

Análisis de datos para la gestión de carteras castigadas en Contact Centers Virtuales

Data Analysis for the Management of Delinquent Portfolios in Virtual Contact Centers

Camilo A. Leuro¹   Juan P. Ospina¹ 

¹ Escuela de Ciencias Exactas e Ingeniería Universidad Sergio Arboleda Bogotá, Colombia.

How to cite?

Leuro, C.A., Ospina, J.P., Análisis de datos para la gestión de carteras castigadas en Contact Centers Virtuales. Ingeniería y Competitividad, 2025, 27(1) e-20714327

<https://doi.org/10.25100/iyc.v27i1.14327>

Recibido: 9-07-24

Evaluado: 20-09-24

Aceptado: 10-02-25

Online: 20-02-25

Correspondence

camilo.leuro01@usa.edu.co

Resumen

Introducción: los contact center virtuales que brindan servicios BPO han experimentado un crecimiento exponencial en los sectores de atención al cliente, ventas y cobranzas en los últimos años. Este crecimiento ha impulsado una búsqueda continua de mayores niveles de eficacia y eficiencia operativa en la gestión de usuarios, especialmente en el contexto de la recuperación de carteras y la optimización de tiempos de gestión.

Objetivo: el objetivo de este estudio es analizar datos históricos mediante técnicas de análisis de datos exploratorios y modelos de aprendizaje automático para identificar estrategias que mejoren la efectividad operativa, específicamente en términos de la cantidad de carteras recuperadas y el tiempo necesario para completar las tareas de gestión.

Metodología: la metodología sigue el marco del ciclo de vida de los datos para proyectos de aprendizaje automático, abarcando seis etapas: desde la adquisición de datos hasta la implementación del modelo. Se aplicaron análisis exploratorios para comprender los patrones en los datos y luego se implementaron modelos de aprendizaje automático para prever y mejorar el rendimiento de la gestión de carteras.

Resultados: los resultados obtenidos se compararon con el modelo basado en reglas utilizado actualmente por la empresa y un enfoque de gestión manual basado en la experiencia de los analistas. Los resultados muestran una mejora significativa del 21,8% en la eficacia respecto a la gestión manual y una mejora del 519,51% en comparación con el modelo basado en reglas existente.

Conclusiones: el estudio demuestra que la implementación de modelos de aprendizaje automático puede mejorar considerablemente la eficiencia operativa en los contact center virtuales, superando significativamente los enfoques tradicionales basados en reglas y la gestión manual. Estos resultados destacan el potencial de la inteligencia artificial para transformar la gestión de usuarios en el ámbito de los servicios BPO, mejorando tanto la recuperación de carteras como los tiempos de ejecución de las tareas.

Palabras clave: Recuperación de Deuda, Cartera Castigada, Machine Learning, Contact Center Virtual

Abstract

Introduction: virtual contact centers providing BPO services have experienced exponential growth in the customer service, sales, and collections sectors in recent years. This expansion has driven a continuous search for greater operational efficiency and effectiveness in user management, especially in the context of portfolio recovery and optimizing management task times.

Objective: the objective of this study is to analyze historical data using exploratory data analysis and machine learning models to identify strategies that improve operational effectiveness, specifically in terms of the number of portfolios recovered and the time required to complete management tasks.

Methodology: the methodology follows the data lifecycle framework for machine learning projects, covering six stages: from data acquisition to model implementation. Exploratory analysis was applied to understand patterns in the data, and machine learning models were implemented to predict and improve portfolio management performance.

Results: the results were compared with the rule-based model currently used by the company and a manual management approach based on the analysts' experience. The results demonstrate a 21.8% improvement in effectiveness compared to manual management and a 519.51% improvement over the existing rule-based model.

Conclusions: the study shows that the implementation of machine learning models can significantly enhance operational efficiency in virtual contact centers, greatly surpassing traditional rule-based approaches and manual management. These results highlight the potential of artificial intelligence to transform user management in the BPO services industry, improving both portfolio recovery and task execution times.

Keywords: Debt Recovery, Punished Portfolio, Machine Learning, Virtual Contact Center.



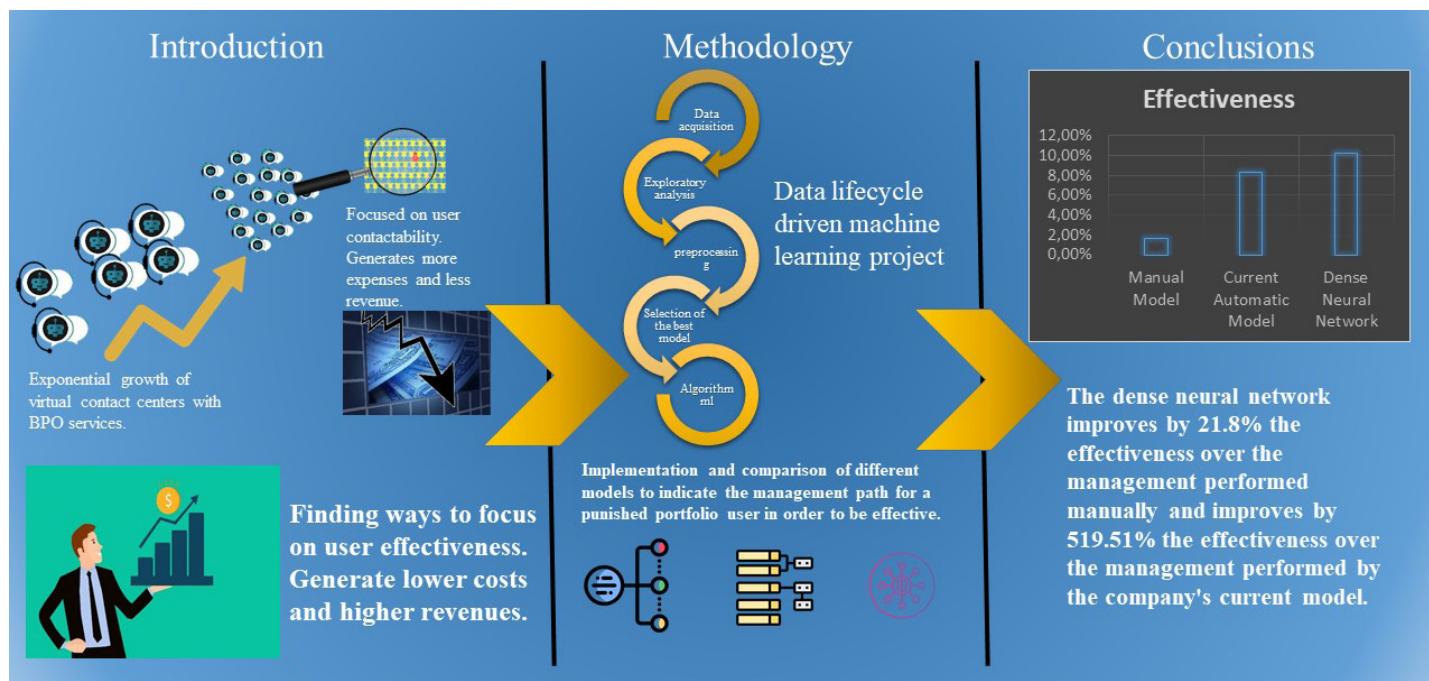
Contribución a la literatura

¿Por qué se realizó?

