

MODELO DE MACHINE LEARNING PARA PREDECIR LA CARGA FINANCIERA MÁXIMA DE CLIENTES EN NEQUI



PRACTICANTE: Esteban Caro Peláez

PROGRAMA: Ingeniería Industrial

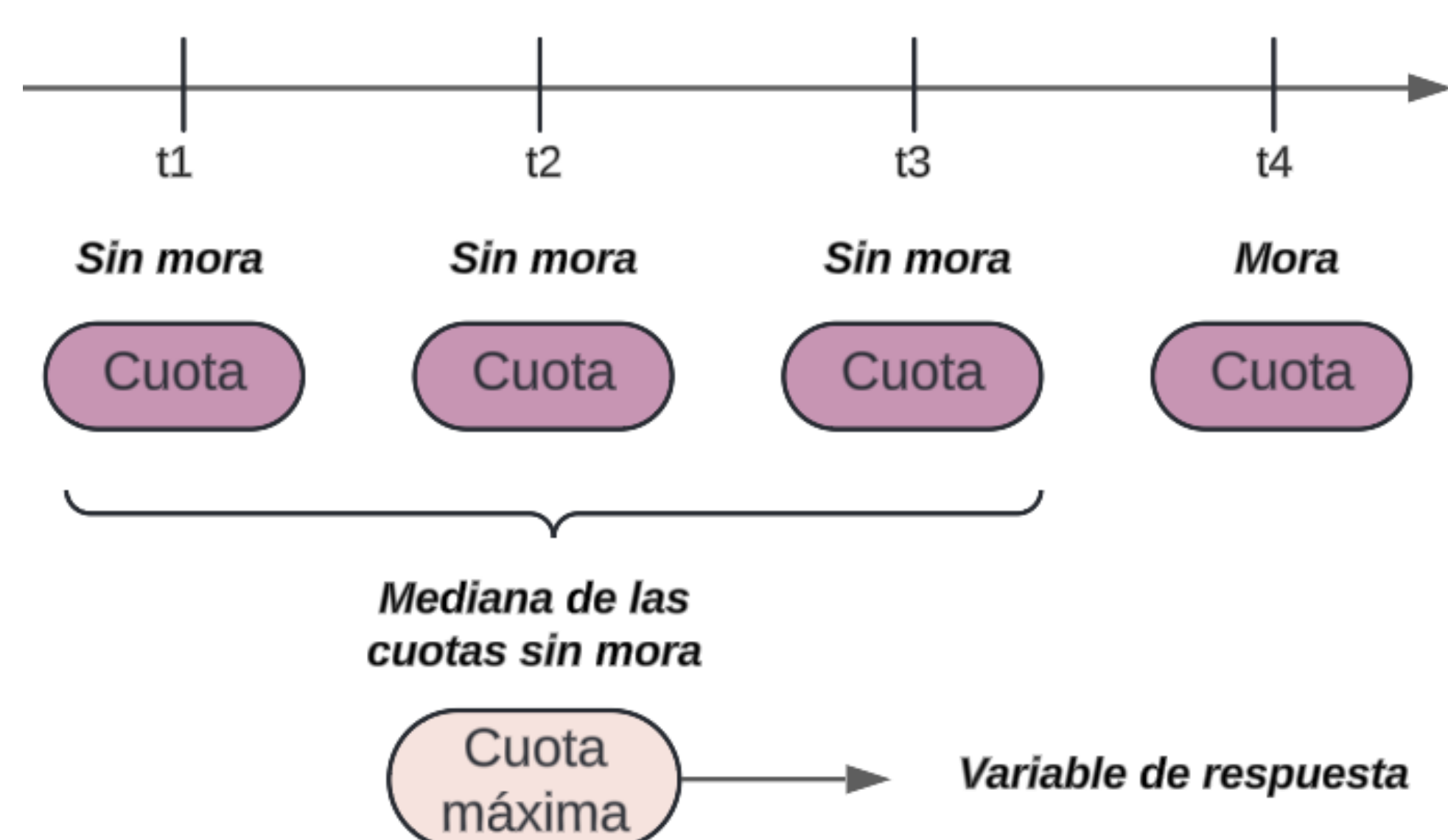
ASESORES: Claudia Sofía Correa Puerta, Stiven Cadavid Cataño

MODALIDAD DE LA PRÁCTICA: Presencial

Nequi es una compañía de financiamiento totalmente digital. Actualmente, la empresa ofrece créditos de bajo monto a sus usuarios. La Gerencia de Riesgo de Crédito determina qué usuarios serán elegibles para recibir créditos y cuáles serán los montos a prestar.

