



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 propuesto se enfoca en el estudio de datos de ventas y clientes con el fin de detectar patrones en la conducta del consumidor y dividir a los clientes en grupos relevantes a través de la aplicación de algoritmos de aprendizaje no supervisado, especialmente el método de agrupación K-Means. Esta segmentación nos permite brindarle a Arca Continental la posibilidad de personalizar estrategias de marketing, optimizar recursos y entender de manera más profunda las necesidades de sus clientes.

2. Objetivos

Objetivo general: Identificar clientes con alto potencial de éxito para el lanzamiento de un nuevo producto mediante la segmentación basada en IA. Para lograr alcanzar dicho objetivo, se analizarán patrones de comportamiento y características de los clientes con el uso de algoritmos de clustering no supervisado, específicamente K-Means, con la finalidad de agrupar a los clientes en segmentos homogéneos que faciliten la selección de los grupos más adecuados para una estrategia de lanzamiento de producto.

Objetivos específicos:

- Generar grupos homogéneos de clientes que compartan patrones de comportamiento similares a través del análisis de sus datos históricos de compra.
- Seleccionar los segmentos de clientes más prometedores para el lanzamiento del nuevo producto, priorizando aquellos con mayor probabilidad de adopción exitosa.
- Analizar la rentabilidad potencial de cada segmento, determinando su capacidad para maximizar los ingresos.

- Identificar patrones de consumo en cada segmento.
- Validar el indicador de éxito desarrollado, evaluando su precisión y eficacia utilizando los datos proporcionados por los socios comerciales de Arca Continental.

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 K-Means**, 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 y scikit-learn.

- 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: Importación de los datos desde archivos CSV utilizando la biblioteca pandas.
- Limpieza: Eliminación de columnas irrelevantes y manejo de valores nulos en atributos clave como *ingreso_promedio_300m* y *gasto_promedio_300m*.
- Transformación: Normalización y estandarización de datos para asegurar una escala consistente.
- Generación de modelos: Implementación del algoritmo K-Means para segmentar los datos en tres clusters, basado en el análisis del método del codo.
- Evaluación: Cálculo de la puntuación de silueta para medir la coherencia de los clusters formados.
- Análisis de resultados: Agrupación de los clientes y análisis de la proporción de gasto respecto al ingreso en cada grupo.

6. *Resultados*

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.

7. *Conclusiones y trabajo a futuro*

El algoritmo K-Means, utilizado para segmentar a los clientes, posibilitó la identificación de tres grupos diferentes basados en sus patrones de ingreso y gasto. Esta categorización proporciona datos

de gran valor para elaborar estrategias de negocio a medida que se ajusten a las particularidades de cada segmento. Comprendiendo de manera más profunda el comportamiento de los clientes, se consiguen maximizar los recursos y centrar esfuerzos y atención en los grupos con mayor posibilidad de éxito para la introducción de un producto nuevo. Resultado que ayuda a incrementar la eficacia en la toma de decisiones comerciales y estratégicas. La segmentación de clientes a través del algoritmo K-Means posibilitó la identificación de tres grupos diferentes basados en sus patrones de ingreso y gasto. Esta clasificación proporciona datos de gran relevancia para el diseño de diseños.

Para entregas futuras, buscamos contar con los siguientes avances:

- Integrar más variables, como datos demográficos y preferencias de compra:
- Aplicar técnicas de clustering jerárquico para comparar resultados.
- Realizar un análisis de sensibilidad para validar la estabilidad de los clusters formados.
- Probar Modelos Predictivos de Éxito de Producto

8. 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
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. Unesco.org. 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