Vol 27 (1)

Ingeniería y Competitividad
Vol 27 (1) 2025 doi: 10.25100/iyc.v27i1.14327

ISSN
0123-3033
e- 2027-8284

Original Article

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

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

Camilo A. Leuro[1] Juan P. Ospina[1]

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

How to cite?

Leuro, C.A., Ospina, J.P., Data Analysis for the Management of Delinquent Portfolios in Virtual Contact Centers. 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

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.

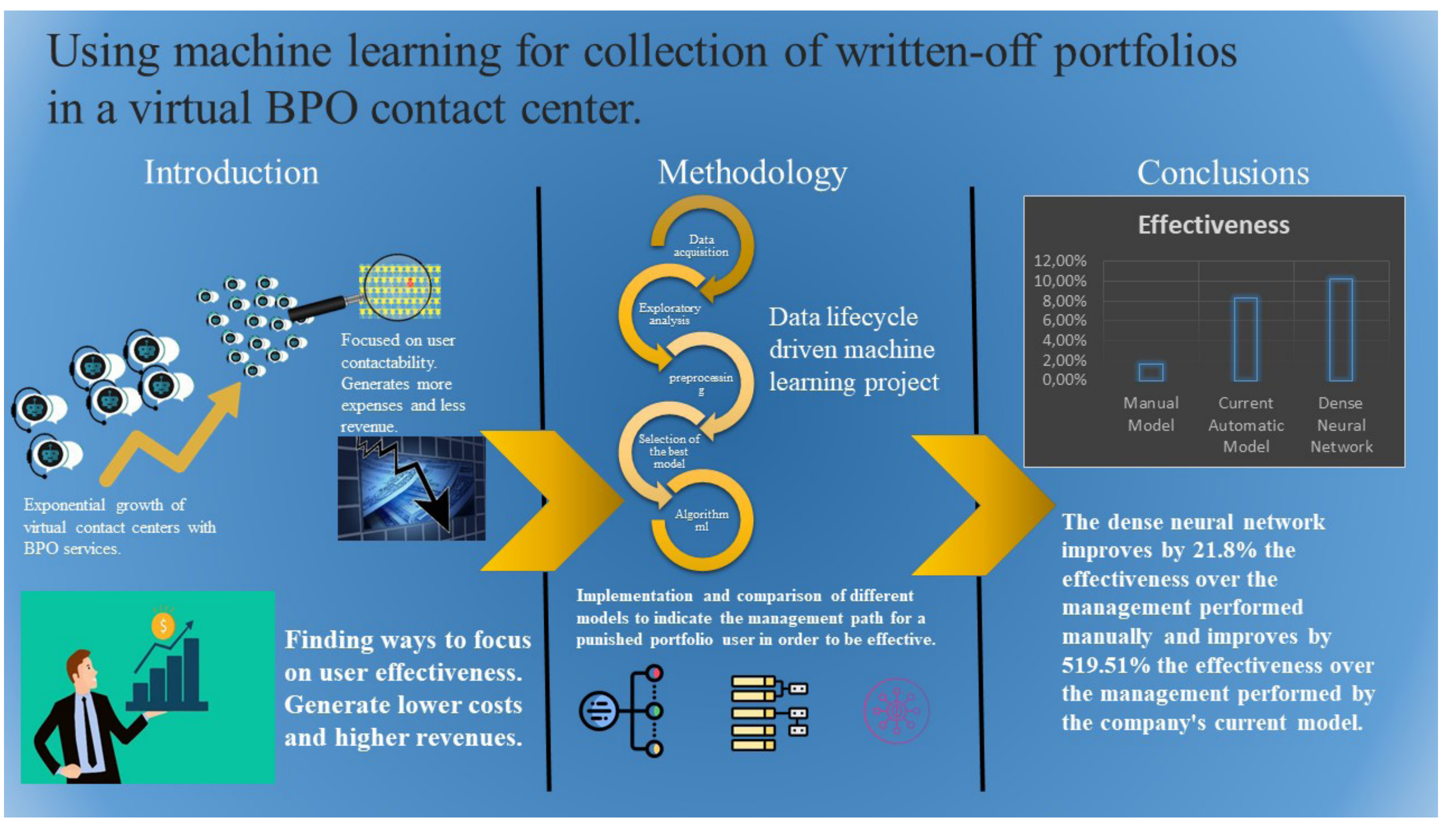
Resumen

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

Contribution to the literature

Why was it done?
This study was motivated by the opportunities offered by machine learning to optimise and automate business processes. In particular, we applied these methods in the context of tasks related to the recovery of punished portfolios through virtual contact services, where many decisions are traditionally based on the personal experience of those responsible for executing the recovery operation. Analysing historical data to identify patterns and trends provides an opportunity to improve the process by systematically approaching decision making.
What were the most relevant results?
The results show that the use of analytical methods improves the effectiveness of the recovery process by 21.8% compared to operations carried out by human managers. This represents an opportunity to improve resources, time efficiency and profitability for the process.



doi: 10.25100/iyc.v27i1.14327

Introduction

Virtual contact centers emerge as an innovative solution for companies seeking to enhance operational efficiency related to user communication. These centers leverage various technologies to automate user management across multiple channels, including customer service, sales, and collections, enabling faster and more efficient user engagement. Virtual contact centers are reshaping the landscape of the business process outsourcing (BPO) industry. Today, BPO services are powered by artificial intelligence, which helps optimize communication processes, data processing, behavior prediction, and other critical factors driving productivity in this sector.

