Electronic Commerce Research and Applications 67 (2024) 101428

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/elerap

Profit-sensitive machine learning classification with explanations in credit risk: The case of small businesses in peer-to-peer lending[☆]

Miller-Janny Ariza-Garzón [a,*], Javier Arroyo [b,c], María-Jesús Segovia-Vargas [d], Antonio Caparrini [a]

[a] Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, 28040 Madrid, Spain
[b] Departamento de Ingeniería del Software e Inteligencia Artificial, Universidad Complutense de Madrid, 28040 Madrid, Spain
[c] Instituto de Tecnología del Conocimiento, Universidad Complutense de Madrid, 28223 Madrid, Spain
[d] Departamento de Economía Financiera y Actuarial y Estadística, Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid, 28223 Madrid, Spain

ARTICLE INFO

Keywords:
Credit risk
P2P lending
Small business loans
Cost-sensitive models
Profit-sensitive learning
Extreme gradient boosting
Explainability
Shapley values

ABSTRACT

We propose a comprehensive profit-sensitive approach for credit risk modeling in P2P lending for small businesses, one of the most financially complex segments. We go beyond traditional and cost-sensitive approaches by including the financial costs and incomes through profits and introducing the profit information at three points of the modeling process: the estimation of the learning function of the classification algorithm (XGBoost in our case), the hyperparameter optimization, and the decision function. The profit-sensitive approaches achieve a higher level of profitability than the profit-insensitive approach in the small business case analyzed by granting mostly lower-risk, lower-amount loans. Explainability tools help us to discover the key features of such loans. Our proposal can be extended to other loan markets or other classification problems as long as the cells of the misclassification matrix have an economic value.

1. Introduction

Technological advancements and financial crises have given rise to peer-to-peer (P2P) lending, a new model that allows lenders and borrowers to interact directly, bypassing traditional intermediaries like banks. Borrowers can secure credit at rates lower than traditional alternatives, while investors often see returns that match or surpass other investment options (Emekter et al., 2015). Importantly, P2P lending extends credit access to individuals typically excluded from conventional financial products and services (Maskara, 2020). Thus, P2P is seen as a complementary market to traditional banking (Milne and Parboteeah, 2016).

However, P2P lending comes with risks, primarily due to increased information asymmetry (Serrano-Cinca and Gutierrez-Nieto, 2016) and moral hazard (Cummins et al., 2019). These risks are intensified as lenders typically bear the credit risk (Serrano-Cinca and Gutierrez-Nieto, 2016). Moreover, P2P lending is susceptible to systemic financial market events associated with risk factors such as falling income, rising unemployment, and the tightening of lending requirements by traditional entities, such as the COVID-19 pandemic. Moreover, some platforms prioritize attracting borrowers by adjusting interest rates, leading to risky loans and increased default probabilities (Klein et al., 2021). Therefore, robust risk management is crucial in the P2P market, including proper credit and financial risk management (Gao et al., 2020), and the implementation of novel, transparent, and tailored regulation policies (Expert Group on Regulatory Obstacles to Financial Innovation, 2019).

To address this, machine learning methods are emerging for default prediction in P2P lending, offering better performance than traditional methods. Various models are used, including tree-based models (Serrano-Cinca and Gutierrez-Nieto, 2016; Misheva et al., 2018), Support Vector Machines (Rodrigues et al., 2018; Bastani et al., 2019), artificial neural networks (Byanjankar et al., 2015; Zang et al., 2015; Lyócsa et al., 2022) or deep learning (Duan, 2019; Fu et al., 2019; Van-Sang et al., 2019; Li et al., 2020). Among these, XGBoost (Chen and Guestrin, 2016), a gradient boosting technique, has gained attention for

---

[☆] This research was supported by European Cooperation in Science and Technology [COST Action 19130] and by the Spanish Ministry of Science and Innovation [Grant PID2020- 115700RB-I00].

[*] Corresponding author.

E-mail addresses: millerar@ucm.es (M.-J. Ariza-Garzón), javier.arroyo@fdi.ucm.es (J. Arroyo), mjsegovia@ccee.ucm.es (M.-J. Segovia-Vargas), acaparri@ucm.es (A. Caparrini).

https://doi.org/10.1016/j.elerap.2024.101428
Received 8 September 2023; Received in revised form 9 May 2024; Accepted 14 June 2024

its superior performance (Boiko Ferreira et al., 2017; Xia et al., 2017; Li et al., 2018; Ariza-Garzón et al., 2021; Chang et al., 2022).

However, the use of machine learning alone is not enough to solve all risk assessment problems. A common problem in credit risk is that the misclassification cost varies widely. In particular, classifying a loan that will default as non-defaulting represents a loss to the lender. However, classifying a non-defaulting loan as a default may represent a lost business opportunity, but not a cost.

This problem can be addressed with cost-sensitive approaches which differently weigh the possible errors of the prediction (false positives and false negatives). In doing so, they not only deal with the class imbalance problem, which is crucial in credit risk where the default class is a minority, but they also serve to incorporate a business sense into the classification methods. As a result, cost-sensitive models provide credit-granting decisions based on misclassification costs, facilitating their use in real life. As noted by Ariza-Garzón et al. (2021), most models for the P2P market focus on predicting default or estimating the probability of default, and they are still evaluated with traditional performance measures that are either unaware of or insensitive to the costs of miss-classification. Few works have studied credit risk using cost-sensitive alternatives, and of those that do, a significant percentage have done it for the P2P lending market (Malekipirbazari and Aksakalli, 2015; Boiko Ferreira et al., 2017; Xia et al., 2017; Wang et al., 2018; Ye et al., 2018; Cho et al., 2019). A further refinement to the approach includes not only the classification costs but also the classification profits. That way, we have a profit-sensitive approach which has a more comprehensive business sense as it aims to maximize the profit and minimize the costs. Only Verbraken et al. (2014) and Herasymovych et al. (2019) do so in the credit market in general and Ye et al. (2018) in the P2P lending market, as we will review later. In our work, we propose using a profit-sensitive approach for the well-known XGBoost classifier to delve into this research avenue.

The literature has explored other ways of including profits in credit modeling, such as profit scoring models (Serrano-Cinca and Gutierrez-Nieto, 2016; Bastani et al., 2019; Babaei and Bamdad, 2020; Lyócsa et al., 2022; Wang et al., 2022). These models work with granted loans, and the target variable corresponds to a continuous quantitative variable of profitability per obligation (e.g. internal rate of return or the annualized rate of return). In other words, this approach usually separates risk (decision on default) and profit modeling, while profit-sensitive models like ours combine them in a single modeling process and have hardly been investigated not only in P2P lending but also in credit risk.

In profit-sensitive models, the inclusion of profits and costs in the classifier can be done directly or indirectly. Following the terminology introduced by Xia et al. (2017), in a direct approach, the values are introduced in the classifier to guide its learning process. On the other hand, in an indirect approach, these values are included to provide a sense of cost (or profit) to other steps of the modeling process. For example, during pre-processing (e.g. by applying sampling or re-sampling techniques on the initial dataset), during the model tuning (e.g. guiding the search for optimal hyperparameters, which affects the classifier learning, but in an indirect manner), or in the decision phase by modifying the outputs of a classifier (e.g. by changing the threshold or cutoff in a probabilistic model to classify as default or non-default category in the decision process). In our approach, we propose to incorporate this information throughout the process. In particular, we will include it at three points:

- In the tuning of the machine learning algorithm, as an expected profit function to be maximized in the hyperparameter optimization process.
- In the inner learning process of the XGBoost. In particular, we guide the XGBoost learning to minimize the inverse of our expected profit function using its gradient and Hessian.
- At the decision stage, we will take the probability provided by the XGBoost algorithm and explore two profit-sensitive decision methods to define prediction.
    - Setting the optimal decision threshold in terms of profit as other works did (Verbraken et al., 2014; Herasymovych et al., 2019).
    - A Bayesian decision process following the ideas of Bahnsen et al. (2014).

As a result, we propose a comprehensive procedure that uses a profit function (using profit and cost information) throughout the classification modeling process. Our approach uses the same profit function to ensure consistency throughout the process. We will investigate whether including profit information at the three modeling points is financially beneficial or not.

In addition, we will investigate whether the inclusion of profits and costs changes the behavior of the classifier and the way it uses the input variables. To do so, we will use the SHAP values (SHapley Additive exPlanations) proposed by Lundberg and Lee (2017a) and Lundberg et al. (2020). The SHAP values provide transparency and explainability to machine learning classification models at both local and global levels. In credit risk models, SHAP values help to determine how risk factors participate individually by obligation and at an aggregated level in the models (Ariza-Garzon et al., 2020; Bussmann et al., 2020). Their use also brings some transparency and control to machine learning credit risk models, which is required by regulators (Expert Group on Regulatory Obstacles to Financial Innovation, 2019) and welcome by users. In our work, we will extend the SHAP values to be used in a profit-sensitive classification context and compare them with the traditional SHAP values from the classification algorithm without considering the profits. In this way, we aim to understand what factors drive profitable loans and if they differ from those driving the traditional classification problem in credit risk.

For the empirical evaluation of our proposal, we will use the dataset from Lending Club, which has been extensively used in other works in recent years (Boiko Ferreira et al., 2017; Durovic, 2017; Xia et al., 2017; Bastani et al., 2019; Calabrese et al., 2019; Cho et al., 2019; Duan, 2019). As a novelty, we will focus on the segment of small businesses, which has been understudied in the literature (Nowak et al., 2018). The P2P market is a funding source for small businesses, which often struggle to access financing (Mach et al., 2014; Sharma et al., 2024). However, the small business sector is high-risk due to information opacity (Mills and McCarthy, 2016), making traditional risk models unsuitable (Chang et al., 2022). We will test our proposal in this challenging sector that demands better risk management models.

The rest of this document is organized as follows. Section 2 introduces the family of cost-sensitive measures and models and reviews the main works in P2P lending. Section 3 presents our proposal of profit-sensitive measures derived from the misclassification matrix, and Section 4 our profit-sensitive machine learning approach with explainability. Section 5 describes the empirical setting, and Section 6 analyzes the empirical results. Section 7 summarizes the main findings.

2. Cost-sensitive methods in loan evaluation

2.1. Classification errors and cost-sensitive methods

Standard machine learning and statistical methods typically use the so-called 0–1 loss function, assigning 0 to a correctly classified instance and 1 to an incorrect one. Since the 0–1 loss function uses the same cost associated with a miss-classification for all classes considered, it is highly susceptible to skewed class distributions (Fernández et al., 2018) and a non-real valuation of the associated costs.

In credit risk modeling, we typically have a binary classification with two categories: Default $y_i = 1$ and Non-default $y_i = 0$. In a cost-sensitive approach, the 2×2 confusion matrix summarizes the costs of each classification decision for each instance. The matrix entries are given by the costs associated with miss-classification $C(k,l)$ for each instance, where k represents the predicted label $\hat{y}_i$ and l the actual label y_i. In a typical cost-sensitive approach, the only cost considered is the cost of a false negative, $C(0,1)$, which is the cost of classifying a bad loan as non-default because it represents a financial cost for the company. On the other hand, $C(1,0) = C(0,0) = C(1,1) = 0$, as these outcomes do not suppose a cost strictly, even if they impact the financial result. However, in standard machine learning and statistical approaches, the loss function typically considers $C(1,1) = C(0,0) = 0$ and $C(1,0) = C(0,1) = 1$.

From a business perspective, it makes sense to broaden the focus to include profits or even other costs, such as opportunity costs, resulting in a profit-sensitive confusion matrix. In this setting, $C(1,0)$ is the cost of miss-classifying a good loan as a default (false positive, FP), and $C(0,1)$ is the cost of classifying a bad loan as non-default (false negative, FN). $C(1,1)$ corresponds to the profit of achieving a true positive (TP), that is, correctly predicting a default, and $C(0,0)$ the profit of obtaining a true negative (TN), that is, correctly predicting a good loan.

If we use costs and profits to measure the performance of a model, it is also natural to include such financial values to learn how to classify with a business sense. The set of techniques that use this approach is known as cost-sensitive methods (or profit-sensitive methods if they include profits). While those that assume the same miss-classification costs for all cases are known as cost-insensitive methods (Ling and Sheng, 2010)

In the literature, cost-sensitive methods can be classified into direct and indirect (Ling and Sheng, 2010; Fernández et al., 2018). The direct approach introduces the miss-classification costs into a classifier's training and learning procedure. In the indirect approach (Xia et al., 2017), also known as meta-learning, cost-insensitive classifiers become cost-sensitive by pre-processing (e.g., sampling or resampling) the training data or by post-processing by modifying the outputs of a classifier during the classification phase (e.g. by changing the threshold or cutoff in a probabilistic model to classify as default or non-default category).

Other taxonomies of algorithms dealing with the class imbalance and Cost-Sensitive Methods are proposed by He and Garcia (2009) and Petrides and Verbeke (2022). In the first work, the authors consider a category where methods incorporate cost-sensitive functions directly into the classification algorithms. Another category includes an expected cost function in the decision criteria for the best modeling alternative. The second work categorizes the cost-sensitive models considering the misclassification costs and defines categories of pre-training, training, and post-training methods. It includes a category based on sampling in the pre-training stage, where the over-sample or under-sample techniques are determined by the ratio between false negative and false positive costs and is called sampling-based. In our work, we include this category in the indirect methods.

