

RESEARCH

Open Access

Estimation of default and pricing for invoice trading (P2B) on crowdlending platforms



Cristian Marques Corrales^{1*} , Luis Alberto Otero González¹ and Pablo Durán Santomil¹

*Correspondence:
cristian.marques@rai.usc.es

¹ Department of Financial Economics and Accounting, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

Abstract

This study developed several machine learning models to predict defaults in the invoice-trading peer-to-business (P2B) market. Using techniques such as logistic regression, conditional inference trees, random forests, support vector machines, and neural networks, the prediction of the default rate was evaluated. The results showed that these techniques can effectively improve the detection of defaults by up to 56% while maintaining levels of specificity above 70%. Unlike other studies on the same topic, this was performed using sampling techniques to address the imbalance of classes and using different time periods for the training and test datasets to ensure intertemporal validation and realistic predictions. For the first-time, default explainability in the invoice-trading market was studied by examining the impact of macroeconomic factors and invoice characteristics. The findings highlighted that gross domestic product, exports, trade type, and trade bands are significant factors that explain defaults. Furthermore, the pricing mechanisms of P2B platforms were evaluated with the observed and implicit probabilities of the default to analyze the price risk adjustment. The results showed that price reflects a significantly higher implicit probability of default than observed default, which in turn suggests that underlying factors exist besides the borrowers' probability of default.

Keywords: Crowdlending, Factoring, P2B, Machine learning, Pricing, Default

Introduction

Crowdlending is a trending topic that has emerged primarily in the last decade and is set to continue growing in the future for FinTech-related studies (Liu et al. 2020). Online invoice trading is a subfield of crowdlending, a digital market that has experienced exponential growth (Ziegler et al. 2017) and helps businesses finance their working capital. It is a niche segment of the broader P2B market and consists of the financial discount on an invoice via a platform (lender) in exchange for the payment of commissions or fees. Usually, invoices are financed by many investors (crowdlenders), and the platform analyzes the risk of the transaction, establishing a rating and price. The pricing mechanism can be decided in an auction style or by fixed prices set by the platforms. Most of these platforms have evolved towards a fixed-price regime (Dorfleitner et al. 2017). Estimating the default probability is essential for investors and the pricing mechanism of securities (Carmichael 2014). The increased availability of open-source datasets and advances

in new techniques have boosted interest in default predictions (Turiel and Aste 2020). Unlike peer-to-peer (P2P) platforms, which provide a great deal of information about debtors, the main invoicing platform (Kriya) offers more limited information, complicating external default evaluation. Additionally, invoice trading has essential characteristics mainly dedicated to short-term financing that differentiate it from other lending operations in terms of risk.

Most research on crowdlending has focused on P2P financing (Carmichael 2014; Serrano-Cinca et al. 2015; Zhu et al. 2019), whereas few studies have addressed the P2B market. Therefore, the first objective of this study is to help investors determine their probability of default (PD) by developing models using publicly available information. To the best of our knowledge, only Dorfleitner et al. (2017) have published a study that focused on estimating the default of an invoice trading platform (Kriya) using logit and Tobit models. Unlike Dorfleitner et al. (2017), our approach focuses on rectifying the imbalance of classes using sampling techniques that allow for the correct classification of defaults with unbiased models. Any model that does not consider this problem is biased towards the majority class (Kotsiantis and Pintelas 2004; Bastani et al. 2019) making it unsuitable for default prediction. Furthermore, greater robustness of the models was provided by ensuring intertemporal validation. This enabled the generation of realistic predictions for future test samples. To the best of our knowledge, this study represents a pioneering application of machine-learning methodologies in the invoice-lending market. We compared the performance of logistic regression, conditional inference trees, conditional inference forests, random forests, support vector machines, and neural networks for predicting defaults, similar to Li et al. (2020) for credit ratings. This study is also the first to examine the influence of new macroeconomic factors and invoice characteristics on the default rate in the invoice trading market. By presenting this innovative approach, a new set of variables that are significant determinants for predicting default was identified: gross domestic product (GDP), exports, trade type, and trade band. The current findings demonstrated that machine-learning techniques can improve default detection by over 56%. Furthermore, the use of sampling techniques to address the imbalance between the two classes produces good results in the detection of loan default, with sampling techniques correctly predicting default by more than 50% compared with scenarios without sampling.

Price setting by the platform, using the implicit and observed probability of default (IPD and OPD, respectively) to analyze the overcharging of sellers, is also discussed for the first time in this study. As such, this study assessed whether the platform's prices correspond to the credit risk inherent to transactions in the context of a diversified retail portfolio. If the charges are not adjusted to the risk of the transactions, this indicates that companies are willing to pay a high interest rate because they are unable to obtain competitive financing from other financial sources, or that they greatly value the flexibility of these platforms. The results revealed a significant distortion between OPD and IPD.

Sect. "[Literature review](#)" provides a literature review on the structure of this study. Sect. "[Dataset and descriptive analysis](#)" provides a descriptive analysis and details of the data preprocessing performed. Sect. "[Predicting invoice-lending default](#)" discusses the determinants used to predict default using logistic regression, including an evaluation comparing the predictive capabilities of the model across various sampling techniques.

