



# Improving debt collection via contact center information: A predictive analytics framework

Catalina Sánchez<sup>a</sup>, Sebastián Maldonado<sup>b,c</sup>, Carla Vairetti<sup>a,c,\*</sup>

<sup>a</sup> Universidad de los Andes, Facultad de Ingeniería y Ciencias Aplicadas, Chile

<sup>b</sup> Department of Management Control and Information Systems, School of Economics and Business, University of Chile, Santiago, Chile

<sup>c</sup> Instituto Sistemas Complejos de Ingeniería (ISCI), Chile

## ARTICLE INFO

### Keywords:

Debt collection  
Contact center  
Call center  
Predictive analytics  
Data integration

## ABSTRACT

Debt collection is a very important business application of predictive analytics. This task consists of foreseeing repayment chances of late payers. In this sense, contact centers have a central role in debt collection since it improves profitability by turning monetary losses into a direct benefit to banks and other financial institutions. In this paper, we study the influence of contact center variables in predictive models for debt collection, which are combined with the financial information of late payers. We explore five different variants of three predictive analytics tasks: (1) the probability of successfully contacting a late payer, (2) the probability of achieving a contact that results in a promise to pay a debt, and (3) the probability that a defaulter repays his/her arrears. Four research questions are developed in the context of debt collection analytics and empirically discussed using data from a Chilean financial institution. Our results show the positive impact of the combination of the two data sources in terms of predictive performance, confirming that valuable information on late payers can be collected from contact centers.

## 1. Introduction

Risk management is an important application domain for predictive analytics. Analytical models have been considered in many important decisions for banks and financial institutions in order to assess and control threats to their capital and earnings [3]. For example, credit scoring models have been used for several decades in order to assess the creditworthiness of a customer [26].

Despite the plethora of predictive models designed for risk management and, in particular, for credit scoring [19], debt collection has received little attention in the academic community. Debt collection models not only allow making accurate projections of future cash flows and receivables, but it also improves the success of collection operations [9]. For example, the workforce can be redirected towards those customers that are most likely not to repay their obligations [16], or new collection policies can be designed by distinguishing between debtors that are likely to repay or not [32].

Thanks to predictive analytics, the role of contact centers has evolved dramatically, becoming a strategic differentiator for companies instead of basic service providers [6,15]. They also generate important information for improving business decisions. Some of the relevant tasks that

can be tackled with this source of data are staffing, which consists in finding the optimal number of agents required to deal with the demand [14], agent shift scheduling [14,16], or understanding the customer experience to optimize customer-agent interactions [17].

The sheer volume and variety of data that face current analytical systems have challenged academics and practitioners, and business analytics is not oblivious to this ‘Big Data’ trend [2,10]. In this sense, one of the most important challenges for companies is data integration [10,21]. The fusion of different sources of information can boost the performance of predictive models, extracting relevant insights for decision-making to make improvements to the products or services provided by companies [10,30].

In this study, we propose a data integration framework designed to incorporate contact center data in the debt collection process of a financial institution. The late payment prediction problem is addressed from three different angles via binary classification: (1) we estimate the probability of successfully contacting a late payer through the call center, (2) we calculate the probability of contacting a late payer, and the debtor promise to pay as a result of this contact, and (3) we predict successful late payment. Notice that the first two tasks have not been discussed previously in the academic literature, to the best of our

\* Corresponding author at: Universidad de los Andes, Chile.

E-mail address: [cvairetti@uandes.cl](mailto:cvairetti@uandes.cl) (C. Vairetti).

<https://doi.org/10.1016/j.dss.2022.113812>

Received 18 August 2021; Received in revised form 25 November 2021; Accepted 10 May 2022

Available online 14 May 2022

0167-9236/© 2022 Elsevier B.V. All rights reserved.

knowledge. Therefore, the main novelty and contribution of this study rely on these two predictive approaches.

This study explores the following research questions:

**(RQ1):** Can we predict late payment accurately to improve collection strategies?

**(RQ2):** Is contact center information a valuable data source for late payment prediction?

**(RQ3):** Does a promise of late payment have an impact on the debt collection policies?

**(RQ4):** Can we address first time contacts and follow-ups with two different prediction models?

Furthermore, our study discusses the following aspects that are relevant for the state of the art in debt collection:

- We propose a framework for combining contact center and financial information from customers for the debt collection problem to address RQ1 and RQ2. We show empirically how the combination of the two sources leads to successful classifiers in terms of predictive performance, in contrast to using only one data source for feature construction. This approach provides a new application for academics in the data fusion and data integration domains.
- In order to address RQ3 and RQ4, we propose five different predictive approaches that can be implemented for a successful debt collection system, which are variants of the three aforementioned tasks: (1) successful contact prediction based on the full dataset of debtors (Task 1), (2) promise-to-pay prediction based on the full dataset of debtors (Task 2a), (3) promise-to-pay prediction based on the contacted customers only (Task 2b), (4) late payment prediction based on the full dataset of debtors (Task 3a), and (5) late payment prediction based on contacted customers that promised to pay only (Task 3b). We thoroughly discuss why we believe tasks 2a and 3b are the most relevant ones for first time contacts and follow-ups, respectively.
- The proposed framework includes a theoretical discussion on how the predictive tasks can be used as an input for several debt collection decisions, such as customer segmentation, agent assignment, staffing, identification of self-cure customers, and Value-at-risk.
- We collect a rich dataset from a Chilean financial institution, consisting of 1,023,418 call records, 26,466 promises of repayment, and 54,349 effective debt payments. This dataset allowed us to validate our analysis and understand the hierarchical decision scheme based on the debt collection process via contact centers.

The remainder of this paper is structured as follows: [Section 2](#) provides an overview of prior studies on contact center analytics and debt collection. The proposed framework for debt collection is presented in [Section 3](#). [Section 4](#) discusses the main results obtained by using our dataset with a wide variety of statistical and machine learning classifiers, providing also relevant insights for decision-making. The main conclusions are presented in [Section 5](#), addressing possible directions for future work.

## 2. Prior work on debt collection analytics

Several studies have been proposed in the last decades designed to improve debt collection via predictive analytics and operational research. This problem was considered for the first time by Mitchner and Peterson [22], which optimized the time at which an agent must contact a late payer to collect his/her debt based on the cost that it generates.

[Section 2.1](#) is devoted to predictive models for late payment (RQ1). Next, [Section 2.2](#) discusses the studies that incorporate contact center information into the debt collection process. Finally, [Section 2.3](#) provides a summary of the main studies on debt collection analytics and discusses the research gaps.

### 2.1. Predicting late repayment via analytics models

Research on debt collection analytics focuses mostly on RQ1, predicting whether the customers will pay their outstanding debt in order to optimize the collection process, generate new policies, and make projections in cash flows, among others [4]. Statistical techniques have been considered traditionally, however, the rise of machine learning and artificial intelligence has provided a new set of tools that has been proved effective in debt collection [4,28].

Zurada and Lonial [32] predicted debt payment in health systems, a task in which there is a large rate of unpaid obligations. They explored five different classification models: artificial neural networks (ANNs), decision trees, logistic regression, memory-based reasoning, and ensemble methods. They concluded that ANNs, logistic regression, and ensembles were the best strategies for debt payment prediction, which is characterized by having limited financial data from the individuals.

Abe et al. [1] presented a debt collection optimization process that aims at finding the best actions that lead to maximum recovery. To do this, they proposed an approach based on the restricted Markov decision process (MDP) through restricted reinforcement learning. They applied this methodology in the tax and finance department of New York, concluding that restricted reinforcement learning is better at combining data modeling and restricted optimization than other alternative approaches.

In the same direction, Chehrizi et al. [8] proposed a stochastic programming approach for dynamic credit-collections optimization. Although this approach does not consider data-driven modeling, they show in simulated examples that their stochastic optimal control model is able to reduce losses, concentrating the collection policies at the best possible time.