Also, Petrides and Verbeke (2022) draw attention to a category where records are weighted by their costs and used with any algorithm. For us, it corresponds with a direct method. These costs can also be used in ensemble methods, which the authors call cost-sensitive ensemble methods. They also mention the Cost-Sensitive Post-training methods, which include the definition of a threshold for the probability, which is estimated based on a process of minimization of an expected function of classification costs, which we call and use as an indirect method.

2.2. Review of cost-sensitive and profit-sensitive methods for loan evaluation

We can find cost-sensitive methods in diverse financial applications, such as credit card fraud (Correa Bahnsen et al., 2016; Akila and Reddy, 2018; Correa Bahnsen et al., 2018; Nami and Shajari, 2018; Almhaithawi et al., 2020), credit card business (Zhang et al., 2020), business failure prediction (Zou et al., 2022) or retail credit (Oreski and Oreski, 2018). As mentioned by Xia et al. (2017), cost-sensitive loan evaluation models have received little attention not only in P2P lending but also in traditional loan evaluation, and, when used, most of them are indirect methods. In this section, we complement the review by Xia et al. (2017) analyzing the cost-sensitive loan evaluation approaches proposed recently. We will not delve into proposals on indirect pre-processing methods, such Bastani et al. (2019), Duan (2019), or (Zhu et al., 2019), because our proposal focuses on the learning and the decision processes and does not take into account the pre-processing stage. Similarly, we will not review the works that deal with the class imbalance problem but do not use a function that depends on the confusion matrix (Calabrese et al., 2019; Cho et al., 2019; Jadwal et al., 2020). Table 1 summarizes the proposals analyzed.

Regarding indirect methods, Malekipirbazari and Aksakalli (2015) uses random forest (RF) as a base model and incorporates a cost-weighted matrix that increases the cost of miss-classification associated with bad borrowers through a cost-sensitive function. The study includes accuracy (ACC) and the area under the curve (AUC), among other traditional measures for performance evaluation.

Wang et al. (2018) compare the incorporation of different non-financial cost matrices indirectly on different classification methods such as C4.5 Decision Tree (DT C4.5), Naive Bayes (NB), Logistic regression (LR) and Support Vector Machine (SVM). For the comparison, they use non-financial cost measures plus other typical performance measures such as ACC.

Herasymovych et al. (2019) propose a dynamic reinforcement learning approach to optimize the acceptance threshold of a credit score. They maximize an expected profit function similar to the one proposed by Verbraken et al. (2014), which modifies the threshold of the classifiers to achieve incremental financial profit, taking into account Loss Given Default (LGD), the interest rate, and other financial components across the entire loan data set. In particular, they develop a dynamic reinforcement learning system that adapts the threshold in response to the changing data distribution to maximize profits. From our point of view, they develop a dynamic approach of an indirect thresholding method. They use as an objective function an expected classification profit per loan for a threshold value and assume $C(1,0) = C(1,1) = 0$. Their evaluation is performed using the convergence of the optimization method, along with AUC and total profits. This work incorporates the profits of each cell, aggregates them in the function, and optimizes them, considering a maximization problem; thus, it can be considered a profit-sensitive model.

Wang et al. (2021) estimate the different credit risk grades for a data set from Lending Club. They use a cost-sensitive matrix with financial elements like LGD, returns interest rate, and the probability of default associated with the different risk grades and transitions. They apply an indirect cost-sensitive approach to DT C4.5, RF, LR, and SVM methods. They use ACC and financial measures like total cost, average cost, and cost-saving rate (decreased cost after using cost-sensitive) to measure performance. Although the ACC is low in the cost-sensitive proposals, their results show that costs are reduced, obtaining a more financially efficient method. This work highlights the importance and benefits of using financial information to construct the cost matrix.

Regarding direct methods, Bahnsen et al. (2015) develop a direct cost-sensitive decision tree that introduces cost-sensitive elements in the impurity measure and the pruning criterion. They apply their method in three contexts. First, in direct marketing for classifying customers more likely to accept an offer to open a bank long-term deposit account. Second, in credit card fraud, and lastly, the one with the most interest for this work, in loan evaluation. In particular, they assume $C(k,l) = 0$ when $k = l$ and incorporate financial elements such as LGD for the case of false negatives. They also consider the opportunity cost determined by the loss in profit (e.g. the interest rate) by rejecting what would have been a good loan, netting the expected profit that the lenders can obtain by using the money in other loans, for the case of false positives. The evaluation uses traditional performance measures such as ACC or the F1-Score but includes a financial measure called cost savings. Their proposal is compared with other alternatives containing indirect cost-sensitive methods such as Bayes Minimum Risk (Bahnsen et al., 2014), showing the profits of business-oriented decision-making. It is worth mentioning that, although its proposal includes the value of profits in some cells, it is still oriented towards cost minimization; thus, it still can be considered a cost-sensitive approach.

Maldonado et al. (2017) address the class imbalance problem but incorporates alternative information acquisition costs into the classification problem. They predict loan default using Support Vector Machines (SVM), including these costs in the objective function within the optimization process, and using constraints that aid in the feature selection process. The proposal can be considered direct since the costs are included in the algorithm optimization. However, it does not deal with the costs in the classification matrix.

Xia et al. (2017) propose a cost-sensitive XGBoost model that uses both direct and indirect approaches. However, they incorporate the financial information only in the indirect approach, more precisely for the hyperparameterization of their methods, where they maximize the average annualized return (ARR). Note that this information is not derived from the misclassification matrix. In addition, they compare their proposal against different cost-insensitive alternatives (LR, RF) converted into cost-sensitive options through thresholding, also maximizing the ARR. Their main proposal is a direct cost-sensitive model based on the XGBoost model, although it does not directly incorporate financial costs and profits from the misclassification matrix, as we will do. The incorporation of cost information is done using an alternative cost function whose gradient and Hessian are already developed. Specifically, they use the structure of a cost-insensitive loss function and transform it into a cost-sensitive one through the cost-sensitive logistic loss proposed by Masnadi-Shirazi and Vasconcelos (2011), assuming costs $C(k,l)$ with $k \neq l$ and 0 when $k = l$, incorporated through two parameters that relate these costs. For the decision of the model, they use 0.5 as the threshold after a calibration process of the probabilities with Platt's scaling. At a later stage, the results of their model are used to construct Markowitz-type P2P lending investment portfolios, whose purpose is to find the most profitable loans with the lowest risk. In this proposal, the inclusion of cost information in the classification matrix is done using an alternative cost function whose gradient and Hessian are already developed.

Ye et al. (2018) consider for each obligation not only the costs but also the profits derived from the confusion matrix results and incorporate them directly in the objective function. Thus, they present a profit-sensitive model. Specifically, they use genetic algorithms to establish the optimal weights that maximize the resulting profit function in a Random Forest voting process. Having the profits and not only the costs as an objective function is one of the few cost-sensitive works incorporating information for $C(k,l)$ when $k = l$. Their performance measure and objective function is a profit function that mainly incorporates the principal and interest of each loan. They also evaluate the performance using ACC and AUC.

In turn, Rao et al. (2020) propose a direct cost-sensitive RF model called Syncretic Cost-sensitive Random Forest (SCSRF). They consider several aspects in their cost function, such as the sampling distribution of the risk categories by feature and the sampling distribution of the features in the different random samples of the algorithm, and assume a factor for the class imbalance problem. However, none of these aspects is purely financial. In addition, within the random forest, they use a weighted voting system associated with the degree of accuracy of each decision tree. The evaluation of the proposal focuses on traditional performance measures like AUC and ACC, among others, and reveals that the SCSRF obtains better results than other RF cost-sensitive approaches and classical LR or DT methods.

Chen et al. (2021) propose a direct cost-sensitive logistic regression, but without incorporating financial values. They evaluate their cost-sensitive alternative against class imbalance approaches such as oversampling, undersampling, and the SMOTE method in models such as RF and Artificial Neural Networks (ANN). However, they do not find marked or definitive differences. They use metrics such as ACC, Recall, F1, and G-mean for this.

Finally, the recent work by Wu et al. (2022) proposes a systematic loan evaluation framework called COst Sensitive Loan Evaluation (COSLE). It includes a direct method on DT and RF with different options of misclassification cost matrix, one of them with financial information by obligation. It compares its results with other cost-sensitive options found in the literature and cost-insensitive options, showing superior performance in alternatives that optimize financial information. They use metrics such as AUC, Err (prediction error rate), and LER (Loan Evaluation-specific lender's Return), a proposed alternative measure based on misclassification costs, to evaluate the profitability of the model for lenders.

As a result, as Table 1 shows, we can conclude that although some proposals include the financial costs associated with the risk of default, those that include information on the profits are scarce. Similarly, most works propose indirect approaches, but few studies have employed direct ones. Surprisingly, not all the works evaluate the measures using financial measures. Finally, it is important to remark that the proposals occasionally come from traditional markets, probably due to the increasing interest in P2P lending, its need for accurate risk assessment, and data availability.

2.3. *Contextualizing our profit-sensitive proposal within the literature*

Our work covers a research gap by proposing a profit-sensitive method that incorporates the profit information directly and indirectly at different points of the modeling process and analyzing its performance from the business and the classification perspective.

For the direct approach, we consider the XGBoost algorithm as Xia et al. (2017) did. However, we incorporate financial information to guide the learning of the XGBoost, while they just use a cost function with no financial meaning. We incorporate the profits and financial costs in the misclassification matrix and use the resulting matrix in the learning function. Furthermore, we estimate the costs and profits using the actual financial costs and profits estimated from the available Lending Club data by obligation in the hits and misses of the misclassification matrix. Our profit function is based on the one proposed by Verbraken et al. (2014). As they did, we aim to optimize a profit function by considering the average loss-given default (LGD), the annualized return on investment (ROI) on the loans and other financial elements. We also propose the use of the same function to tune the model. In particular, we will use it as the function to be maximized in a hyperparameter search process. Very few works use a profit-sensitive indirect approach for hyperparameter optimization, e.g. Ye et al. (2018) use genetic algorithms in RF and, Xia et al. (2017) use Bayesian hyperparameter optimization in RF and XGBoost.

Finally, we also use the profit information to transform the default probability provided by our classifier into a decision indicator for granting new credits. For such purpose, we consider two profit-sensitive alternatives: use the profit information to optimize the decision threshold (Verbraken et al., 2014; Herasymovych et al., 2019), and a decision-making method inspired by the binary Bayes Minimum Risk classifier (BMR) proposed by Bahnsen et al. (2014), Bahnsen et al. (2015), which we will call Bayes Maximum Profit (BMP).

It is important to note that to include the financial information throughout our proposal, we have consistently defined several profit-sensitive functions, which are also used for evaluation purposes. We present them in the next section.

3. A proposal of profit-sensitive measures

This section presents a profit-sensitive confusing matrix, specifying the loss and profits in each cell. Then, we introduce the profit functions that can be derived from it, which will be used in the modeling process or for evaluation purposes.

Table 1
Cost and profit-sensitive models research.

Research	Base model	Function	Approach	Includes financial Values	Classification Metrics/Financial Metrics	Scope
Bahnsen et al. (2014)	LR	Cost-sensitive	Direct	Yes	None/Cost Savings	Credit
Bahnsen et al. (2015)	DT	Cost-sensitive	Direct	Yes	ACC, F1/Cost Savings	Credit
Malekipirbazari and Aksakalli (2015)	RF	Cost-sensitive	Indirect	No	ACC, AUC/None	P2P
Maldonado et al. (2017)	SVM	Cost-sensitive	Indirect	No	AUC/Total Information Acquisition Cost	Credit
Xia et al. (2017)	XGB	Cost-sensitive	Indirect/Direct	Yes/No	AUC/Annualized Return Rate (ARR)	P2P
Wang et al. (2018)	DT C4.5, NB,LR and SVM	Cost-sensitive	Indirect	No	ACC, MAE, Cost Measures/None	P2P
Ye et al. (2018)	RF	Profit-sensitive	Indirect	Yes	ACC, AUC/ Total profit function	P2P
Herasymovych et al. (2019)	Heterogeneous ensemble (LR, SVM, ANN, RF, BDT)	Profit-sensitive	Indirect	Yes	AUC/Total Profit	Credit
Wang et al. (2021)	DT C4.5, RF, LR, and SVM	Cost-sensitive	Indirect	Yes	ACC/Total Cost, Average Cost, and Cost-Saving Rate	P2P
Rao et al. (2020)	RF	Cost-sensitive	Direct	No	ACC, AUC, Precision, Recall/None	P2P
Chen et al. (2021)	LR	Cost-sensitive	Direct	No	ACC, Recall, F1 and G-mean/None	P2P
Wu et al. (2022)	DT, RF	Cost-sensitive	Direct	Yes	AUC, Err/LER	P2P

Table 2
Confusion matrix with miss-classification costs and profits.

	Actual positive	Actual negative
	$y_i = 1$	$y_i = 0$
Predicted positive $\hat{y}_i = 1$	$\pi_{TP_i} = \pi(1,1)$	$\pi_{FP_i} = \pi(1,0)$
Predicted negative $\hat{y}_i = 0$	$\pi_{FN_i} = \pi(0,1)$	$\pi_{TN_i} = \pi(0,0)$

3.1. A profit-sensitive confusion matrix

A profit-sensitive function includes the profit for each lending decision once the outcome is observed. Specifically, we include the profits and the losses derived from the main risk factors associated with the grant process. The confusion matrix entries will be denoted by $\pi(k,l)$ as a representation of the profits; regardless, they can be positive or negative. Table 2 shows the resulting confusion matrix.