Currently, the most widely used artificial intelligence (AI) techniques in the BPO industry are natural language processing (NLP) models, primarily employed to enhance customer contactability. These models use automated analysis to process audio, aiming to understand user intent. In the case of chatbots, they leverage text processing to enable more human-like interactions (1), (2). It is important to note that, for deteriorated or overdue debt portfolios, the effectiveness of debt recovery is an even more critical indicator than contactability. While contactability is essential for user communication and represents the most implemented type of model in BPOs (3), the effectiveness of debt recovery offers greater value. It reduces production costs, accelerates return on investment, and increases monetary revenues for both the client and the BPO (4), (5).

This research aims to implement data analytics techniques to enhance the effectiveness of debt recovery compared to the company's current model. The existing model is based on a set of rules designed to replicate the manual management processes performed by the company's analysts. To improve effectiveness through historical data analysis, the study will adopt the data-driven project lifecycle methodology. This approach includes data cleaning and transformation, exploratory data analysis of the collected data, and the implementation of machine learning models to identify the optimal management strategies for users with defaulted debt portfolios. As a validation strategy, the performance of various models will be evaluated and compared against the metrics of the company's current model (6), (7), (8).

The article is structured as follows: Section 2 provides a detailed explanation of the methodology based on the data-driven project lifecycle approach. Chapter 3 presents the results of the exploratory analysis and the evaluation of various predictive models. Chapter 4 discusses the outcomes obtained through experimentation. Finally, Chapter 5 presents the conclusions.

Methodology

The research will be conducted on a desktop computer equipped with a 12th-generation Intel Core i5 processor, 32GB of DDR5 RAM operating at 4800MT/s, 1.5TB of M.2 storage, and an 8GB Nvidia GeForce RTX 3060 TI GPU. The methodology follows the data lifecycle framework for machine learning projects, encompassing six stages, from data acquisition to model implementation.

doi: 10.25100/iyc.v27i1.14327

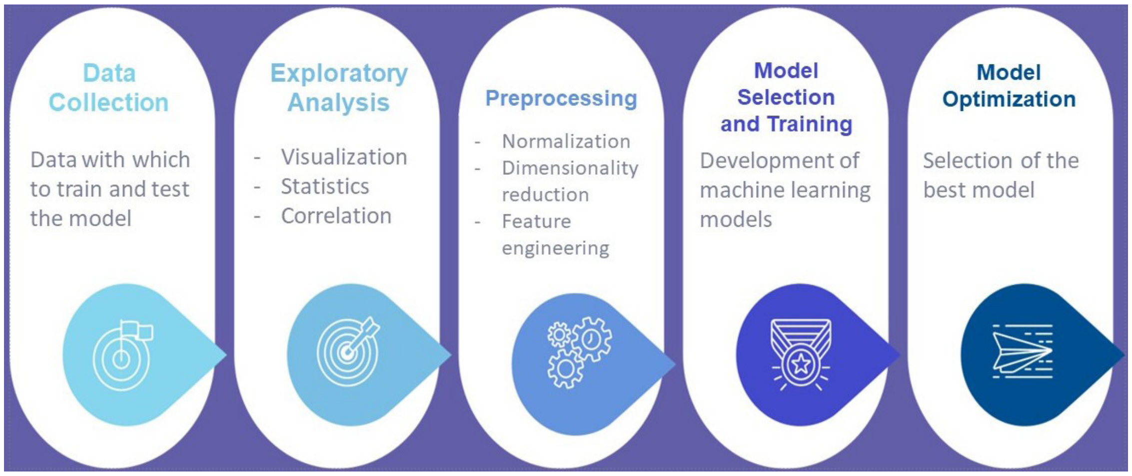


Figure 1. Lifecycle framework for machine learning projects

Data acquisition

Data was extracted from the management activities conducted on the users of a single debt collection client served by the BPO. The dataset spans the period from June 2022 to March 2023 and includes 23 columns containing client information. Additionally, a table of first and last names was extracted, specifying the associated gender for each name. Furthermore, a table was generated describing the types of strategies employed by the BPO for debt recovery for each client, indicating which strategies were effective and which were not. The extracted data was loaded into a local environment database for validation and cleaning. A table containing all the transactions was created, resulting in a total of 263,865,057 records, along with a table for names containing 740,362 records, and another table for homologations with 492 records.

Table 1. Selected features

Name	Description	Data Type
id	Identification homologation.	Integer
attempt	Attempt in which the transaction was made.	Integer
name	User's full name.	String
interaction_date	Transaction date.	Date YYYY-MM-DD
f5	Date and time of the transaction.	Date YYYY-MM-DD HH:MM:SS
intearction_time	Start time of the transaction.	Date HH:MM:SS
answer_time	Start time of the interaction.	Date HH:MM:SS
interaction_endtime	End time of the transaction.	Date HH:MM:SS
id_result	Transaction typification.	Integer

doi: 10.25100/iyc.v27i1.14327

payment date1	Possible payment date 1.	Date YYYY-MM-DD
payment date2	Possible payment date 2.	Date YYYY-MM-DD
payment_agreement	Amount to be paid in the transaction.	Float
debt	User's debt.	Float
overdue_days	Days in arrears of the debt	Integer
token	Unique alphanumeric ID assigned to each user.	String
url	URL to be sent via SMS.	String
id_allocation	Number assigned by the system to the allocation.	Integer
v11	Previous transaction pool.	String
v13	Current transaction pool.	String
v14	Maximum payment date.	Date YYYY-MM-DD
chanel	Channel through which the transaction was made.	String
max_attempts	Number of attempts the user had at the time of the transaction	Integer
Installments	Installments offered by the user in an agreement.	Interger