Este estudio surgió de las oportunidades que ofrece el aprendizaje automático para optimizar y automatizar los procesos de negocio. En particular, aplicamos estos métodos en el contexto de las tareas relacionadas con la recuperación de carteras castigadas mediante servicios de contacto virtual, donde muchas decisiones se basan tradicionalmente en la experiencia personal de los responsables de la operación de recuperación. El análisis de datos históricos para identificar patrones y tendencias ofrece la oportunidad de mejorar el proceso mediante un enfoque sistemático en la toma de decisiones.

¿Cuáles fueron los resultados más relevantes?

Los resultados muestran que el uso de métodos analíticos mejora la eficacia del proceso de recuperación en un 21,8 % en comparación con las operaciones realizadas por gestores humanos. Esto representa una oportunidad para mejorar los recursos, la eficiencia del tiempo y la rentabilidad del proceso.



Introducción

Los centros de contacto virtuales surgen como una solución innovadora para las empresas que buscan mejorar la eficiencia operativa relacionada con la comunicación con los usuarios. Estos centros aprovechan varias tecnologías para automatizar la gestión de usuarios en múltiples canales, incluidos el servicio al cliente, las ventas y las cobranzas, lo que permite una participación más rápida y eficiente de los usuarios. Los centros de contacto virtuales están remodelando el panorama de la industria de la externalización de procesos empresariales (BPO). Hoy en día, los servicios de BPO están impulsados por la inteligencia artificial, que ayuda a optimizar los procesos de comunicación, el procesamiento de datos, la predicción del comportamiento y otros factores críticos que impulsan la productividad en este sector.

Actualmente, las técnicas de inteligencia artificial (IA) más utilizadas en la industria de BPO son los modelos de procesamiento del lenguaje natural (NLP), empleados principalmente para mejorar la contactabilidad con el cliente. Estos modelos utilizan el análisis automatizado para procesar el audio, con el objetivo de comprender la intención del usuario. En el caso de los chatbots, aprovechan el procesamiento de texto para permitir interacciones más humanas (1), (2). Es importante tener en cuenta que, en el caso de las carteras de deuda deterioradas o vencidas, la eficacia de la recuperación de la deuda es un indicador aún más crítico que la contactabilidad. Si bien la contactabilidad es esencial para la comunicación con el usuario y representa el tipo de modelo más implementado en BPO (3), la efectividad de la recuperación de deudas ofrece un mayor valor. Reduce los costos de producción, acelera el retorno de la inversión y aumenta los ingresos monetarios tanto para el cliente como para el BPO (4), (5).

Esta investigación tiene como objetivo implementar técnicas de análisis de datos para mejorar la efectividad de la recuperación de deudas en comparación con el modelo actual de la empresa. El modelo existente se basa en un conjunto de reglas diseñadas para replicar los procesos de gestión manual realizados por los analistas de la empresa. Para mejorar la eficacia a través del análisis de datos históricos, el estudio adoptará la metodología del ciclo de vida del proyecto basada en datos. Este enfoque incluye la limpieza y transformación de datos, el análisis exploratorio de los datos recopilados y la implementación de modelos de aprendizaje automático para identificar las estrategias de gestión óptimas para los usuarios con carteras de deudas impagadas. Como estrategia de validación, se evaluará el rendimiento de varios modelos y se comparará con las métricas del modelo actual de la empresa (6), (7), (8).

El artículo está estructurado de la siguiente manera: La sección 2 proporciona una explicación detallada de la metodología basada en el enfoque del ciclo de vida del proyecto basado en datos. En el capítulo 3 se presentan los resultados del análisis exploratorio y la evaluación de diversos modelos predictivos. En el capítulo 4 se analizan los resultados obtenidos a través de la experimentación. Finalmente, en el capítulo 5 se presentan las conclusiones.

Metodología

La investigación se llevará a cabo en un ordenador de sobremesa equipado con un procesador Intel Core i5 de 12^a generación, 32 GB de RAM DDR5 que funciona a 4800 MT/s, 1,5 TB de

almacenamiento M.2 y una GPU Nvidia GeForce RTX 3060 TI de 8 GB. La metodología sigue el marco del ciclo de vida de los datos para los proyectos de aprendizaje automático, que abarca seis etapas, desde la adquisición de datos hasta la implementación del modelo.



Figura 1. Marco de ciclo de vida para proyectos de aprendizaje automático

Adquisición de datos

Los datos se extrajeron de las actividades de gestión realizadas sobre los usuarios de un único cliente de cobro de deudas atendido por el BPO. El conjunto de datos abarca el período de junio de 2022 a marzo de 2023 e incluye 23 columnas que contienen información del cliente. Además, se extrajo una tabla de nombres y apellidos, especificando el género asociado para cada nombre. Además, se generó una tabla que describe los tipos de estrategias empleadas por el BPO para la recuperación de deudas de cada cliente, indicando qué estrategias fueron efectivas y cuáles no. Los datos extraídos se cargaron en una base de datos del entorno local para su validación y limpieza. Se creó una tabla con todas las transacciones, dando como resultado un total de 263.865.057 registros, junto con una tabla de nombres que contiene 740.362 registros, y otra tabla para homologaciones con 492 registros.

Tabla 1. Características seleccionadas

Nombre	Descripción	Tipo de dato
identificación	Homologación de identificación.	Entero
intento	Intento en el que se realizó la transacción.	Entero
nombre	Nombre completo del usuario.	Cuerda
interaction_date	Fecha de la transacción.	Fecha AAAA-MILÍMETRO-DD
F5	Fecha y hora de la transacción.	Fecha AAAA-MM-DD HH:MM:SS
intearction_time	Hora de inicio de la transacción.	Fecha HH:MM:SS
answer_time	Hora de inicio de la interacción.	Fecha HH:MM:SS
interaction_endtime	Hora de finalización de la transacción.	Fecha HH:MM:SS
id_result	Tipificación de transacciones.	Entero
Fecha de pago1	Possible fecha de pago 1.	Fecha AAAA-MILÍMETRO-DD
Fecha de pago2	Possible fecha de pago 2.	Fecha AAAA-MILÍMETRO-DD
payment_agreement	Monto a pagar en la transacción.	Flotar
deuda	Deuda del usuario.	Flotar
overdue_days	Días de mora en el pago de la deuda	Entero
seña	ID alfanumérico único asignado a cada usuario.	Cuerda
URL	URL a enviar por SMS.	Cuerda
id_allocation	Número asignado por el sistema a la asignación.	Entero
v11	Grupo de transacciones anteriores.	Cuerda
v13	Grupo de transacciones actual.	Cuerda
v14	Fecha máxima de pago.	Fecha AAAA-MILÍMETRO-DD
Chanel	Canal a través del cual se realizó la transacción.	Cuerda
max_attempts	Número de intentos que tuvo el usuario en el momento de la transacción	Entero
Cuotas	Cuotas ofrecidas por el usuario en un contrato.	Entierro