In particular, $\pi_{TP_i} = \pi(1,1)$ represents the financial outcome that corresponds with a true positive (TP), that is, correctly predicting a default, which implies avoiding a financial loss. On its part, $\pi_{TN_i} = \pi(0,0)$ corresponds with the profit of obtaining a true negative (TN), that is, correctly predicting a good loan which represents a successful investment decision. On the other side, $\pi_{FP_i} = \pi(1,0)$ represents the opportunity cost of miss-classifying a good loan as a Default (false positive, FP), while $\pi_{FN_i} = \pi(0,1)$ is the loss of classifying a bad loan as Non-default (false negative, FN).

Different approximations for estimating the financial values in the confusion matrix might exist. They mostly depend on the availability of information associated with the business or market analyzed in the data at hand. The more information you have about the financial implications of each obligation (that is, its actual costs, losses, and profits), the more accurate the estimation of the values of each cell of the matrix and, consequently, the more business sense will have the granting model and the more helpful it will be to take business decisions.

Each platform or entity should be able to assess the values of the matrix and use profit-sensitive alternatives such as the ones we propose. Below, we will review how these values can be approached or have been approached in the literature and how we will represent them. A more detailed specification of these values in our paper will be given in the next section once we have introduced our data set and the information it includes.

For the case of a correctly predicted default, π_{TP_i}, different works have assumed that the profit is zero (Verbraken et al., 2014; Bahnsen et al., 2015; Herasymovych et al., 2019); i.e., there is no profit from the correct identification of a default obligation. Alternatively, Ye et al. (2018) considers that, in that case, the classifier avoids the loss of principal for the lender; thus, it treats this element as a potential return.

Regarding π_{FP_i}, it represents the financial impact of not granting a loan the borrower would have returned. Some works consider that the impact is zero, no loss, and no profit (Verbraken et al., 2014; Herasymovych et al., 2019). On the other hand, Bahnsen et al. (2015) considers that the loss of not granting a loan is reduced by the average profit that can be earned by investing the money in other borrowers. This approach assumes that the money is not idle, and requires some investment assumptions to properly value this component.

For π_{FN_i}, we can estimate the loss the lender can achieve by considering the expected loss from the obligation falling into default. For doing that, we can include components that estimate the risk of potential losses based on Basel guidelines, with critical risk parameters like the probability of default (PD), LGD (Verbraken et al., 2014; Bahnsen et al., 2015) and exposure at default (EAD), which are used to obtain Expected Loss as the product $PD \cdot LGD \cdot EAD$. The P2P lending market inherently has high PD and LGD relative to the traditional lending context due to information asymmetry, non-professional lenders, and uncollateralized lending (Xia et al., 2017). If we had the information available, we could even incorporate International Financial Reporting Standards (IFRS) in the profit and loss assessment, improving the estimate for this matrix cell.

Finally, for π_{TN_i}, we can associate this value with the expected annualized ROI of having placed an obligation correctly (Verbraken et al., 2014). Ye et al. (2018) determines this value in a coarse way as the product of the principal and the rate of interest.

3.2. Profit-sensitive measures

The profit-sensitive measures used in the modeling process, decision process, and evaluation of the model must be consistent with the values defined in the confusion matrix. However, at some point, they could differ as the objective at each point or the information required demands other interests.

For example, if our classification method provides the default probability estimated by the models F_i, we could use a profit function incorporating the probabilistic information. While in other parts of the process, we need to use profit measures that reflect the impact of the classification decision ($\hat{y}$). Below we make this distinction for presenting a set of profit-sensitive measures based on the confusion matrix in Table 2.

3.2.1. Expected profit-sensitive measures

It is possible to propose probabilistic profit measures that measure the expected profits for the classification methods that yield a probability, such as the XGBoost. In particular, we can associate F_i, representing $F_i(x_i)$, with the adjusted distribution function associated with default event per loan, obtained as a function of the feature vector x_i of the ith loan or credit obligation and being y_i its outcome (default or not). In particular, F_i could follow a sigmoid function as in logistic regression. In general, we can define the expected profit (EP) function per loan application as in Eq. (1):

$$\pi_i^E(F_i) = y_i \left(F_i \pi_{TP_i} + (1 - F_i)\, \pi_{FN_i} \right) + (1 - y_i) \left(F_i \pi_{FP_i} + (1 - F_i)\, \pi_{TN_i} \right) \tag{1}$$

Using the previous definition, we can estimate the Expected Total Profit (ETP), similar to that proposed by Verbraken et al. (2014), as in Eq. (2):

$$\pi_T^E(F) = \sum_{i=1}^{N} \pi_i^E(F_i) = \sum_{i=1}^{N} \Big(y_i \Big(F_i \pi_{TP_i} + \ (1 - F_i)\, \pi_{FN_i} \Big) + (1 - y_i) \left(F_i \pi_{FP_i} + (1 - F_i)\, \pi_{TN_i} \right) \Big) \tag{2}$$

It is more convenient to consider the expected average profit instead of the expected total profit as it has a clearer business interpretation. Thus, we propose in Eq. (3) the average expected profit rate per obligation:

$$R\pi_T^E = \frac{\pi_T^E(F)}{N} \tag{3}$$

3.2.2. Profit-sensitive measures based on the predictions

For determining the predicted outcome $\hat{y}_i$, we will use some decision functions based on thresholding and Bayes Minimum Profit methods (BMP), as we will explain in the 4.3 section. Once we have predicted the outcome $\hat{y}_i$, we can define profit functions that measure the financial impact of the classification decisions. We can use such functions to assess the proposed methods' financial performance and guide their modeling process.

In particular, We can define the obligation profit as in Eq. (4).

$$\pi_i(\hat{y}_i) = y_i \left(\hat{y}_i \pi_{TP_i} + (1 - \hat{y}_i)\, \pi_{FN_i} \right) + (1 - y_i) \left(\hat{y}_i \pi_{FP_i} + (1 - \hat{y}_i)\, \pi_{TN_i} \right). \tag{4}$$

As for the case of the expected profit-sensitive measures, we can define a total profit function expressed in the currency used in the data set as in Eq. (5):

$$\pi_T = \sum_{i=1}^{N} \pi_i(\hat{y}_i). \tag{5}$$

Again, to provide a clearer business interpretation, we can define the average profit rate per obligation considering the prediction as in Eq. (6):

$$R\pi_T = \frac{\pi_T}{N} \tag{6}$$

4. A comprehensive profit-sensitive machine learning approach

In our approach, we incorporate the profit-sensitive information in three points: model tuning, model learning, and decision function. Fig. 1 represents the whole approach. The left side represents the process of tuning and learning the algorithm. We use a profit-sensitive machine learning algorithm that incorporates information on the expected profits to guide its inner learning. To fine-tune the algorithm (e.g. choose the best hyperparameters), we seek to maximize expected profit instead of a classification performance measure, that is, we look for different algorithm configurations and choose the best one in terms of profit. The best machine learning algorithm in the tuning and learning process is then used in the decision process (right side of Fig. 1). The algorithm produces a probability of default and this probability is used as input for a decision method that also takes into account the profits of the decisions. With this information, the decision method produces the final classification decision, which is a granting decision in our case.

It is important to remark that all the profit-sensitive measures used throughout the process should be based on the same profit-sensitive confusion matrix. Still, the measures used at each point could differ since each point of the modeling process may have different needs. In the following subsections, we will present how we will incorporate profit-sensitive information in three points of the modeling process.

4.1. Profit-sensitive model learning

This section explains how to introduce a profit function as a learning function of the eXtreme Gradient Boosting (XGBoost) method proposed by Chen and Guestrin (2016). To do that, we take advantage of the fact that XGBoost is generalizable to different loss functions. However, it demands the learning function to be continuously differentiable, as it happens with other boosting methods and some neural network models. Thus, the profit function defined must comply with this requirement.

We use the average expected profit function in Eq. (3). The learning or estimation method will determine the function values of F that maximize the profit, that is, $argmax_F R\pi_T^E(F)$. Since machine learning algorithms typically minimize a loss function, we will consider as loss function our profit function in negative, i.e., $-R\pi_T^E(F)$.

In the Python implementation of the XGBoost model used,[1] the inclusion of a different loss function just requires the definition of the gradient and the Hessian of each instance for the optimization process. This implementation uses a diagonal approximation of the Hessian matrix. It incorporates this approximation into a second-order Taylor expansion of the objective loss function, optimized using Newton's method, considering a diagonal approximation of the Hessian matrix makes the optimization process more efficient, reduces the computational load, and gains convergence speed without neglecting the precision in searching for optimal values.

For our loss function $-R\pi_T^E(F)$, the gradient and the Hessian for each instance are defined as $g(x_i)$ and $h(x_i)$, with $g(x_i)$ as in Eq. (7) and $h(x_i)$ as in Eq. (8), respectively.

$$g(x_i) = F_i(x_i) \left(F_i(x_i) - 1 \right) \left(y_i \left(\pi_{TP_i} - \pi_{FN_i} \right) + (1 - y_i) \left(\pi_{FP_i} - \pi_{TN_i} \right) \right) \tag{7}$$

$$h(x_i) = F_i(x_i) \left(F_i(x_i) - 1 \right) \left(1 - 2F_i(x_i) \right) \Big(y_i \left(\pi_{TP_i} - \pi_{FN_i} \right) + \ (1 - y_i) \left(\pi_{FP_i} - \pi_{TN_i} \right) \Big) \tag{8}$$

4.2. Profit-sensitive model tuning

We can use a profit function as the guiding criteria for the model-tuning process, which can include hyperparameter configuration, feature selection, or feature engineering. In our case, we will use it for hyperparameter configuration using the Bayesian optimization method with the support of the open-source package of Python *hyperopt*, which is based on Bergstra et al. (2013).

For model tuning, we can use any profit function. However, we will use a probabilistic one because our subsequent decision process will use the probabilities derived from the model adjusted with the best parameter set found.

In particular, in our experiment, we will use the annualized average expected profit rate per obligation $R\pi_T^E$ as in Eq. (3), as a criterion to select the best set of hyperparameters, regardless of whether or not

---

[1] https://xgboost.readthedocs.io/en/latest/python/python_intro.html

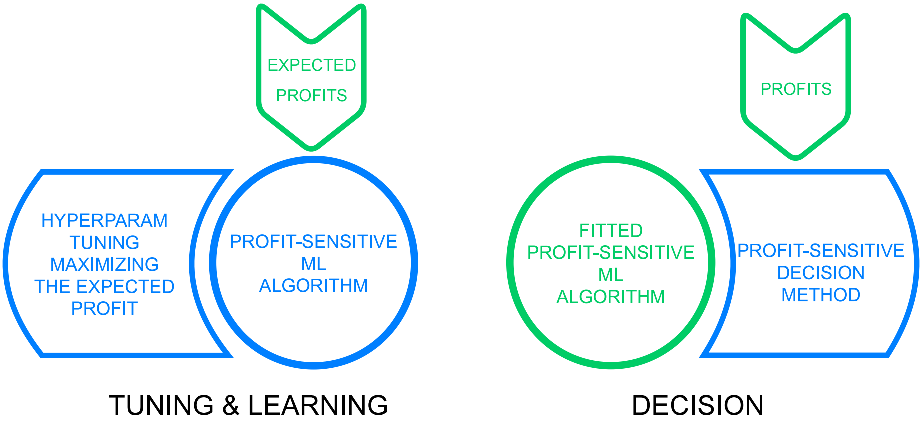


Fig. 1. Profit-sensitive modeling approach. The left side describes the tuning and learning of the machine learning algorithm and the right side the decision process once the machine learning algorithm has been trained.

the learning process incorporates the standard learning function or a profit-sensitive one.

We chose the annualized average expected profit rate because it is the same measure used in the direct estimation. It also aligns with the measure used in the decision function, which measures observed rather than expected returns.

In short, the chosen objective function allows us to estimate the probabilities of default that guarantee a higher expected net profit per financial liability which is consistent with our aim in the decision process of obtaining a higher average net profit.

4.3. *Profit-sensitive decision function*

Some machine learning and statistical methods yield a probability as an output, and such probability needs to be transformed into a classification decision $\hat{y}_i$. This process can be done in different ways with the help of a profit-sensitive decision method. Specifically, in this work, we will consider two of them: thresholding and Bayes Maximum Profit (BMP).

Regarding the thresholding process, many works use the default threshold (0.5) on F_i and pay no attention to setting an adequate threshold value to improve the performance of the models. However, approaches such as those from Verbraken et al. (2014) and Bahnsen et al. (2014) define a decision rule based on an acceptance threshold or cut-off t on F_i that minimizes a loss function (L), i.e., $\text{argmin}_t\, L$, as we did. Thus, the decision rule for defining the prediction $\hat{y}_i$ is set by a sign function such as $sign(F_i - t)$.

As our work aims to maximize the profits and not the classification performance, we will use $\text{argmax}_t\, \pi(t)$, being $\pi(t)$ a function that recognizes the profits derived from a correct or incorrect classification for each of the cells found in the matrix in Table 2. We will determine the threshold that maximizes the profit in the training data and evaluate it in the validation sets.

We could consider any profit function based on the default categories' final prediction $\hat{y}$, and in our case, we use the function $R\pi_T$ as in Eq. (6) that represents the net profit per financial liability and provides a clear financial interpretation.