Once the data was uploaded, a data cleaning process was conducted. During this process, duplicate entries resulting from overlapping monthly database queries were removed. Records with a value of 0 in the attempt variable were excluded, as they represent initial states. Additionally, the columns url, id, v14, max_attempts, installments, and token were removed because they are custom variables defined by the manager, randomly generated, and irrelevant to the debt recovery analysis. The answer_time column was removed, because the dataset already included the interaction_date and interaction_endtime which are critical values for determining successful user interactions. Subsequently, new features were added to the dataset. The records were cross-referenced with the name column to add the gender and with the id column to incorporate the effectiveness and management variables. It is important to note that, in this context, management refers to an action performed to recover the debt, while an interaction_date occurs only when the end-user responds to that action.

Afterward, the resulting 76,203,786 records were stored in a new table, as the focus was on the current transaction with the end-user. Similarly, the columns id and name were excluded, as they were only relevant for extracting gender and effectiveness. The remaining columns were renamed as follows: v11 to previous_bag, v13 to current_bag and payment_agreement to amount_paid, The resulting columns were: attempt, payment_date1, payment_date2, amount_paid, debt, chanel, installments, current_bag id, gender, effectiveness, managment, interaction_time and interaction_endtime.

From the resulting data, the SMS channel was removed as it consistently proved ineffective and provided no relevant information. Records lacking effectiveness, those with atypical bag values (such as empty or zero), and records where the management column had a False value were also

doi: 10.25100/iyc.v27i1.14327

eliminated, as they did not contribute meaningful insights due to the lack of interaction. Finally, a new table was created by extracting only the effective records and their corresponding ineffective counterparts. The managment and chanel columns were removed, as they were no longer useful. This process resulted in a total of 12,178,681 records available for exploratory analysis.

Exploratory Data Analysis

For the exploratory analysis, six questions were formulated to better understand the extracted data and identify the most important features for addressing the problem and subsequently training the models. These questions were developed based on observations of effectiveness during the manual debt recovery process. The first question is as follows: *Is there seasonality in the data related to effectiveness over time*? This question was designed to evaluate the applicability of a time series model. To answer this question, a time series plot was created using 1,059,572 entries, and the results are presented in Figure 2. A simple visual analysis revealed no discernible pattern in the data over time, as it does not exhibit periodicity. This finding suggests that time series models are not suitable for this problem.

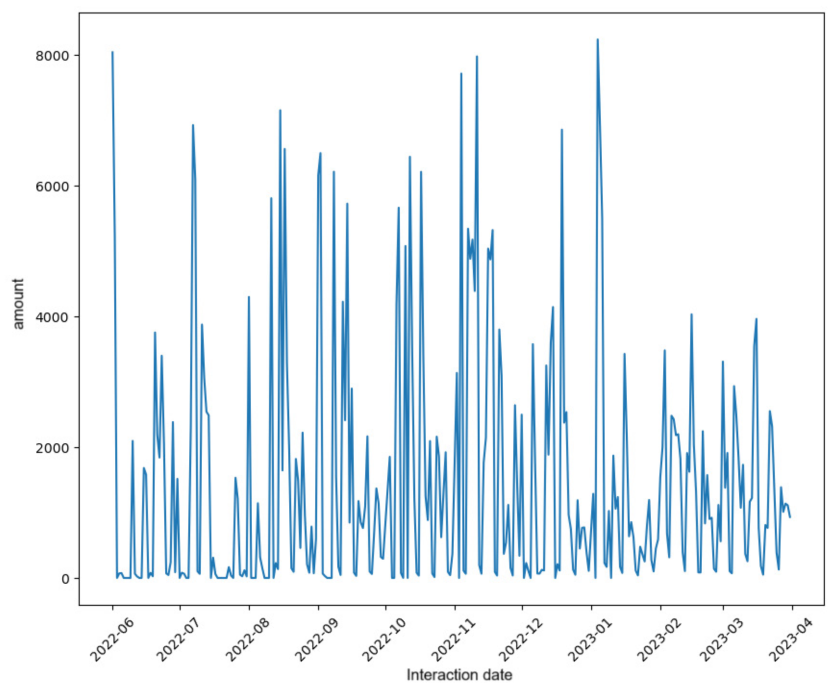


Figure 2. Data Regarding Effectiveness.

