



CARRERA DE ESPECIALIZACIÓN EN INTELIGENCIA ARTIFICIAL

MEMORIA DEL TRABAJO FINAL

Modelos de predicción de *overdue* y rotación para la redistribución del crédito financiero

Autor:

Lic. Paola Mariscal Zegarra

Director:

Ing. Yoel Yamil López (FIUBA)

Jurados:

Marcos Maillot (pertenencia)

Hanes Sciarrone (pertenencia)

Pablo Martín Gómez (pertenencia)

*Este trabajo fue realizado en la Ciudad Autónoma de Buenos Aires,
entre abril de 2022 y diciembre de 2022.*

Resumen

La presente memoria describe el diseño e implementación de un algoritmo predictivo de mora y de rotación del crédito desarrollado para una importante compañía de consumo masivo que otorga créditos a sus clientes para la venta de sus productos. El sistema permite distinguir clientes que por su comportamiento podrían entrar en mora o aquellos que usen en mayor proporción el crédito asignado, lo que permitirá a la compañía mejorar la distribución del presupuesto destinado a créditos.

Para poder abordar este trabajo se aplicaron conceptos de estadística, bases de datos y aprendizaje automático que permitieron llevar a cabo la implementación de un modelo estadístico basado en machine learning desde la extracción de datos hasta la generación de dos scores: riesgo y rotación

Agradecimientos

A mi familia por apoyarme en cada nuevo desafío.

A mi director, Ing. Yoel Yamil López por su acompañamiento, orientación y por los conocimientos compartidos a lo largo de la carrera.

Índice general

Resumen	I
Agradecimientos	III
1. Introducción general	1
1.1. Inteligencia artificial en el ámbito financiero	1
1.2. Estado del arte: caso Nosis	2
1.3. Motivación	3
1.4. Objetivos y alcance	3
1.4.1. Objetivos	3
1.4.2. Alcance	3
2. Introducción específica	5
2.1. Requerimientos asociados al desarrollo e implementación de/los modelo/s	5
2.1.1. Requerimientos asociados al desarrollo	5
2.1.2. Requerimientos asociados a la implementación	5
2.2. Contexto de negocio	5
2.3. Modelos de inteligencia artificial utilizados	6
3. Diseño e implementación	7
3.1. Arquitectura del sistema completo	7
3.2. Generación y tratamiento de los datos	7
3.3. Automatización e implementación	7
4. Ensayos y resultados	9
4.1. Descripción del banco de pruebas	9
4.2. Modelos probados	9
4.3. Descripción del banco de pruebas	9
4.4. Descripción de la salida	9
4.5. Caso de uso	9
5. Conclusiones	11
5.1. Conclusiones generales	11
5.2. Trabajo futuro	11

Índice de figuras

1.1. Relación entre el <i>score</i> y el riesgo.	2
1.2. Diagrama en bloques del sistema.	3
2.1. Proceso previo a la automatización.	6

Índice de tablas

Dedicado a... [OPCIONAL]

Capítulo 1

Introducción general

En este capítulo se realiza una introducción a los conceptos básicos del aprendizaje automático y financieros. Asimismo, se menciona el estado del arte en un buró de crédito argentino, y por último se explica la motivación, alcance y objetivos del presente trabajo.

1.1. Inteligencia artificial en el ámbito financiero

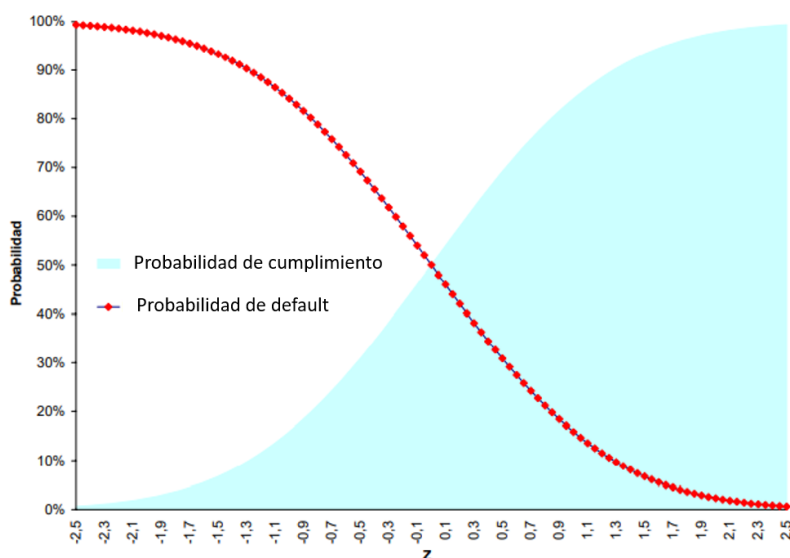
El sector financiero fué y sigue siendo uno de los pioneros a la hora de implementar inteligencia artificial, una de las tecnologías más revolucionarias de los últimos tiempos.

La inteligencia artificial se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas, mejorando iterativamente a partir de la recopilación de información. Su objetivo es mejorar significativamente las capacidades humanas.