Dentro de la Gerencia, se ha identificado que la principal causa de rechazo de créditos a los usuarios es su capacidad de pago. Por esta razón, se ha reconocido la necesidad de estimar la carga financiera máxima aplicable a cada cliente.

Construcción variable de respuesta



Introducción

En el presente poster se presenta el desarrollo de un modelo de machine learning como alternativa estimación de carga financiera máxima haciendo uso de la metodología CRISP – DM. Este modelo pretende predecir la carga financiera máxima de cierto segmento de clientes para el trimestre siguiente a su fecha de ejecución. Con este proyecto se espera refinar la estimación de la capacidad de pago de cada cliente, de manera tal que se refleje su realidad financiera y permita a la compañía ofrecerles créditos que se ajusten a sus necesidades individuales, optimizando la gestión del riesgo.

Objetivos

Desarrollar un modelo de machine learning para estimar la carga financiera máxima de los egresos financieros de clientes Nequi, con el fin de mejorar la precisión en la evaluación de su capacidad de pago en el flujo de preaprobados al interior del área de riesgos de crédito.

Establecer las métricas de negocio y los requerimientos del área de riesgos de crédito en relación con la estimación de cuotas de tarjetas de crédito, para definir el impacto del proyecto en el proceso de flujo de preaprobados

Explorar y analizar los datos históricos comprados a la central de información financiera, con el objeto de identificar variables relevantes, evaluar calidad de los datos y detectar tendencias que puedan afectar el rendimiento del modelo.

Implementar el modelo de machine learning de regresión para la estimación de la carga financiera máxima

Evaluar el mejor modelo de machine learning en función de las métricas de desempeño, con el objetivo de cumplir con los requerimientos mínimos exigidos por la mesa de riesgo de crédito

Seleccionar el mejor modelo en función de las métricas de desempeño de modelos de machine learning y de negocio, a fin de garantizar que la solución propuesta dé respuesta a las necesidades específicas del negocio

Documentar los hallazgos encontrados y los hiperparámetros del modelo de machine learning, para facilitar el despliegue del modelo en caso de ser aprobado por la mesa de riesgos de Nequi

Documentar el modelo y los hallazgos encontrados durante la ejecución del proyecto, a fin de tener información que soporte la toma de decisiones relacionadas a capacidad de pago ante la Superintendencia Financiera y MRM de Bancolombia.