Una vez cargados los datos, se llevó a cabo un proceso de limpieza de datos. Durante este proceso, se eliminaron las entradas duplicadas resultantes de la superposición de consultas mensuales a la base de datos. Se excluyeron los registros con un valor de 0 en la variable intento, ya que representan estados iniciales. Además, se eliminaron las columnas url, id, v14, max_attempts, cuotas y token porque son variables personalizadas definidas por el administrador, generadas aleatoriamente e irrelevantes para el análisis de recuperación de deudas. La columna answer_time se quitó, porque el conjunto de datos ya incluía los interaction_date y interaction_endtime que son valores críticos para determinar las interacciones correctas del usuario. Posteriormente, se agregaron nuevas características al conjunto de datos. Los registros se cruzaron con la columna de nombre para agregar el género y con la columna de identificación para incorporar las variables

de efectividad y gestión. Es importante tener en cuenta que, en este contexto, la gestión se refiere a una acción realizada para recuperar la deuda, mientras que una interaction_date se produce solo cuando el usuario final responde a esa acción.

Posteriormente, los 76.203.786 registros resultantes se almacenaron en una nueva tabla, ya que la atención se centró en la transacción actual con el usuario final. Del mismo modo, se excluyeron las columnas id y name, ya que solo eran relevantes para extraer el género y la efectividad. Las columnas restantes se renombraron de la siguiente manera: v11 a previous_bag, v13 a current_bag y payment_agreement a amount_paid, Las columnas resultantes fueron: intento, pago_date1, payment_date2, amount_paid, deuda, canal, cuotas, current_bag id, género, efectividad, gestión, interaction_time y interaction_endtime.

A partir de los datos resultantes, se eliminó el canal de SMS, ya que se mostró sistemáticamente ineficaz y no proporcionó información relevante. También se eliminaron los registros que carecían de efectividad, aquellos con valores de bolsa atípicos (como vacío o cero) y los registros en los que la columna de administración tenía un valor **Falso**, ya que no aportaban información significativa debido a la falta de interacción. Por último, se creó una nueva tabla extrayendo solo los registros efectivos y sus correspondientes contrapartes ineficaces. Se eliminaron las columnas de gestión y canal, ya que ya no eran útiles. Este proceso dio como resultado un total de 12.178.681 registros disponibles para análisis exploratorio.

Análisis exploratorio de datos

Para el análisis exploratorio se formularon seis preguntas para comprender mejor los datos extraídos e identificar las características más importantes para abordar el problema y posteriormente entrenar los modelos. Estas preguntas se desarrollaron a partir de las observaciones de la eficacia durante el proceso manual de recuperación de deudas. La primera pregunta es la siguiente: ¿Existe estacionalidad en los datos relacionados con la efectividad a lo largo del tiempo? Esta pregunta fue diseñada para evaluar la aplicabilidad de un modelo de series temporales. Para responder a esta pregunta, se creó un gráfico de serie temporal utilizando 1.059.572 entradas, y los resultados se presentan en la Figura 2. Un simple análisis visual no reveló ningún patrón discernible en los datos a lo largo del tiempo, ya que no exhiben periodicidad. Este hallazgo sugiere que los modelos de series temporales no son adecuados para este problema.

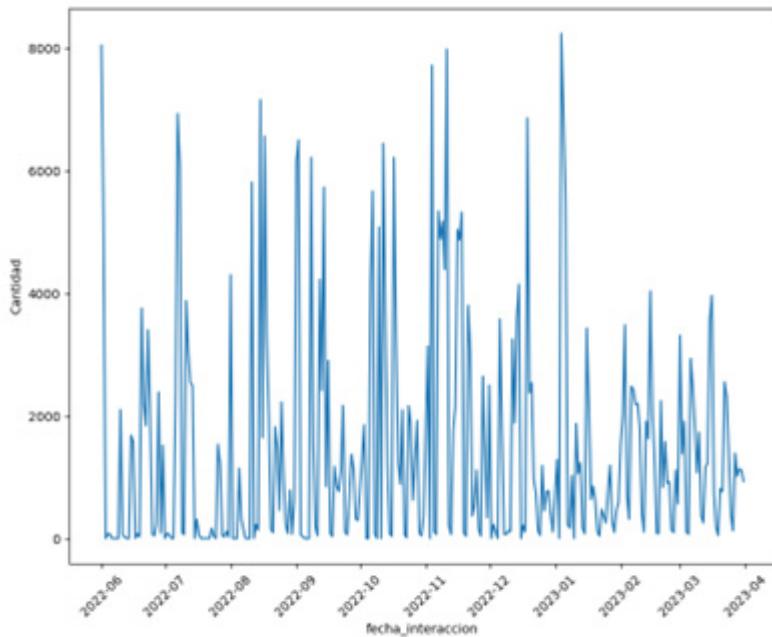


Figura 2. Datos relativos a la efectividad.

La segunda pregunta fue: ¿Qué grupos tienen la mayor efectividad? Para responder a esto, los registros se agruparon por categoría, se contó el número de registros en cada grupo y se excluyeron los grupos con menos de 1.000 muestras para garantizar resultados significativos. Al analizar la efectividad de los grupos, se encontró que de 76 grupos con efectividad medible, solo 37 fueron significativos, como se muestra en la Figura 3.

