

Juan José Pardo
AV Vertical Data

Una incursión a la

Toma de decisiones basadas en datos.

- Analisis de la información
- Tecnicas de Machine Learning
- Técnicas de Inteligencia Artificial
- Conclusiones numéricas del análisis
- Recomendaciones para la organización

Juan José Pardo -
Data Analytics Specialist



Optimización de procesos de cobranza: Un enfoque basado en Machine Learning para mejorar la eficiencia y el retorno de inversión en la cuenta Colsubsidio

2 Introducción

4 Descripción de la solución

8 Metodología

20 Resultados

26 Impacto en el negocio

35 Conclusiones

35 Recomendaciones

Al interior de Hevarán, call center especializado en cobranza se desarrolló un proyecto centrado en machine learning e inteligencia artificial, con el objetivo de brindar un proceso de mejora en términos de retornos de inversión y eficiencia operacional.

Introducción

En un mundo que presenta cada vez más niveles de competitividad, las organizaciones se ven obligadas a enfrentar desafíos complejos para mantener su posición en el mercado y lograr sus objetivos financieros.

Es por esta razón que Hevarán, como organización no debe ser la excepción. En un esfuerzo por mejorar la eficiencia en su modelo de gestión de cobranza para su cliente Colsubsidio, se propuso la creación de un modelo de inteligencia artificial para priorizar las cuentas que presentan un mayor nivel de prioridad, representadas como aquellas encargadas de presentar un mayor nivel de retorno económico.

En este documento, presentamos detalles de la metodología utilizada para la recolección y procesamiento de datos, las técnicas de machine learning aplicadas, el análisis de los resultados y las conclusiones obtenidas.

Este proyecto busca no solo mejorar la eficiencia en la operación de Hevarán como organización especializada en modelos de cobranza sino también servir como un ejemplo en demostración de cómo la inteligencia artificial puede convertirse en una herramienta de gran valor para la toma de decisiones en el ámbito empresarial.

Descripción de la solución

- Partiendo de la base del problema que la organización encuentra "en términos de identificar aquellas categorías de la cuenta de Colsubsidio que pueden representar un mayor y mejor potencial de ingresos" se desarrolló un modelo de solución pensado en modelación con algoritmos de inteligencia artificial y machine learning bajo métodos supervisados, bajo su eficiencia, fue posible presentar aquellos elementos que representarían una ventaja en términos monetarios para la cobranza de forma eficiente.

Análisis de cohortes: El valor de la clusterización de datos para identificar probabilidades de pago

Como primer modelo del proyecto, realizamos un proceso de clusterización de los datos, permitiéndonos identificar, por medio del análisis de cohortes aquellos grupos de edad que podrían presentar una ventaja económica en términos de recuperación de cobranza, presentando una mejora en términos de adquisición. Bajo su categoría, fue posible identificar que aquellos clientes que presentaban edades avanzadas (Ubicados entre los 66-100 años) cuentan con una mayor disposición en términos de pagos, presentando, ante todo una prioridad en términos de recuperación.

"Uniendo la inteligencia artificial y la experiencia humana, logramos maximizar el retorno económico y la satisfacción del cliente en el proceso de gestión de cobranza de Colsubsidio"

Tomando decisiones con árboles de decisión: Una solución efectiva para la gestión de cobranza.

En términos sencillos, la solución propuesta para el problema que presentaba Hevarán se basó en la utilización de árboles de decisión. Esta técnica de machine learning permitió identificar a los clientes de colsubsidio que presentaban una mayor probabilidad para realizar pagos, su implementación ayudó a la empresa a centrar sus esfuerzos en aquellos clientes que presentaban una mayor posibilidad de generar un mayor y mejor retorno económico, afectando directamente el ROI organizacional.

Además, los árboles de decisión permitieron identificar las variables más importantes que influían en la probabilidad de pago, lo que facilitó la toma de decisiones en términos de enfoque en la estrategia de cobranza.

En resumen, la aplicación de los árboles de decisión fueron de gran utilidad en su método de implementación, dado que le permitieron a Hevarán optimizar la gestión de cobranza de Colsubsidio y aumentar su eficiencia en términos de recuperación de cartera.

"Los datos son el nuevo petróleo. Así como el petróleo impulsó la economía industrial, los datos impulsarán la economía digital"
-Michael Chui, Boston Consulting Group

Desbloqueando el potencial de recuperación de cartera de Hevarán con regresión logística

El modelo de regresión logística utilizado en este proyecto es una técnica de aprendizaje automático que permite predecir la probabilidad que un evento ocurra, en este caso, el pago de un cliente de Colsubsidio.