Regarding BMR, we obtain $\hat{y}_i$ as proposed by Bahnsen et al. (2014) and Bahnsen et al. (2015). More precisely, as a function of risk R, as in Eq. (9).

$$\hat{y}_i = 1 \text{ if } R(\hat{y}_i = 1 \mid x_i) < R(\hat{y}_i = 0 \mid x_i). \tag{9}$$

In particular, given a profit-sense function $R\pi$, we define its process as in Eq. (10). Since the model's decision is obtained by maximizing profit and not minimizing risk, we will name it as Bayes Maximum Profit (BMP).

$$\hat{y}_i = 1 \text{ if } R\pi(\hat{y}_i = 1 \mid x_i) > R\pi(\hat{y}_i = 0 \mid x_i), \tag{10}$$

where

$$R\pi(\hat{y}_i = 1 \mid x_i) = \pi_{TP_i} F_i + \pi_{FP_i}(1 - F_i), \text{ and} \tag{11}$$

$$R\pi(\hat{y}_i = 0 \mid x_i) = \pi_{TN_i}(1 - F_i) + \pi_{FN_i} F_i. \tag{12}$$

To realistically apply this strategy, given that we will not have all the matrix values from the application, we have estimated the misclassification values from the average data of the defaulted and non-defaulted obligations. For example, for π_{FN_i}, we used the average LGD of the defaulted records and the mean term to annualize the profit. However, it could be possible to estimate these matrix values in other ways, resulting in different applications of this prediction method.

It is important to remark that the indirect approaches can also be applied to other methods, such as logistic regression, regardless of whether or not the learning or estimation process incorporates the profit function. This is precisely what we will do in our experiment.

5. Experimental setting

5.1. *Dataset and model variables*

We will use the public data set from Lending Club,[2] which includes loans issued by the company between 2007 and 2018. The Lending Club (LC) data set is widely used in credit risk publications and the most used when addressing the P2P market (Ariza-Garzón et al., 2021).

It is important to note that the Lending Club dataset is of loans granted. This affects the experiment in several ways. One is that the dataset is biased towards a priori less risky loans since many of the a priori riskier loans were likely not granted. This makes it more difficult for a granting model that only uses the information available at the time of application, such as the one we propose, to determine which loans defaulted. Another is that financial performance information is available for all loans in the dataset, which allows us to use realistic data to propose a profit-based model, such as the one we propose in this paper.

Since we aim to develop a credit granting model, we create a target variable based on the credit's final resolution. For this purpose, we use the loan status variable, representing the loan state at the end of the data set time window. Our target variable has two values

---

[2] https://www.kaggle.com/wordsforthewise/lending-club

Table 3
Description of the input variables.

Variable	*Description*
Categorical variables	
emp_length	Employment length. LC represents the current employment time in years into 12 categories, including the no information category.
home_ownership_n	The homeownership status provided by the borrower during the registration process. Categories defined by the entity: mortgage, rent, own, other (which includes other, none, and any).
Quantitative variables	
revenue	Yearly income of the applicant (and co-borrowers, if any) self-reported in the registration process.
dti_n	Debt ratio per obligation. Debts are provided by the credit bureau (mortgage debts are excluded). The indicator is monthly. The income is the self-reported monthly income of the applicant (and co-borrowers, if any).
loan_amnt	Amount of credit requested by the applicant.
fico_n	Credit bureau score of the applicant. Defined between 300 and 850, reported by Fair Isaac Corporation as a summary risk measure based on historical credit information reported at the time of application.

representing the loan's final state: the default category corresponds to a loan eventually charged off, and the non-default one represents fully paid loans. Our data set excludes loans with other values in the loan status variable since we cannot determine their final state.

As previously mentioned, many credit companies and banks consider small business credit as a specific product that deserves particular treatment, such as collecting specific information to assess the risk, granting, and eventual conditions. In the LC data set, the loans of the *small business* category represent around 1% of the loans (15,575 loans) and present the highest default rate, 29.86%, being the default rate of 19,86% in the whole data set. Other works have already drawn attention to the credit risk of this segment (Serrano-Cinca et al., 2015; Havrylchyk and Verdier, 2018; Polena and Regner, 2018; Ariza-Garzon et al., 2020) and some of them even have focused their analysis and modeling efforts on it (Nowak et al., 2018).

As we explained, we also consider that this type of credit needs more careful risk management and will focus our analysis on it. We recognize that its segment would need to be better characterized financially, including the type of investment made and its projected indicators. Nevertheless, we do not have such variables.

Table 3 shows our models' categorical and quantitative variables. These variables are taken directly from the credit applications. Below, we explain the decisions and transformations we took for some variables.

In the *home_ownership_n* variable (see Table 3), we merged the categories 'other', 'none' and 'any' in the category 'other' since we did not differentiate these options with clarity and they exhibit similar behavior for the percentage of default. The other options of this categorical variable remain the same. We consider the employment length (*emp_length* in Table 3) as another categorical variable since it includes the category of no information and the category of more than ten years in the current employment.

Regarding the Fair Isaac Corporation credit bureau (FICO) variable (*fico_n* in Table 3), in the original data set, the information from FICO is given by two values that define the credit scoring interval of the applicant to the bureau, that is, a minimum and maximum range of limits to which the borrower's FICO belongs at loan origination. However, our FICO variable (*fico_n*) is estimated as the average of these two values to have a single indicator of the creditworthiness of potential borrowers.

For the case of the debt variable (*dti_n*), it is estimated from the original dataset variables as the ratio calculated from the total debts of the co-borrowers over the total debt obligation divided by the combined monthly income of the co-borrowers.

5.2. Estimation of the profit functions

Table 4 lists the financial variables used to calculate the values of our profit-based confusion matrix. The information available in the data set conditions how such values are estimated. It is worth clarifying that we have not included operative costs related to the loan management and granting process, use of the platforms, evaluation of models, or study of loan applications. It is also important to remark that our matrix considers the financial impact on the lender and not on the platform. The values we propose for the cells of the confusion matrix are explained below.

The value π_{FP_i} will be zero in our case. We assume that if the loan is not granted, it has no financial impact, neither positive nor negative, on the lender's side. We also assume that π_{TP_i} is zero, which again implies that the outcome has no financial impact on the lender. In this case, we are simply avoiding a financial loss for an individual lender.

For π_{TN_i}, we define it as annualized "recovered" ROI (ROI_rec_i) of having classified obligation i correctly as in Eq. (13)

$$\pi_{TN_i} = (ROI\_rec_i + 1)^{\frac{12}{term_i}} - 1, \tag{13}$$

where ROI_rec_i is as in Eq. (14)

$$ROI\_rec_i = \frac{total\_pymnt_i}{funded\_amnt_i} - 1, \tag{14}$$

where $total\_pymnt_i$ includes the principal recovered, interest, and late fees received to date.

For π_{FN_i} case, we use an estimation of annualized profit achieved by a lender, taking into account the estimation of expected loss arising from the obligation falling into default as in Eq. (15):

$$\pi_{FN_i} = \left(\frac{(1 - LGD_i)(funded\_amnt_i - total\_rec\_prncp) + total\_pymnt_i}{funded\_amnt_i} \right)^{\frac{12}{term_i}} - 1. \tag{15}$$

Based on Basel guidelines, the π_{FN_i} value includes components that estimate potential loss risk. In particular, critical risk parameters, such as the probability of default (PD), LGD, and exposure at default (EAD), are used to obtain the Expected Loss ($PD \times LGD \times EAD$) in credit risk management. In our case, the value of EAD corresponds mainly to the outstanding loan amount of the obligation at the time of our evaluation.

For the case of LGD, according to Zhou et al. (2018), there are mainly three approaches for its estimation: workout LGD, market LGD, and Implied Market LGD. For our data, we follow the ideas of Papouskova and Hajek (2019) and use a workout LGD estimation as in Eq. (16):

$$LGD_i = 1 - \frac{recoveries_i - collection\_recovery\_fee_i}{outstanding\_loan\_amount_i} \tag{16}$$

where $outstanding\_loan\_amount$ is estimated as $(funded\_amnt - total\_rec\_prncp)$.

Table 4
Description of variables.

Variable	*Description*
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
funded_amnt	The total amount committed to that loan.
total_payment	Payments received to date for total amount funded.
total_rec_prncp	Principal received to date.
recoveries	Post charge off gross recovery.
collection_recovery_fee	Post charge off collection fee.

5.3. Models considered

In this section, we describe the different approaches we will use for modeling the problem of granting loans, including profit-sensitive elements. As explained in Sections 4.1–4.3, we can introduce such elements at three points of the modeling stage, namely:

1. The profit function used in the learning of the XGBoost algorithm, $R\pi_T^E$ as in Eq. (3) from which we obtain the gradient and hessian.
2. The profit function to guide the model tuning, in our case, using Bayesian optimization, $R\pi_T^E$ as in Eq. (3).
3. The function used for deciding on the credit granting. In this case, we will consider two alternatives:
    - BMP, where we compare the profits associated with predicting an obligation as Default or as not-Default, as in Eq. (10).
    - Thresholding, where we find the threshold that maximizes the function $R\pi_T$ as in Eq. (6).

We advocate including the profit function at all three points in the modeling process. The opposite approach is not to consider it in any of them and use instead the following elements:

1. The standard learning function used by the method considered, parametrized to deal with the unbalanced nature of our data set.
2. The balanced accuracy measure (BACC) guides the hyperparameter optimization. This measure averages the accuracy obtained from both the minority and majority classes.
3. The use of the standard threshold or cut-off $t = 0.5$.

However, it is possible to consider including the profit functions only at some of the points of the modeling stage. Table 5 shows all the modeling approaches that we will compare in our experiment. The table also shows which function is used at each stage (learning, optimization, and decision). We will explain the abbreviations used to refer to them. Each abbreviation consists of three parts separated by '-'. The first is the classification method (XGBoost named as XGB for all the cases), the second is the optimization approach (Bayesian optimization or BO for all the cases), and the third is the decision method. For the decision stage, we will represent the thresholding method by t and the binary Bayes Maximum Profit classifier by BMP, while the standard threshold is denoted by 0.5. The inclusion of π in the abbreviation represents that the profit function is considered at that stage, while its omission means that no profit function is used. For example, the approaches that include all the profit-sense approaches proposed in this paper can be found in rows 8 and 9, identified by the abbreviations πXGB-πHO-t and πXGB-πHO-BMP, respectively. On the other hand, the profit-insensitive approach in row 1 is identified by XGB-HO-0.5.

5.4. Model tuning

We used a stratified evaluation k-fold cross-validation with k = 4 on the training set to avoid over-fitting in the machine learning models. We use the k-fold to search for the best hyperparameter combination using Bayesian Optimization, provided by the Python open-source package *hyperopt*.[3] We define the initial space search for the optimization assuming that the hyperparameters are uniformly distributed.

The hyperparameters considered for the XGBoost are the number of trees (*n_estimators*), the maximum tree depth (*max_depth*), the subsample ratio of the training instances in every boosting iteration (*subsample*), the proportion of features when constructing each tree (*colsample_bytree*), the proportion of features considered by level (*colsample_bylevel*), the learning rate (*eta*), and the regularization hyperparameters (*alpha*, *lambda* and *gamma*). Furthermore, for the profit-insensitive and the indirect profit-sensitive algorithms, we also optimized the *scale_pos_weight* parameter to deal with the class-imbalance problem.

For the XGBoost we will use the Python implementation of XGBoost[4] and the Python library Scikit-learn[5] (Pedregosa et al., 2011).

5.5. Performance metrics

Regarding performance, the different approaches will be evaluated both in financial and classification terms. Among the financial measures, we will use the profit functions included in the learning as $R\pi_T^E$ defined in Eq. (3), which is a probabilistic measure that represents the expected return, and a criterion that measures the financial impact of the decision taken as $R\pi_T$ defined in Eq. (6). We will also provide the loan acceptance rate (LAR) that measures each method's percentage of loans granted.

As classification metrics, we will show measures such as the aforementioned AUC and Kolmogorov–Smirnov (KS) and metrics on the observed classification such as the aforementioned Balanced Accuracy ($BACC$).

We will also use inferential methods to evaluate whether the profit-sensitive methods generate better financial performance than the traditional alternatives. We want to assess whether profit models are selected in the set of models with better statistical performance per profit. For doing so, we will use the model confidence set (MCS) methodology proposed by Hansen et al. (2011) and used in other credit risk works (Lyócsa et al., 2022). The MCS methodology selects the best models based on a given significance level (α). The higher α, the lower the confidence level, and the more models tend to be selected in the top set of models. We will use inferential criteria through a non-parametric test for multiple comparisons and paired samples to compare alternatives. The algorithm is a model selection process that consists of performing a hypothesis testing process that starts with all models and eliminates options as long as the null hypothesis is rejected. The process stops when the null hypothesis is not rejected or when only one model option remains. We will use the R library MCS[6] developed by Bernardi and Catania (2014).

In our case, we define the loss function as the additive inverse of the profit per model and per obligation i ($l_{i,m} = -R_{\pi_{i,m}}$), similar to

---

[3] http://hyperopt.github.io/hyperopt/

[4] https://xgboost.ai/

[5] https://scikit-learn.org/stable/

[6] http://cran.r-project.org/web/packages/MCS/index.html

Table 5
Modeling approaches used in the paper according to the method considered, the learning optimization and decision functions used, and the rule used for decision making.

