Propuesta de Proyecto

Seminario de Analítica y Ciencia de Datos

# Presentación del Proyecto

Este proyecto utiliza la implementación de métodos de clustering para el perfilamiento de clientes, utilizando técnicas avanzadas de análisis de datos para segmentar la base de clientes de la empresa Los Olivos, en grupos homogéneos basados en comportamientos y características clave. A través de algoritmos de machine learning, identificamos patrones que nos permiten generar perfiles, permitiendo el desarrollo de estrategias de marketing personalizadas y mejorando la experiencia del cliente. Esta herramienta ofrece una visión profunda de las preferencias y necesidades de los clientes, impulsando la toma de decisiones informadas y maximizando el retorno de inversión en las campañas comerciales.

# Objetivo del Proyecto

Desarrollar un modelo basado en algoritmos de machine learning, para la identificación de patrones de comportamiento de clientes de la UEN-Previsión de la empresa Los Olivos; contribuyendo a la formulación de acciones para la fidelización y rentabilidad de los productos ofrecidos.

# Contexto del Problema

La UEN-Previsión de la empresa Los Olivos, presenta características similares a las de una compañía de seguros; dado que ofrece productos en forma de pólizas. Por eso, es fundamental la observación y medición de los clientes de forma que se puedan considerar todos los diferentes factores, que dado el caso, representan o no un riesgo para el negocio. [1] Es así, que cómo procedimiento de observación y medición, se monitorea la cantidad de clientes-persona que contratan póliza.

En primera instancia, es pertinente definir los productos que hacen parte del portafolio disponible, dando la claridad que, debido al modelo de negocio existente, hay un portafolio diversificado que responde a las necesidades de los diferentes segmentos de venta, siendo los más rentable del negocio el plan familiar integral y el club de amigos. Estos segmentos de venta donde se comercializan son: el sector solidario (cooperativas), convenios con empresas del sector real, fincas productoras y particulares (personas).

Precisada la importancia de la medición de los clientes, cómo lo menciona Zurawski [2], el monitoreo se ha convertido en la herramienta utilizada para ese seguimiento y la comprensión de los comportamientos de los consumidores. De ahí que analizando los registros de la medición con respecto al crecimiento cliente-persona de la empresa, se identificó una tendencia estática o invariable. Es decir, la cantidad de clientes que no continúan con la póliza versus los que se vinculan como nuevos titulares de póliza, tienen efecto cero en términos de crecimiento del negocio.

Otro aspecto a considerar cómo parte de la problemática, son las fuentes de captación de clientes para la UEN-Previsión. Y es que, el principal ingreso de ventas-clientes, es a través de venta cruzada con el sector solidario durante la colocación de créditos. En el sector real y agrícola, cómo convenio de protección de las empresas y las fincas. Por último, las personas particulares que conocen la marca Los Olivos y se les ofertan los otros productos de la línea de previsión.

Con base en lo anterior, se determina que el desafío que representa mayor importancia para este contexto es la retención de los clientes actuales. Debido a que la evidencia a través de los datos es una muy baja tasa de fidelización o en el recaudo del pago de las primas de productos de la UEN-Previsión.

En conjunto, los factores mencionados tienen una implicación sobre el crecimiento y la rentabilidad de la UEN-Previsión. De ahí que se podrían materializar riesgos operativos y representar un decrecimiento en la rentabilidad del negocio en el futuro.

Por estas consideraciones, se hace necesaria la implementación de métodos analíticos para la identificación del perfil de cliente fidelizado, realizar una segmentación o agrupación de los diferentes tipos de clientes y a partir de ello brindar herramientas que permitan la estructuración de estrategias para la disminución de “retiros” o “cancelaciones” a través de la profundización del portafolio y la permanencia en el mismo.

# Propuesta de Solución y Alcance

La ciencia de datos ofrece soluciones valiosas para la optimización de recursos y la toma de decisiones estratégicas en diversos escenarios de negocio. En el contexto específico de la fidelización de clientes para la compra de pólizas, se recomienda la implementación de la metodología CRISP-DM [3]. Esta metodología se basa en un entendimiento profundo del negocio y los datos para desarrollar soluciones ajustadas a las necesidades específicas [4].

Un componente esencial para el éxito en la aplicación de CRISP-DM es la preparación adecuada de los datos. Esto incluye la recopilación y organización coherente de la información, que servirá como base para la creación y entrenamiento de modelos predictivos [5]. La evaluación final de estos modelos debe determinar si efectivamente resuelven las preguntas planteadas desde la perspectiva del negocio [6].

Para implementar CRISP-DM de manera efectiva, es fundamental utilizar herramientas y técnicas avanzadas de análisis de datos. Entre las herramientas de clustering más relevantes se encuentran K-means, el Algoritmo de Agrupamiento Jerárquico y DBSCAN. Estas técnicas permiten la segmentación de clientes en grupos homogéneos basados en características comunes, facilitando la personalización de la oferta y mejorando la efectividad de las estrategias de marketing [7][8]. K-means, por ejemplo, es ampliamente utilizado por su simplicidad y eficacia en clustering basado en centroides [9]. Por otro lado, DBSCAN es útil para identificar clusters de forma arbitraria y manejar el ruido en los datos [10].

Para que las metodologías como CRISP-DM generen valor real en el análisis de problemas de negocio, es crucial utilizar herramientas tecnológicas y técnicas estadísticas adecuadas. En el caso de Los Olivos, esto implica identificar las características clave de los clientes mediante el análisis de datos. Se deben evaluar las variables presentes en los datos, particularmente aquellas que influyen en el momento decisivo de compra de las pólizas [11]. Además, es importante considerar los canales a través de los cuales se ofrece la oferta de productos [12].