Actualmente la inteligencia artificial mejora el rendimiento y la productividad de las empresas mediante la automatización de los procesos y tareas que antes requerían esfuerzo humano o expertos dedicados a ello. Hoy en día, muchas aplicaciones de inteligencia artificial son utilizadas tanto por instituciones tradicionales como por *fintech*.

Un *score* crediticio que utiliza inteligencia artificial es mucho mas completo y sofisticado comparado con criterios tradicionales de score de crédito. El score o scoring es un método estadístico para predecir comportamientos futuros o probabilidad de *default*. Asigna un número de tres cifras que también puede ser determinado como un porcentaje de probabilidad de ocurrencia de un evento futuro y es uno de los factores que analizan los prestamistas ante solicitudes de crédito.

En la figura 1.1 se observa que la relación entre el *score* y el riesgo no es lineal, por lo que el cambio en el riesgo derivado de un cambio en el *score* depende de los valores que este último tome. Para valores del *score* muy bajos, un aumento en el mismo produce una rápida subida en la probabilidad de cumplimiento y una rápida disminución de la probabilidad de default, mientras que para valores del *score* altos, una mejora en el mismo hace que la probabilidad de cumplimiento aumente poco y genera una leve caída en el riesgo. Es decir, cuanto mayor es el *score*, menor es la caída en el riesgo derivada de un aumento en el primero.

FIGURA 1.1. Relación entre el *score* y el riesgo.

Los bancos digitales y fintech utilizan algoritmos de aprendizaje automático o *machine learning* cuya importancia radica en la objetividad y la eliminación de sesgos para la toma de decisiones. Asimismo, la capacidad de cómputo de este tipo de sistemas, permite manejar grandes cantidades de datos en poco tiempo y la computación cognitiva ayuda a administrar datos estructurados y no estructurados. Los algoritmos analizan historiales de transacciones e identifican a tiempo signos de futuros problemas.

1.2. Estado del arte: caso Nosis

Nosis es una empresa fundada en la década del 80 con el objeto de brindar información de antecedentes comerciales, mercados financieros en línea y comercio exterior para aportar herramientas analíticas que faciliten la toma de decisiones. Como buró, cuenta con bases de datos exclusivas, información compartida por mas de 100 entidades, información pública tanto del BCRA, ANSES, AFIP, entre otros e innovadoras técnicas analíticas con actualización constante de datos.

Una de sus principales unidades de negocio trabaja sobre informes comerciales que tienen como variable principal un score desarrollado por la misma empresa. El score de Nosis es un ejemplo de un score de riesgo. Brinda información sobre la probabilidad de default o mora del cliente consultado desde el momento de la consulta propiamente dicha y 12 meses hacia adelante. Detrás de este número único o calificación, hay mas de 70 variables de información dentro del algoritmo desarrollado por la empresa que puso en consideración el endeudamiento histórico de varias fuentes de información y mas de 600 atributos de datos. El algoritmo se procesa en el momento de la consulta e indica que cuanto más alto sea el score, existe una menor probabilidad de default o mora. Las bases de datos con las que cuenta Nosis, cubren el 99.99 % del endeudamiento de una persona o empresa con el sistema financiero total a nivel país.

1.3. Motivación

La principal motivación para la realización de este trabajo fue la eliminación de subjetividades en procesos de asignación de límite crediticio, la reducción de la deuda generada por los clientes morosos y la optimización del capital empleado (*receivables*) para la generación de mejores resultados, otorgando un mejor servicio de crédito a los clientes, junto a la adopción de tecnologías relacionadas al análisis predictivo en una empresa tradicional de muchos años y con procesos establecidos adversos al cambio.

El presente trabajo fue el primer algoritmo desarrollado por el área de *data analytics* y el principal caso de uso dentro de la compañía, rompiendo así la brecha entre las áreas de negocio y de TI (tecnología de la información).

1.4. Objetivos y alcance

1.4.1. Objetivos

El propósito de este trabajo consistió esencialmente en el diseño, pruebas e implementación de una solución de inteligencia artificial que permite estimar la mora y la rotación del crédito de los clientes activos de la compañía, asegurando a una mejor distribución de los recursos crediticios, considerando distintos aspectos del cliente en cuanto a compras, pagos atrasados, deuda total vigente, cheques rechazados, antigüedad e historial de pagos, entre otras variables.

En la figura 1.2. se observa un diagrama general de la solución y las etapas que la componen.



FIGURA 1.2. Diagrama en bloques del sistema.

En el capítulo 3 se detalla el diseño y la implementación de las etapas que componen este sistema.