The second question was: *Which groups have the highest effectiveness?* To answer this, the records were grouped by category, the number of records in each group was counted, and groups with fewer than 1,000 samples were excluded to ensure significant results. Analyzing the effectiveness of the groups, it was found that out of 76 groups with measurable effectiveness, only 37 were significant, as shown in Figure 3.

doi: 10.25100/iyc.v27i1.14327

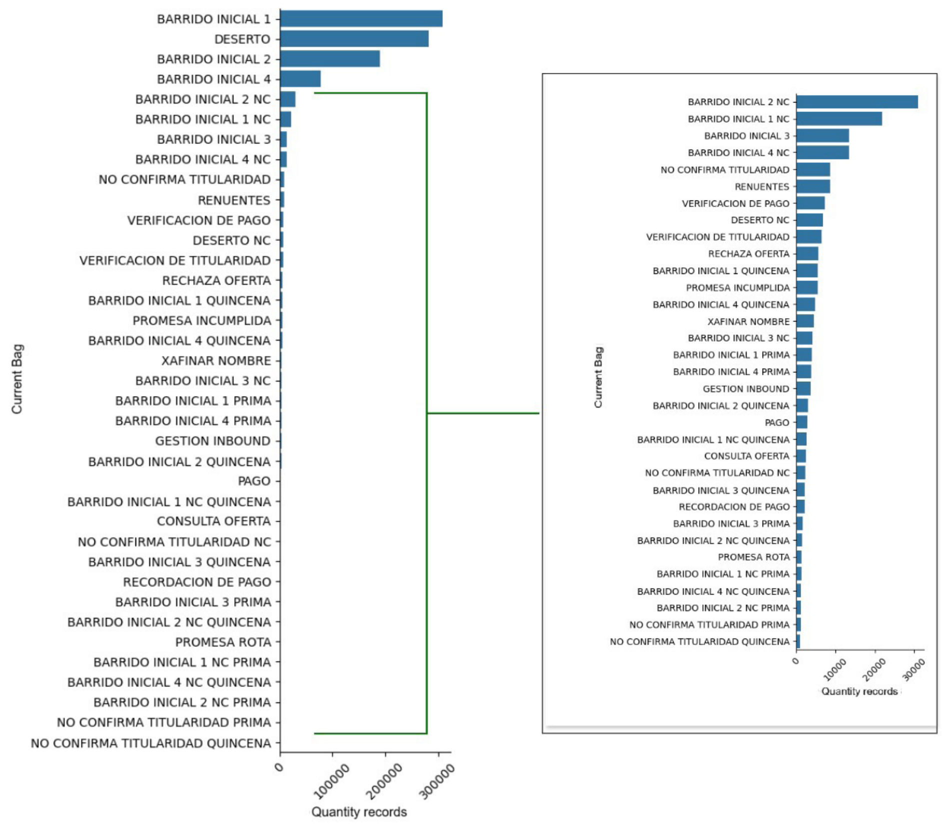


Figure 3. Effectiveness of a bag

The third question was: *Is there a relationship between debt ranges and installments days?* This question was proposed to explore the relationship between a user's delinquency days and debt by grouping them into four ranges corresponding to the percentiles: 0-25, 25-50, 50-75, and 75-100. Both debt and delinquency days were categorized using these ranges, which are based on best collection practices employed by the BPO. The results are presented in Figure 4. It can be observed that most effective users fall within debt ranges 1 and 2. Additionally, there is a strong relationship between these debt ranges and the day ranges.

doi: 10.25100/iyc.v27i1.14327

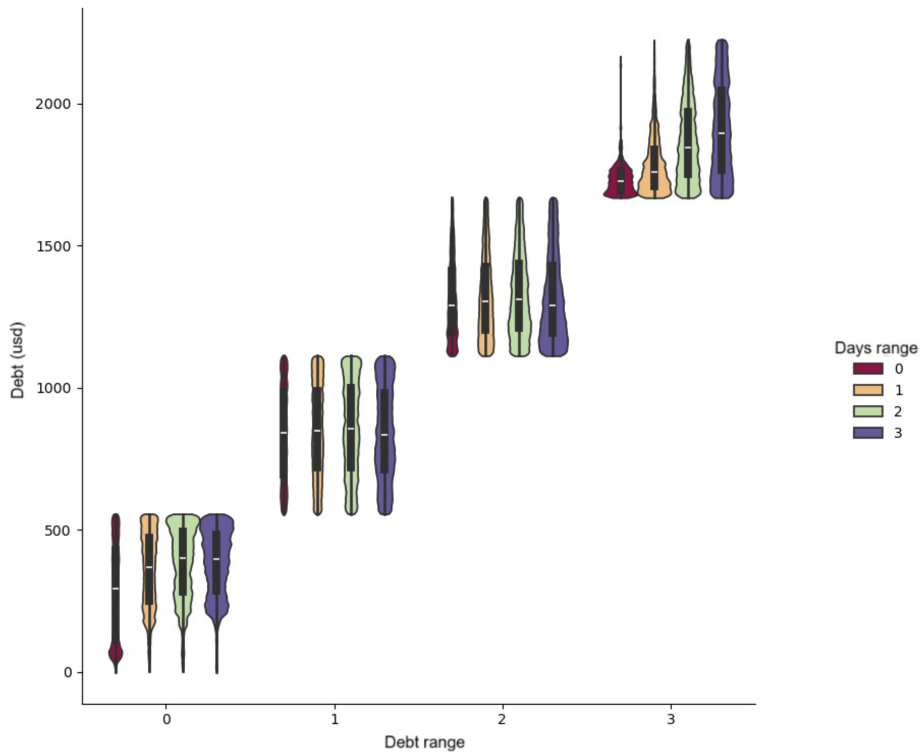


Figure 4. Violin plots of debt ranges versus debt in dollars and day ranges.

The fourth question was: *Is the gender variable relevant in the data for training a model?* To address this, the proportion of effective users by gender was calculated, revealing no significant difference in effectiveness based on gender. The results are shown in Figure 5.