Una vez identificadas las características significativas de los clientes y su relación con las decisiones de compra, se debe proceder al entrenamiento y prueba de la influencia de cada característica sobre el comportamiento de los clientes. Este análisis permitirá desarrollar un modelo que segmente a los clientes en grupos con perfiles similares, utilizando herramientas de clustering para optimizar la segmentación. Esta segmentación permitirá un enfoque más efectivo en la oferta de productos, minimizando el riesgo de sesgo en la oferta de productos [13].

Finalmente, al segmentar eficazmente a los clientes en grupos con perfiles homogéneos, se podrán ajustar los modelos para mejorar la precisión en el momento de la venta, teniendo en cuenta las características descriptivas de cada grupo. Esto contribuirá a una oferta de productos más alineada con las necesidades de los clientes y a una estrategia de marketing más efectiva [14].

# Definición del Alcance

En función de los análisis propuestos y las metodologías establecidas para el tratamiento de los datos, se busca proporcionar herramientas basadas en los resultados y conclusiones obtenidos, que sean relevantes para el crecimiento del negocio. El enfoque está en profundizar el portafolio según los diferentes perfiles de clientes identificados, con el fin de garantizar la sostenibilidad y rentabilidad a largo plazo.

A partir de los resultados del análisis, se espera desarrollar un modelo que destaque los aspectos clave para definir el perfil óptimo del cliente, así como identificar patrones de comportamiento que faciliten la expansión del portafolio y aumenten la rentabilidad.

Un aspecto adicional importante es la agrupación de perfiles de clientes con características similares. No obstante, dentro de los límites del proyecto, no se incluye el diseño o la implementación de estrategias derivadas de los conocimientos obtenidos. El objetivo del análisis es proporcionar las herramientas necesarias para esbozar soluciones, pero no para desarrollarlas o implementarlas directamente.

# Riesgos e Impacto del Negocio

La solución planteada tiene el potencial de mejorar la tasa de conversión y la fidelización de los clientes, optimizando los recursos de marketing, fortaleciendo las alianzas estratégicas. facilitando la toma de decisiones más precisas y fundamentadas.

Un ejemplo destacado es Amazon, que, mediante su algoritmo de recomendación, selecciona un conjunto reducido de productos que podrían interesar a cada cliente en función de su contexto y comportamiento [15]. Este algoritmo, iniciado en los años 90, ha evolucionado constantemente con el mercado, permitiéndole a la compañía no solo mantener la fidelidad de sus clientes mediante una experiencia de compra personalizada, sino también atraer a nuevos usuarios, impulsando su crecimiento y consolidando su liderazgo en el comercio electrónico.

Sin embargo, es crucial considerar algunos riesgos asociados a este proyecto. Uno de los principales es la calidad de los datos disponibles. Actualmente, Los Olivos utiliza una base de datos transaccional donde los registros se sobrescriben continuamente, lo que limita el acceso a información histórica completa. Además, trabajar con archivos planos descargados implica el riesgo de sesgos en los datos debido al filtrado, lo que podría comprometer la representatividad de los análisis.

Por último, los cambios en el entorno del mercado, como la evolución en el comportamiento del consumidor o la aparición de nuevos competidores, pueden reducir la eficacia y utilidad de la solución a largo plazo si no se contempla un ajuste continuo en función de estas nuevas condiciones.

Bibliografía

[1] Montserrat Guillén, Jens Perch Nielsen y Ana M. Pérez-Marín. (2008). La necesidad de monitorizar la fidelidad de los clientes y el riesgo empresarial en el sector asegurador europeo. *Los Documentos de Ginebra sobre riesgos y seguros. Problemas y práctica*, 33(2), 207-218. Springer Nature.

[2] Zurawski, Nils. (2011). Consumer Surveillance. *Surveillance & Society*, vol. 8, no. 4, pp. 472-486. JSTOR.

[3] J. W. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Elsevier, 2012.

[4] L. C. Jain, *Data Mining: A Computational Intelligence Approach*, Springer, 2016.

[5] R. S. T. M. Jayaraman, "Data preprocessing and preparation," in *Handbook of Statistical Analysis and Data Mining Applications*, S. M. U. S. S. (Eds.), Academic Press, 2013, pp. 215-244.

[6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Morgan Kaufmann, 2016.

[7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer, 1981.

[8] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226-231.

[9] T. J. D. M. Hartigan, "Clustering Algorithms," Wiley, 1975.

[10] M. M. Rahman and A. M. T. Khan, "Implementing CRISP-DM methodology for data mining project: A case study on loan prediction," *Journal of Computer and Communications*, vol. 9, no. 4, pp. 19-28, 2021.

[11] Y. Chen, Y. Xu, and X. Yang, "Predicting customer purchase behavior with machine learning techniques," *IEEE Access*, vol. 9, pp. 84528-84540, 2021.

[12] J. M. P. A. B. O. M. Johnson, "Channel analysis and customer segmentation in marketing," *Journal of Marketing Research*, vol. 53, no. 2, pp. 151-164, 2016.

[13] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2018.

[14] S. B. K. M. J. Smith, "Effective data-driven marketing strategies using customer segmentation," *Journal of Business Analytics*, vol. 5, no. 3, pp. 225-238, 2020.

[15] Smith, B., & Linden, G. (2017). Two decades of recommender systems at Amazon.com. *IEEE*