1.4.2. Alcance

El alcance de este trabajo estuvo orientado a desarrollar una solución de software que permite cubrir aspectos que giran en torno a los siguientes ejes:

1. Entendimiento del problema:
 - *Assesment* del proceso de asignación manual del límite de crédito.

- Resultados esperados a partir de la implementación de la solución.
- 2. Aspectos relacionados al entendimiento del negocio:
 - Entendimiento de las restricciones asociadas al proceso.
 - Entendimiento de aspectos generales del negocio.
- 3. Aspectos relacionados a la adquisición y comprensión de datos:
 - Assesment de las fuentes de datos (modelos en *datawarehouse* y otras).
 - Exploración de los datos para determinar la calidad de la información.
- 4. Aspectos relacionados al modelado:
 - Diseño de características: generación de variables adicionales a partir de los datos sin procesar para facilitar el entrenamiento del modelo.
 - Entrenamiento del modelo: elección del modelo que responda a la pregunta de negocio con la máxima precisión, evaluando métricas de éxito.
- 5. Aspectos relacionados al despliegue:
 - Implementación del/los modelo/s en un entorno de producción o similar para el consumo por el usuario final o aplicaciones.

El alcance de este trabajo no cubre:

- El mantenimiento de la base de datos en la cual se aloja la información generadora de la entrada para el modelo de inteligencia artificial.
- La instalación y mantenimiento del *hardware* necesario para el procesamiento de información del modelo desarrollado.
- La disponibilización de la salida del algoritmo directamente en alguna aplicación de la compañía.

Capítulo 2

Introducción específica

En este capítulo se detallan los requerimientos asociados al desarrollo y la implementación de los modelos, así como el contexto y las necesidades de negocio. Finalmente se especificarán las tecnologías aplicadas en el desarrollo del trabajo, destacando los componentes de inteligencia artificial seleccionados para la predicción de la mora y la rotación.

2.1. Requerimientos asociados al desarrollo e implementación de/los modelo/s

2.1.1. Requerimientos asociados al desarrollo

1. Los códigos deben desarrollarse con herramientas de Microsoft Azure que es el servicio de computación en la nube utilizado por el área de data analytics.
2. El equipo de desarrollo tiene la potestad de utilizar el lenguaje de su conocimiento para el desarrollo del código del modelo.
3. Se utiliza GIT como repositorio para el control de versionado de código.

2.1.2. Requerimientos asociados a la implementación

1. Se utiliza Azure Databricks como herramienta tanto para el entrenamiento de los modelos como para su implementación.
2. Debe existir un archivo de ejecución principal llamado *main.py* y un *job* de ejecución automática de ese archivo en Databricks para la implementación *batch*, es decir sin supervisión.
3. Utilización de MLflow, una plataforma de código abierto para la administración del ciclo de vida de los modelos de *machine learning*, para el despliegue en producción y el seguimiento de los desarrollos dentro de la compañía.

2.2. Contexto de negocio

Uno de los motivos principales por los cuales fue necesario abordar el presente trabajo fue la no existencia de una herramienta que permita objetivamente asignar límite de crédito a cada cliente, lo que significaba un alto riesgo asociado a

la mora o incobrabilidad. De hecho, durante el año 2021 el *bad debt* o deuda incobrable ascendía casi al 30 % de la cartera de clientes con crédito. Por otro lado, la compañía no estaba cobrando interés alguno por los créditos cedidos y el P&L de crédito (*profit and loss*) arrojaba resultados negativos por un porcentaje considerable de clientes en mora sin un ROIC (*return on invested capital*) asociado.

En referencia a la opinión de los clientes en temáticas de crédito, se obtuvo un bajo NPS (*net promoter score*) en las encuestas de satisfacción de servicio enviadas. Sumado a lo anteriormente mencionado se encontraban a menudo errores manuales en análisis o asignación de crédito por causa de la no automatización de procesos. Existía la dependencia de personal específico solo para realizar un análisis de riesgo de cada cliente a partir de reglas de negocio definidas.

En la figura 2.1 se observa el proceso previo al desarrollo del presente trabajo.

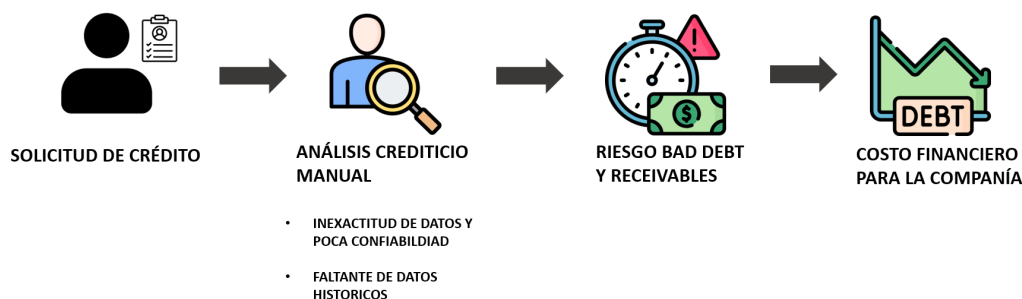


FIGURA 2.1. Proceso previo a la automatización.

2.3. Modelos de inteligencia artificial utilizados

Capítulo 3

Diseño e implementación

3.1. Arquitectura del sistema completo

3.2. Generación y tratamiento de los datos

3.3. Automatización e implementación

Capítulo 4

Ensayos y resultados

- 4.1. Descripción del banco de pruebas**
- 4.2. Modelos probados**
- 4.3. Descripción de la salida**
- 4.4. Caso de uso**

Capítulo 5

Conclusiones

5.1. Conclusiones generales

5.2. Trabajo futuro