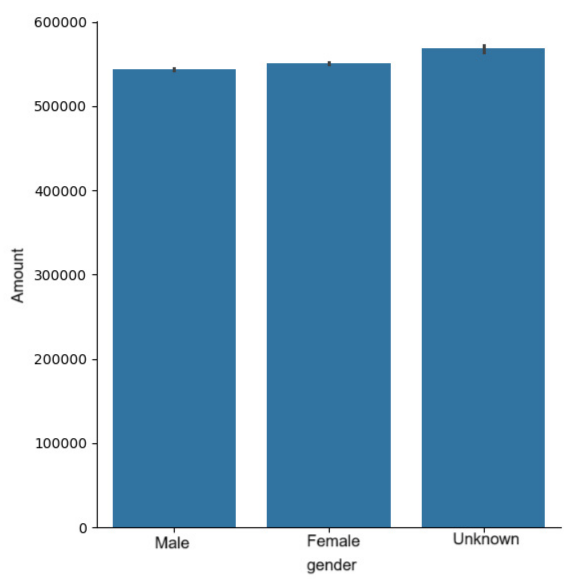


Figure 5. Effectiveness by Gender

The fifth question was: *Which variables are correlated with each other?* To answer this question, ordinal features were transformed into numerical values, and pairwise correlations between columns (excluding the columns for hours and dates) were calculated using Pearson's correlation coefficient.

doi: 10.25100/iyc.v27i1.14327

This metric measures the statistical relationship between two numerical variables and can take values between -1 and +1. A value greater than 0 indicates a positive association between the variables, meaning that as one increases, the other does as well. Conversely, a value less than 0 indicates a negative association, where the relationship weakens as one variable increases. Figure 6 presents the results. The analysis revealed that delinquency days, debt, and effectiveness are correlated with each other.

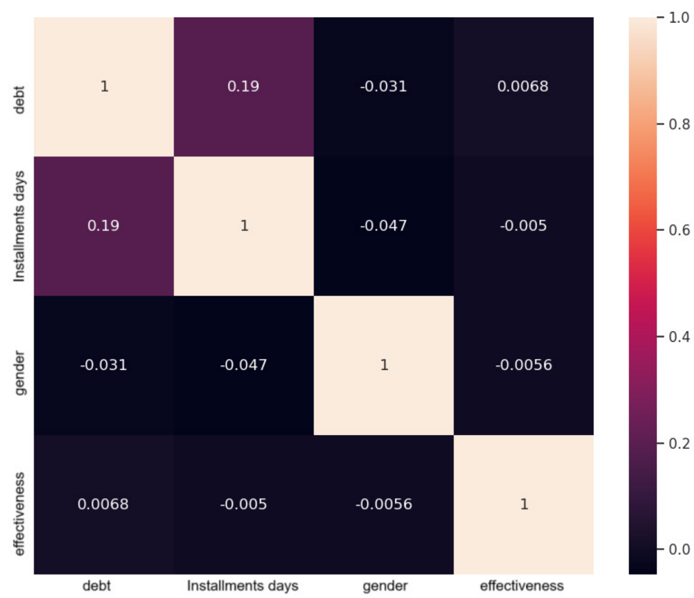


Figure 6. Pearson Correlation Matrix

Finally, the following question was proposed: *Is there a relationship between groups to set a management pathway?* To address this, a frequency analysis model was applied, and association rules were used to identify relationships between the groups. The model used was FPGrowth from the Apache Spark ML library in Python, as it offers stronger mathematical foundations for generating association rules compared to other models and supports grouping. The configuration for the model included a minSupport of 0.0001 to identify very rare patterns, that is, those occurring in a very small proportion of the total transactions, as the EFFECTIVE cases are few compared to the NON-EFFECTIVE ones. A *minConfidence* of 0.6 was set to ensure reliability greater than random, and *numPartitions* was set to 96, as clustering was performed using 96 logical processors. Table 2 presents the results obtained.

doi: 10.25100/iyc.v27i1.14327

Table 2. Sample of Association Rule Results

Antecedent	Consequent	Confidence	Lift	Support
[DESERTO NC, BARRIDO INICIAL 4, BARRIDO INICIAL 2, BARRIDO INICIAL 1]	[EFFECTIVE]	1	1.89579257	2.44E-04
[BARRIDO INICIAL 2 CIERRE, BARRIDO INICIAL 2, DESERTO, BARRIDO INICIAL 1]	[EFFECTIVE]	1	1.89579257	1.23E-04
[BARRIDO INICIAL 3 QUINCENA, BARRIDO INICIAL 2, DESERTO, BARRIDO INICIAL 1]	[EFFECTIVE]	1	1.89579257	2.53E-04
[BARRIDO INICIAL 4 NC, BARRIDO INICIAL 2 NC, BARRIDO INICIAL 4, BARRIDO INICIAL 2, DESERTO]	[EFFECTIVE]	1	1.89579257	2.85E-04
[DESERTO NC, BARRIDO INICIAL 4, BARRIDO INICIAL 2, DESERTO]	[EFFECTIVE]	1	1.89579257	3.23E-04

The association results show that a relationship between the groups, and management pathways can be created, indicating which groups have the highest probability of being effective. This conclusion is based on the "lift" variable, which measures the strength of the relationship between the antecedent and the consequent compared to their independent occurrence. A lift value greater than 1 indicates that the occurrence of the antecedent is positively associated with the occurrence of the consequent. The higher this value, the stronger the relationship.