### 2.2. Contact center information for debt collection

Some previous studies considered contact center data for improving debt collection. One example is the paper by Kim and Kang [16] that proposed a system to assign debtors to agents of a contact center in an efficient and fair manner. They developed a machine learning-based late payment prediction approach, comparing decision trees, ANNs, Support Vector Machines (SVMs), and random forest. The predictive models act as an input for deriving scoring rules for a fair allocation of customers to call center agents. They showed that predictive analytics generate better scoring rules in relation to existing heuristics.

Recently, Bellotti et al. [5] took a different predictive analytics approach, forecasting the recovery rates of non-performing loans. They applied this strategy to the retail banking industry, comparing 20 different statistical and machine learning methods. They concluded that rule-based ensembles such as gradient boosting, random forest, and cubist achieved the best predictive performance. Although the authors do not predict late payment, they include monthly information on calls, visits and contacts between the financial institution and the borrowers, and therefore this study can be partially linked to RQ2.

Along the same lines, van de Geer et al. [28] proposed a methodology that combines machine learning and dynamic programming, following the reasoning behind the paper by Abe et al. [1]. Instead of improving actions to maximize recovery, however, the methodology by van de Geer et al. [28] optimizes the calls in such a way that those customers who are more likely to pay their debts are prioritized. They consider gradient boosting as a predictive approach, achieving excellent results in terms of debt collection in a real-world application. This study considers basic information on past contacts with the customers, and can also be related to RQ2.

Finally, Bahrami et al. [4] considered a behavioral scoring approach for invoice payment prediction considering their due time. They compared logistic regression with one rule (OneR), and SVMs, showing that the former is able to achieve excellent results on a large dataset of 1.6 million customers. Among the variables included in the model, they

consider historic records of the various actions made by invoice-issuing companies towards customers to encourage payment, including e-mails, short message service (SMS), and phone calls.

### 2.3. Summary and research gaps

The most relevant debt collection studies are summarized in Table 1. We categorize each article according to five relevant aspects of this study: (1) whether it involves machine learning, (2) whether it performs late payment prediction, (3) whether the predictive task is addressed as a classification problem, (4) whether contact center information is considered for predictive purposes, and (5) whether it performs either successful contact prediction or promise of late payment prediction (or both). The goal of this classification is to present research gaps. This table also includes the best modeling strategy reported in the various studies.

Based on Table 1, we identified important gaps in the literature, which are addressed with our proposal. First, the state of the art does not consider estimating the probability of a successful contact with the customer, which is key in order to define the best contacting policies and communication channels, nor the probability that a customer commits to pay their outstanding debt on a given due time. This commitment is extremely important in the debt collection process since it dramatically increases the probability that a customer pays its outstanding debt, which is the main goal of the task. Notice that existing studies focus only on this predictive task.

A second gap consists in the use of contact center variables in the debt collection process. Our proposal formalizes their inclusion in the predictive models, providing guidelines for feature generation and showing their relevance for the various tasks. Most studies either ignore this information and consider only financial variables, while others limit themselves to counting variables of past contacts (see e.g. [5,28]). These studies also do not address the predictive task as a classification problem.

In summary, most of the existing debt collection analytics literature focuses on RQ1, addressing the late payment prediction as either a classification or a regression problem. All these studies show that this task can be successful when considering machine learning methods such as ensembles or ANNs. Regarding RQ2, only two studies incorporate information of past contacts with the debtor as covariates, and only one in the context of late payment prediction (RQ1 + RQ2). Both studies show that these variables can improve predictive performance. However, the contact center information is very limited and does not consider other valuable resources such as *best time to call* (BTTC) variables and previous contact attempts, which are discussed in this paper. In relation to RQ3 and RQ4, we did not find any study that addresses promise to pay and differentiated approaches to first time contact and follow-up.

**Table 1**  
Summary of relevant debt collection studies.

Publication	Task	Best methods	(1)	(2)	(3)	(4)	(5)
Mitchner and Peterson [22]	Best time to call	Stochastic modeling	✗	✗	✗	✗	✗
Chehrizi et al. [8]	Debt collection optimization process	Stochastic programming	✗	✗	✗	✗	✗
Abe et al. [1]	Debt collection optimization process	Reinforcement learning	✓	✗	✗	✗	✗
Zurada and Lonial [32]	Late payment in health systems	ANNs, logit, ensembles	✓	✓	✓	✗	✗
Kim and Kang [16]	Fair assignment of agents in call centers	Ensembles and trees	✓	✓	✓	✗	✗
Bellotti et al. [5]	Forecasting recovery rates	Ensembles for regression	✓	✗	✗	✓	✗
van de Geer et al. [28]	Debt collection optimization + prediction	Ensembles for regression	✓	✓	✗	✓	✗
Bahrami et al. [4]	Invoice payment prediction	Logit + clustering	✓	✓	✓	✓	✗
This study	Late payment prediction	ANNs	✓	✓	✓	✓	✓

<sup>1</sup>Use of machine learning.

<sup>2</sup>Late payment prediction.

<sup>3</sup>Classification task.

<sup>4</sup>Contact center data for predictive purposes.

<sup>5</sup>Successful contact / promise of Late payment prediction.

### 3. Proposed data integration framework for debt collection

In this section, we formalize a framework based on predictive analytics to extract knowledge from contact center data and financial information of late payers in financial institutions. Our objective is to merge these two sources to create accurate classifiers for debt collection decision-making, addressing RQ1 and RQ2. We also propose novel predictive tasks, such as successful contact prediction and promise-to-pay prediction, going beyond late payment prediction and addressing RQ3 and RQ4.

The proposed framework is presented next. This is a customization of the well-known processes for knowledge discovery in databases, such as KDD [11] and CRISP-DM [7]. Step 1 consists of data preparation, the definition of the predictive tasks, and the data integration. Step 2 corresponds to feature engineering, in which we design variables that can be relevant for the given task. In Step 3, the best-performing classifier is selected, exploring the predictive performance of different techniques. Step 4 consists of model evaluation, interpretation of the mined patterns, and feature selection. In this case, we embed a feature selection strategy in the model training process in order to remove irrelevant and/or redundant variables. This strategy not only leads to a better prediction, but also allows a better understanding of the process that generated the data. Finally, step 5 discusses the decision-making process behind debt collection, which can be benefited from the novel predictive tasks presented in this paper.

#### 3.1. Steps 1 & 2: task definition, data preparation, and feature engineering

We propose a simple data integration step, which consists of generating features from two different sources: financial information from bank customers or any other financial institution, and contact center data. Our main research question is whether late payment behavior can be better addressed by combining these two sources, rather than considering a single source.

We propose performing three different classification tasks to improve decision-making in the debt collection process through a contact center. These three tasks are aligned with the main objective of the process, which is that a debtor pays its debts. These three tasks lead to five different modeling approaches, which are described next.

- **Successful Contact:** Our first challenge is to be able to predict whether a debtor will answer the call made through the contact center, which represents the debt collection effort. In this case, we foresee that financial variables would have little or no influence in this predictive task as it is expected that this problem is inherently dependent on the call center setting. Notice that a successful contact is defined as a direct one, which occurs only when the debtor answers the call and

not a third party. Although this task is strongly linked to the BTTC literature (see e.g. [31]), this is a novel contribution of our study in the context of debt collection. Despite the fact that this study is limited to phone calls, this task can be extended to multi-channel debt collection, resulting in an interesting avenue for future research. We refer to the successful contact prediction problem as Task 1.

- *Promise of late payment:* The second challenge is to predict whether a debtor will make a commitment to pay his/her obligations when contacted through the contact center. This is also a novel application, which we believe can be tackled adequately by combining the two sources of information. Having in mind RQ4, we consider two different approaches depending on the sample used for model training: we used either the full dataset (Task 2a) or the sample of the contacted customers only (Task 2b).
- *Late payment prediction:* The final challenge is to be able to predict accurately whether a debtor will pay his/her debt. We consider that this payment is made between zero and five days after the call. Similar to the previous task, we suggest addressing this challenge considering two different samples of debtors: the full set (Task 3a) and only the ones that made a promise to pay (Task 3b). This approach takes into consideration the four research questions indicated in the first section.

