Segmentación de Clientes para Créditos de Vivienda

Seminario de Analítica y Ciencia de Datos

# Resumen Descriptivo del Proyecto

Este proyecto tiene como objetivo desarrollar un modelo de Machine Learning de clasificación no supervisado con el fin de segmentar la base de clientes de una compañía dedicada a facilitar el financiamiento para la adquisición de vivienda a colombianos residentes en el exterior. Al concluir este proceso de segmentación, la empresa estará en la capacidad de adaptar y personalizar sus productos, servicios y estrategias de marketing para satisfacer de forma más precisa las necesidades particulares de cada segmento identificado.

El proyecto se llevará a cabo siguiendo la metodología CRISP-DM y se empleará un algoritmo de clustering, técnica reconocida por su eficacia en la resolución de este tipo de problemas. Además, se realizará un análisis y comparación de distintos algoritmos de agrupamiento para determinar cuál es el más adecuado según las necesidades y características específicas del negocio. Se espera que esta segmentación ayude a la empresa objeto de estudio a potenciar su capacidad de innovación y a impulsar un crecimiento sostenible a largo plazo.

Basado en el Data Product Canvas1:

## Problema.

* 1. ¿Cuál es mi problema alrededor de mi proyecto?

El mercado de créditos de vivienda presenta ciertas particularidades y desafíos que lo hacen complejo para las empresas que se dedican a otorgar créditos de este tipo. Para destacarse en este entorno competitivo y dinámico, estas empresas deben adoptar un enfoque estratégico, innovador y orientado al cliente. Lamentablemente, la empresa objeto de este proyecto aún no ha logrado comprender completamente las necesidades, comportamientos y características de sus clientes.

Una de las metas prioritarias establecidas para 2024 fue la mejora continua de los procesos internos, la optimización de la experiencia del cliente y el perfeccionamiento de las estrategias de marketing, sin descuidar, por supuesto, la maximización de la rentabilidad. Sin embargo, la falta de comprensión sobre sus clientes ha llevado a la empresa a invertir significativas sumas de dinero en campañas publicitarias que no generan el retorno esperado. Indiscutiblemente en un mercado donde el ciclo de compra puede extenderse durante años, es fundamental garantizar la satisfacción del cliente, no solo para asegurar transacciones exitosas sino también para fomentar recomendaciones y lealtad hacia la marca. Actualmente, la empresa aún tiene trabajo por hacer en este aspecto, ya que sus índices de fidelización y retención no alcanzan los promedios de la industria.

* 1. ¿Por qué se manifiesta el problema?

La adquisición de un crédito de vivienda se trata de un proceso de larga duración, un proceso donde el ciclo de compra suele durar años y donde la mayoría de trámites son complejos. Esta naturaleza del ciclo de compra distingue a este mercado, ya que los clientes raramente vuelven a solicitar otro crédito de vivienda en un corto periodo de tiempo.

Además, el mercado inmobiliario está influenciado por factores macroeconómicos que pueden influir significativamente en la demanda de créditos hipotecarios. Por ejemplo, las fluctuaciones en las tasas de interés pueden afectar la capacidad de los clientes para acceder al financiamiento, así como la disposición de subsidios por parte de entes gubernamentales.

Captar clientes para las empresas que se dedican a otorgar créditos de vivienda puede ser todo un desafío, pero además retenerlos y fidelizarlos añade una complejidad adicional. Por esto, es crucial entender a fondo a los clientes: sus necesidades, características y patrones de compra. Este conocimiento profundo permite a las empresas optimizar sus procesos y diseñar estrategias efectivas no solo para retener a los clientes existentes, sino también para atraer a nuevos prospectos de manera eficiente.

* 1. ¿De quién es el problema?

Aunque indiscutiblemente, como lo hemos observado, el mercado inmobiliario presenta características propias que añaden complejidad, el problema recae en la empresa misma. Es esencial que la empresa se dedique activamente a entender a sus clientes. En este contexto, el departamento de Customer Success y el equipo de Marketing deben liderar estas iniciativas.

Estos departamentos tienen el papel crucial de desarrollar estrategias que optimicen los procesos de marketing y perfeccionen las etapas involucradas en la adquisición de un crédito hipotecario. Deben centrarse en comprender las necesidades de los clientes, ofrecer soluciones adecuadas y mejorar continuamente la experiencia del cliente. Al hacerlo, no solo se mejorará la satisfacción, sino que también se fomentará la recomendación positiva, lo que puede atraer a nuevos prospectos de manera orgánica.