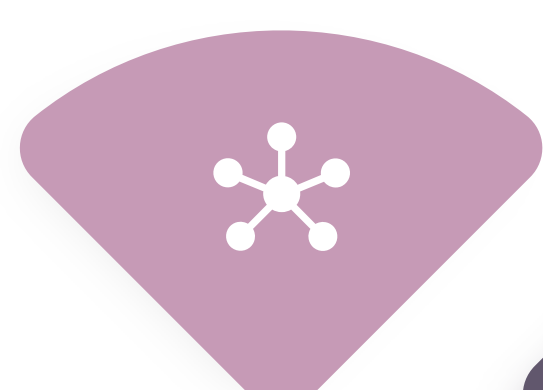
Metodología

Métricas de negocio

Implementación del modelo

Selección del mejor modelo

Documentación del modelo



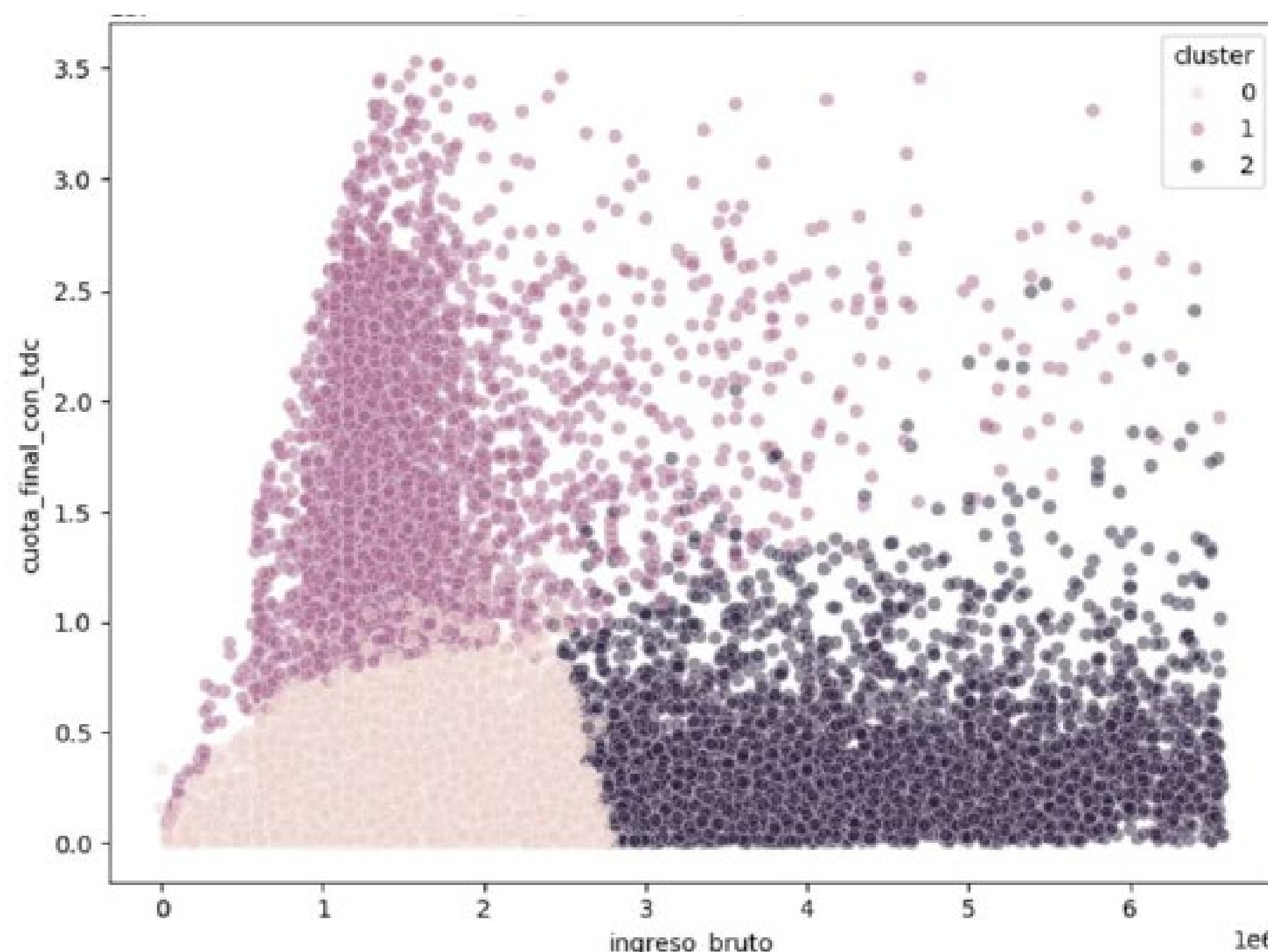
Análisis exploratorio

Evaluación del modelo

Documentación de hiperparámetros

Resultados

1. Se construyó un modelo con un alcance de 700.000 clientes y un MAPE de 33.33%
2. Se formaron tres clústers de clientes; clientes con capacidad de pago alta, baja y media
3. Se implementó un modelo XGBoost
4. Se obtuvo un mejor modelo con un MAPE del 33.33% en OOT.
5. Selección del XGBoost como mejor modelo
6. Documentación de los mejores hiperparámetros para modelo XGBoost
7. Documentación parcial de hallazgos encontrados durante la ejecución del proyecto



Conclusiones

- La estimación de la cuota financiera máxima de los clientes juega un papel fundamental en el proceso de flujo de preaprobados.
- Los clústers por carga financiera máxima son una solución viable descubierta durante el EDA
- El módulo de información del buró es más explicativo para la variable de respuesta.
- El ajuste de hiperparámetros afecta la calibración del modelo, específicamente en entrenamiento, testeo y OOT.
- El algoritmo XGBoost tiene el mejor comportamiento para el caso de negocio
- Los mejores hiperparámetros tienen un balance entre capacidad predictiva y generalización
- La documentación del modelo es importante ya que permite la reproducibilidad del mismo en el tiempo