We believe that the promise of payment prediction task (Task 2a) is the most valuable one, because it aims at predicting to which customer our *action* of contacting them will *trigger* a positive response (a promise of payment). This is the main goal of causal inference and causal machine learning [29]. In contrast, predicting a successful contact focuses exclusively on reaching out to the customers, but the target variable includes individuals that are not willing to make a payment despite being contacted. For these customers, our action has no effect on them. They are labeled as ‘lost causes’ in the causal inference literature [29].

It is important to notice that the three tasks we are addressing are related to the following goal: identifying those debtors from the whole customer database that should be contacted in order to maximize the return. In this sense, we recommend Task 2a for implementation when the goal is to achieve the first successful contact with the customer. Task 1 is interesting for gaining insights on when and how to contact a customer, but we believe it is more valuable to aim at achieving not only a successful contact but a promise of late payment in our first attempt at reaching a debtor.

Task 2b suggests that we should limit ourselves to only the customers that have been successfully contacted, but the decision to make a promise of late payment occurs at the same time the customer answers the phone. In other words, it is interesting to analyze the variables that make a customer commit to a late payment, however, it is less actionable than tasks 1 and 2a. We can apply Task 2b to the whole dataset of debtors, but then our target set for prediction differs completely from the training set. We recommend Task 2b only for the purpose of understanding what triggers a promise of late payment to the contacted customers.

Notice that tasks 1 and 2a are designed to predict the success of first time calls, considering the whole dataset of debtors for model training (noncontacted debtors included). Notice that Task 2a is not actionable if Task 1 is implemented since it leads to selection bias. Task 2b, however, can be used for follow-ups as it includes only contacted customers.

The late payment prediction task (Task 3a) is the traditional approach for debt collection, but we also believe it is suboptimal as a first contact in comparison with predicting a promise of payment. Similar to the successful contact prediction task, it includes customers to whom the action has no effect on the target variable. In this case, we consider as a success ( $y = 1$ ) individuals that pay their debts even when they were not contacted (self-cure customers). They are labeled as ‘sure things’ in the causal inference literature [29]. In contrast, Task 3b does provide an interesting actionable model in the sense that we can call

those customers that already made a promise again as a reminder. This way we focus only on those debtors that are more likely to fulfill the promise.

The definition of the various tasks provides a comprehensible conceptual framework to discuss the research questions:

- In order to assess whether late payment prediction can be performed accurately (RQ1), the predictive performance of the various machine learning methods proposed in the next section is computed for the variants of Task 2 and Task 3.
- Regarding the combination of the financial and contact center data sources (RQ2), we explore the predictive performance of each task with different combinations of variables (financial variables only, contact center variables only, and the combination of both sources), discussing the differences between models.
- In relation to the promise of late payment (RQ3), we analyze this information as a predictive task (Task 2) and as a covariate (number of past promises to pay made by a debtor) for the various tasks.
- Finally, we propose predictive tasks designed for a first contact with the debtors (tasks 1, 2a, and 3a) and for follow-ups (tasks 2b and 3b). The performance of the various tasks can be linked to RQ4.

Regarding the definition of the target variable for tasks 3a and 3b (late payment prediction), we suggest a ten-day period for evaluating the outcome of the contact. If no payment was made within this ten-day period, we consider the contact as unsuccessful. The contact itself usually had a seven-day payment notice from a bank or other financial institution. This is consistent with the debt collection literature (see e.g. [28]).

We recommend the application of the predictive model developed in Task 3b when performing sequential attempts on the debtors that made a promise to pay. Since the customers usually commit to pay within the seven-day period of the notice, we suggest a subsequent call should be made 10–14 days after the successful contact.

The second step of the proposed framework corresponds to the construction of the variables that can be useful for the three aforementioned tasks. We refer to a set of variables constructed based on financial information from the bank as *financial data* (FD), while we use *contact center data* (CCD) for those features constructed from historical information collected through the contact center. The various sets of suggested variables are described next.

- *FD1.Debt:* Variables that characterize the outstanding debt, such as the unpaid balance, the amount of delinquent debt, and the number of days in arrears. Such variables are well-known in tasks such as behavioral scoring (see e.g. [3,26]).
- *FD2.Debtor:* Variables that describe the debtor, such as the risk of the customer in terms of his/her record of past debts, the average and standard deviation of previous payments, and any other socio-demographic information available. These variables are grounded in previous debt collection studies, such as [4,28,32].
- *CCD1.BTTC:* Variables that allow the identification of the best day and time to call a customer. Although the scientific literature on this topic is rather scarce and outdated (see e.g. [31]), there are several business reports and other resources that indicate variables that can be constructed. Some examples are the hour of the day in which the call starts, whether the call is made during the morning, whether it is made between Monday through Thursday, or whether it is made either at the beginning (first 10 days), the middle, or the end of the month (last 10 days).
- *CCD2.Attempts:* Variables that describe recent and past attempts to reach the customer, for example the number of attempts to contact the debtor or the maximum number of past attempts until a successful contact. These variables are grounded in previous contact center analytics studies, such as [4,5].



- **CCD3.Contacts:** The history of past contacts can be considered to characterize the debtor in terms of his/her behavior with the contact center. We propose constructing variables such as the number of successful contacts, the number of promises of payment made in the past, or the number of promises paid by the debtor.

After the variables are constructed, there are some important pre-processing aspects that need to be considered in any analytics project, especially when dealing with customer data. Following the CRISP-DM and KDD processes [7,11], we suggest at least the following steps: missing data imputation, outlier removal, transformation of categorical variables, and feature selection.

### 3.2. Step 3: classification and data resampling

Based on the literature review, it is quite evident that no method outperforms others for this application, and therefore we propose exploring the predictive performance of various classifiers using a suitable measure.

The class-imbalance issue is an important challenge in predictive analytics, in which the class distribution is too skewed. This problem causes classifiers to fail in the sense that they are unable to predict the minority class accurately [25]. This is an important issue in business analytics tasks, such as churn prediction, because the minority class is the one with the highest misclassification costs [20].

In order to deal with the class-imbalance issue, we propose to combine two resampling techniques: random undersampling (RUS) and the synthetic minority oversampling technique (SMOTE). First, we suggest a 100% oversampling, doubling the sample size in the minority class. Next, RUS is applied to the majority class, downsizing it in such a way that both training patterns achieve the same size. This strategy has been successfully applied in several classification tasks (see e.g. [25]). Alternatively, different resampling combinations can be explored, such as 200% or 300% oversampling.

The size of the dataset is the main reason why RUS and SMOTE are suggested. Intelligent undersampling and oversampling techniques, such as ADASYN, Borderline SMOTE, or Safe-level SMOTE, are usually too time-consuming when facing large datasets [27]. This is because they consider the majority class in the computation of distances. On the one hand, RUS is a very fast strategy that does not compute distances. On the other hand, standard SMOTE only considers the minority class. The same reasoning applies to cost-sensitive strategies, which are algorithmic solutions that deal with the class-imbalance problem during model training [27].

Based on the literature review discussed in the previous section, several classifiers can be considered for the debt collection task. We propose to evaluate the performance of at least the following five methods:

- **Logistic regression (LR):** The logistic regression is arguably the best-known classification model since it provides good predictive results while identifying the effect of each covariate on the target variable [2]. This method constructs a weighted sum of the independent variables, estimating a vector of coefficients  $\beta$  that represents the slopes of the hyperplane.
- **Decision trees (DT):** This hierarchical strategy selects the most relevant variable, generating branches from it. The best variable for the branching process is usually defined using measures such as entropy or the Gini index. Each generated node is split further in the same manner unless a pruning criterion is met. Pruning strategies are important since they reduce the risk of overfitting, limiting the size of the tree. The parameters of the pruning strategy can be tuned using a validation subset extracted from the training set [13].
- **Random forest (RF):** This ensemble algorithm creates several decision trees, combining them using majority voting. The subset of training instances that are considered for training each tree is defined using

bootstrapping, while the subset of variables is selected randomly. This strategy allows creating very different decision trees, conferring robustness to the approach. This strategy usually leads to better predictive results in comparison to a single decision tree. However, it sacrifices the interpretability that this latter method confers [13].