After concluding the exploratory data analysis, the relevant columns were extracted, and a new table was created with the filtered data, eliminating duplicates based on the id and current_bag columns. This resulted in a total of 2,766,220 records and six columns: id, amount_paid, debt, installments, current_bag, and effectiveness, for preprocessing and use in training artificial intelligence models.

Data Preproccesing

During data preprocessing, the distribution of data for each group was analyzed, revealing that the groups "BARRIDO INICIAL 1", "BARRIDO INICIAL 2", "DESERTO", and "BARRIDO INICIAL 1 NC" were not comparable with other groups and introduced bias into the model. This problem arises because all users pass through these groups regardless of the configuration of other parameters, preventing the model from effectively differentiating records among them. Consequently, these groups were removed. Additionally, the 'PAGO' group was excluded, as it contained only the effective class. This process resulted in 1,956,403 records, which were used to create separate groups, one for each of the remaining categories. The features amount_paid, debt, and overdue_days were used, with effectiveness as the label. This approach enabled the training of a separate model for each group, with the goal of developing a customized multi-objective model. This procedure is shown in Figure 7.

doi: 10.25100/iyc.v27i1.14327

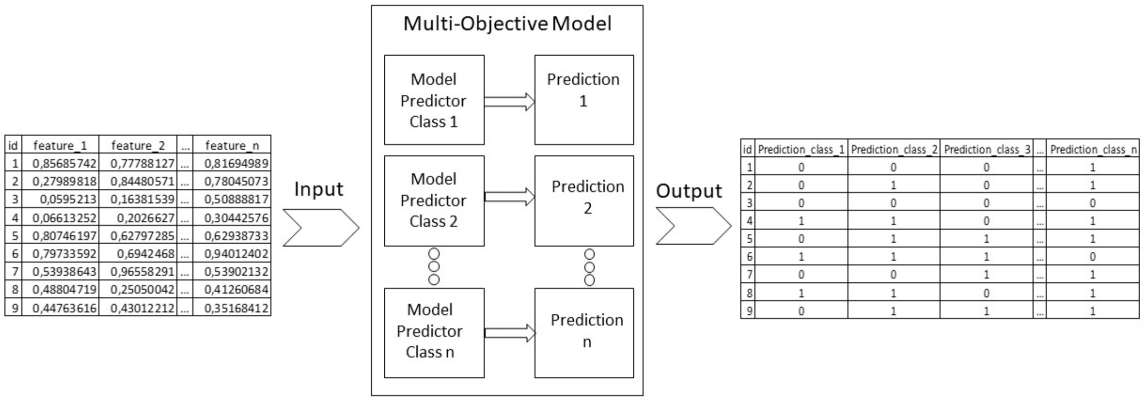


Figure 7. Multi-Objective Model Architecture

A subsampling of the data was performed for each of the generated groups, followed by a split of the transformed data into training and testing sets in a 70/30 ratio: 70% for training data and 30% for testing data. Each group was subjected to the subsampling method to balance the "EFFECTIVE" and "NON-EFFECTIVE" classes. For this, the Python library *ImbalancedLearn* was used, specifically the *RandomUnderSampler* method, which randomly selects samples from the majority class until a defined quantity is reached (9). In this case, the 'NON-EFFECTIVE' class was reduced to three times the quantity of the 'EFFECTIVE' class, ensuring that the 'EFFECTIVE' class would not exceed its current quantity. These divisions were applied to achieve an adequate balance of the data for each group, ensuring that the model does not excessively prefer one class but is instead equitable across them.

To transform the data, a pipeline was created. Each model's pipeline included two methods: one for filling missing data and another for scaling the data. For the first method, used to fill missing data, the *SimpleImputer* function from the preprocessing module of Scikit-learn was employed. This function uses basic statistics, such as the mean, median, or most frequent value, to fill missing values for each feature. In our case, we used the mean according to the data distribution. For the second method, the *MinMaxScaler* function from the Scikit-learn preprocessing library was utilized. This function is a transformer that scales each feature individually to fit within the range specified in the training set, which is by default between zero and one.

Experimental Method

The first trained model is Random Forests, defined by Breiman as a combination of tree predictors, in which each tree depends on the values of a random vector sampled independently and with the same distribution for all the trees in the forest. The generalization error of a forest of classification trees depends on the strength of each tree in the forest and the correlation between them. Internal estimations control the error, strength, and correlation and are used to evaluate the model's response to increasing the number of features used in splitting (10). The pipeline was executed using GridSearchCV, which determined that the best hyperparameters for the model are 200 n_estimators, the Gini criterion, and a max_depth of 20. Results are shown in Table 3.

doi: 10.25100/iyc.v27i1.14327

Table 3. Average results of the Random Forest model.

Metric	Training Value	Testing Value
Accuracy	73.65%	73.65%
Precision	63.22%	62.61%

The results indicate that the model performs satisfactorily in terms of accuracy but slightly less so in precision. This may be attributed to cases where the model struggles to discriminate effectively due to the nature of the data. The second model was developed using a Support Vector Machine (SVM) with a linear kernel (LinearSVC) for classification. This approach works by identifying a hyperplane that maximizes the distance between the closest data points of different classes, known as support vectors (11). The model was executed using GridSearchCV, which confirmed that the default hyperparameters were optimal. Results are shown in Table 4. We observe that the accuracy of this model is higher than that of Random Forests; however, its precision is significantly lower. As with the previous model, there is no overfitting, but the precision remains considerably low.