Row	Abbreviation	Learning function	Optimization function	Decision function	Decision rule
1	XGB-HO-0.5	Standard	BACC	None	$t = 0.5$
2	XGB-HO-t	Standard	BACC	$R\pi_T$	t
3	XGB-HO-BMP	Standard	BACC	$R\pi_T$	BMP
4	XGB-πHO-0.5	Standard	$R\pi_T^E$	None	$t = 0.5$
5	XGB-πHO-t	Standard	$R\pi_T^E$	$R\pi_T$	t
6	XGB-πHO-BMP	Standard	$R\pi_T^E$	$R\pi_T$	BMP
7	πXGB-πHO-0.5	$R\pi_T^E$	$R\pi_T^E$	None	$t = 0.5$
8	πXGB-πHO-t	$R\pi_T^E$	$R\pi_T^E$	$R\pi_T$	t
9	πXGB-πHO-BMP	$R\pi_T^E$	$R\pi_T^E$	$R\pi_T$	BMP

what Lyócsa et al. (2022) did. Thus, a higher profit translates into a lower loss. The difference between losses of the m and n models is given by $d_{i,mn} = l_{i,m} - l_{i,n}$, with m, n models under evaluation. Under the null hypothesis $E(d_{i,mn}) = 0$ for all m, n models under evaluation.

We will use T_{max}, see Hansen et al. (2011) for an explanation, as a test statistic. For the inferential process and selection of the best set of models, the method uses a bootstrapping procedure, which, in our case, will consider 3000 bootstrap samples.

5.6. Model explanation method

In this work, we will use the SHAP values (Lundberg and Lee, 2017a,b) to provide explainability to the proposed models. The SHAP values are mainly based on the definition of Shapley values developed in game theory, specifically in cooperative games (Shapley, 1953). In general, the SHAP values aim to explain a prediction through the marginal contribution of each feature. The feature values of an instance of the dataset behave like actors in a coalition. The SHAP values explain the prediction function as a sum of each input feature's effects allowing for an equitable distribution of the payoff. This methodology also uses other additive feature attribution methods, such as Local Interpretable Model-Agnostic Explanations (LIME) by Ribeiro et al. (2016) and Learning Important FeaTures (DeepLIFT) by Shrikumar et al. (2017), among others. We will rely on the open-source Python *shap* package[7] for this. However, for the SHAP values of categorical values, we will use the adjustment proposed by Ariza-Garzon et al. (2020) that considers the dependence relationships between the categories of the variables.

The SHAP values allow us to estimate the contribution of the variables in a model's prediction, to know the variables' relative importance, to identify the monotonicity and dependence relationships, and even to identify nonlinear relationships derived exclusively from machine learning proposals. In our work, we will use the SHAP values to identify the default drivers. In the case of the direct profit-sensitive approaches, we will analyze the drivers of the default weighted by the costs and profits derived from the misclassification errors.

The SHAP values will be used to observe the changes in the feature importance and the dependency relationships that can be observed when including profit information in the XGBoost approach and to identify the key feature values of the resulting model.

6. Results

6.1. Preliminary descriptive and inferential analysis

In this section, we analyze the data set: the independent variables used, those that characterize the financial aspects of the obligations and their impact on the profit function, and those used to predict the outcome. These results will help us better contextualize the discussion of the results in the following sections.

Tables 6 and 7 present a descriptive summary of the data set we are using, including some inferential tests to identify the potential of the chosen variables to predict default. In particular, we use the Kolmogorov–Smirnov test for quantitative variables to compare the empirical cumulative probability distributions between the Default and Non-default records. On the other hand, for categorical variables, we use the chi-square test to evaluate the dependence of each of these variables on the target variable (Default and Non-default).

Table 6 shows that the behavior of quantitative variables differs significantly for default and non-default obligations. Defaulted obligations are characterized by lower *revenue*, higher indebtedness (*dti_n*), lower FICO (*fico_n*) and higher requested amount (*loan_amnt*). These results are consistent with what is expected for credit risk.

Table 7 shows the result for qualitative variables where we find a significant dependence between the target variable and the variables that describe the homeownership status (*home_ownership_n*) and the employment length (*emp_length*). In the *home_ownership_n* variable, the *RENT* and *OWN* categories show the highest risk (33.1% and 30% default rates, respectively). In *emp_length* variable, the categories that denote a shorter length or no information are among those with higher risk, all over 30%. In general, we see that employment length is not perfectly ordered with the default rate, finding, for example, significant default rates also in 7 years of employment length (32.5%), which supports the inclusion of this variable in the model through one-hot encoding.

Table 8 shows the financial characterization of the loans. The table shows that the small business segment has a high default rate, and the annualized profit rate of the defaulted loans exhibits a mean loss of 0.21065 (the loans have terms of 36 or 60 months). The total distribution is also slightly skewed to the left with a negative mean and positive median, suggesting the possibility of obtaining a small profit from selecting some of the obligations.

The mean annualized profit rate is −0.02505, with high volatility, and also the default is high. All these aspects reinforce the idea of the high risk of the small business segment. These results demand a risk prediction model from granting but show the complexity of achieving it, as we expected from the literature review of this financial market segment. Our profit-sensitive approaches will validate whether it is possible to increase the naive profit, find positive values, and, thus, reduce risk for lenders.

6.2. Performance evaluation

In this section, we report the results of the different approaches in the k-fold cross-validation setting. The tables report the average of the measures across the four validation folds together with their respective standard deviation (denoted as sd).

---

[7] https://github.com/shap/shap

Table 6
Descriptive Analysis. Quantitative variables.

Variable	Statistic	Non-Default	Default	Total
revenue	Mean	$ 93,443.10	$ 90,129.74	$ 92,453.88
	Median	$ 75,000.00	$ 74,000.00	$ 75,000.00
	SD	$ 78,535.02	$ 97,266.65	$ 84,573.41
	KS D-Test			0.04***
dti_n	Mean	14.10	15.71	14.58
	Median	13.06	14.88	13.60
	SD	8.74	9.57	9.03
	KS D-Test			0.07***
fico_n	Mean	708.54	698.69	705.60
	Median	702.00	692.00	697.00
	SD	38.61	32.18	37.08
	KS D-Test			0.10***
loan_amnt	Mean	$ 15,032.92	$ 17,039.00	$ 15,631.85
	Median	$ 12,150.00	$ 15,000.00	$ 13,750.00
	SD	$ 9,702.54	$ 9,722.70	$ 9,751.56
	KS D-Test			0.10***

*** Significant at the 0.01 level.

Table 7
Descriptive Analysis. Categorical variables.

Variable	Category	Count	Rel. Frec.	Default Rate	CHI Test
home_ownership_n	MORTGAGE	6,923	44.4%	26.5%	71.648***
	OTHER	20	0.1%	25.0%	
	OWN	1,674	10.7%	30.0%	
	RENT	6,958	44.7%	33.1%	
emp_length	< 1 year	1,354	8.7%	31.4%	25.338***
	1 year	1,203	7.7%	33.2%	
	2 years	1,707	11.0%	29.7%	
	3 years	1,545	9.9%	30.4%	
	4 years	1,179	7.6%	32.2%	
	5 years	1,140	7.3%	28.9%	
	6 years	779	5.0%	28.0%	
	7 years	750	4.8%	32.5%	
	8 years	713	4.6%	29.0%	
	9 years	513	3.3%	27.9%	
	10+ years	4,254	27.3%	27.9%	
	NI	438	2.8%	32.0%	

*** Significant at the 0.01 level.

Table 8
Financial description of loans.

Variable	Statistic	Non-Default	Default	Total
Loans	Count	10 925	4650	15 575
	Rate	70.144%	29.856%	100.000%
Annualized profit rate	Mean	0.05395	−0.21065	−0.02505
	Median	0.05411	−0.19699	0.03559
	SD	0.02935	0.15859	0.15091

Table 9 shows the performance of the estimated probability function for the XGBoost alternatives. At this point, we only have three alternatives: the profit-insensitive one, the one that uses indirect learning of the profits (in the hyperparameter optimization), and the one that uses both direct and indirect learning of the profits (in the learning function of the XGBoost and the hyperparameter optimization). The only alternative that obtains a positive expected profit rate is the πXGB-πHO, which uses the profit directly and indirectly, while the worst is the profit-insensitive alternative (XGB-HO). Furthermore, all the proposed methods perform much better than the naive method that grants all the loans. However, if we look at the naive method that considers the observed default rate (29.86%), it outperforms the profit-insensitive alternative. This result shows the potential to improve profitability by using a reasonable estimate of a probability associated with default, as it can balance both risk and profitability in a unified model. The better this estimate is, the greater the chances of obtaining good expected profitability results, which enhances potential investments in small business loan purposes and helps identify better loans.

Regarding the classification measures (AUC and KS), the worst method is, again, the profit-insensitive alternative. It suggests that trying to guide the learning using profit information can be useful for more accurately estimating the probabilities of the classification function.

Table 10 shows the performance of the models once the classification decision is made. We have three alternatives for each of the methods considered in the previous step: profit-insensitive (using the default threshold at 0.5) and profit-sensitive (thresholding and BMP). The business measures include the average profits of the prediction options $R\pi_T$ the loan acceptance rate (LAR), and the default rate (DR), while as classification measures, we show the balanced accuracy $BACC$.

First, we can see that all the methods outperform the mean annualized profit rate of the dataset, which was −0.02505. Thus, learning with either profit-sensitive or -insensitive approaches is financially beneficial. A closer look reveals that the approaches with profit-sensitive learning generally perform best. In particular, using the MCS procedure, we identified the methods that yield the highest annualized average profits and we can see that all the methods selected use profit-sensitive learning; the best ones include the profit directly in the learning, that is, the πXGB-πHO with the three decision-making strategies. Next, two of the indirect-learning options, XGB-πHO-0.5 and XGB-πHO-BMP, are also included in the model confidence set (MCS). In

Table 9

Expected profit and score-based classification measures in validation.

Model	$R\pi_T^E$	sd	AUC	sd	KS	sd
Naive by granting all loans	−0.02505	0.00122	–	–	–	–
Naive default rate	−0.00748	0.00037	–	–	–	–
XGB-HO	−0.01196	0.00091	0.5595	0.0070	0.0946	0.0047
XGB-πHO	−0.00361	0.00035	**0.6151**	0.0069	**0.1785**	0.0119
πXGB-πHO	**0.00059**	0.00019	0.5848	0.0173	0.1312	0.0265

Bold numbers represent the best values for each metric.

Table 10

Profit and classification measures for the decisions in validation.

Model	$R\pi_T$	sd	LAR	sd	DR	sd	$BACC$	sd
XGB-HO-0.5	−0.01195	0.00116	0.60411	0.0086	0.27271	0.00377	0.53727	0.00554
XGB-HO-t	−0.01754	0.00177	**0.79904**	0.01515	0.28423	0.00293	0.52737	0.00598
XGB-HO-BMP	−0.00716	0.00040	0.43634	0.00767	0.25634	0.00150	0.54399	0.00222
XGB-πHO-0.5	0.00041[a]	0.00038	0.06388	0.00350	0.13505	0.01960	0.52484	0.00217
XGB-πHO-t	−0.00081	0.00085	0.24597	0.02013	0.19118	0.01154	**0.56282**	0.00600
XGB-πHO-BMP	0.00021[a]	0.00009	0.01393	0.00303	**0.10929**	0.05830	0.50625	0.00222
πXGB-πHO-0.5	**0.00077**[a]	0.00026	0.08706	0.01652	0.16941	0.00916	0.52665	0.00421
πXGB-πHO-t	0.00051[a]	0.00040	0.10684	0.02056	0.17699	0.01871	0.53048	0.00443
πXGB-πHO-BMP	0.00069[a]	0.00032	0.05849	0.00889	0.16097	0.01686	0.51916	0.00352

[a] Indicates model that belongs to the set of top models for a given α level (0.10), with MCS procedure. Bold numbers represent the best values for each metric.

either of the best methods, the LAR rates are low (ranging from 24.6% to 1.4%). Thus, these methods have a selective behavior that results in higher returns. If we look at the balanced accuracy, we can see that the three methods with the highest value do not belong to the selection by the MCS procedure, which denotes no clear correlation between classification performance (using a balanced measure) and profitability, as already reported in other works (Verbraken et al., 2014; Xia et al., 2017; Wang et al., 2021).

Another remarkable conclusion is that the methods that include the profit in the learning function and the hyperparameter optimization are better than those that only include it in the hyperparameter optimization for each of the three decision-making methods. Similarly, the methods that include the profit in the hyperparameter optimization are better than the profit-insensitive ones for each of the three decision-making methods. This suggests that including profit information for guiding the learning, both directly or indirectly, is good for maximizing profits.

However, if we look at the decision-making alternatives, the results are not conclusive. The profit-insensitive decision-making strategy (i.e. considering the default threshold at 0.5) is the best strategy for the πXGB-πHO and XGB-πHO cases. In contrast, the BMP is the best decision-making strategy for the profit-insensitive case (XGB-HO), and the thresholding strategy is the worst for all the cases. Interestingly, while the profit-sensitive approaches produced mixed results in validation, they were better than the profit-insensitive counterpart during the training. Another remarkable fact is that the decision-making strategy greatly alters the profit and classification measures of a given classifier. These facts could be due to the complexity of the small business segment. In any case, our results do not support that profit-sensitive decision-making contributes in financial terms, at least under the three decision-making proposals studied. This finding points to an area that demands further study.