- **k-nearest neighbors (k-NN):** This technique assigns a label to a target sample based on its  $k$  nearest neighbors from the training set. The neighborhood is usually computed using the Euclidean or the Manhattan distance. The only parameter of this method is  $k$ , the number of neighbors. Despite its simplicity, this model can lead to good results. However, it weighs all covariates similarly, which is a strong assumption that may cause an important loss in performance [12].
- **Neural networks (ANN):** A shallow ANN model defines a network structure with one hidden layer, which confers the flexibility to construct nonlinear decision surfaces. Flexibility, however, is a double-edged sword because it increases the risk of overfitting [12]. It is important to note that, despite the success of deep learning architectures, they usually do not bring significant gains in tabular data applications [29].

Notice that we focus on well-known classifiers of different nature that have been considered previously in business analytics, and risk analytics in particular [29]. The debt collection task can be time-consuming in terms of training times as it may face large datasets. In such cases, some machine learning methods become intractable due to high computational costs.

### 3.3. Step 4: interpretation and feature selection

The final step consists of analyzing the influence of the different variables in the final solution to gain additional insights. For the feature selection and interpretation step, we propose the TreeSHAP approach [18], which has become one of the main tools for eXplainable Artificial Intelligence (XAI). Several studies have supported its positive performance in recent years, including its usage in operations management and prescriptive analytics [24].

The TreeSHAP method computes the Shapley values, a well-known measure from coalitional game theory, as an estimation of the 'payouts' that provide the various attributes [23]. The performance of the model is then disaggregated in a set of payouts that contribute to the success of the prediction, and the Shapley values represent a suitable strategy for distributing these payouts among the 'players', which are the variables in the model. Although the exhaustive computation of the Shapley values is computationally very demanding, the TreeSHAP method is a very efficient heuristic for assessing feature importance. This model also allows the construction of appealing visual representations of the variables and their relevance [18].

TreeSHAP can be used as a feature ranking method since it assesses the contribution of each variable within a black-box model. This ranking can be considered as an input for a backward elimination process, in which the  $p$  most relevant variables are selected, re-training the machine learning model with different subsets of variables. The value of  $p$  is user-defined and depends on the relevance of the variables. The TreeSHAP plots can be useful for this decision.

Model evaluation is also a very important step. We propose using the area under the ROC curve (AUC) as the performance measure. The AUC provides an adequate balance between sensitivity and specificity, making it suitable under a class-imbalance condition [13]. Furthermore, it is independent of the decision threshold as it integrates over all possible model solutions. This is desirable since we would like to identify the best classifier first, leaving the threshold decision to the financial institution based on the capacity of the contact center.

It is suggested to perform an initial training/test partition using stratified sampling, assigning 70% of the data to train the models and the remaining 30% to the final model evaluation. Parameter selection should be performed on an inner validation loop within the training set

[13]. This can be done using e.g. ten-fold cross-validation. This procedure guarantees that the test set remains unseen during the model calibration step.

### 3.4. Step 5: decision-making process

The final step of the model consists of the decision-making process. The output of the predictive models is the vector of probabilities of successful contact, promise of payment, and successful late payment. This information is useful for the following decisions:

- **Budget decisions and Staffing:** Understanding the probability of successful payments and the recovery rates leads to data-driven budgeting strategies. The predictive models could suggest that it is profitable to contact more debtors, which could lead to adjustments in the staffing plans in order to allocate more agents to this task.
- **Identification of self-cure customers:** Some of the debtors are likely to pay their debts without the need of being contacted. The identification of self-cure debtors can be done with the models proposed in this framework. For example, it may not be necessary to contact a customer with a probability of successful payment close to one. Current practices in banks to address self-cure identification include simple models and rules of thumb [9]. However, the use of mathematical models can increase collector capacity substantially by reassigning agents to the complex collection cases.
- **Two-step contacting approach:** The models discussed in this framework can be combined to improve debt collection efforts. The goal of the first contact is to reach debtors that are likely to make a promise of payment (Model 2a). In case the commitment made by the customer was not fulfilled, a second call can be made based on the predictive task designed for debtors that made a promise (Model 3b). The bank can prioritize customers that are likely to fulfill the promise, but they are late on the deadline of the commitment. This second call can be a simple reminder, but also an opportunity to present a new offer to reduce charge-offs.
- **Optimizing pre- and post-charge-off offers:** Financial institutions design offers for debtors to avoid late-stage delinquency. Some strategies include adjusting loan characteristics, re-amortizing the term, or consolidating debts [9]. These offers can be tailored based on the customer attributes and the probability of late payment.
- **Segmentation and Allocation of Agents:** In the same direction, debtors can be segmented according to their risk levels, and assigned to the corresponding agents based on the various risk categories. For example, low-risk customers (e.g. self-cure customers) can be assigned to inexperienced agents, while debtors with high risk can be assigned to skilled agents [9]. The risk segmentation can be defined based on the customer value of the debtor and the probability of late payment, which can be obtained with our models. A different approach is the one presented in Kim and Kang [16], in which analytics is used for a fair allocation of agents. This is also an interesting objective for a predictive task, however, it differs from the approach developed in this framework.
- **Value-at-risk definition:** Finally, an adequate debt collection model allows the quantification of recovery rates, which can be useful for a sophisticated Value-at-risk measure. The goal is to quantify the level of financial risk within the financial institution, going beyond the traditional measures based on delinquency time.

## 4. Experimental results

We applied the proposed framework for debt collection via contact center analytics on a dataset from a Chilean financial institution. Section 4.1 describes this dataset and the experimental setting. Next, the results are presented in sections 4.2, 4.3, and 4.4 for the contact, promise, and payment prediction tasks, respectively. Finally, a result summary is presented in Section 4.5, discussing managerial implications for

decision-making. This latter section also provides an analysis of the research questions stated in the introduction.

### 4.1. Experimental setting and dataset

The contact center dataset encompasses 4,409,437 call attempts. However, most of these records did not have financial information available, and therefore it was not possible to use them as the main purpose of this study is to evaluate the combination of the two sources. A total of 3,386,019 rows were discarded for this reason, leading to a final dataset of 1,023,418 records. The data was collected from November 2019 to January 2020.

Each record is a call attempt to a debtor. Although there could be several attempts to reach the same customer, each row is different since the number of past call attempts and the BTTC variables change for each call. Notice that a customer can appear multiple times in the dataset since it is common that retail debtors make a late payment on a given month, but they are in arrears in the following months again. Furthermore, customers are contacted multiple times until a call is successful according to our definition.

The metadata is reported in Table 2, including the sample size, the number of unique customers, and the fraction of calls in the positive class. From the total number of calls, we observe that only 75,595 are successful contacts, 26,466 of them lead to a promise of payment, and in 54,349 calls the debtor pays the delinquent debt regardless of whether a promise was made or not. As a consequence, some classification tasks are class-imbalanced.

From all the customers in our dataset, only 5.3% of them paid their debts within the ten-day period after the call. However, we observe that successfully contacted customers are about 10 times more likely to pay their dues than non-contacted ones, and those customers that made a promise of late payment are 1.6 times more likely to pay their debts than contacted customers that do not commit to pay their dues.

Notice that each row represents a call, and a customer can be called several times. Therefore, a debtor can be represented several times in the dataset. Every call to a customer is treated independently, however, this does not mean that we have duplicate rows. Every call made to a given debtor tells a different story in terms of previous contact attempts, previous promises of repayment, outstanding debt, etcetera. It makes sense to consider them in the modeling process. This approach is also followed in other risk analytics tasks, such as application and behavioral scoring [3].

Regarding data pre-processing, we replaced missing values with the mean (mode) for the numerical (nominal) variables. For each column of the final dataset, less than 1% of the values were missing. We did not find inconsistencies nor outliers that needed to be removed. Regarding data transformation, we applied binary encoding to the nominal variables.