A través de diversas variables el modelo se desarrollo en el entorno de programación y visualización de datos llamado R. A través del análisis de diversas variables, el modelo identifica patrones y relaciones que pueden afectar la probabilidad de pago y genera una predicción basada en estos hallazgos.

En el contexto de Hevarán, este modelo se convirtió en una pieza de gran utilidad ya que permite que la empresa canalizar sus esfuerzos en aquellos clientes que cuentan con una mayor probabilidad de pago, lo que traduce en una mayor fuente de recuperación de cartera y eficiencia en la gestión de cobranzas para la organización.



Data Shortcuts

Algunas ilustraciones clave que se tuvieron en cuenta para realizar la modelación en inteligencia artificial en Hevarán

Claves de éxito: Nuestra metodología de análisis de datos para una gestión de cobranza efectiva en Hevarán.

Bajo el modelo de trabajo, se realizó una metodología rigurosa en términos de recolección, exploración, visualización y modelado de datos para generar un modelo de inteligencia artificial con un rendimiento óptimo que permita presentar resultados eficientes en la operación del negocio, desarrollando la información necesaria para una toma de decisiones estratégicas en la operación.

La metodología utilizada permitió una mejor comprensión de los datos, lo que resultó en una mayor precisión en el modelo y una mejor capacidad del pronóstico para el negocio. La exploración de los datos permitió descubrir patrones y relaciones eficientes entre variables, lo cual ayudó a identificar oportunidades para mejorar el rendimiento de la empresa.

La visualización de los datos permitió una comprensión más profunda de los resultados del modelo y facilitó la presentación de información valiosa a los tomadores de decisiones.

En resumen, la metodología utilizada permitió una toma de decisiones más informada y efectiva para la empresa.

La metodología utilizada para la modelación en inteligencia artificial se centró en el uso de R para la creación de modelos de árboles de decisión y regresión logística, con una atención rigurosa a la selección de variables predictoras y la

validación cruzada del modelo. Además, para garantizar la calidad de la información se implementaron técnicas de pre procesamiento de datos, lo que contribuyó significativamente a la precisión del modelo.

La elección de estas técnicas tanto de modelado como de procesamiento de datos fue fundamentada en la amplia literatura y la experiencia previa del equipo de expertos de Hevarán en el área de inteligencia artificial y análisis de datos.

Resultados

A partir del análisis generados en los árboles de decisión, se logró identificar los segmentos de clientes que cuentan con una mayor probabilidad de pago, lo que permitió a Hevarán enfocar sus recursos de manera más efectiva en aquellos clientes con una mayor probabilidad de cumplimiento.

A partir del análisis de establecen los siguientes criterios de insights e indicadores de mayor valor reflejados por el árbol de decisión.

Grupo de consumo no libranza: El modelo identificó que los clientes de este grupo tienen una alta probabilidad de pago (referente a un 83%). En esta categoría, el principal factor que influye en su comportamiento de pago es su edad, siendo los clientes mayores de 40 años los que presentan una mayor propensión a pagar a tiempo.

Grupo de consumo castigo: En este grupo, el modelo identificó que la probabilidad de pago es del 67%, siendo la antigüedad del cliente el principal factor que influye en su comportamiento de pago. En ese sentido, se observó que los clientes que cuentan con más de 2 años como clientes de colsubsidio cuentan con una mayor probabilidad de pago.

Grupo de Cupo de Crédito: Los clientes de este grupo representan una probabilidad de pago del 75%. El modelo identificó que el factor más influyente en su comportamiento de pago es su nivel de endeudamiento y saldos relacionados con la cartera.

Grupo de consumo no libranza: Los clientes de este grupo representan la probabilidad de pago más baja, con un 56%. En este punto, el modelo identificó que el factor más influyente en su comportamiento de pago es el historial de pago negativo afectando el nivel de probabilidad de pago referente a esta causa.

A partir de los resultados generados en el modelo de inteligencia artificial representado por la regresión logística, se lograron tomar en cuenta las siguientes conclusiones como factores de decisión.

Canalización de esfuerzos en segmentos de clientes de mayor edad: En términos de datos, observamos que la cohorte más grande, en términos de saldo total es la de 1977 la cual representa el 31.3% del saldo total general.

En segunda instancia, la cohorte a la cual se le debe generar un mayor nivel de prioridad es aquella que es representada por clientes nacidos en 1967, la cuál cuenta con un 21.9% de importancia a partir de la raíz de la proporción.