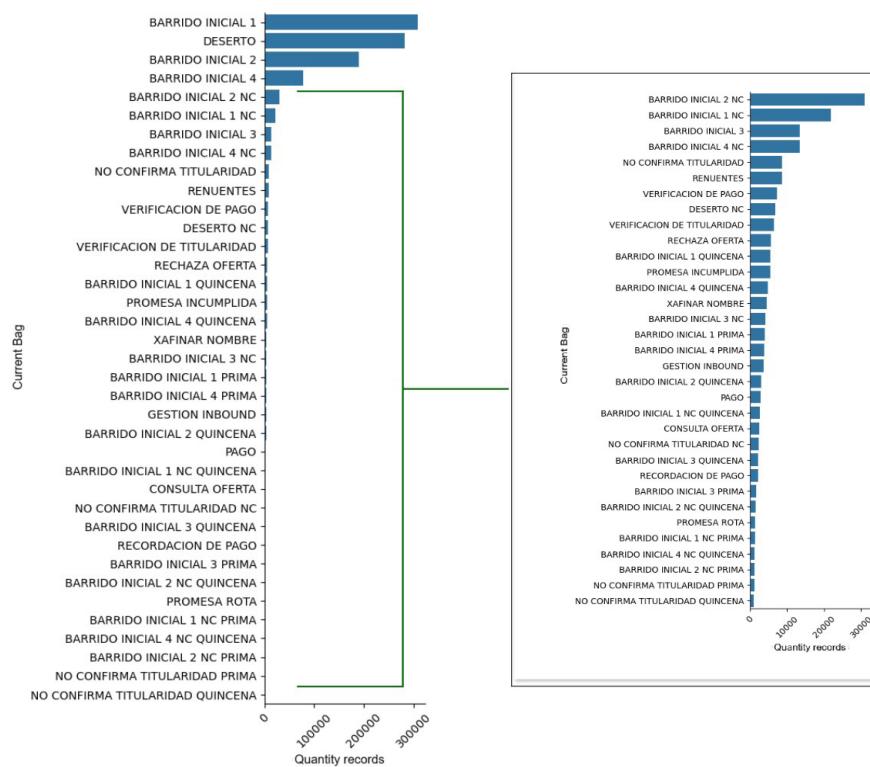


Figura 3. Eficacia de una bolsa

La tercera pregunta fue: ¿Existe relación entre los rangos de deuda y los días de cuotas? Esta pregunta se propuso para explorar la relación entre los días de morosidad y la deuda de un usuario agrupándolos en cuatro rangos correspondientes a los percentiles: 0-25, 25-50, 50-75 y 75-100. Tanto los días de deuda como los de morosidad se clasificaron utilizando estos rangos, que se basan en las mejores prácticas de cobro empleadas por el BPO. Los resultados se presentan en la Figura 4. Se puede observar que los usuarios más efectivos se encuentran dentro de los rangos de deuda 1 y 2. Además, existe una fuerte relación entre estos rangos de deuda y los rangos diarios.

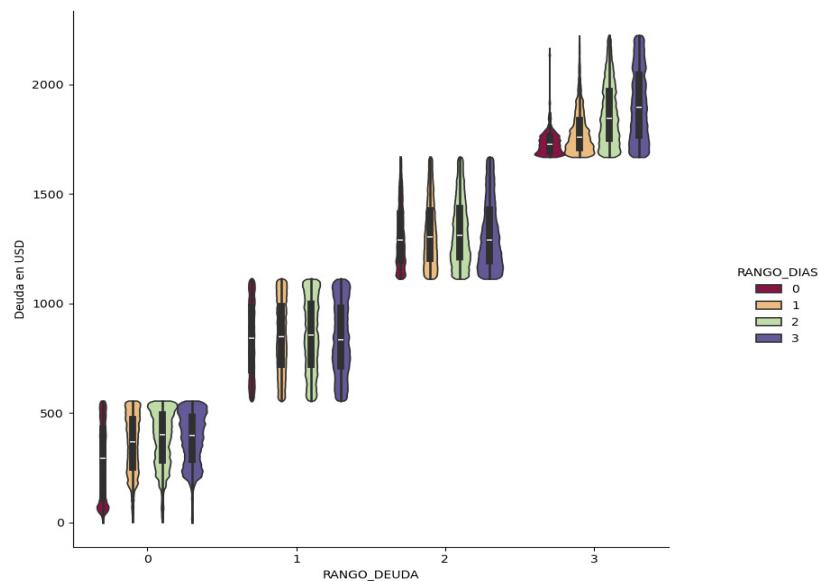


Figura 4. Gráficos de violín de rangos de deuda versus deuda en dólares y rangos de días.

La cuarta pregunta fue: ¿Es relevante la variable género en los datos para entrenar un modelo? Para abordar esto, se calculó la proporción de usuarios efectivos por género, sin revelar diferencias significativas en la efectividad basada en el género. Los resultados se muestran en la Figura 5.

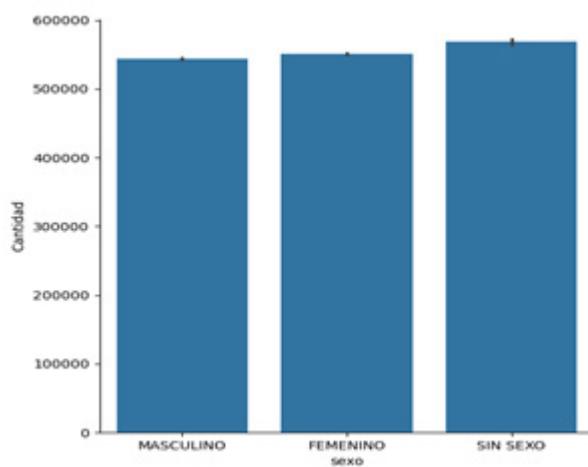


Figura 5. Eficacia por género

La quinta pregunta fue: ¿Qué variables están correlacionadas entre sí? Para responder a esta pregunta, las características ordinales se transformaron en valores numéricos y se calcularon las correlaciones por pares entre columnas (excluyendo las columnas de horas y fechas) utilizando el coeficiente de correlación de Pearson. Esta métrica mide la relación estadística entre dos variables numéricas y puede tomar valores entre -1 y +1. Un valor mayor que 0 indica una asociación positiva entre las variables, lo que significa que a medida que una aumenta, la otra también lo hace. Por el contrario, un valor inferior a 0 indica una asociación negativa, en la que la relación se debilita a medida que aumenta una variable. En la Figura 6 se presentan los resultados. El análisis reveló que los días de morosidad, la deuda y la efectividad están correlacionados entre sí.

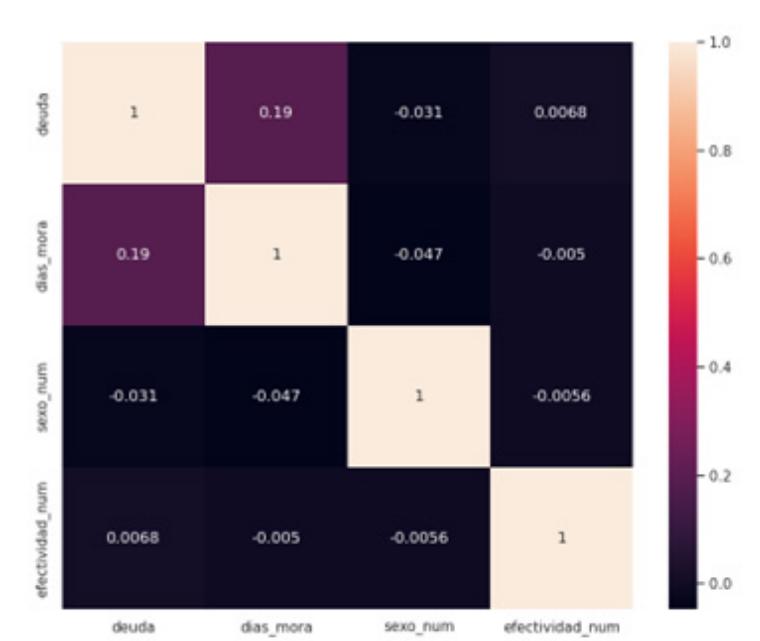


Figura 6. Matriz de correlación de Pearson

Finalmente, se planteó la siguiente pregunta: ¿Existe una relación entre los grupos para establecer un camino de gestión? Para abordar esto, se aplicó un modelo de análisis de frecuencia y se utilizaron reglas de asociación para identificar las relaciones entre los grupos. El modelo utilizado fue FP-Growth de la biblioteca Apache Spark ML en Python, ya que ofrece bases matemáticas más sólidas para generar reglas de asociación en comparación con otros modelos y admite la agrupación. La configuración del modelo incluyó un minSupport de 0,0001 para identificar patrones muy raros, es decir, aquellos que ocurren en una proporción muy pequeña del total de transacciones, ya que los casos EFECTIVOS son pocos en comparación con los NO EFECTIVOS. Se estableció un minConfidence de 0,6 para garantizar una fiabilidad mayor que la aleatoria, y numPartitions se estableció en 96, ya que la agrupación en clústeres se realizó con 96 procesadores lógicos. En la Tabla 2 se presentan los resultados obtenidos.