In Step 2, we constructed variables from the two data sources, which are classified in five groups: *FD1.Debt* (7 variables), *FD2.Debtor* (16 variables), *CCD1.BTTC* (5 variables), *CCD2.Attempts* (2 variables), and *CCD3.Contacts* (4 variables). The feature engineering step then results in a total of 34 attributes from the two sources. The full list of variables is reported in the Appendix A of the supplementary material. We refer to the resulting set of variables as *Unified*. For completeness, we also explore the performance of the *Financial* set ( $FD1.Debt \cup FD1.Debtor$ )

**Table 2**  
Summary of the metadata for the five predictive tasks.

Task	Description	Sample size	Customers	% class +
1	Successful contact (full)	1,023,418	89,909	7.39%
2a	Promise-to-pay (full)	1,023,418	89,909	2.59%
2b	Promise-to-pay (contacted)	75,595	34,584	35.0%
3a	Late payment (full)	1,023,418	89,909	5.31%
3b	Late payment (promise)	26,466	20,170	8.04%

and the *Call Center* set ( $CCD1.BTTC \cup CCD2.Attempts \cup CCD3.Contacts$ ) independently.

In Step 3, we explored several combinations of different parameters for the classifiers and the resampling techniques, which are reported in Table B.1. The results for the best parameter configuration are then presented in the following sections. All methods were implemented in Python using the scikit-learn library for machine learning. For the resampling approach, we explore different combinations: 100%, 200%, and 300% oversampling via SMOTE, while RUS is applied to the majority class to match the size of the two classes. In this sense, the training set is fully balanced as the proportion of the minority class (successful contact, promise of late payment, and successful late payment) is the same in relation to the majority class. Best results were obtained with 100% oversampling in combination with RUS. The range of the various parameters associated with the classifiers and the optimal parameter configuration are reported in the Appendix B of the supplementary material. Finally, the TreeSHAP strategy was considered in Step 4 for feature selection. The  $n \in \{5, 10, 20\}$  most relevant variables are selected for each classification method and each of the predictive tasks. We refer as  $n^*$  to the optimal number of selected attributes, i.e. the one that leads to the best predictive performance.

#### 4.2. Results for the successful contact prediction task

Table 3 shows the results in terms of AUC, F1 (harmonic mean of the precision and recall), accuracy, sensitivity, and specificity for each model and set of variables. This table also provides the cardinality of the optimal set of variables  $n^*$ . We observe that the performance of using only the financial variables is much worse than with the remaining sources. The predictive performance of the *Unified* set is only slightly better than of the *Call Center* set, allowing us to conclude that a successful contact depends mostly on the call center information.

Another important conclusion that can be drawn from Table 3 is that the best predictive performance is achieved with the ANN method with both sources combined, reaching an AUC of 0.808. This performance can be considered as very good for a business analytics task [2]. This confirms the virtues of the data integration process for this task, and also the capabilities of neural networks.

Step 4 of the framework suggests that additional insights can be gained by analyzing the influence of the variables in the final classifier via TreeSHAP. Fig. 1 shows the results obtained with this method when the two sources of variables are considered. Those variables that appear at the top of the chart can be considered as the most important ones.

Fig. 1 suggests that the most valuable variables are from different sets: number of past calls and past contacts ( $CCD3.Contacts$ , top variables 1 and 3), whether the call is made during the first 10 days of the month ( $CCD1.BTTC$ , top variable 4), and, interestingly, two financial variables

**Table 3**

Predictive results for all methods and sets of variables. Successful contact prediction (Task 1). The best result (largest AUC) is highlighted in bold.

Data	CM	$n^*$	AUC	F1	Accu	Sens	Spec
Unified	LR	20	0.775	0.35	0.8	0.75	0.8
	DT	20	0.8	0.42	0.84	0.76	0.84
	RF	20	0.805	0.42	0.84	0.76	0.85
	k-NN	10	0.767	0.34	0.79	0.74	0.79
	ANN	20	<b>0.808</b>	0.38	0.82	0.8	0.82
Financial	LR	20	0.603	0.18	0.55	0.66	0.55
	DT	20	0.594	0.17	0.56	0.63	0.56
	RF	20	0.61	0.18	0.57	0.66	0.56
	k-NN	15	0.629	0.2	0.66	0.6	0.66
	ANN	20	0.607	0.18	0.55	0.67	0.54
Call Center	LR	8	0.761	0.37	0.84	0.67	0.85
	DT	8	0.784	0.41	0.85	0.7	0.87
	RF	8	0.784	0.42	0.86	0.69	0.87
	k-NN	8	0.686	0.36	0.88	0.46	0.91
	ANN	8	0.783	0.43	0.87	0.68	0.89

from  $FD1.Debt$ : number of days in arrears (Days-Debt-Assignment, top variable 2) and whether the debtor has a delay of more than 30 days in the payment of his debt (Campaign-Punishment, top variable 5). We can conclude that, even when call center information is the main data source, financial variables that describe the debt can be useful in combination with information of past contacts and best time to call variables in order to predict whether a contact is successful or not.

#### 4.3. Results for the promise of late payment prediction task

Table 4 shows the results in terms of the various performance metrics and  $n^*$  for Task 2a (full dataset of calls). In this case, results are relatively similar to the previous task in the sense that the best source is clearly the *Call Center* set. However, the predictive performance of the *Unified* set is slightly better, achieving an AUC of 0.788 for the RF classifier. We can conclude that variables such as BTTC and past contacts are again more relevant than information related to the debt and the debtor.

Next, the results for Task 2b (contacted customers only) are reported in Table 5. In this case, the predictive performance is quite poor, and the best result is achieved with RF using the *Unified* set (AUC = 0.616). The *Financial* set is slightly more relevant in this task, however, we cannot recommend the implementation of this model. This is somewhat expected since it is very hard to predict an attitude towards a commitment of payment with the available variables.

The feature ranking via TreeSHAP is illustrated in Fig. 2 for Task 2a (*Unified* set). The TreeSHAP for Task 2b is omitted due to the poor performance of all classifiers. Similar to the first task, the most valuable attributes are from different sets: the number of past calls, past promises, and past contacts ( $CCD3.Contacts$ ), whether the call is made during the first 10 days of the month ( $CCD1.BTTC$ ), the number of days in arrears, whether the debtor has a delay of more than 30 days in the payment of his debt, and the amount of delinquency debt  $FD1.Debt$ ). We can conclude that the information of the current debt has a stronger influence on the promise of late payment since three of the seven relevant variables belong to this source. We can also observe the consistency between the two predictive tasks since the five variables that are relevant in predicting a successful contact are also relevant in predicting a promise of late payment.

#### 4.4. Results for the debt payment prediction task

Table 6 presents the results of the various performance metrics and  $n^*$  for Task 3a. Similar to the previous tasks, the *Unified* set achieves better results when compared to the ones obtained with the *Call Center* set, which are both better than the ones achieved with the *Financial* set. Best results are obtained again with the RF classifier using the two sources, leading to an AUC of 0.834, which is excellent for a business analytics task.

Finally, the results for Task 3b (contacted customers that made a promise of payment only) are reported in Table 7. In this case, results are better than Task 2b but worse than Task 3a, achieving an AUC of 0.705 with RF and both sources.

Fig. 3 illustrates the results for the TreeSHAP model of Task 3a. From the variables selected by the backward elimination algorithm, seven of them generated the largest impact in the prediction: the number of past calls and past contacts ( $CCD3.Contacts$ ), whether the call is made during in the middle of the month (days 10 to 20,  $CCD1.BTTC$ ), the number of days in arrears and the amount of unpaid debt ( $FD1.Debt$ ), and the amount of previous payments made by the debtor and whether his/her risk level is high according to the classification of the bank based on previous payments ( $FD1.Debtor$ ).

Similarly, Fig. 4 presents the TreeSHAP model for Task 3b. In contrast to the analysis made for the two previous tasks, we can observe that more variables are relevant from the financial source than from the *Call Center* source. We also observe different relevant variables, including some that characterize the debtor ( $FD1.Debtor$ ).