We can see that the XGB-HO methods have the highest acceptance rates (all of them over 40%) but at the same time the highest default rates (all of them over 25%, while the default rate in the whole dataset, as shown in Table 8, was close to 30%). The rest of the methods, which use the profit for learning directly or indirectly, obtain much lower default rates, but also much lower acceptance rates. It means that, in this case, including profit in the learning process produces classifiers that cut risk by reducing the default rate, which in turn is achieved by applying more restrictive filtering of the loans exploiting their financial information.

Table 11

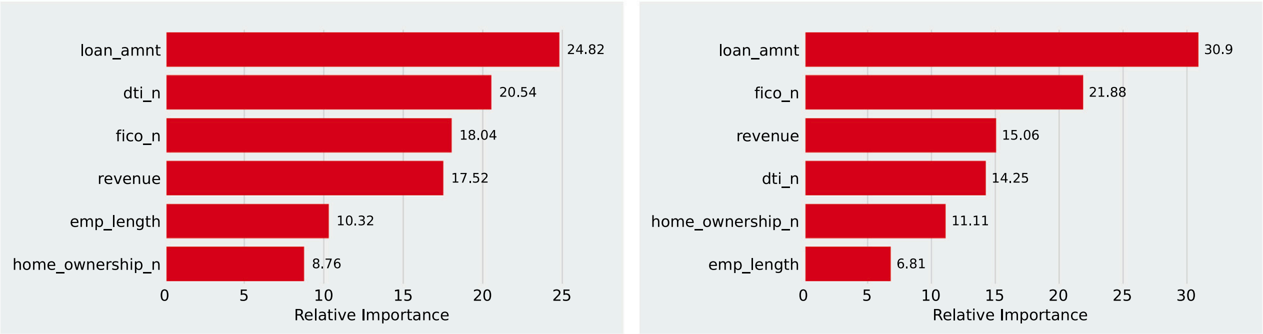
Profit values of the misclassification matrix in validation.

Model	π_{TN}	sd	π_{FN}	sd
XGB-HO-0.5	0.05179	0.00018	−0.21066	0.00511
XGB-HO-t	0.05291	0.00028	−0.21054	0.00637
XGB-HO-BMP	0.05065	0.00047	−0.21098	0.00372
XGB-πHO-0.5	0.03982	0.00087	−0.20536	0.01468
XGB-πHO-t	0.04578	0.00050	−0.21017	0.00626
XGB-πHO-BMP	0.03394	0.00308	−0.15038	0.03983
πXGB-πHO-0.5	0.05168	0.00157	−0.19842	0.01890
πXGB-πHO-t	0.05088	0.00203	−0.20516	0.01698
πXGB-πHO-BMP	0.05111	0.00174	−0.19281	0.03462

Finally, in Table 11, we look at the values of our profit-sensitive misclassification matrix that can take a value different from zero according to our assumptions to define the profit function in Section 5.2. Those values are the profit obtained by the true negatives (π_{TN}), i.e., the correctly identified non-default loans, and the loss obtained by the false negatives (π_{FN}), i.e. the accepted loans that will default. As a reference, we will use the mean annualized profit rates of the non-defaulted and defaulted loans in the whole dataset (see Table 8), which are 0.05395 and −0.21065, respectively. We can see that in the methods with profit-insensitive learning (i.e. XGB-HO), the values of the mean profit and the mean loss are very similar to those in the whole dataset. It means that although these methods can learn to classify loans correctly to some extent, that improvement does not translate into financial terms. However, the methods that use indirect profit-based learning (i.e. XGB-πHO) can obtain smaller losses on average at the expense of cutting the average profits (values below 0.05). Interestingly, the methods that present better financial behavior are those that use both direct and indirect profit-based learning (i.e. πXGB-πHO) because they can reduce the losses on average without greatly reducing the average profits (all of them over 0.05). Thus, we can see that introducing profit information in the learning process produces classifiers that simultaneously look at cutting risks and obtaining profits. We will deepen into the behavior of the three approaches in the next section.

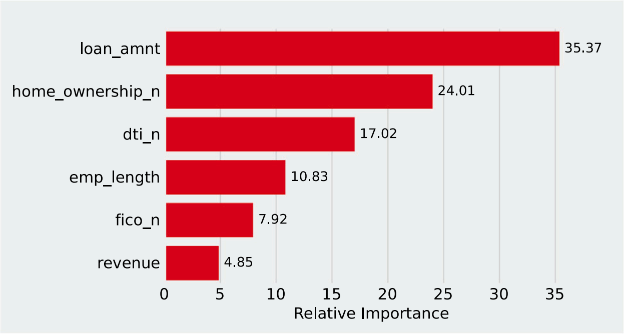
6.3. Model explainability

In this section, we will analyze the three models' explainability. Fig. 2 shows the relative feature importance.



(a) Aggregated relative importance XGB-HO

(b) Aggregated relative importance XGB-πHO



(c) Aggregated relative importance πXGB-πHO

Fig. 2. Aggregated relative importance by XGBoost model.

For the profit-insensitive model in Fig. 2(a), we see that quantitative variables are the most important, followed by the length of employment *emp_length* and *home_ownership_n*. It is important to highlight that the order of the variables for this model is completely different from that reported by Ariza-Garzon et al. (2020) for the same model and the whole Lending Club dataset. It suggests that the segment of small businesses demand their own credit risk models.

In the case of the indirect profit-sensitive model in Fig. 2(b), the most important variables are again the quantitative ones, but the *loan_amnt* is followed by the FICO (*fico_n*) variable, and together account for more of the 50% of the relative importance. At the same time, the *dti_n* variable loses prominence compared to the profit-insensitive option. On the qualitative side, the variable *home_ownership_n* takes on greater importance, with a share of over 10%, while *emp_length* significantly reduces its share.

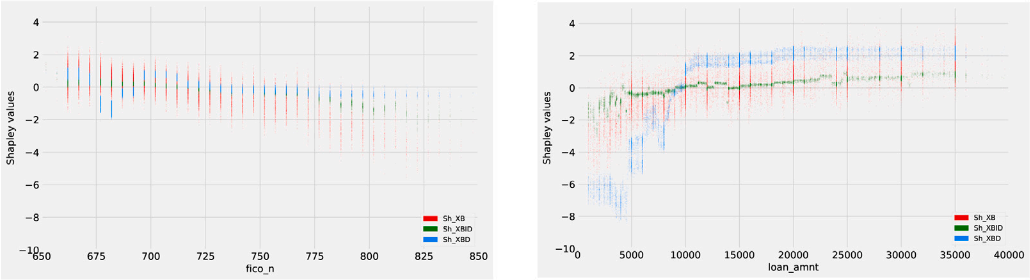
Finally, Fig. 2(c) shows the variable importance of the direct and indirect profit-sensitive model. In this case, the loan amount variable remains the most relevant variable, with even higher relative importance, which means it is even more important for detecting profits than defaults. It is followed by the *home_ownership_n* variable, which stands out from the other variables. It means that its role is more important in a profit-sensitive model than in a model that just predicts default. The *dti_n* indebtedness maintains its prominence, and the *emp_length* increases it. The FICO (*fico_n*) and *revenue* variables lose strength, which means that they are more relevant for predicting the default rather than for seeking profits. FICO (*fico_n*) recognizes prior financial experience but does not guarantee higher average profit. Similarly, for the case of the *revenue* variable. Furthermore, revenues in the small-business segment combine business and applicant revenues and are self-reported. This variable does not allow for a precise risk prediction in combination with the expected profit per obligation. This last variable would be more relevant if this small business market were to specialize, recognizing the specific income of the business alternative with an adequate verification process.

In Fig. 3, we see the dependence plots of the quantitative variables for the three models (the insensitive approach in red, the indirect approach in green, and the direct and indirect approach in blue). We can appreciate that the dependency relationships of the variables are different for the three models. Interestingly, the dispersion of the profit-insensitive model is higher for the four variables. It means that when predicting default, each loan can contribute differently for the same value of a given value. However, when predicting default informed by profit, the variation of the contribution is reduced. It means that considering profit reduces the uncertainty of the variable's role in the model.

If we look at the FICO variable (*fico_n*) in Fig. 3(a), we see, for example, that its highest values contribute much more to detecting non-default in the profit-insensitive model (red dots) than in the ones that include profit, which is closer to zero. If we look at the model that considers profit both directly and indirectly. We see a break on 680 as if detecting a niche of profitable non-defaulted loans at slightly lower values is possible. At values above 750, we also identify profitable non-defaulted loans.

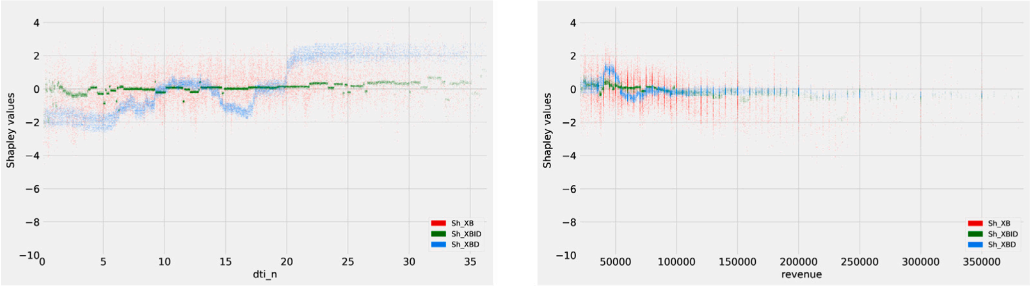
Loan amount, shown in Fig. 3(b) was the most relevant variable for all the models. We can see how having a direct profit-sensitive proposal substantially changes the structure and linearity of the variable's relationship for default prediction. Although the higher the loan amount, the higher the risk, we can see that the model that considers profit both directly and indirectly has a more marked relationship and takes more extreme values. This model considers that the loans below \$5000 are extremely interesting (SHAP values around −6), and those over \$10,000 are the ones the model tries to avoid (SHAP values around 2). This behavior suggests a market opportunity for microloans.

The variable *dti_n* is shown in Fig. 3(c). Again, this variable's role and contribution are more marked for the model that considers both profit directly and indirectly. However, the relationship shows several breaks as if identifying niches of different behavior. Loans with values of *dti_n* below 6% are the ones more interesting (SHAP values around



(a) *fico_ n* dependence graph

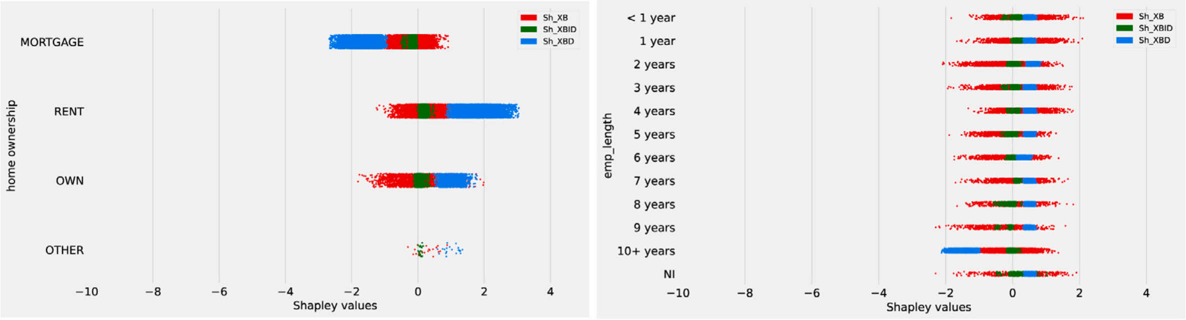
(b) *loan_ amnt* dependence graph



(c) *dti_ n* dependence graph

(d) *revenue* dependence graph

Fig. 3. Dependence plots of quantitative variables.



(a) *home ownership* dependence graph

(b) *emp_ length* dependence graph

Fig. 4. Dependence plots of categorical variables.

-2), while those over 20 are the ones that should be avoided (SHAP values around 2).

In Fig. 3(d) we see the dependency of the revenue variable. The profit-insensitive model allowed for more extreme values and hinted at an inverse relationship (the higher the revenues, the less probability of default). However, such a relationship is less clear for the profit-sensitive models, where we also see how the SHAP values tend to concentrate around zero, corroborating the loss of relevance of this variable.

The dependency relationship of the categorical variables is shown in Fig. 4. Interestingly, the model that considers profit information both directly and indirectly determines the contribution of each category to the prediction more clearly, as it can assign SHAP values that do not include zero to each category. It can be seen in the *home_ownership_n* variable in Fig. 4(a), where this model considers MORTGAGE as the lowest risk category and RENT and OWN as the riskier categories. However, the role of the categories is much less marked in the other models. It reveals that this variable can offer more valuable information when looking for profit. In particular, it could be the case that when you have a mortgage, you tend to take better care of the budget, and your decisions are more measured.

The case of the *emp_length* variable in Fig. 4(b) is similar. The model that directly and indirectly considers the profits can assign more significant values to the categories and is much less dispersed than the profit-insensitive alternative. In particular, it considers that all the categories are similar in risk, except for the case of those loans from people who held their employment for more than 10 years, a niche that the model considers that reduces the default risk weighted by profit information.

The explainability tools have helped us find that profit-sensitive ones use the input variables differently than those profit-insensitive ones. In addition, we have seen that in this case, profit information serves to determine the variable contribution more precisely, that is, with less variability. Furthermore, for the model that considers profit both directly and indirectly, the explainability tools have helped us to determine a clear profile of the most interesting loans in terms of default weighted by profitability for the small business segment. These loans should be below $5000, and the borrower should have a DTI below 5%, a mortgage, and hold his or her employment for more than ten years, while the FICO score and the revenues are much less relevant.

