

Tarea3_2

November 10, 2024

1 Proyecto de ML para identificar el riesgo de un credito

1.1 Introduccion

El Banco CloseBank necesita identificar de un grupo de clientes que tienen actualmente cuales de estos es probable que puedan no pagar un credito. Para esto crearemos un modelo que analizara la situacion y cararacteristicas de los cientes y nos dira cuales de estos tienen mayor probabilidad de no pagar el crédito.

1.2 Objetivos del proyecto

1. ¿Cuáles son los objetivos del negocio?

Crear un modelo que pueda que nos permita saber cuales de los clientes tienen una probabilidad alta de no pagar un crédito.

2. ¿Qué decisiones o procesos específicos desea mejorar o automatizar con ML?

Un proceso a mejorar seria poder tener de manera automática una lista clientes que potencialmente pueden no pagar un crédito y de esta manera tenerlos en foco y poder hacer un seguimiento adecuado.

3. ¿Se podría resolver el problema de forma no automatizada?

Si, pero este sería un proceso muy largo y seguramente no tendríamos la precisión que nos puede arrojar el modelo propuesto.

1.3 Metodologia propuesta

Para crear el modelo utilizaremos un algoritmo de aprendizaje no supervisado, en este caso utilizaremos K-Means. K-Means es buena opción para grupos bien definidos y es interpretable en cuanto al número de clusters. Adicionalmente este algoritmo nos puede arrojar una segmentacion clara que es lo que buscamos ya que de esta manera podemos ver patrones claros y diferenciados.

La idea seria que uno de los clusters generados sea que segun sus características poder tener un cluster donde agrupemos a las personas que tiene mayor probabilidad de no pagar un credito.

Variables a usar: + age (edad) + balance (saldo de la cuenta) + housing (tiene hipoteca o no) + loan (tiene préstamo o no)

1.4 Datos disponibles

5. ¿Qué datos están disponibles para abordar este problema?

El banco nos ha suministrado una lista de clientes completa que nos permitira crear el modelo.

1.5 Metrica de salida

Nuestra Metrica de salida seria una lista de clientes, estos seran los clientes que tienen mayor probabilidad de no pagar un credito.

1.6 Responsabilides eticas y sociales

El uso de modelos de ML para analizar y segmentar a personas con base en sus características demográficas y financieras trae consigo responsabilidades éticas y sociales importantes como:

1. Evitar la Discriminación y el Perjuicio » Hay un riesgo de que estos clusters refuercen o incluso exacerbén prejuicios existentes, la entidad debe asegurarse de que las variables utilizadas en el análisis no introduzcan sesgos que resulten en discriminación.
2. Respeto a la Privacidad y Protección de Datos

El clustering involucra datos personales que pueden revelar información sensible sobre las personas, como sus finanzas o su ocupación, lo cual tiene implicaciones de privacidad, la entidad debe proteger los datos durante y después del análisis manteniendolos anonimizados en todo momento.

3. Uso Ético en la Toma de Decisiones Financieras

Los clusters generados por K-Means podrían usarse para tomar decisiones significativas, como establecer condiciones de crédito o tasas de interés, que pueden afectar directamente a la vida de las personas, la entidad debe usar el clustering como una herramienta de apoyo, no como la única base para decisiones críticas y ver cada caso de manera individual considerando múltiples fuentes de información.

4. Responsabilidad Social en el Impacto General

Un mal uso de los clusters podría contribuir a ampliar brechas de desigualdad y afectar negativamente a ciertos sectores de la sociedad, especialmente si el clustering resulta en decisiones que dificultan el acceso al crédito o a servicios financieros para ciertos grupos, la entidad debe tener en cuenta el impacto social de los resultados de clustering, especialmente en sectores sensibles de la población se deben evaluar cómo sus decisiones afectan a diferentes segmentos de clientes.

La utilización de modelos de ML para segmentar clientes conlleva una serie de responsabilidades éticas y sociales que deben ser priorizadas en todo momento. Estas incluyen la prevención de sesgos y discriminación, la protección de la privacidad de los datos, la minimización de daños, y la transparencia en el uso e interpretación de los resultados.