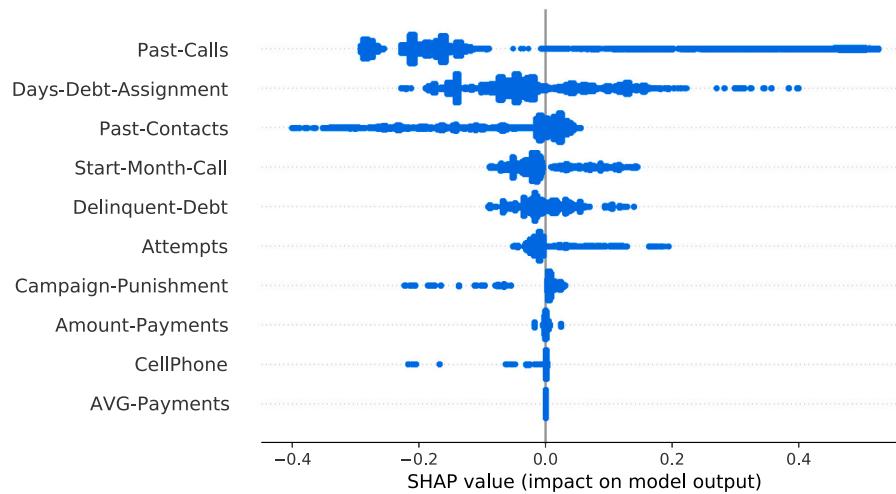


Fig. 1. Feature importance with TreeSHAP. Successful contact prediction (Task 1). The variables appear in order of relevance.

Table 4

Predictive results for all methods and sets of variables. Promise of payment prediction (Task 2a). The best result (largest AUC) is highlighted in bold.

Data	CM	n*	AUC	F1	Accu	Sens	Spec
Unified	LR	20	0.743	0.14	0.74	0.74	0.74
	DT	20	0.783	0.17	0.78	0.79	0.78
	RF	20	<b>0.788</b>	0.17	0.78	0.8	0.78
	k-NN	15	0.739	0.15	0.76	0.71	0.76
	ANN	20	0.787	0.16	0.77	0.81	0.76
Financial	LR	10	0.626	0.08	0.58	0.68	0.57
	DT	10	0.621	0.08	0.53	0.71	0.53
	RF	20	0.627	0.09	0.6	0.66	0.6
	k-NN	15	0.599	0.08	0.64	0.56	0.64
	ANN	20	0.63	0.08	0.57	0.69	0.57
Call Center	LR	10	0.717	0.15	0.79	0.64	0.8
	DT	10	0.755	0.16	0.79	0.72	0.79
	RF	10	0.755	0.17	0.81	0.7	0.81
	k-NN	10	0.701	0.23	0.91	0.48	0.92
	ANN	10	0.754	0.17	0.8	0.7	0.8

Table 5

Predictive results for all methods and sets of variables. Promise of payment prediction (Task 2b). The best result (largest AUC) is highlighted in bold.

Data	CM	n*	AUC	F1	Accu	Sens	Spec
Unified	LR	20	0.596	0.55	0.59	0.63	0.56
	DT	10	0.595	0.54	0.59	0.61	0.58
	RF	20	<b>0.616</b>	0.57	0.61	0.65	0.58
	k-NN	20	0.571	0.51	0.57	0.56	0.58
	ANN	15	0.608	0.57	0.6	0.67	0.55
Financial	LR	20	0.564	0.52	0.55	0.62	0.51
	DT	10	0.567	0.5	0.57	0.55	0.58
	RF	20	0.58	0.53	0.57	0.61	0.55
	k-NN	15	0.562	0.5	0.57	0.54	0.58
	ANN	20	0.574	0.52	0.57	0.58	0.56
Call Center	LR	15	0.584	0.53	0.58	0.6	0.56
	DT	10	0.588	0.56	0.57	0.58	0.6
	RF	10	0.601	0.56	0.59	0.66	0.56
	k-NN	10	0.578	0.52	0.58	0.59	0.56
	ANN	10	0.599	0.55	0.59	0.65	0.55

#### 4.5. Result summary and managerial insights

We derive the following insights and recommendations based on the results obtained with our case study:

- The predictive task that achieves the best results is payment prediction (Task 3a). However, the performances of the three tasks are

relatively similar and very good when the two sources are considered for at least one of the variants. In other words, the three tasks can be implemented to complement the debt collection policies. It can be also concluded that the combination of the two sources leads to best results in all the three tasks, showing the importance of the proposed data integration approach.

- The *successful contact* prediction task is an excellent strategy to understand which customers we should contact and which day and when. Variables collected from the contact center, such as past call attempts and contacts and the moment of the month the contact was made are of utmost importance here. Notice that, for this task, the decision tree model based on the contact center variables achieves the third-best performance among all methods (see Table 3). This strategy can provide interesting rules for when in the month and through which channel a customer should be contacted.
- The *promise-to-pay* prediction tasks achieved dissimilar performance: On the one hand, Task 2a showed great accuracy (AUC = 0.788), confirming it as a relevant actionable model. We suggest the implementation of Task 2a for a first contact with the customers as it incorporates the commitment of the debtor in the contacting process. On the other hand, Task 2b fails at achieving positive results, and therefore it cannot be considered for managerial decisions.
- The *payment* prediction tasks also achieved dissimilar performance: the largest AUC values are 0.834 and 0.705 for Task 3a and 3b, respectively. Although the former task leads to best results, we suggest Task 3b for seeking a second contact for customers that committed to pay and did not fulfill the promise, as suggested in Section 3.4. Financial behavior prior to the contact is of utmost importance in this task, with variables such as the amount of previous payments and the days in debt being the most relevant ones for this task (top variables 1 and 2 according to the TreeSHAP approach reported in Fig. 4).
- Regarding the classifiers, random forest showed best predictive performance, achieving the largest AUC in four of the five tasks, followed by ANNs. The proposed feature selection methodology based on TreeSHAP also showed benefits as the best performance is usually achieved with either 15 or 20 of the 34 variables (see Tables 3 to 7). Feature selection not only is able to improve predictive performance in terms of AUC, but also provides a valuable strategy to gain insights into the application. We have identified and discussed the top variables using TreeSHAP (see Figs. 1 to 3), conferring interpretability to the approach. The interpretation of the predictive models is key to improve debt collection policies and to design the script of the calls.



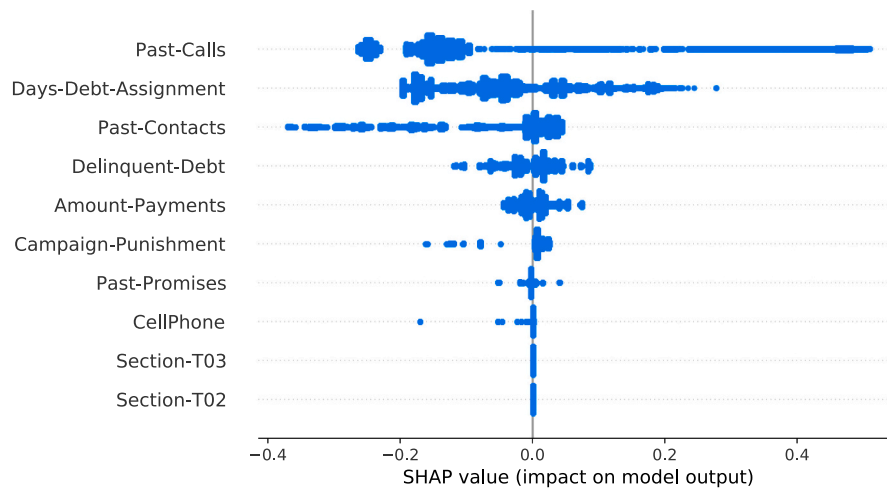


Fig. 2. Feature importance with TreeSHAP. Promise of payment prediction (Task 2a).

Table 6

Predictive results for all methods and sets of variables. Late payment prediction (Task 3a). The best result (largest AUC) is highlighted in bold.