Table 4. Average results of the SVM model

Metric	Training Value	Testing Value
Accuracy	74.47%	74.51%
Precision	55.56%	55.61%

The third model implemented was Extreme Gradient Boosting, commonly known as XGBOOST. This model improves classification predictions by sequentially building shallow decision trees using Gradient Boosting, in which each new tree is fitted to the residuals (errors) of the previous one. XGBOOST employs optimization techniques to achieve faster convergence and prevent overfitting (12). Executing the pipeline with GridSearchCV determined that the best hyperparameters for the model were: max_depth=6, n_estimators=200, learning_rate=0.1, subsample=0.8, and colsample_ bytree = 0.8. The results are presented in Table 5. The metrics for this model are similar to those of Random Forests, with slightly higher accuracy and precision, though not significantly impactful in practical applications.

Table 5. Average results of XGBOOST model.

Metric	Training Value	Testing Value
Accuracy	74.25%	74.28%
Precision	64.69%	64.73%

Lastly, a Dense Neural Network was used as the final model. This deep learning model consists of a series of interconnected nodes, known as neurons. Each neuron has inputs, an activation function, weights, and an output. The training process involves optimizing the weights of the neurons to achieve the best possible result. The network is organized into layers, designed to mimic the structure of the human brain. The proposed neural network architecture is detailed in Table 6.

doi: 10.25100/iyc.v27i1.14327

Table 6. Dense Neural Network Configuration.

Layer (type)	Output shape
dense (dense)	(None, 512)
dense_1 (dense)	(None, 256)
dense_2 (dense)	(None, 128)
dense_3 (dense)	(None, 64)
dense_4 (dense)	(None, 1)

The activation function for the hidden layers is ReLU, while the output layer uses a sigmoid activation. The neural network was trained over 30 epochs with the default batch size. The Adam optimizer was employed with a learning rate of 0.001, binary cross-entropy as the loss function, and accuracy as the evaluation metric. Table 7 presents the results. We observe that the metrics for the Dense Neural Network are similar to those of the XGBOOST model, with slightly lower values, though not significantly so. As with the previous models, there is no overfitting; however, the precision remains lower for unseen data.

Table 7. Average results of Dense Neural Network model.

Metric	Training Value	Testing Value
Accuracy	74.67%	73.56%
Precision	61.39%	57.99%

After validating the results of the models, they were optimized by removing the variable *paid_ amount*, which contains similar data in different cases for classification. This may have caused the models to fail to discriminate properly. Additionally, the effectiveness metric was evaluated, as it holds high relevance for this research by indicating whether the model is significant compared to the company's current effectiveness results. The training results of the models without this variable are presented in Figure 8 and Table 8.

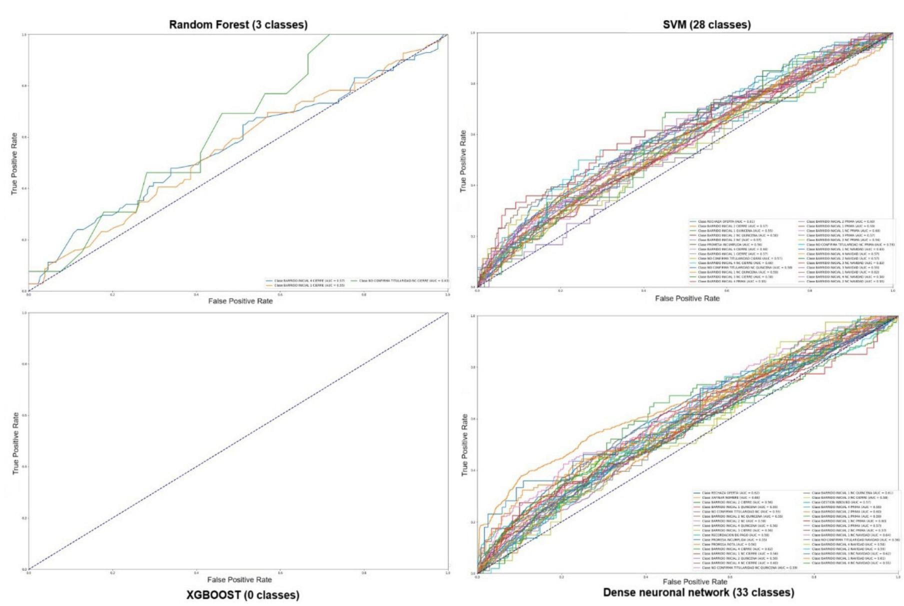


Figure 8. Comparison of ROC Curves with AUC > 0.55 for Optimized Models

doi: 10.25100/iyc.v27i1.14327

Table 8. Comparison of optimized models.

		With the feature valor_pago			Without the feature valor_pago		
Model	Number of bags with AUC > 0.55	Accuracy	Precision	Effectiveness	Accuracy	Precision	Effectiveness
Random Forest	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%
Dense Neural Network	33	73.56%	57.99%	9.21%	73.39%	57.44%	10.17%

From the results presented in Table 8, it is possible to observe that removing the "valor_pago" feature caused a slight decrease in the precision of all models, a significant drop in the effectiveness of Random Forests and XGBOOST, and an increase in the effectiveness of the Dense Neural Network. Based on these results, it can be concluded that the Dense Neural Network is the best-performing model. It has a metric superior to random and achieves effectiveness that surpasses the company's current average effectiveness for both automatic and manual models. Particularly, the Dense Neural Network, with 32 adequately discriminated classes, achieves an effectiveness of 10.17%, which is significantly higher than the 1.64% average of the automatic model and the 8.35% average of the manual processes in the BPO.