Es importante considerar la priorización de esfuerzos en términos de la recuperación en este punto, entendiendo que, dentro de la ley de Pareto, son aquellos los que representan mayor importancia en relación con los saldos y en términos de recuperación de cartera.

Para la modelación del proceso de regresión lineal, seleccionamos la cohorte representativa que se enfocaba en el 22% de los datos, gracias a su uso, fue posible contar con la probabilidad de pago según la edad de nacimiento de los usuarios segmentados en un rango de (65-100 años) bajo los registros en donde se tiene acceso al interior de la base de datos.

Decisiones basadas en edades: Basado en el modelo de probabilidad de pagos, observamos que en una proporción general, las edades que presentan una proporción de edad entre los 63 a los 87 años de edad, cuentan con un 23.57% mayor en términos de pagos en contraste con las medias de otras cohortes.

El modelo cuenta con una efectividad superior al 80% de los datos, relacionando su efectividad para medir la probabilidad de pago de un usuario con respecto a la edad de sus clientes.

Impacto en el negocio

Se espera entonces que, bajo la implementación de este modelo de inteligencia artificial se tenga un impacto significativo en la eficiencia de la gestión de cobranza de la empresa Hevarán y su cliente Colsubsidio.

Bajo el uso de la utilización de técnicas relativas con Machine Learning y su implementación del modelo de regresión logística y árboles de decisión se permita una mejor automatización en términos de segmentación de clientes y en los procesos de gestión de cobranza. Con esta implementación, será posible reducir los costos de operación y aumentar la capacidad de la empresa para manejar grandes volúmenes de datos.

En general, se espera que la implementación de este modelo permita a Hevarán mejorar su capacidad para apoyar a los clientes en la gestión de cobranza y de esta manera mantener una relación sólida y productiva con su cliente actual, dado que el análisis de datos juega un papel fundamental para el desarrollo de proyectos tanto presentes como futuros en el negocio per sé.

Conclusiones

A partir del proyecto trabajado, consideramos los siguientes elementos como fundamentales para la toma de decisiones corporativas que afectan, tanto a Hevarán como a Colsubsidio como su cliente final.

Los modelos de inteligencia artificial como el árbol de decisiones y la regresión logística pueden convertirse en herramientas poderosas para mejorar la eficiencia y la efectividad de

los procesos de gestión de cobranzas.

En otro punto, elementos importantes relacionados con la segmentación de los clientes según sus características y comportamientos de pago pueden ayudar a los clientes de cobranza a priorizar las cuentas que presentan una mayor probabilidad de pago, y en este sentido, es posible mejorar significativamente el rendimiento de los resultados de la empresa.

Elementos importantes como la validación cruzada y la evaluación de la precisión del modelo son técnicas esenciales para garantizar la confiabilidad de los resultados del modelo.

La metodología de recolección, limpieza y transformación de datos utilizada es fundamental para asegurar la calidad de los datos en el proceso de transformación garantizando la efectividad en las predicciones generadas por el modelo.

Adicionalmente, para sacar el máximo beneficio de la información es recomendable para la empresa continuar optimizando y actualizando el modelo para la gestión de cobranzas.

Recomendaciones

Se espera que la organización, de manera continua se encuentre en la disposición de actualizar el modelo, esto con el fin de proveer de recursos suficientes en el volumen de datos. Al convertirse en un modelo de aprendizaje automático por método supervisado, el mismo cuenta con la capacidad de generar valor a partir de los estímulos de las bases de datos .csv y las actualizaciones que se puedan hacer.

El proyecto puede tener un alcance de mejora y expansión a la nube de la organización mediante la integración de servicios de la nube de AWS o Microsoft Azure, permitiendo a los diferentes miembros del equipo trabajar en el mismo y escalarlo de forma remota.

Es importante incluir un proceso de actualización constante de la base de datos: Mantener esta información actualizada y completa es fundamental para mejorar la precisión del modelo y la información que genera.

Continuar con el modelo en la implementación diaria: Para este punto, es vital que la organización pueda contar con utilizarlo de manera diaria para identificar valores de predictibilidad diarios.

Reforzar la capacitación de los agentes de cobranza: En este punto, es importante que los agentes cuenten con las habilidades y conocimientos necesarios para gestionar efectivamente las situaciones de impago y garantizar una experiencia positiva para los clientes.

Realizar una evaluación constante del modelo: Es importante que el modelo pueda ser evaluado de manera constante, con el objetivo de identificar posibles áreas de mejora.