## Datos.

* 1. Describa la fuente de datos (Sea más específico que en documento pasado)

Viventa LLC posee la información de sus clientes potenciales y efectivos en los servicios de administración y almacenamiento de información prestados por Zoho Corporation. Los datos a tratar provienen de fases iniciales del ciclo de vida de los clientes en la empresa, es decir, la información proviene del historial de clientes que han ingresado a la compañía.

La base de datos contiene información detallada sobre los clientes, incluyendo campos como edad, ciudad de residencia, score de crédito, ocupación, entre otros. Esta información se obtiene en las fases iniciales del ciclo de vida del cliente en la empresa, porque al inicio del proceso, después de que el departamento de Customer Success hace el contacto inicial con el cliente y le brinda una presentación comercial explicando todo el proceso, le envia un link para que el cliente diligencie la encuesta donde plasma su información inicialmente y continúa el proceso.

* 1. ¿Cuál es la calidad que posee dicha fuente?

La calidad de la información suministrada es regular según el análisis y necesidad de la misma. El proyecto pretende tener el resultado de la clusterización lo más próximo a la obtención del diligenciamiento de la encuesta para tener la clusterización del cliente rápidamente. En este contexto, la base de datos contiene sesgos en cuanto a las cifras exactas informadas por los clientes, dado que no existen procesos estandarizados del cómo diligenciar los campos y la base exige un preprocesamiento adecuado de los datos suministrados.

* 1. ¿Es accesible y disponible?

La información es accesible y disponible ya que se cuenta con acceso a ella en tiempo real, gracias a que la administración de la relación con los clientes y las fases clave para el proceso se realizan a través de Zoho CRM, plataforma donde los empleados dejan visible su gestión e información de todos los clientes.

## Hipótesis.

* 1. ¿Qué es lo que pretende probar?

Segmentar la base de clientes puede ser una herramienta que ayude a la empresa a comprender mejor las necesidades, comportamientos y características de sus usuarios. Con una segmentación precisa, la empresa puede diseñar estrategias de marketing más efectivas al identificar los canales de comunicación adecuados y crear mensajes personalizados dirigidos a cada segmento de clientes. Además, al comprender mejor a sus clientes, la empresa puede mejorar la satisfacción, fortalecer la fidelización y fomentar recomendaciones positivas hacia otros potenciales clientes.

* 1. ¿Cuáles pueden ser las respuestas esperadas a lo que pretende probar?

1. La segmentación es una herramienta clave que mejora la experiencia del cliente y optimiza las estrategias del departamento de Customer Success y el equipo de Marketing.

2. La segmentación de clientes no genera respuestas positivas para mejorar la experiencia del cliente, ni facilita el diseño de estrategias de marketing efectivas y personalizadas.

* 1. ¿Qué acciones debo tomar para cada respuesta anterior?

1. En una evaluación posterior, validar las características de cada cluster, comportamientos similares y resultado de sus procesos, esto permitirá a la empresa crear perfiles de clientes más precisos, personalizar las recomendaciones de productos, establecer frecuencias de contacto y demás estrategias que el departamento de Customer Success y el equipo de Marketing establezcan. Todo esto siempre enfocado a la mejora continua del servicio prestado. Al obtener estas respuestas a través del modelo de clusterización, la empresa podrá tomar decisiones más informadas, ofreciendo un valor añadido significativo tanto para clientes como para socios bancarios.

2. Evaluar los posibles inconvenientes y fallos en la aplicación del proyecto para proponer a la empresa procesos de mejora en metodologías de recolección de información, gestión de herramientas ofimáticas u otra estrategia según sea el caso.

## Solución.

* 1. Tipo. (¿Qué tipo de algoritmo de ML es? ¿Qué técnica es?
* Modelo de aprendizaje automático no supervisado.
* Técnica: Clustering
* Se busca probar con diferentes técnicas de clustering para determinar el modelo que ofrezca los mejores resultados. Se considerarán métodos como K-means, DBSCAN, Hierarchical Clustering, Spectral Clustering, entre otros.
  1. ¿Cómo debe ser implementado?

Posterior al desarrollo del modelo a través de la metodología CRISP-DM, este será implementado en un equipo local con el cargue en lote de información de clientes nuevos y con una frecuencia determinada, acorde a la personalización de estrategias que cada departamento determine.

* 1. ¿Qué resultados se esperan?

El modelo identificará diferentes segmentos de clientes con características similares en términos de comportamiento, necesidades financieras, perfil demográfico, etc. Cada uno de estos segmentos identificados tendrá un perfil único que le permitirá a la empresa entender mejor las características y necesidades específicas de ese grupo. Esto facilitará la personalización de servicios para cada segmento y optimizará las estrategias de marketing.

## KPI’s

* 1. ¿Cómo espera evaluar el modelo?

