Propuesta de Proyecto

Seminario de Analítica y Ciencia de Datos

# Presentación del Proyecto

**‘Segmentación de Clientes para Créditos de Vivienda’**

El proyecto tiene como objetivo principal la segmentación de la base de clientes de una compañía dedicada a facilitar el financiamiento para la adquisición de vivienda a colombianos residentes en el exterior. Se busca enriquecer la experiencia de los clientes mediante su clasificación en grupos específicos y homogéneos, basándose en criterios esenciales para el negocio, tales como edad, género, nivel de ingresos, localización geográfica, intereses, entre otros. 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. Esta estrategia no solo permitirá ajustar la experiencia y los puntos de contacto para cada grupo de clientes, sino que también permitirá la identificación de aquellos grupos que presentan un mejor o peor rendimiento. De esta forma, el proyecto buscará destacar los grupos de clientes más lucrativos, optimizando el enfoque y la eficiencia de la compañía.

Objetivo del Proyecto

Desarrollar un modelo de Machine Learning de clasificación no supervisado con el fin de segmentar la base de clientes de una empresa especializada en otorgar créditos para la adquisición de vivienda, dirigida específicamente a ciudadanos colombianos que residen en el exterior y requieren un proceso de desembolso adaptado a sus necesidades.

Contexto del Problema

Viventa LLC es un bróker de crédito hipotecario dedicado a prestar servicios de asesoría y acompañamiento a ciudadanos colombianos residentes en países del exterior, que desean comprar inmuebles en Colombia a través de instituciones financieras autorizadas para operar en el país, tales como Bancolombia, Davivienda, Banco Unión y Banco de Occidente. El objetivo al finalizar los ciclos del negocio es que los clientes puedan adquirir su vivienda en Colombia con un desembolso cuya garantía de la deuda es el inmueble a favor de la institución financiera que concede el crédito o leasing.

Actualmente, Viventa administra los datos generados de sus procedimientos a través de la empresa Zoho Corporation, empresa que ofrece productos que facilitan la administración, tratamiento y visualización de datos. 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. Zoho Analytics es la plataforma donde se realizan los informes y dashboards solicitados por los directores de las diferentes áreas.

Es de destacar que se cuenta con información de clientes desde el año 2012 en bases de datos estructuradas y se cuenta además con un conocimiento empírico de la historia de los datos por parte de los líderes y dueños de la empresa con una trayectoria dentro de la organización, logrando así tener conceptos desarrollados y objetivos ya planteados.

Sin embargo, actualmente Viventa carece de una segmentación de su base de clientes y emplea estrategias de marketing genéricas que no se adaptan a las necesidades particulares de cada grupo de usuarios. Además, como no logra identificar las necesidades y preferencias de cada segmento, los indicadores de retención de clientes y satisfacción de los mismos son bajos en comparación con el sector. La falta de segmentación, al parecer, ha obstaculizado la capacidad de Viventa para innovar y generar un crecimiento sostenible a largo plazo.

Propuesta de Solución y Alcance

En este proyecto se implementará, a través de la metodología CRISP-DM, un modelo de aprendizaje automático no supervisado para segmentar la base de clientes de la empresa. Específicamente se utilizará un algoritmo de clustering, reconocido por su efectividad en resolver este tipo de problemas. No obstante, se llevará a cabo una evaluación y comparación de diversos algoritmos de agrupamiento para identificar el más compatible con las necesidades específicas y características particulares del negocio. La solución propuesta comprende las siguientes etapas y acciones:

1. **Preprocesamiento de los datos**:

* Recopilación de todos los datos requeridos de los clientes.
* Limpieza de los datos para eliminar inconsistencias o errores.
* Realizar un análisis exploratorio de datos (EDA) para entender mejor las características y distribución de los mismos.
* Normalización de características para asegurar que todas estén en la misma escala y evitar sesgos en el modelo.
* Selección de características, si resulta necesario para el modelo, con el fin de optimizar su desempeño.

1. **Modelo de Clustering:**

* Seleccionar el algoritmo o los algoritmos de clusterización más adecuados según los datos y objetivos del proyecto, considerando la posibilidad de emplear un ‘Ensamble’ para mejorar la precisión del modelo.
* Entrenamiento del modelo seleccionado utilizando el conjunto de datos preprocesado.

1. **Evaluación de métricas:**

* Evaluar el rendimiento y la calidad de la segmentación del modelo entrenado, calculando las métricas seleccionadas en el conjunto de datos de prueba.
* Evaluar la cohesión y separación entre grupos para determinar la calidad de la segmentación.
* Analizar los resultados del modelo de clustering para determinar su efectividad y comprender las características distintivas de cada grupo de clientes.

1. **Iteración y refinamiento del modelo**:

* En caso de que los resultados no sean satisfactorios, identificar áreas de mejora en el modelo, tales como el ajuste de parámetros, incluyendo el número de clusters, la selección de características, la inclusión de nuevas variables, etc.
* Aplicar los cambios y/o ajustes propuestos y volver a entrenar el modelo.

1. **Evaluación del modelo:**

* Validar que los resultados del modelo sean interpretables y accionables para el negocio, en términos de su capacidad para generar acciones concretas y significativas.
* Documentación y comunicación de resultados a los stakeholders del negocio.