Tabla 2. Ejemplo de resultados de reglas de asociación

Antecedente	Consistente	Confianza	Levantar	Apoyo
[Deserto nc, barrido inicial 4, barrido inicial 2, barrido inicial 1]	[efectivo]	1	1.89579257	2.44e-04
[Barrido inicial 2 cierre, barrido inicial 2, deserto, barrido inicial 1]	[efectivo]	1	1.89579257	1.23e-04
[Barrido inicial 3 quincena, barrido inicial 2, deserto, barrido inicial 1]	[efectivo]	1	1.89579257	2.53e-04
[Barrido inicial 4 nc, barrido inicial 2 nc, barrido inicial 4, barrido inicial 2, deserto]	[efectivo]	1	1.89579257	2.85e-04 3.23e-04
[Nc desert, comenzando barrido 4, comenzando barrido 2, desert]	[efectivo]	1	1.89579257	

Los resultados de la asociación muestran que se puede crear una relación entre los grupos y vías de gestión, indicando qué grupos tienen la mayor probabilidad de ser efectivos. Esta conclusión se basa en la variable "elevación", que mide la fuerza de la relación entre el antecedente y el consecuente en comparación con su ocurrencia independiente. Un valor de elevación mayor que 1 indica que la ocurrencia del antecedente está asociada positivamente con la ocurrencia del consecuente. Cuanto mayor sea este valor, más fuerte será la relación.

Una vez concluido el análisis exploratorio de los datos, se extrajeron las columnas correspondientes y se creó una nueva tabla con los datos filtrados, eliminando los duplicados basados en las columnas id y current_bag. Esto dio como resultado un total de 2.766.220 registros y seis columnas: id, amount_paid, deuda, cuotas, current_bag y efectividad, para su preprocesamiento y uso en el entrenamiento de modelos de inteligencia artificial.

Preprocesamiento de datos

Durante el preprocesamiento de los datos se analizó la distribución de los datos para cada grupo, revelando que los grupos "BARRIDO INICIAL 1", "BARRIDO INICIAL 2", "DESERTO" y "BARRIDO INICIAL 1 NC" no eran comparables con otros grupos e introducían sesgos en el modelo. Este problema surge porque todos los usuarios pasan por estos grupos independientemente de la configuración de otros parámetros, lo que impide que el modelo diferencie eficazmente los registros entre ellos. En consecuencia, estos grupos fueron eliminados. Además, se excluyó el grupo 'PAGO', ya que contenía solo la clase efectiva. Este proceso dio como resultado 1.956.403 registros, que se utilizaron para crear grupos separados, uno para cada una de las categorías restantes. Se utilizaron las características amount_paid, deuda y overdue_days, con la efectividad como etiqueta. Este enfoque permitió el entrenamiento de un modelo separado para cada grupo, con el objetivo de desarrollar un modelo multiobjetivo personalizado. Este procedimiento se muestra en la Figura 7.

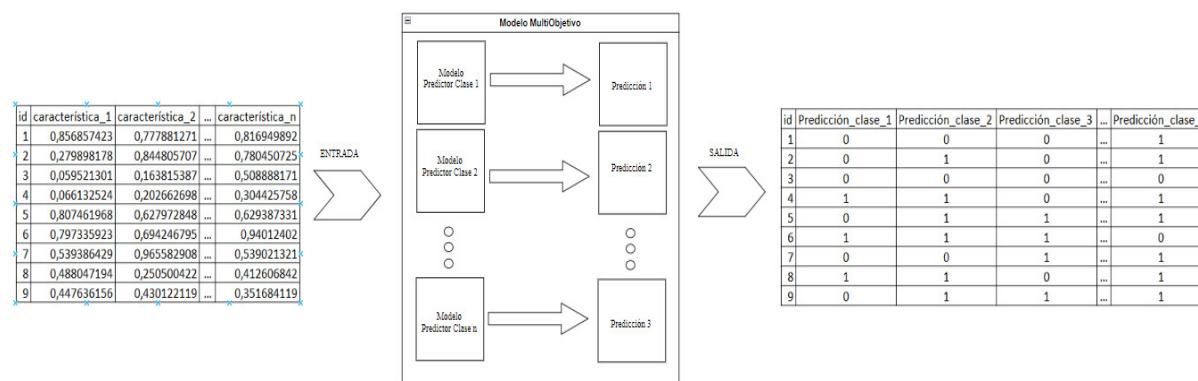


Figura 7. Arquitectura de modelos multiobjetivo

Se realizó un submuestreo de los datos para cada uno de los grupos generados, seguido de una división de los datos transformados en conjuntos de entrenamiento y prueba en una proporción de 70/30: 70% para los datos de entrenamiento y 30% para los datos de prueba. Cada grupo fue sometido al método de submuestreo para equilibrar las clases "EFFECTIVO" y "NO EFFECTIVO". Para ello, se utilizó la librería de Python ImbalancedLearn, concretamente el método RandomUnderSampler, que selecciona aleatoriamente muestras de la clase mayoritaria hasta alcanzar una cantidad definida (9). En este caso, la clase 'NON-EFFECTIVE' se redujo a tres veces la cantidad de la clase 'EFFECTIVE', asegurando que la clase 'EFFECTIVE' no excediera su cantidad actual. Estas divisiones se aplicaron para lograr un equilibrio adecuado de los datos para cada grupo, asegurando que el modelo no prefiera excesivamente a una clase, sino que sea equitativo entre ellas.

Para transformar los datos, se creó una canalización. La canalización de cada modelo incluía dos métodos: uno para llenar los datos que faltaban y otro para escalar los datos. Para el primer método, utilizado para llenar los datos faltantes, se empleó la función SimpleImputer del módulo de preprocessamiento de Scikit-learn. Esta función utiliza estadísticas básicas, como el valor medio, la mediana o el valor más frecuente, para llenar los valores que faltan para cada entidad. En nuestro caso, se utilizó la media según la distribución de los datos. Para el segundo método, se utilizó la función MinMaxScaler de la biblioteca de preprocessamiento Scikit-learn. Esta función es un transformador que escala cada característica individualmente para que se ajuste al rango especificado en el conjunto de entrenamiento, que está de forma predeterminada entre cero y uno.