7. Conclusions

We have proposed a machine learning model for credit risk that consistently uses profit information throughout the modeling process. In particular, it uses it directly in the learning function of the classification algorithm (the popular XGBoost in our case), and indirectly for tuning the model and for making the granting (i.e. classification) decision. Our approach has been successfully applied to develop a granting model in the small business segment of the Lending Club dataset, which is a particularly complex and risky market segment. Finally, with the help of explainability tools, we have found that by including profit information, the models' behavior changes greatly and also changes how the input variables are used.

Our profit-based approach goes beyond other approaches by including both costs and profits in the misclassification matrix at the three modeling steps mentioned above. For incorporating profits in the XGBoost learning function, we defined the profit-based gradient and the Hessian for each instance using the cell values in the matrix. For model tuning, we used Bayesian optimization to find the hyperparameters that maximized the $R\pi_T^E$ as in Eq. (3). We used both thresholding and Bayes Maximum Profit (BMP) for the decision method. In the first case, we aim to maximize the $R\pi_T$, as in Eq. (6), while in the BMP, we compare the profits associated with predicting an obligation as default or as not-default, as in Eq. (10).

Our results show that including profit information in the learning function and optimizing the hyperparameters translate into higher financial returns. However, the profit-based approaches for analyzing the classification decision did not obtain more profits than using the standard threshold (0.5). This result is surprising and might be due to the complexity of the small business segment. It remains as future research considering other competing methods (Verbraken et al., 2014; Xia et al., 2017; Herasymovych et al., 2019) in other datasets to find further evidence about profit-based decision methods.

We also observed that profitability and classification performance do not necessarily correlate, as seen in other works (Verbraken et al., 2014; Xia et al., 2017; Wang et al., 2021). Our results extend this notion as they show that the models obtained by maximizing the classification performance and maximizing profit are different in terms of how they exploit the information of the variables. In particular, the explainability analysis revealed that including profit information in the model changes the variables' role. This was observed in the different relative importance assigned by the models to each feature, but also in changes in the functional relationships of the quantitative variables with the target variable or in the different roles that played some categories of the categorical variables. Furthermore, we also observed that in the profit-based models, the contribution of a variable for a given value has less variance, which means that including the profit consideration reduced the uncertainty of the variable's role present in the profit-insensitive models.

These facts support the idea that credit risk modeling must include profit information to be successfully applied in real life and improve profit-risk management. We have proposed a methodology that shows that risk and profit can be included in a single modeling process, unlike profit scoring models that do it separately with a model for each target (Serrano-Cinca and Gutierrez-Nieto, 2016; Lyócsa et al., 2022) or in stages, with a model for predicting default and then selecting the most profitable obligations (Xia et al., 2017; Bastani et al., 2019). As can be seen, there are several alternatives to manage the profit-risk dichotomy. Therefore, it is necessary to investigate them further to find the best ways to tackle this problem in lending models.

Going deeper into the results of our experiment from the business perspective, we observed that the best returns corresponded with models with low acceptance rates, ranging from 11% to 1% (i.e. models that grant very few loans). This result suggests that obtaining profits in the small-business segment implies filtering out most of the loans, granting only those with the lowest expectations of default weighted by profit. At this point, it is important to remember that the loans in the Lending Club dataset were granted; that is, the loans in the dataset were a priori less risky than those not granted (and not part of the dataset). These aspects make it more challenging to build a granting model, as we have seen in the low classification rates obtained, and make it more surprising the result of the profitable models are those with extremely low acceptance rates. Our work can be extended by using reject inference to decrease the bias of dealing with granted loans.

A limitation of the work is that we only analyze a segment of the Lending Club dataset. However, we believe the main conclusion will hold in the other segments and also in other datasets because it is aligned with the rationale of cost-sensitive machine learning. Namely, learning the profits and costs of correct and incorrect classifications enables the algorithm to make more informed decisions that improve its performance from the "business" perspective.

Regarding the use of explainability tools, they helped us identify a prototypical loan profile with low risk weighted by profitability for the model that considers profit directly and indirectly. In particular, the best opportunities to grant credit for small business purposes are mainly found in low indebtedness below 6% and loan amounts below $5000, which greatly reduce the risk. Regarding home ownership, the mortgage category is the only one that reduces the risk, as it happens with more than ten years of employment. The income and FICO variables are not very relevant to our profit-based model. In the case of income, a verified net income of the business (instead of the applicant's self-reported income) would probably produce a different result. We can see that the prototypical loan, according to our model, is very specific.

All these aspects reinforce the idea of the complexity not only of the risk-profit dichotomy but also of the particular segment where we worked: the small business segment in P2P lending. In the Lending Club dataset, this segment has a high default rate (close to 30%) and high potential losses (a mean annualized profit rate of −0.025 with a high variability). When small businesses use P2P lending, they look to an alternative market to cover their financial needs not satisfied by traditional markets due to risk conditions. Our methods have proven useful for selecting the most promising financial obligations regarding risk and profit in this setting. However, the average return and loan acceptance rates are very low. This fact reinforces the idea that the intervention of the state or investors who wish to support this type of proposition is needed. Small businesses tend to have low returns below the interest rates charged by traditional financing alternatives, increasing their risk of default. Thus, offering financing tools to support the productive segment of small businesses becomes vital. While P2P lending can be a financing alternative for some small businesses, the segment demands its own regulatory and leverage conditions and tailor-made application forms asking for specific information about the business object of the loan. Of course, the segment would also benefit from using profit-based approaches like the ones proposed in this paper. In this way, it will be possible to offer small businesses better financing alternatives with fairer rates and, in turn, serve as an investment and support scenario for those who wish to leverage new small business ventures. All this would contribute to mitigating the gap between the demand and supply of financial products and services for small businesses, which is one of the implicit objectives of the United Nations Sustainable Development Goals to reduce the access gap in financial services (Carpentier and Braun, 2020; Liguori and Bendickson, 2020).

Finally, from a classification perspective, as Sun et al. (2009) point out for class-imbalance problems, financial default is a concept with such a degree of complexity that the patterns of each class in the feature space may overlap, and discriminatory rules become difficult to induce. However, our work has shown that it is possible to successfully combine the search for risk minimization and high profitability in one single model, resulting in better financial and risk management. Furthermore, our work demonstrates that explainability tools can help to understand better the main drivers of risk-weighted by profit, appreciate the differences from profit-insensitive alternatives, and find good profiles of loans with business sense. Although we worked in the case of small businesses in P2P lending, the approach presented in our paper can be easily replicated in other credit markets or classification problems where misclassification matrix values have a measurable impact.

CRediT authorship contribution statement

Miller-Janny Ariza-Garzón: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Javier Arroyo:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Conceptualization. **María-Jesús Segovia-Vargas:** Writing – review & editing, Supervision, Conceptualization. **Antonio Caparrini:** Writing – review & editing, Software.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Akila, S., Reddy, S.U., 2018. Cost-sensitive risk induced Bayesian inference bagging (RIBIB) for credit card fraud detection. J. Comput. Sci. 27, 247–254. http://dx.doi.org/10.1016/j.jocs.2018.06.009.

Almhaithawi, D., Jafar, A., Aljnidi, M., 2020. Example-dependent cost-sensitive credit cards fraud detection using SMOTE and Bayes minimum risk. SN Appl. Sci. 2 (9), http://dx.doi.org/10.1007/s42452-020-03375-w.

Ariza-Garzon, M., Arroyo, J., Caparrini, A., Segovia-Vargas, M.J., 2020. Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending. IEEE Access 8, http://dx.doi.org/10.1109/ACCESS.2020.2984412.

Ariza-Garzón, M.-J., Camacho-Miñano, M.-D.-M., Segovia-Vargas, M.J., Arroyo, J., 2021. Risk-return modelling in the p2p lending market: Trends, gaps, recommendations and future directions. Electron. Commer. Res. Appl. 49, 101079. http://dx.doi.org/10.1016/j.elerap.2021.101079.

Babaei, G., Bamdad, S., 2020. A multi-objective instance-based decision support system for investment recommendation in peer-to-peer lending. Expert Syst. Appl. 150, 113278. http://dx.doi.org/10.1016/j.eswa.2020.113278.

Bahnsen, A.C., Aouada, D., Ottersten, B., 2014. Example-dependent cost-sensitive logistic regression for credit scoring. In: 2014 13th International Conference on Machine Learning and Applications. pp. 263–269. http://dx.doi.org/10.1109/ICMLA.2014.48.

Bahnsen, A.C., Villegas, S., Aouada, D., Ottersten, B., 2015. Example-dependent cost-sensitive decision trees. Expert Syst. Appl. 42 (19), 6609–6619. http://dx.doi.org/10.1016/j.eswa.2015.04.042.

Bastani, K., Asgari, E., Namavari, H., 2019. Wide and deep learning for peer-to-peer lending. Expert Syst. Appl. 134, 209–224. http://dx.doi.org/10.1016/j.eswa.2019.05.042.

Bergstra, J., Yamins, D., Cox, D., 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In: International Conference on Machine Learning. PMLR, pp. 115–123.

Bernardi, M., Catania, L., 2014. The model confidence set package for R. arXiv preprint arXiv:1410.8504.

Boiko Ferreira, L.E., Barddal, J.P., Enembreck, F., Gomes, H.M., 2017. Improving credit risk prediction in online Peer-to-Peer (P2P) lending using imbalanced learning techniques. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence. In: Proceedings-International Conference on Tools With Artificial Intelligence, pp. 175–181. http://dx.doi.org/10.1109/ictai.2017.00037.

Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J., 2020. Explainable AI in fintech risk management. Front. Artif. Intell. 3, 26. http://dx.doi.org/10.3389/frai.2020.00026.

Byanjankar, A., Heikkila, M., Mezei, J., 2015. Predicting Credit Risk in Peer-to-Peer Lending: A Neural Network Approach. 2015 IEEE Symposium Series on Computational Intelligence. pp. 719–725. http://dx.doi.org/10.1109/ssci.2015.109.

Calabrese, R., Osmetti, S.A., Zanin, L., 2019. A joint scoring model for peer-to-peer and traditional lending: A bivariate model with copula dependence. J. Roy. Statist. Soc. Ser. A 182 (4), 1163–1188. http://dx.doi.org/10.1111/rssa.12523.

Carpentier, C.L., Braun, H., 2020. Agenda 2030 for sustainable development: A powerful global framework. J. Int. Council Small Bus. 1 (1), 14–23. http://dx.doi.org/10.1080/26437015.2020.1714356.

Chang, A.H., Yang, L.K., Tsaih, R.H., Lin, S.K., 2022. Machine learning and artificial neural networks to construct P2P lending credit-scoring model: A case using Lending Club data. Quantit. Finance Econ. 6 (2), 303–325. http://dx.doi.org/10.3934/QFE.2022013.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, ACM, New York, NY, USA, pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Chen, Y.R., Leu, J.S., Huang, S.A., Wang, J.T., Takada, J.I., 2021. Predicting default risk on peer-to-peer lending imbalanced datasets. IEEE Access 9, 73103–73109. http://dx.doi.org/10.1109/ACCESS.2021.3079701.

Cho, P., Chang, W., Song, J.W., 2019. Application of instance-based entropy fuzzy support vector machine in Peer-to-Peer lending investment decision. IEEE Access 7, 16925–16939. http://dx.doi.org/10.1109/access.2019.2896474.

Correa Bahnsen, A., Aouada, D., Stojanovic, A., Ottersten, B., 2016. Feature engineering strategies for credit card fraud detection. Expert Syst. Appl. 51, 134–142. http://dx.doi.org/10.1016/j.eswa.2015.12.030, URL https://www.sciencedirect.com/science/article/pii/S0957417415008386.

Correa Bahnsen, A., Villegas, S., Aouada, D., Ottersten, B., 2018. Fraud detection by stacking cost-sensitive decision trees. In: Security Science and Technology. World Scientific, pp. 251–266. http://dx.doi.org/10.1142/9781786345646_012.

Cummins, M., Lynn, T., Mac an Bhaird, C., Rosati, P., 2019. Addressing information asymmetries in online Peer-to-Peer lending. In: Disrupting Finance: FinTech and Strategy in the 21st Century. Springer International Publishing, pp. 15–31. http://dx.doi.org/10.1007/978-3-030-02330-0_2.

Duan, J., 2019. Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction. J. Franklin Inst. 356 (8), 4716–4731. http://dx.doi.org/10.1016/j.jfranklin.2019.01.046.

Durovic, A., 2017. Estimating probability of default on Peer to Peer market - Survival analysis approach. J. Cent. Bank. Theory Pract. 6 (2), 149–167. http://dx.doi.org/10.1515/jcbtp-2017-0017.

Emekter, R., Tu, Y., Jirasakuldech, B., Lu, M., 2015. Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. Appl. Econ. 47 (1), 54–70. http://dx.doi.org/10.1080/00036846.2014.962222.

Expert Group on Regulatory Obstacles to Financial Innovation, 2019. Thirty Recommendations on Regulation, Innovation and Finance. Final Report to the European Commission. Technical Report, European Commission, URL https://ec.europa.eu/info/files/191113-report-expert-group-regulatory-obstacles-financial-innovation_en.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from Imbalanced Data Sets, vol. 10, Springer.