Definición del Alcance

Este proyecto pretende llegar hasta el desarrollo y evaluación del rendimiento del modelo de machine learning, evidenciando los grupos de clasificación y las particularidades de cada uno. Basado en este resultado, se recomienda dar continuidad al despliegue del modelo hasta llegar a evidenciar el resultado de la clasificación en el CRM de la empresa, logrando así que diferentes equipos de la organización puedan ver dicha clasificación y ejecutar procesos según sea el caso. Para lo cual, se recomienda seguir los siguientes pasos:

* Despliegue en la nube: una vez que el modelo ha sido desarrollado y evaluado, se despliega en un entorno de computación en la nube. Esto permite que el modelo esté disponible para su uso en tiempo real y escalable a medida que aumenta la demanda, Viventa LLC cuenta con servicios de computación de la nube de AWS.
* Integración con el CRM: el modelo se integra con el Zoho CRM de la empresa Viventa a través de una API o un conector personalizado. Esto permite que los resultados de la clasificación del cliente se entreguen directamente en un campo del CRM.
* Automatización de pipelines: se automatiza los pipeline de datos y el proceso de entrenamiento del modelo para que se ejecute de forma regular y sin intervención manual. Esto garantiza que el modelo esté actualizado con los datos más recientes y que los resultados sean consistentes.
* Monitorización y mantenimiento: se supervisa el rendimiento del modelo y el pipeline de datos en tiempo real, y se realizan tareas de mantenimiento y actualización periódicas para garantizar que el sistema funcione sin problemas y con la máxima precisión.

En cuanto a recomendaciones conjuntas o complementarias para el despliegue total del proceso, se pueden considerar la implementación de una estrategia de versionado para el modelo y el pipeline de datos, así como la implementación de un proceso de validación y verificación de los resultados.

Riesgos e Impacto del Negocio

La implementación de una estrategia de segmentación de clientes representa una oportunidad significativa para impactar de manera positiva el negocio. Esta estrategia le permitirá a la empresa personalizar de manera efectiva sus servicios, estrategias de marketing, tácticas comerciales y enfoques de servicio al cliente, lo que puede generar un aumento en las ventas de la compañía, una mayor retención de clientes o una mejora significativa en la atención al usuario, lo que se traduce directamente en una mayor satisfacción del mismo. De igual forma, si el modelo de segmentación es eficaz y logra clasificar a los clientes con la mayor precisión posible, la empresa estará en la capacidad de focalizar los esfuerzos de su equipo comercial hacia los segmentos de mayor valor, optimizando los costos y mejorando la rentabilidad de su negocio.

Un ejemplo del valor que la segmentación de clientes puede aportar a una empresa es el éxito de Coca Cola. Según se destaca en el artículo de El Tiempo, “La segmentación, el lado Coca-Cola de la vida”, la empresa ha aplicado la segmentación geográfica para adaptar sus productos y mensajes de marketing a diferentes regiones geográficas, la clave como lo menciona en el artículo Marco Llinás, gerente de asuntos Públicos de Coca Cola, es “pensar siempre en diferenciación, es decir, proponer algo distinto y novedoso para el destinatario” (El Tiempo, 2007).

Esto demuestra que la implementación de herramientas de machine learning ha sido fundamental para numerosas empresas en diversas industrias, permitiéndoles alcanzar una segmentación de clientes más eficiente. Especialmente pequeñas empresas y fintechs, a pesar de enfrentar desafíos relacionados con la calidad de los datos, han sabido aprovechar las ventajas de la tecnología para automatizar procesos y tomar decisiones basadas en datos. Un ejemplo de esta tendencia es el estudio realizado por María Moya para la Universidad de los Andes, donde se analizó un caso de estudio en una fintech latinoamericana. Este análisis no solo generó valor para la empresa mediante la caracterización y los insights obtenidos sobre los clientes, sino que también contribuyó al fortalecimiento del área de riesgos (Moya, 2020).

Sin embargo, a pesar de las numerosas ventajas, esta solución también conlleva varios riesgos, entre los cuales se encuentra el hecho de que el modelo de clustering no logre capturar con precisión las características de cada grupo de clientes, lo que podría generar toma de decisiones erróneas por parte de las directivas de la empresa. De igual forma, en algunas ocasiones los modelos de aprendizaje automático pueden ser difíciles de interpretar o comprender para los usuarios finales lo cual puede dificultar la aceptación y adaptación de la solución por parte de los usuarios.

Finalmente, aunque el alcance del proyecto no llega a un despliegue en el entorno de producción, la implementación de un modelo de Machine Learning puede requerir por parte de la empresa recursos significativos en términos de infraestructura, tecnología y recurso humano especializado. Lo cual puede ser un problema para empresas con recursos limitados.

**Referencias**

El Tiempo. (2007, 05 31). La segmentación, el lado Coca-Cola de la vida. *El Tiempo*. https://www.eltiempo.com/archivo/documento/MAM-2513369

Moya, M. J. (2020, 01 01). *Aprendizaje no supervisado en el perfilamiento de clientes para profit scoring*. Repositorio Uniandes. Retrieved April 3, 2024, from https://repositorio.uniandes.edu.co/server/api/core/bitstreams/7fb9cbec-6760-478d-9962-c699b61bc1e1/content