo CatBoostClassifier demostró una capacidad notable para identificar clientes con potencial de éxito en nuevos lanzamientos, con un recall del 70% que permite detectar la mayoría de los casos exitosos. Esta herramienta ofrece a la empresa una perspectiva innovadora para priorizar y optimizar sus estrategias de comercialización.

Sin embargo, el modelo actual presenta desafíos importantes que requieren atención. La baja precisión, cercana al 5%, genera un número significativo de falsos positivos, lo que podría resultar en inversiones de marketing ineficientes. Esta limitación subraya la necesidad de un refinamiento continuo del modelo, mediante la recolección de más datos, el ajuste de hiper parámetros y la exploración de técnicas de machine learning más avanzadas.

El análisis de clustering reveló insights valiosos sobre el comportamiento de los clientes, identificando tres segmentos con patrones de gasto sorprendentemente similares. Esta información, combinada con la capacidad predictiva del modelo, posiciona a Arca Continental en una posición privilegiada para personalizar sus estrategias de marketing, optimizar la asignación de recursos y mejorar significativamente la efectividad de los lanzamientos de nuevos productos.

A futuro, el proyecto plantea un camino prometedor para la integración de inteligencia artificial en la toma de decisiones empresariales. La metodología desarrollada no solo ofrece una herramienta de predicción, sino que establece un marco para un análisis más profundo y estratégico del mercado, equilibrando la innovación tecnológica con consideraciones éticas de privacidad y uso responsable de datos.

## *Referencias*

- Peer, D. (2024, August 21). Implementación de la agrupación K-means para la segmentación de clientes. Retrieved November 2, 2024, from peerdh.com website: <https://peerdh.com/es/blogs/programming-insights/implementing-k-means-clustering-for-customer-segmentation-1>
- Sanz, F. (2020, November 26). Algoritmo K-Means - Clustering y cómo funciona. Retrieved November 2, 2024, from The Machine Learners website: <https://www.themachinelearners.com/k-means/>
- Skiller Education. (2024a, March 11). *IA Para Análisis De Big Data: Avances Y Aplicaciones - Skiller Academy*. Skiller Academy. <https://skiller.education/ia-para-analisis-de-grandes-volumenes-de-datos/>
- Skiller Education. (2024b, March 11). *Impacto de la IA en la toma de decisiones: Datos y eficiencia - skiller academy*. Skiller Academy. <https://skiller.education/ia-y-su-impacto-en-la-toma-de-decisiones-basada-en-datos/>
- Skiller Education. (2024c, March 12). *Ética y privacidad en IA: Claves para el análisis de datos - skiller academy*. Skiller.education. <https://skiller.education/etica-y-privacidad-en-ia-para-analisis-de-datos/>
- Torrubi, A. (2024, February 6). *Inteligencia artificial: Desde 2019 hasta 2024 y más allá*. Fundación Innovación Bankinter. [https://www.fundacionbankinter.org/noticias/inteligencia-artificial/?\\_adin=02100180067](https://www.fundacionbankinter.org/noticias/inteligencia-artificial/?_adin=02100180067)
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. Unesco.org. [https://unesdoc.unesco.org/ark:/48223/pf0000380455\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa). Document code: SHS/BIO/REC-AIETHICS/2021.
- UNESCO. (2024). Retrieved November 8, 2024, from Unesco.org website: [https://unesdoc.unesco.org/ark:/48223/pf0000380455\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000380455_spa)