Método Experimental

El primer modelo entrenado es Random Forests, definido por Breiman como una combinación de predictores de árboles, en el que cada árbol depende de los valores de un vector aleatorio muestreado de forma independiente y con la misma distribución para todos los árboles del bosque. El error de generalización de un bosque de árboles de clasificación depende de la fuerza de cada árbol en el bosque y de la correlación entre ellos. Las estimaciones internas controlan el error, la fuerza y la correlación, y se utilizan para evaluar la respuesta del modelo al aumentar el número de características utilizadas en la división (10). La canalización se ejecutó mediante GridSearchCV, que determinó que los mejores hiperparámetros para el modelo son 200 n_estimators, el criterio de Gini y un max_depth de 20. Los resultados se muestran en la Tabla 3.

Tabla 3. Resultados promedio del modelo de bosque aleatorio.

Métrico	Valor de la formación	Valor de la prueba
Exactitud	73.65%	73.65%
Precisión	63.22%	62.61%

Los resultados indican que el modelo se comporta satisfactoriamente en términos de exactitud, pero ligeramente menos en precisión. Esto puede atribuirse a casos en los que el modelo tiene dificultades para discriminar de manera efectiva debido a la naturaleza de los datos. El segundo modelo se desarrolló utilizando una máquina de vectores de soporte (SVM) con un kernel lineal (LinearSVC) para la clasificación. Este enfoque funciona mediante la identificación de un hiperplano que maximiza la distancia entre los puntos de datos más cercanos de diferentes clases, conocidos como vectores de soporte (11). El modelo se ejecutó mediante GridSearchCV, que confirmó que los hiperparámetros predeterminados eran óptimos. Los resultados se muestran en la Tabla 4. Observamos que la precisión de este modelo es mayor que la de Random Forests; sin embargo, su precisión es significativamente menor. Al igual que con el modelo anterior, no hay sobreajuste, pero la precisión sigue siendo considerablemente baja.

Tabla 4. Resultados promedio del modelo SVM

Métrico	Valor de la formación	Valor de la prueba
Exactitud	74.47%	74.51%
Precisión	55.56%	55.61%

El tercer modelo implementado fue el Extreme Gradient Boosting, comúnmente conocido como XGBOOST. Este modelo mejora las predicciones de clasificación mediante la construcción secuencial de árboles de decisión poco profundos mediante el aumento de gradiente, en el que cada nuevo árbol se ajusta a los residuos (errores) del anterior. XGBOOST emplea técnicas de optimización para lograr una convergencia más rápida y evitar el sobreajuste (12). La ejecución de la canalización con GridSearchCV determinó que los mejores hiperparámetros para el modelo eran: max_depth=6, n_estimators=200, learning_rate=0.1, subsample=0.8 y colsample_bytree = 0.8. Los resultados se presentan en la Tabla 5. Las métricas de este modelo son similares a las de Random Forests, con una exactitud y precisión ligeramente superiores, aunque no tienen un impacto significativo en aplicaciones prácticas.

Tabla 5. Resultados medios del modelo XGBOOST.

Métrico	Valor de la formación	Valor de la prueba
Exactitud	74.25%	74.28%
Precisión	64.69%	64.73%

Por último, se utilizó una Red Neuronal Densa como modelo final. Este modelo de aprendizaje profundo consiste en una serie de nodos interconectados, conocidos como neuronas. Cada neurona tiene entradas, una función de activación, pesos y una salida. El proceso de entrenamiento consiste en optimizar el peso de las neuronas para conseguir el mejor resultado posible. La red está organizada en capas, diseñadas para imitar la estructura del cerebro humano. La arquitectura de red neuronal propuesta se detalla en la Tabla 6.

Tabla 6. Configuración de redes neuronales densas.

Capa (tipo)	Forma de salida
dense (dense)	(Ninguno, 512)
dense_1 (denso)	(Ninguno, 256)
dense_2 (denso)	(Ninguno, 128)
dense_3 (denso)	(Ninguno, 64)
dense_4 (denso)	(Ninguno, 1)

La función de activación para las capas ocultas es ReLU, mientras que la capa de salida utiliza una activación sigmoide. La red neuronal se entrenó durante 30 épocas con el tamaño de lote predeterminado. Se empleó el optimizador Adam con una tasa de aprendizaje de 0,001, con la entropía cruzada binaria como función de pérdida y la precisión como métrica de evaluación. En la Tabla 7 se presentan los resultados. Observamos que las métricas de la Red Neuronal Densa son similares a las del modelo XGBOOST, con valores ligeramente más bajos, aunque no significativamente. Al igual que con los modelos anteriores, no hay sobreajuste; Sin embargo, la precisión sigue siendo menor para los datos no vistos.

Tabla 7. Resultados medios del modelo de Red Neuronal Densa.

Métrico	Valor de la formación	Valor de la prueba
Exactitud	74.67%	73.56%
Precisión	61.39%	57.99%

Después de validar los resultados de los modelos, se optimizaron eliminando la variable paid_amount, que contiene datos similares en diferentes casos para su clasificación. Esto puede haber causado que los modelos no discriminaran correctamente. Además, se evaluó la métrica de efectividad, ya que tiene una alta relevancia para esta investigación al indicar si el modelo es significativo en comparación con los resultados de efectividad actuales de la empresa. Los resultados de entrenamiento de los modelos sin esta variable se presentan en la Figura 8 y la Tabla 8.

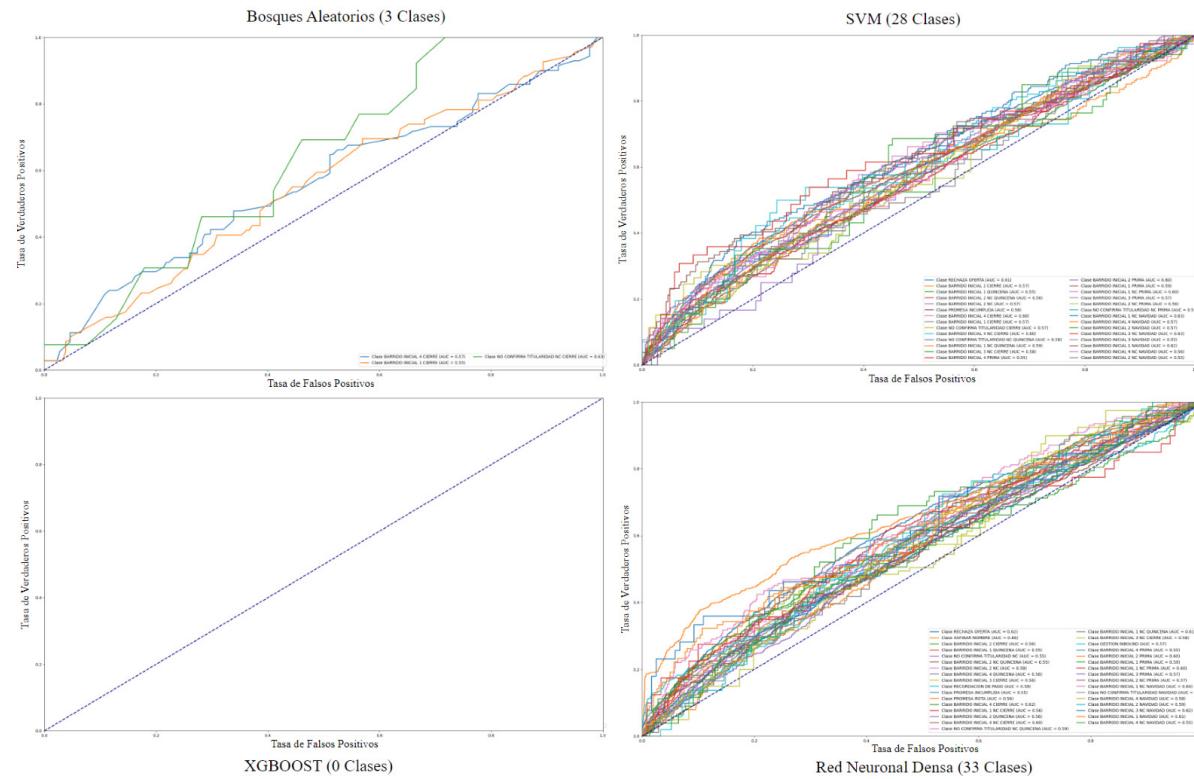
Figura 8. Comparación de curvas ROC con $AUC > 0.55$ para modelos optimizados