Data	CM	n*	AUC	F1	Accu	Sens	Spec
Unified	LR	20	0.796	0.28	0.78	0.81	0.78
	DT	20	0.817	0.29	0.79	0.85	0.78
	RF	20	<b>0.834</b>	0.32	0.81	0.86	0.81
	k-NN	20	0.794	0.29	0.8	0.79	0.8
	ANN	20	0.831	0.31	0.8	0.86	0.8
Financial	LR	15	0.708	0.19	0.69	0.73	0.68
	DT	10	0.712	0.17	0.59	0.85	0.58
	RF	20	0.723	0.19	0.63	0.83	0.62
	k-NN	20	0.693	0.19	0.7	0.69	0.7
	ANN	20	0.72	0.18	0.61	0.84	0.6
Call Center	LR	10	0.742	0.28	0.83	0.64	0.84
	DT	10	0.769	0.3	0.84	0.7	0.84
	RF	10	0.77	0.31	0.84	0.69	0.85
	k-NN	10	0.727	0.31	0.87	0.57	0.89
	ANN	10	0.769	0.31	0.85	0.68	0.86

Table 7

Predictive results for all methods and sets of variables. Late payment prediction (Task 3b). The best result (largest AUC) is highlighted in bold.

Data	CM	n*	AUC	F1	Accu	Sens	Spec
Unified	LR	15	0.68	0.3	0.67	0.7	0.66
	DT	15	0.69	0.3	0.64	0.76	0.62
	RF	15	<b>0.705</b>	0.32	0.68	0.74	0.67
	k-NN	10	0.636	0.27	0.66	0.61	0.66
	ANN	20	0.7	0.31	0.67	0.74	0.66
Financial	LR	15	0.66	0.28	0.62	0.71	0.61
	DT	10	0.662	0.28	0.63	0.71	0.62
	RF	15	0.665	0.29	0.66	0.68	0.65
	k-NN	15	0.613	0.25	0.63	0.59	0.64
	ANN	20	0.664	0.28	0.62	0.72	0.61
Call Center	LR	15	0.596	0.23	0.62	0.56	0.63
	DT	15	0.592	0.23	0.61	0.57	0.62
	RF	15	0.596	0.24	0.67	0.5	0.69
	k-NN	15	0.562	0.21	0.69	0.41	0.72
	ANN	15	0.597	0.23	0.59	0.61	0.58

Based on our previous analysis, we next discuss the research questions stated in the introductory section:

**(RQ1):** Can we predict late payment accurately to improve collection strategies? The answer is clearly yes. All the predictive methods achieve a positive performance, with the exception of Task 2b. Late payment can be either addressed by identifying the customers that are likely to commit to pay their debts (Task 2a) or the ones that eventually pay (Task

3a). Both predictive tasks achieved an AUC above 0.78.

**(RQ2):** Is contact center information a valuable data source for late payment prediction? The answer is also yes. In most cases, the combination of both sources leads to dramatically better results in relation to using one of the sources. Some tasks are strongly related to one of the sources, such as Task 1 and contact center data or Task 3b and financial data. However, the tasks designed to address the late payment prediction challenge as a first contact (Task 2a and Task 3a) require both data sources to achieve an optimal performance.

**(RQ3):** Does a promise of late payment have an impact on the debt collection policies? Yes. As indicated before, customers that made a promise of payment are 1.6 times more likely to pay their debts than contacted customers that do not commit to pay. Furthermore, the information on past promises leads to relevant predictors in some of the tasks. Finally, we argue that predicting a promise of late payment can be an excellent approach for first time contact to avoid self-cure debtors. In Task 3a, for example, we have customers that repay their debts even when the contact was not successful. Task 3a is the traditional approach to late payment prediction, in which information of previous contacts is usually ignored. Task 2a and 3b, in contrast, do not consider a success ( $y = 1$ ) when a customer pays their debts without being contacted properly.

**(RQ4):** Can we address first time contacts and follow-ups with two different prediction models? We propose the use of Task 2a and Task 3b for first time contacts and follow-ups, respectively. These two tasks lead to very different models and can be combined in order to reach different customers and enhance the debt collection policies.

Since Task 2a and Task 3b are considered the most relevant ones in this study for an adequate debt collection policy, we present the two resulting decision trees associated with these tasks in the Appendix C of the supplementary material. Notice that the performance of this classifier is relatively similar to the top methods (RF and ANN), and therefore can be considered when interpretability is an important requirement in the modeling process. Nevertheless, the most relevant attributes in the trees coincide with the ones in the TreeSHAP plots, and therefore the same conclusions can be drawn from the use of this classifier.

## 5. Conclusions

This paper presents a predictive analytics approach for addressing the debt collection task. Although there are some studies that aim at identifying the right customers to target with collection efforts (see e.g. [4,28,32]), our proposal is the first formal framework designed for integrating contact center and financial information, to the best of our knowledge. Our proposal also extends the existing literature by

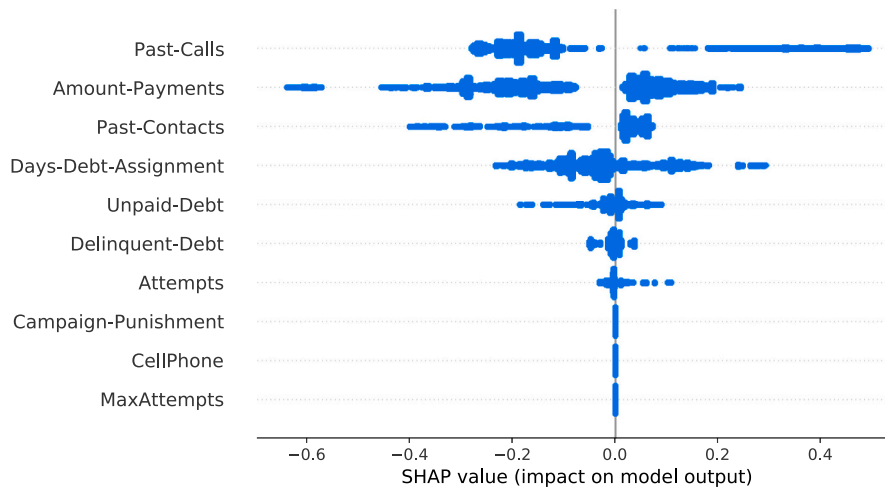


Fig. 3. Feature importance with TreeSHAP. Late payment prediction (Task 3a).

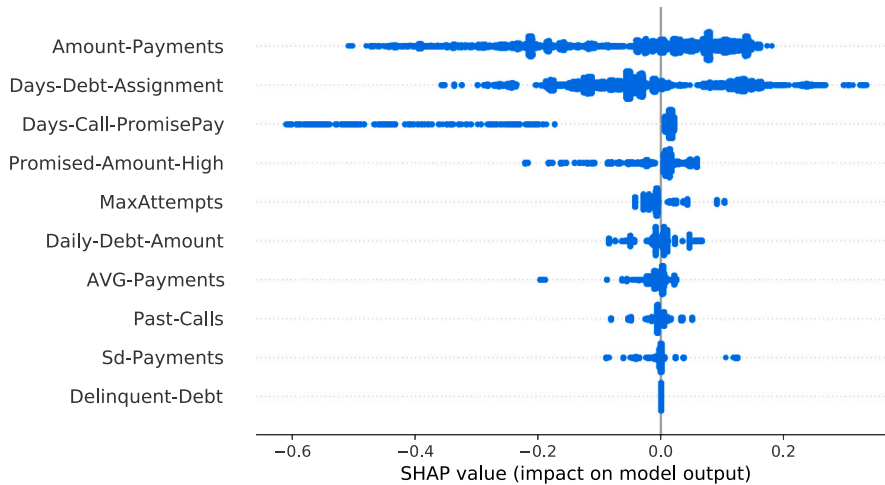


Fig. 4. Feature importance with TreeSHAP. Late payment prediction (Task 3b).

proposing two novel predictive tasks: successful contact and promise of payment prediction. We strongly believe the latter task is more suitable for debt collection than predicting those customers that will pay their dues independent of being contacted. A total of five different tasks are discussed in this study, addressing the four research questions.

Our approach is based on recent explainable methods such as TreeSHAP in order to extract knowledge from the predictive methods. We can conclude that historical information from previous successful contacts and payments are the most variable source of information for the debt collection task, followed by variables that describe the debt itself, such as the amount of delinquency debt, number of days in arrears, and whether there is a delay of more than 30 days in the payment. BTTC variables such as when in the week or the month a customer should be contacted also improved predictions.