Results and Discussion

When using machine learning models to improve operational effectiveness in a BPO, no models have been specifically proposed or applied for this purpose. Instead, most existing models focus on improving contactability or segmenting users to assess the risks of granting loans to individuals (13), (14). However, these models utilize variables similar to those in the proposed research, enabling a performance comparison based on the data rather than the problem's context, as the approaches differ. Additionally, some models employ a metric similar to the effectiveness used in this research (15), (16), (17).

Beginning with feature selection, Martín Calero et al. (18) proposed a model to identify the most relevant features for determining which users are most likely to repay or default. Their results showed that out of 59 features (9 of which pertain to machine management), only 5 were identified as relevant according to the best-performing model. Among these, the most notable is days in arrears (referred to in this research as "days of installments"), which holds an 85% importance score. This finding highlights the significance of days in arrears as a key feature, a conclusion further supported by Arango Correa et al. (19), who conducted a correlation analysis using Pearson's method. By selecting features above the 66th percentile, they reduced the initial 33 features to the

doi: 10.25100/iyc.v27i1.14327

most significant ones, including the highest days in arrears and the average for the past 4 months. While monetary amount variables are mentioned in both studies, no analysis is provided to explain why they were excluded. In both models, the precision exceeds 88%.

Swati Tyagi (20) begins with a dataset containing 110 variables, of which 80 were discarded during the initial data-cleaning phase, leaving 30 variables. These included debt delinquency days, debt balances (such as initial loan amounts, payments made, and current debt), and other features related to credit operations. The best-performing model achieved a precision of 89%. Unlike other studies that rely on standard measures such as ROC/AUC, Tyagi also employed ROI as a performance metric, which is particularly crucial in debt collection contexts. The best-case ROI reported was 3.8%. This metric can be compared to the effectiveness metric used in the present research, as higher effectiveness correlates with lower investment and higher revenue. Debt amount and delinquency days are critical variables in any financial model, whether for user classification or loan amount prediction. Various datasets in this field incorporate these variables, such as the "GiveMeSomeCredit" dataset by Credit Fusion and Will Cukierski (21), which was used in a 2011 Kaggle competition. The challenge in this competition was to develop a model to assist in loan decision-making.

It is therefore possible to compare various aspects of this research and validate its quality. The first aspect highlights that, while companies may provide numerous variables, not all are useful for training a model. Various studies demonstrate that feature selection can significantly reduce the initial set of variables. The second aspect confirms the relevance of debt and delinquency variables for debt collection and financial models, as supported by the findings of other researchers. The third aspect is that model performance metrics can remain high when applied to both training data and previously unseen data, with precision rates exceeding 65%. Finally, although trained models may achieve good performance metrics, an additional step is necessary to validate their functionality. This is exemplified by Swati Tyagi's (20) work, where ROI is used to evaluate results beyond the dataset.

Conclusions

In the present research, a thorough analysis of the collected data was conducted, from its acquisition to its exploration, with the aim of deeply understanding the characteristics and trends in the operations carried out by operations managers and the company's current model. This analysis made it possible to identify patterns and relationships relevant to the portfolio management process, confirming the importance of the debt amount and days of delinquency variables, as also highlighted by other researchers in the BPO and debt collection fields.

Subsequently, four machine learning models were implemented, including one deep learning model: Random Forests, SVM, XGBOOST, and a Dense Neural Network. These were designed to identify the most effective approach for managing delinquent portfolio users in a BPO. Various algorithms, preprocessing techniques, and modeling approaches were explored to find the best solution to enhance management effectiveness and minimize production costs.

doi: 10.25100/iyc.v27i1.14327

Once the models were implemented, their results were analyzed using evaluation metrics such as Accuracy, Precision, and the comparison metric with the current management model, Effectiveness, as well as ROC curves to measure the predictive capability and overall performance of each model. This evaluation identified the Dense Neural Network as the most promising model for practical application.

Finally, the results of the selected model were compared with the company's current performance metrics. This comparison revealed that the Dense Neural Network improves management effectiveness by 21.8% over the operations conducted by managers, whose current average effectiveness is 8.35%. It also improves effectiveness by 519.51% over the company's current decision tree model, which has an average effectiveness of 1.64%. This highlights the strengths and advantages of implementing such models. With the application of this model, the company could increase its monthly revenue from $20,000 to $103,902 USD.

CRediT authorship contribution statement

Conceptualization - Ideas: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Data Curation: Camilo Andrés Leuro Camacho. Formal analysis: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Investigation: Camilo Andrés Leuro Camacho. Methodology: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Project Management: Juan Pablo Ospina López. Resources: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Software: Camilo Andrés Leuro Camacho. Supervision: Juan Pablo Ospina López. Validation: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Visualization - Preparation: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Writing - original draft - Preparation: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López. Writing - revision and editing - Preparation: Camilo Andrés Leuro Camacho, Juan Pablo Ospina López.

Financing: no declare.

Conflict of interest: no declare

Etics Implications: no declare

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. Ş. 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

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.jfds.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

doi: 10.25100/iyc.v27i1.14327

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].

doi: 10.25100/iyc.v27i1.14327