Tabla 8. Comparación de modelos optimizados.

Modelo	Número de bolsas con $AUC > 0,55$	Con la función <code>valor_pago</code>			Sin la función <code>valor_pago</code>		
		Exactitud	Precisión	Efectividad	Exactitud	Precisión	Efectividad
Bosque aleatorio	3	73.65%	62.61%	2.38%	74.51%	55.61%	0%
SVM	28	74.51%	55.61%	0%	74.51%	55.61%	0%
XGBOOST	0	74.28%	64.73%	1.38%	74.46%	61.67%	0.23%
Red neuronal densa	33	73.56%	57.99%	9.21%	73.39%	57.44%	10.17%

A partir de los resultados presentados en la Tabla 8, es posible observar que la eliminación de la característica “valor_pago” provocó una ligera disminución en la precisión de todos los modelos, una caída significativa en la efectividad de Random Forests y XGBOOST, y un aumento en la efectividad de la Red Neuronal Densa. Con base en estos resultados, se puede concluir que la Red Neuronal Densa es el modelo de mejor desempeño. Tiene una métrica superior a la aleatoria y logra una efectividad que supera la efectividad promedio actual de la compañía tanto para modelos automáticos como manuales. Particularmente, la Red Neuronal Densa, con 32 clases adecuadamente discriminadas, logra una efectividad de 10.17%, que es significativamente superior al promedio de 1.64% del modelo automático y el promedio de 8.35% de los procesos manuales en el BPO.

Resultados y Discusión

Cuando se utilizan modelos de aprendizaje automático para mejorar la efectividad operativa en un BPO, no se han propuesto ni aplicado modelos específicamente para este propósito. En cambio, la mayoría de los modelos existentes se centran en mejorar la contactabilidad o segmentar a los usuarios para evaluar los riesgos de conceder préstamos a particulares (13), (14). Sin embargo, estos modelos utilizan variables similares a las de la investigación propuesta, lo que permite una comparación del rendimiento basada en los datos y no en el contexto del problema, ya que los enfoques difieren. Además, algunos modelos emplean una métrica similar a la efectividad utilizada en esta investigación (15), (16), (17).

A partir de la selección de características, Martín Calero et al. (18) propusieron un modelo para identificar las características más relevantes para determinar qué usuarios tienen más probabilidades de pagar o incumplir. Sus resultados mostraron que de las 59 características (9 de las cuales pertenecen a la gestión de la máquina), solo 5 se identificaron como relevantes según el modelo de mejor rendimiento. Entre estos, el más notable es el de los días de atraso (denominados en esta investigación como "días de cuotas"), que tiene una puntuación de importancia del 85%. Este hallazgo pone de manifiesto la importancia de los días de mora como característica clave, conclusión que también apoya Arango Correa et al. (19), quienes realizaron un análisis de correlación utilizando el método de Pearson. Al seleccionar características por encima del percentil 66, redujeron las 33 características iniciales a las más significativas, incluidos los días más atrasados y el promedio de los últimos 4 meses. Si bien en ambos estudios se mencionan variables de monto monetario, no se proporciona ningún análisis que explique por qué se excluyeron. En ambos modelos, la precisión supera el 88%.

Swati Tyagi (20) comienza con un conjunto de datos que contiene 110 variables, de las cuales 80 se descartaron durante la fase inicial de limpieza de datos, dejando 30 variables. Estos incluían días de morosidad de deudas, saldos de deudas (como montos iniciales del préstamo, pagos realizados y deudas actuales) y otras características relacionadas con las operaciones de crédito. El modelo con mejor rendimiento alcanzó una precisión del 89%. A diferencia de otros estudios que se basan en medidas estándar como ROC/AUC, Tyagi también empleó el ROI como una métrica de rendimiento, que es particularmente crucial en contextos de cobro de deudas. El ROI reportado en el mejor de los casos fue del 3.8%. Esta métrica se puede comparar con la métrica de efectividad utilizada en la presente investigación, ya que una mayor efectividad se correlaciona con una menor inversión y mayores ingresos. El monto de la deuda y los días de morosidad son variables críticas en cualquier modelo financiero, ya sea para la clasificación del usuario o la predicción del monto del préstamo. Varios conjuntos de datos en este campo incorporan estas variables, como el conjunto de datos "GiveMeSomeCredit" de Credit Fusion y Will Cukierski (21), que se utilizó en un concurso de Kaggle en 2011. El reto de este concurso era desarrollar un modelo que ayudara en la toma de decisiones sobre préstamos.

Por lo tanto, es posible comparar varios aspectos de esta investigación y validar su calidad. El primer aspecto pone de manifiesto que, si bien las empresas pueden aportar numerosas variables, no todas son útiles para entrenar un modelo. Varios estudios demuestran que la selección de características puede reducir significativamente el conjunto inicial de variables. El segundo aspecto

confirma la relevancia de las variables de deuda y morosidad para los modelos financieros y de cobranza, tal como lo apoyan los hallazgos de otros investigadores. El tercer aspecto es que las métricas de rendimiento del modelo pueden permanecer altas cuando se aplican tanto a los datos de entrenamiento como a los datos no vistos anteriormente, con tasas de precisión superiores al 65%. Por último, aunque los modelos entrenados pueden lograr buenas métricas de rendimiento, es necesario un paso adicional para validar su funcionalidad. Esto se exemplifica en el trabajo de Swati Tyagi (20), donde el ROI se utiliza para evaluar resultados más allá del conjunto de datos.

Conclusiones

En la presente investigación se realizó un análisis exhaustivo de los datos recolectados, desde su adquisición hasta su exploración, con el objetivo de comprender en profundidad las características y tendencias en las operaciones realizadas por los gerentes de operaciones y el modelo actual de la empresa. Este análisis permitió identificar patrones y relaciones relevantes para el proceso de gestión de carteras, confirmando la importancia de las variables monto de la deuda y días de morosidad, como también lo han destacado otros investigadores en las áreas de BPO y cobranza de deudas.