Fu, X., Zhang, S., Chen, J., Ouyang, T., Wu, J., 2019. A Sentiment-Aware Trading Volume Prediction Model for P2P Market Using LSTM. IEEE Access 7, 81934–81944. http://dx.doi.org/10.1109/access.2019.2923637.

Gao, Y., Yu, S., Chen, M., Shiue, Y., 2020. A 2020 perspective on "The performance of the P2P finance industry in China". Electron. Commer. Res. Appl. http://dx.doi.org/10.1016/j.elerap.2020.100940.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Havrylchyk, O., Verdier, M., 2018. The financial intermediation role of the P2P lending platforms. Comparat. Econ. Stud. 60 (1), 115–130.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. http://dx.doi.org/10.1109/TKDE.2008.239.

Herasymovych, M., Märka, K., Lukason, O., 2019. Using reinforcement learning to optimize the acceptance threshold of a credit scoring model. Appl. Soft Comput. 84, 105697. http://dx.doi.org/10.1016/j.asoc.2019.105697.

Jadwal, P.K., Jain, S., Agarwal, B., 2020. Spectral clustering and cost-sensitive deep neural network-based undersampling approach for P2P lending data. Int. J. Inf. Technol. Web Eng. 15 (4), 37–52.

Klein, G., Shtudiner, Z., Zwilling, M., 2021. Why do peer-to-peer (P2P) lending platforms fail? The gap between P2P lenders' preferences and the platforms' intentions. Electron. Comm. Res. http://dx.doi.org/10.1007/s10660-021-09489-6.

Li, W., Ding, S., Chen, Y., Yang, S., 2018. Heterogeneous ensemble for default prediction of Peer-to-Peer lending in China. IEEE Access 6, 54396–54406. http://dx.doi.org/10.1109/access.2018.2810864.

Li, W., Ding, S., Wang, H., Chen, Y., Yang, S., 2020. Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China. World Wide Web-Internet Web Inf. Syst. 23 (1), 23–45. http://dx.doi.org/10.1007/s11280-019-00676-y.

Liguori, E., Bendickson, J.S., 2020. Rising to the challenge: Entrepreneurship ecosystems and SDG success. J. Int. Council Small Bus. 1 (3–4), 118–125.

Ling, C.X., Sheng, V.S., 2010. Cost-sensitive learning. In: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning. Springer US, Boston, MA, pp. 231–235. http://dx.doi.org/10.1007/978-0-387-30164-8_181.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. http://dx.doi.org/10.1038/s42256-019-0138-9.

Lundberg, S.M., Lee, S.I., 2017a. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30, 10.

Lundberg, S.M., Lee, S.I., 2017b. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., pp. 4765–4774, URL http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Lyócsa, Š., Vašaničová, P., Hadji Misheva, B., Vateha, M.D., 2022. Default or profit scoring credit systems? Evidence from European and US peer-to-peer lending markets. Finan. Innov. 8 (32), 1–21. http://dx.doi.org/10.1186/s40854-022-00338-5.

Mach, T., Carter, C., Slattery, C., 2014. Peer-to-peer lending to small businesses. FEDS Working Paper. URL https://www.federalreserve.gov/pubs/feds/2014/201410/201410pap.pdf.

Maldonado, S., Pérez, J., Bravo, C., 2017. Cost-based feature selection for support vector machines: An application in credit scoring. European J. Oper. Res. 261 (2), 656–665.

Malekipirbazari, M., Aksakalli, V., 2015. Risk assessment in social lending via random forests. Expert Syst. Appl. 42 (10), 4621–4631. http://dx.doi.org/10.1016/j.eswa.2015.02.001.

Maskara, P.K., 2020. The role of P2P platforms in enhancing financial inclusion in US – An analysis of peer-to-peer lending across the rural-urban divide. http://dx.doi.org/10.2139/ssrn.3715211, SSRN.

Masnadi-Shirazi, H., Vasconcelos, N., 2011. Cost-sensitive boosting. IEEE Trans. Pattern Anal. Mach. Intell. 33 (2), 294–309. http://dx.doi.org/10.1109/TPAMI.2010.71.

Mills, K., McCarthy, B., 2016. The state of small business lending: Innovation and technology and the implications for regulation. Harvard Business School Entrepreneurial Management Working Paper 17–042.

Milne, A., Parboteeah, P., 2016. The business models and economics of Peer-to-Peer lending. In: Centre for European Policy Studies, vol. 17, p. 36.

Misheva, B.H., Giudici, P., Pediroda, V., IEEE, 2018. Network-based models to improve credit scoring accuracy. In: 2018 Ieee 5th International Conference on Data Science and Advanced Analytics. In: Proceedings of the International Conference on Data Science and Advanced Analytics, pp. 623–630. http://dx.doi.org/10.1109/dsaa.2018.00080.

Nami, S., Shajari, M., 2018. Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors. Expert Syst. Appl. 110, 381–392. http://dx.doi.org/10.1016/j.eswa.2018.06.011.

Nowak, A., Ross, A., Yencha, C., 2018. Small business borrowing and peer-to-peer lending: Evidence from lending club. Contemp. Econ. Policy 36 (2), 318–336.

Oreski, S., Oreski, G., 2018. Cost-sensitive learning from imbalanced datasets for retail credit risk assessment. TEM J. 7 (1), 59–73. http://dx.doi.org/10.18421/TEM71-08.

Papouskova, M., Hajek, P., 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decis. Support Syst. 118, 33–45. http://dx.doi.org/10.1016/j.dss.2019.01.002.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Petrides, G., Verbeke, W., 2022. Cost-sensitive ensemble learning: a unifying framework. Data Min. Knowl. Discov. 36 (1), 1–28.

Polena, M., Regner, T., 2018. Determinants of borrowers' default in P2P lending under consideration of the loan risk class. Games 9 (4), http://dx.doi.org/10.3390/g9040082.

Rao, C., Liu, M., Goh, M., Wen, J., 2020. 2-stage modified random forest model for credit risk assessment of P2P network lending to "Three Rurals" borrowers. Appl. Soft Comput. 95, http://dx.doi.org/10.1016/j.asoc.2020.106570.

Ribeiro, M.T., Singh, S., Guestrin, C., Assoc Comp, M., 2016. Why should I trust you? Explaining the predictions of any classifier. In: KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144. http://dx.doi.org/10.1145/2939672.2939778.

Rodrigues, D.S., Brasil, A.R.A., Costa, M.B., Komati, K.S., Pinto, L.A., ACM, 2018. A comparative analysis of loan requests classification algorithms in a Peer-to-Peer lending platform. Proceedings of the 14th Brazilian Symposium on Information Systems. pp. 325–332. http://dx.doi.org/10.1145/3229345.3229390.

Serrano-Cinca, C., Gutierrez-Nieto, B., 2016. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. Decis. Support Syst. 89, 113–122. http://dx.doi.org/10.1016/j.dss.2016.06.014.

Serrano-Cinca, C., Gutierrez-Nieto, B., Lopez-Palacios, L., 2015. Determinants of default in P2P lending. PLoS One 10 (10), http://dx.doi.org/10.1371/journal.pone.0139427.

Shapley, L.S., 1953. A value for n-person games. In: Kuhn, H., Tucker, A. (Eds.), Contributions to the theory of games. In: Annals of Mathematics Studies, vol. 2, (28), Princeton University Press, pp. 307–317. http://dx.doi.org/10.1515/9781400881970-018.

Sharma, S.K., Ilavarasan, P.V., Karanasios, S., 2024. Small businesses and FinTech: a systematic review and future directions. Electron. Commer. Res. 535–575. http://dx.doi.org/10.1007/s10660-023-09705-5.

Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. arXiv e-prints URL https://arxiv.org/abs/1704.02685.

Sun, Y., Wong, A.K.C., Kamel, M.S., 2009. Classification of imbalanced data: A review. Int. J. Pattern Recognit. Artif. Intell. 23 (04), 687–719. http://dx.doi.org/10.1142/S0218001409007326.

Van-Sang, H., Dang-Nhac, L., Choi, G.S., Ha-Nam, N., Yoon, B., 2019. Improving credit risk prediction in online Peer-to-Peer (P2P) lending using feature selection with deep learning. In: 2019 21st International Conference on Advanced Communication Technology. In: International Conference on Advanced Communication Technology, pp. 511–515.

Verbraken, T., Bravo, C., Weber, R., Baesens, B., 2014. Development and application of consumer credit scoring models using profit-based classification measures. European J. Oper. Res. 238 (2), 505–513. http://dx.doi.org/10.1016/j.ejor.2014.04.001.

Wang, H., Kou, G., Peng, Y., 2018. Cost-sensitive classifiers in credit rating a comparative study on P2P lending. In: Dzitac, I., Filip, F.G., Manolescu, M.J., Dzitac, S., Oros, H., Dzitac, D. (Eds.), 2018 7th International Conference on Computers Communications and Control. pp. 210–213.

Wang, H., Kou, G., Peng, Y., 2021. Multi-class misclassification cost matrix for credit ratings in Peer-to-Peer lending. J. Oper. Res. Soc. 72 (4), 923–934. http://dx.doi.org/10.1080/01605682.2019.1705193.

Wang, C., Liu, Q., Li, S., 2022. A two-stage credit risk scoring method with stacked-generalisation ensemble learning in Peer-to-Peer lending. Int. J. Embed. Syst. 15 (2), 158–166. http://dx.doi.org/10.1504/IJES.2022.123312.

Wu, S., Gao, X., Zhou, W., 2022. COSLE: Cost sensitive loan evaluation for P2P lending. Inform. Sci. 586, 74–98. http://dx.doi.org/10.1016/j.ins.2021.11.055, URL https://www.sciencedirect.com/science/article/pii/S0020025521011774.

Xia, Y., Liu, C., Liu, N., 2017. Cost-sensitive boosted tree for loan evaluation in Peer-to-Peer lending. Electron. Commer. Res. Appl. 24, 30–49. http://dx.doi.org/10.1016/j.elerap.2017.06.004.

Ye, X., Dong, L.A., Ma, D., 2018. Loan evaluation in P2P lending based on Random Forest optimized by genetic algorithm with profit score. Electron. Commer. Res. Appl. 32, 23–36. http://dx.doi.org/10.1016/j.elerap.2018.10.004.

Zang, D., Qi, M., Fu, Y., 2015. The credit risk assessment of P2P lending based on BP neural network. In: Lee, G. (Ed.), Industrial Engineering and Management Science. pp. 91–94.

Zhang, W., Xu, W., Hao, H., Zhu, D., 2020. Cost-sensitive multiple-instance learning method with dynamic transactional data for personal credit scoring. Expert Syst. Appl. 157, http://dx.doi.org/10.1016/j.eswa.2020.113489.

Zhou, G., Zhang, Y., Luo, S., 2018. P2P network lending, loss given default and credit risks. Sustainability 10 (4:1010), 105697. http://dx.doi.org/10.3390/su10041010.

Zhu, L., Qiu, D., Ergu, D., Ying, C., Liu, K., 2019. A study on predicting loan default based on the random forest algorithm. In: HerreraViedma, E., Shi, Y., Berg, D., Tien, J., Cabrerizo, F.J., Li, J. (Eds.), Procedia Computer Science, 7th International Conference on Information Technology and Quantitative Management Procedia Computer Science, vol. 162, 503–513.http://dx.doi.org/10.1016/j.procs.2019.12.017,

Zou, Y., Gao, C., Gao, H., 2022. Business failure prediction based on a cost-sensitive extreme gradient boosting machine. IEEE Access 10, 42623–42639. http://dx.doi.org/10.1109/ACCESS.2022.3168857.

Miller Janny Ariza-Garzón holds a Ph.D. in Data Science from the Faculty of Statistical Studies of Universidad Complutense of Madrid (2023). He currently serves as a data scientist in the Liability and Capital Risk Management Department of VidaCaixa, an insurance company. He was previously a researcher for the Fin-Tech Ho2020 European project at the Department of Software Engineering and Artificial Intelligence of Universidad Complutense of Madrid. He has research experience in management and consulting on modeling areas for risk management and time series forecasting in financial institutions. He has been a quantitative and statistical consultant in evaluating financial inclusion projects in Latin America and the Caribbean.

Javier Arroyo received a Ph.D. degree in computer science from Universidad Pontificia Comillas, Madrid, in 2008. He is an Associate Professor at the Department of Software Engineering and Artificial Intelligence, Universidad Complutense de Madrid (UCM), Madrid, and a Researcher with the Instituto de Tecnología del Conocimiento. His research interests include machine learning, time series forecasting, and open collaboration, among others.

María-Jesús Segovia-Vargas received the B.S. and Ph.D. degrees in Economics and Business Administration from Universidad Complutense de Madrid, Spain, in 1994 and 2003, respectively. Currently, she is an Associate Professor at the Department of Financial and Actuarial Economics and Statistics at Universidad Complutense de Madrid. Her main research topic focuses on applying operational research methods to analyze financial problems, especially insurance companies' solvency, internationalization success, bank crisis, and bankruptcy. She has published several monographs and papers in internationally renowned journals. She has also participated in several research projects.

Antonio Caparrini is a project manager at the Department of Information Systems at the Bank of Spain. He is currently pursuing his Ph.D. in Data Science at the Faculty of Statistical Studies of Universidad Complutense de Madrid (UCM). MSc in Data Science, MA in Business Consulting, and BSc in Computer Engineering, he collaborates as an assistant professor at the Universitat Oberta de Catalunya (UOC). His main areas of research interest are machine learning, data engineering, risks, and finance.