One limitation of the case study is that we were not able to obtain rich sociodemographic and financial information of the debtors. The bank was able to provide only a few variables that characterize the customers, and most of them were irrelevant for prediction. We strongly believe that our models can be improved further with an adequate set of financial variables. This information is usually available in banks that apply debt collection policies. A second limitation of our approach is that it is based on a single debt collection dataset. Our conclusions should be considered with caution for this reason. Given the applied nature of this study, we were unable to find additional datasets to

strengthen our findings.

There are some opportunities for future developments. For example, the conundrum of the best predictive task can be analyzed further using causal machine learning methods, such as uplift modeling [29]. With these techniques, we can contrast the differences in terms of payment behavior between the ‘contacted’ and ‘non-contacted’ groups using A/B testing. Additionally, we can extract more information about the contacts and the promises made by the customers using text analytics, which is an important trend in contact center analytics [17]. The content of a successful contact can be transcribed automatically using speech to text techniques, and analyzed with the objective of identifying what triggers a promise of payment. Finally, the use of profit metrics can be valuable for defining the final set of customers to be contacted [20]. The trade-off between the benefits and costs of the debt collection campaign, prioritizing not only those customers that are more likely to make a promise of payment but also those to whom the action leads to the largest profitability.

#### Acknowledgements

The authors gratefully acknowledge financial support from ANID PIA/BASAL AFB180003 and FONDECYT-Chile, grants 1200221 (Sebastián Maldonado) and 12200007 (Carla Vairetti). Catalina Sánchez also acknowledges a grant provided by UANDES-FAI for her Ph.D.

studies. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions for improving the quality of the paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dss.2022.113812>.

## References

- [1] N. Abe, P. Melville, C. Pendus, C. Reddy, D. Jensen, V. Thomas, J. Bennett, G. Anderson, B. Cooley, M. Kowalczyk, M. Domick, T. Gardiner, Optimizing debt collections using constrained reinforcement learning, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010, pp. 75–84.
- [2] B. Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, 1st edition, Wiley Publishing, 2014.
- [3] B. Baesens, D. Roesch, H. Scheule, Credit Risk Analytics: Measurement Techniques, Applications, and Examples in SAS, John Wiley & Sons, 2016.
- [4] M. Bahrami, B. Bozkaya, S. Balcişoy, Using behavioral analytics to predict customer invoice payment, Big Data 8 (1) (2020) 25–37.
- [5] A. Bellotti, D. Brigo, P. Gambetti, F. Vrin, Forecasting recovery rates on non-performing loans with machine learning, Int. J. Forecast. 37 (1) (2021) 428–444.
- [6] J. Binesh, T.R. Ramanan, S.M. Kumar, Big data provenance and analytics in telecom contact centers, in: TENCON 2017–2017 IEEE Region 10 Conference, IEEE, 2017, pp. 1573–1578.
- [7] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, et al., Crisp-dm 1.0: Step-by-Step Data Mining Guide 9, SPSS inc, 2000, p. 13.
- [8] N. Chehrizi, P.W. Glynn, T.A. Weber, Dynamic credit-collections optimization, Manag. Sci. 65 (6) (2019) 2737–2769.
- [9] I. Crespo, A. Govindarajan, The analytics-enabled collections model, in: Technical Report, McKinsey & Company, 2018. April.
- [10] X.L. Dong, T. Rekatsinas, Data integration and machine learning: A natural synergy, in: Proceedings of the 2018 International Conference on Management of Data, 2018, pp. 1645–1650.
- [11] U. Fayyad, Data mining and knowledge discovery: making sense out of data, IEEE Expert 11 (5) (1996) 20–25.
- [12] J. Han, M. Kamber, J. Pei, Data mining, concepts and techniques, in: The Morgan Kaufmann Series in Data Management Systems, third edition, 2012.
- [13] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Science & Business Media, 2009.
- [14] N. Ilk, M. Brusco, P. Goes, Workforce management in omnichannel service centers with heterogeneous channel response urgencies, Decis. Support. Syst. 105 (2018) 13–23.
- [15] N. Ilk, G. Shang, P. Goes, Improving customer routing in contact centers: an automated triage design based on text analytics, J. Oper. Manag. 66 (5) (2020) 553–577.
- [16] J. Kim, P. Kang, Late payment prediction models for fair allocation of customer contact lists to call center agents, Decis. Support. Syst. 85 (2016) 84–101.
- [17] S. Lam, C. Chen, K. Kim, G. Wilson, J.H. Crews, M.S. Gerber, Optimizing customer-agent interactions with natural language processing and machine learning, in: 2019 Systems and Information Engineering Design Symposium (SIEDS), IEEE, 2019, pp. 1–6.
- [18] S. Lundberg, G. Erion, H. Chen, A. DeGrave, J. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable ai for trees, Nat. Mach. Intellig. 2 (2020).
- [19] S. Maldonado, C. Bravo, J. López, J. Pérez, Integrated framework for profit-based feature selection and svm classification in credit scoring, Decis. Support. Syst. 104 (2017) 113–121.
- [20] S. Maldonado, G. Domínguez, D. Olaya, W. Verbeke, Profit-driven churn prediction for the mutual fund industry: a multisegment approach, Omega 100 (2021), 102380.
- [21] T. Meng, X. Jing, Z. Yan, W. Pedrycz, A survey on machine learning for data fusion, Informat. Fusion 57 (2020) 115–129.
- [22] M. Mitchner, R.P. Peterson, An operations-research study of the collection of defaulted loans, Oper. Res. 5 (4) (1957) 522–545.
- [23] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, URL, <https://christophm.github.io/interpretable-ml-book>, 2020.
- [24] P.M. Notz, Explainable subgradient tree boosting for prescriptive analytics in operations management, SSRN Electron. J. (2020) 1–34.
- [25] S. Simsek, A. Dag, T. Tiahrt, A. Oztekin, A bayesian belief network-based probabilistic mechanism to determine patient no-show risk categories, Omega 100 (2020), 102296.
- [26] L. Thomas, J. Crook, D. Edelman, Credit scoring and its applications, second edition, Society for Industrial and Applied Mathematics, 2017.
- [27] I. Triguero, M. Galar, D. Merino, J. Maillo, H. Bustince, F. Herrera, Evolutionary undersampling for extremely imbalanced big data classification under apache spark, in: 2016 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2016, pp. 640–647.
- [28] R. van der Geer, Q. Wang, S. Bhulai, Data-driven consumer debt collection via machine learning and approximate dynamic programming, SSRN Electron. J. (2018) 1–32.
- [29] W. Verbeke, B. Baesens, C. Bravo, Profit Driven Business Analytics, Wiley, 2017.
- [30] S.F. Wamba, D. Mishra, Big data integration with business processes: a literature review, Bus. Process. Manag. J. 23 (2017) 477–492.
- [31] M.F. Weeks, R.A. Kulka, S.A. Pierson, Optimal call scheduling for a telephone survey, Public Opin. Q. 51 (4) (1987) 540–549.
- [32] J. Zurada, S. Lonial, Comparison of the performance of several data mining methods for bad debt recovery in the healthcare industry, J. Appl. Bus. Res. (JABR) 21 (2) (2005).



**Catalina Sánchez** received her B.S. and M.S. degrees from the Universidad de Los Andes, in 2021. She is currently Ph.D. student in Engineering Sciences in the same university. This manuscript is the result of her master thesis, which was supervised by Prof. Carla Vairetti.



**Sebastián Maldonado** received his B.S. and M.S. degrees from the University of Chile, in 2007, and his Ph.D. degree from the University of Chile, in 2011. He is currently Full Professor at the Department of Management Control and Information Systems, School of Economics and Business, University of Chile. His research interests include statistical learning, data mining and business analytics. Sebastián Maldonado has published more than 70 scientific contributions in the last ten years.



**Carla Vairetti** received her B.S. degree in Computer Science in 2000 from the University Nacional de La Plata, Argentina. She also received an M.S. degree in Sciences in 2013 from the Pontificia Universidad Católica, Chile, and the Ph.D. degree in Engineering Sciences in 2016 from the University of Trento, Italy. Currently, she is Assistant Professor of Universidad de Los Andes. Her research interests include classification in imbalanced domains, data science in big data applications, and computational Intelligence (including machine and deep learning).