Dado que se trabajará con datos no etiquetados, se utilizarán medidas que evalúen los resultados de la agrupación sin depender de etiquetas previas. Dentro de estas técnicas de validación interna se incluyen la cohesión dentro de cada cluster y la separación entre los distintos clusters. Además, se llevará a cabo un análisis visual de los clústers utilizando técnicas como el Análisis de Componentes Principales (PCA) para visualizar la distribución de los datos.

* 1. ¿Qué métricas debe usar?

Para evaluar la calidad y eficacia de los clusters generados, se utilizarán las siguientes métricas:

* **Coeficiente de Silueta:** es una métrica utilizada para evaluar la calidad de las agrupaciones en un análisis de clustering, determina tanto la cohesión dentro de los clusters como la separación entre ellos. Un valor cercano a 0 sugiere que los objetos podrían estar en el límite entre dos clusters, y un valor cercano a -1 indica que los objetos están probablemente en el cluster incorrecto (Portland State University, n.d.)
* **Índice Calinski Harabasz:** evalúa la relación entre la dispersión inter-cluster y la dispersión intra-cluster para determinar la calidad de la agrupación. Es una métrica con la que se suele evaluar el grado de agrupación de un conjunto de datos. Indicando una mejor agrupación cuando mayor sea el valor del índice (Rodríguez, 2023).
* **Método de Elbow (Codo):** Es una técnica utilizada para determinar el número óptimo de clusters en los que dividir un conjunto de datos.

## Actores.

* 1. ¿Quién es su cliente?

Son ciudadanos colombianos que residen en el exterior y desean adquirir viviendas en Colombia a través de crédito hipotecario sin abandonar su actual país de residencia, beneficiándose de la tasa de cambio entre monedas y generando una oportunidad de vivienda para sus familias en Colombia o una inversión de su país de origen.

* 1. ¿Quiénes son los interesados (Stakeholders)?
* **Viventa LLC:** La empresa que se beneficiaría directamente de la segmentación, adaptando y personalizando sus servicios de manera eficiente.
* **Departamento de Marketing y Customer Success:** Los profesionales de estos departamentos emplearán los resultados obtenidos del modelo de Machine Learning para diseñar estrategias más efectivas y personalizadas. Además, optimizarán los procesos dirigidos a los clientes, mejorando su experiencia.
* **Departamento de Datos:** Encargado de implementar, mantener y perfeccionar el modelo de Machine Learning.
* **Clientes:** Experimentarán directamente las mejoras en los servicios, conduciendo a una experiencia más satisfactoria y adaptada a sus necesidades.
  1. ¿Quién usará la solución?

Los directores de la empresa serán los principales usuarios de la solución, encargados de analizar los resultados obtenidos para tomar decisiones estratégicas basadas en los hallazgos. De igual forma, los departamentos involucrados en las mejoras que se pretenden lograr con la segmentación, como el departamento de Marketing y el de Customer Success, quienes desarrollarán e implementarán estrategias específicas para mejorar los procesos internos de sus respectivas áreas.

* 1. ¿A quién impactará?

El desarrollo de este modelo de machine learning para la clusterización de clientes tendrá un impacto significativo en varias partes interesadas, incluyendo instituciones financieras aliadas y clientes, por ejemplo:

* **La empresa:** Viventa LLC mejorará sus procesos de gestión crediticia y toma de decisiones al aprovechar la clasificación temprana de clientes.
* **Instituciones Financieras aliadas:** Pueden mejorar el acceso al crédito, aumentar la precisión en las evaluaciones de clientes enviados desde/hacia la empresa y mejorar el rendimiento operativo. Esto puede conducir a un mayor volumen de préstamos, un mejor manejo del riesgo y una mayor satisfacción del cliente.
* **Clientes:** Obteniendo un perfilamiento temprano y diagnóstico de acceso al crédito, logrando corregir tempranamente comportamientos inadecuados como historial crediticio limitado o fuentes de ingresos no convencionales.

**References**

Portland State University. (n.d.). *Machine Learning - 12 Unsupervised Learning*. Portland State University. https://web.pdx.edu/~gerbing/Books/ML/12-cluster.html

Rodríguez, D. (2023, June 16). *Identificar el número de clústeres con Calinski-Harabasz en k-means e implementación en Python*. Analytics Lane. Retrieved April 25, 2024, from https://www.analyticslane.com/2023/06/16/identificar-el-numero-de-clusteres-con-calinski-harabasz-en-k-means-e-implementacion-en-python/



1

[https://medium.com/@leandroscarvalho/data-product-canvas-a-practical-framework-for-building-high-](https://medium.com/%40leandroscarvalho/data-product-canvas-a-practical-framework-for-building-high-) performance-data-products-7a1717f