Posteriormente, se implementaron cuatro modelos de aprendizaje automático, incluido un modelo de aprendizaje profundo: Random Forests, SVM, XGBOOST y una red neuronal densa. Estos fueron diseñados para identificar el enfoque más efectivo para la gestión de usuarios morosos de la cartera en un BPO. Se exploraron varios algoritmos, técnicas de preprocesamiento y enfoques de modelado para encontrar la mejor solución para mejorar la eficacia de la gestión y minimizar los costos de producción.

Una vez implementados los modelos, se analizaron sus resultados utilizando métricas de evaluación como Exactitud, Precisión, y la métrica de comparación con el modelo de gestión actual, Efectividad, así como curvas ROC para medir la capacidad predictiva y el rendimiento general de cada modelo. Esta evaluación identificó la Red Neuronal Densa como el modelo más prometedor para su aplicación práctica.

Finalmente, se compararon los resultados del modelo seleccionado con las métricas de rendimiento actuales de la empresa. Esta comparación reveló que la Red Neuronal Densa mejora la efectividad de la gestión en un 21,8% sobre las operaciones realizadas por los gerentes, cuya efectividad promedio actual es del 8,35%. También mejora la efectividad en un 519,51% con respecto al modelo de árbol de decisión actual de la compañía, que tiene una efectividad promedio del 1,64%. Esto pone de manifiesto las fortalezas y ventajas de la implementación de este tipo de modelos. Con la aplicación de este modelo, la empresa podría aumentar sus ingresos mensuales de \$20,000 a \$103,902 USD.

Declaración de contribución de autoría CRediT

Conceptualización - Ideas: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Curación de datos: Camilo Andrés Leuro Camacho. Análisis formal: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Investigación: Camilo Andrés Leuro Camacho. Metodología: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Dirección de Proyecto: Juan Pablo Ospina López. Recursos:

Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Software: Camilo Andrés Leuro Camacho.

Supervisión: Juan Pablo Ospina López. Validación: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Visualización - Elaboración: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Redacción - borrador original - Elaboración: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Redacción - revisión y edición - Elaboración: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López.

Financiación: no declara.

Conflicto de intereses: no declara.

Implicaciones éticas: no declara.

References

1. L. Wang, N. Huang, Y. Hong, L. Liu, X. Guo, and G. Chen, "Voice-based AI in call center customer service: A natural field experiment," *Production and Operations Management*, vol. 32, no. 4, pp. 1002-1018, 2023. <https://doi.org/10.1111/poms.13953>
2. B. Li and L. Liu, "Does voice-based AI improve call center operational performance? Event study in a working telecommunication company," in *PACIS 2021 Proceedings*, vol. 132, 2021. Available: <https://aisel.aisnet.org/pacis2021/132>.
3. F. Kreuter and G. Müller, "A note on improving process efficiency in panel surveys with paradata," *Field Methods*, vol. 27, no. 1, pp. 55-65, 2015. <https://doi.org/10.1177/1525822X14538205>
4. Z. Zhang, B. Li, and L. Liu, "The impact of AI-based conversational agent on the firms' operational performance: Empirical evidence from a call center," *Applied Artificial Intelligence*, vol. 37, no. 1, 2022. <https://doi.org/10.1080/08839514.2022.2157592>
5. S. Ozan, "Classification of audio segments in call center recordings using convolutional recurrent neural networks," *arXiv e-prints*, Art. no. arXiv:2106.02422, 2021. Available: <https://arxiv.org/pdf/2106.02422.pdf>
6. C. Sánchez, S. Maldonado, and C. Vairetti, "Improving debt collection via contact center information: A predictive analytics framework," *Decision Support Systems*, vol. 159, p. 113812, 2022. <https://doi.org/10.1016/j.dss.2022.113812>
7. M. Pinelis and D. Ruppert, "Machine learning portfolio allocation," *The Journal of Finance and Data Science*, vol. 8, pp. 35-54, 2022. <https://doi.org/10.1016/j.jfds.2021.12.001>
8. K. Sivamayilvelan, et al., "Flexible recommendation for optimizing the debt collection process based on customer risk using deep reinforcement learning," *Expert Systems with Applications*, vol. 256, p. 124951, 2024. <https://doi.org/10.1016/j.eswa.2021.12.001>
9. G. Lemaitre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, pp. 1-5, 2017. Available: <https://inria.hal.science/hal-01516244v1>.

10. L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. <https://doi.org/10.1023/A:1010933404324>
11. D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine Learning*, Academic Press, 2020, pp. 101-121. <https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
12. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, New York, NY, USA: Association for Computing Machinery, 2016, pp. 785-794. <https://doi.org/10.1145/2939672.2939785>
13. Á. F. Pinheiro, D. S. da Silveira, and F. B. de Lima Neto, "Use of machine learning for active public debt collection with recommendation for the method of collection via protest," in *CS & IT Conference Proceedings*, vol. 12, no. 9, May 2022. <https://doi.org/10.5121/csit.2022.120909>
14. E. Dumitrescu, et al., "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178-1192, 2022. <https://doi.org/10.1016/j.ejor.2021.06.053>
15. M. Doumpos, et al., "Operational research and artificial intelligence methods in banking," *European Journal of Operational Research*, vol. 306, no. 1, pp. 1-16, 2023. <https://doi.org/10.1016/j.ejor.2022.04.027>
16. S. Akter, et al., "Algorithmic bias in machine learning-based marketing models," *Journal of Business Research*, vol. 144, pp. 201-216, 2022. <https://doi.org/10.1016/j.jbusres.2022.01.083>
17. A. M. Benavides, et al., *Modelo para la predicción de interacciones recibidas en un centro de atención telefónico*, Tesis de maestría, Facultad de Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia, 2022. [Online]. Available: <http://hdl.handle.net/10554/61764>
18. M. Calero, M. Baldeon Calisto, S. Bonilla, and D. Riofrio, "Application of machine learning algorithms for the prediction of payment by agreement in a debt collection company with the CRISP-DM methodology," in *3rd South American International Conference on Industrial Engineering and Operations Management*, May 2022. <https://doi.org/10.46254/SA03.20220112>
19. D. M. Arango Correa, L. J. Colmenares Colmenares, and I. C. Rave Contreras, "Comparación entre el método tradicional y algunos basados en inteligencia artificial para el estudio del riesgo crediticio en instituciones financieras colombianas," 2018. Available: <http://hdl.handle.net/10784/12983>
20. S. Tyagi, "Analyzing machine learning models for credit scoring with explainable AI and optimizing investment decisions," *arXiv preprint*, arXiv:2209.09362, 2022. Available: <https://arxiv.org/abs/2209.09362>
21. Kaggle, "Give Me Some Credit - Kaggle Competition," Kaggle, Available: <https://www.kaggle.com/competitions/GiveMeSomeCredit> [Accessed: Feb. 19, 2025].