It also compares the performance of different machine learning alternatives in terms of predicting defaults. Sect. "[Pricing mechanism with implicit probability of default](#)" examines the platform's credit-pricing method using IPD and OPD. Finally, Sect. "[Discussion of results](#)" discusses the findings and conclusions are presented at the end of this paper.

Literature review

P2P lending has attracted considerable attention from the academic community over the past decade. Most research on this topic has used logistic regression to predict the probability of default and/or profitability of loans. Most of the research was performed using publicly available data from the Lending Club platform, with the remainder obtained from a small set of platforms.

Early research was conducted by Carmichael (2014), who predicted the probability of default and expected returns from P2P lending. The former was estimated using a dynamic logistic regression, and the log of income, recent credit inquiries, loan purpose, loan amount, credit score, and subgrade were the most significant explanatory factors. Variables gathered from borrowers' loan descriptions, such as whether the description lacked complete sentences or claimed that the author was creditworthy, were also significant in explaining default. The full model performed better than the final club subgrade. Meanwhile, the model without a subgrade was included as an explanatory variable and performed similarly to the Lending Club subgrade. To estimate the expected returns, the probability of early repayment and the principal repaid were calculated given default. The first was modelled with a dynamic logistic regression using the same regressors as for default, whereas the second was performed with ordinary least squares. The expected return on the lowest-risk subgrade (A1) loans was 5%. This increased steadily to a maximum of 11% for mid-risk loans (D2), and then decreased to 10% for the highest-risk loans (E5). Similarly, Li et al. (2016) estimate the probability of prepayment in addition to the probability of default. They used multivariate logistic regression incorporating macroeconomic factors and found that the factors explaining default and prepayment were very similar. Specifically, loan features, macroeconomic factors, and most borrower characteristics are significant. Their results showed that high interest rates are not only associated with higher probabilities of default but also with higher probabilities of prepayment, as borrowers do not want to bear the costs associated with them. Serrano-Cinca et al. (2015) study the determinants of default in P2P lending. Using a hypothesis test and survival analysis, they first determine the most significant factors for estimating default, namely, loan purpose, annual income, current housing situation, credit history, and indebtedness. Subsequently, several logistic regressions are performed to predict the probability of default based on the previously determined factors. Their results showed a clear relationship between the Lending Club subgrade and the probability of default, where the subgrade was the variable with the highest predictive capability. Furthermore, interest rates appeared to depend on the grade assigned: the higher the interest rate, the higher the probability of default. These results are similar to those of Möltenkamp (2017), who used binary logistic regression to investigate the determinants of P2P loan performance based on different credit grades. The results show a positive relationship between credit grade and loan performance, wherein a higher credit

grade was related to a lower probability of default. As for loan performance, credit grade is the most influential factor. Loan amount and annual income were also significant predictors, whereas all other variables lost significance in forecasting. Regarding the determinants of default, Avgeri and Psillaki (2023) explored borrower-related and macroeconomic factors in the US P2P market using logistic regression. Their study suggests that the higher the percentage change in the house price, consumer sentiment, and S&P500 indices, the lower the delinquency. Unemployment and GDP affected the rate of default. Nigmonov et al. (2022) used a probit model and found that for the same market, the higher the interest rates and inflation, the higher the probability of default. Moreover, the effect of interest rates on default is significantly higher for loans with lower ratings.

Other authors have attempted to predict the probability of default using a different set of methodologies. For example, Zhu et al. (2019) compared a random forest model with other statistical techniques and discovered that the former outperforms decision trees, logistic regression, and support vector machines with an outstanding 98% level of accuracy. They used a synthetic minority oversampling technique (SMOTE) to solve the imbalance problem in the sample. Similarly, Malekipirbazari and Aksakalli (2015) compared different statistical techniques—such as random forests, support vector machines, logistic regression, and the nearest neighbor algorithm—to predict the default rate of P2P lending. Their results showed that random forests outperformed Fair Isaac Corporation (FICO) credit scores, the Lending Club subgrade, and other methodologies for identifying good borrowers, with an accuracy level of 78%. Moreover, based on their findings, although this technique was highly suitable for identifying good borrowers, misclassifications existed for some borrowers who were erroneously deemed bad. Borrower status was also studied by Fu (2017), who combined random forests with neural networks and compared the performance of each model alone and together, where the probability of default was estimated by one model and then given to the other for re-estimation. Preprocessing was also performed, where the data were first normalized, and old observations were discarded. This combined technique, along with data preprocessing, considerably improves accuracy and outperforms the Lending Club subgrade. The highest accuracy was achieved by first using a neural network and then using random forests. A similar method was used by Kim and Cho (2018), who proposed a deep dense convolutional network (DenseNet) for default predictions in P2P. Their findings revealed that this model could achieve a relatively high level of accuracy (79.6%) and reduce overfitting compared with other convolutional neural networks.

Furthermore, Turiel and Aste (2020) developed several artificial intelligence models to predict loan rejection and estimate the probability of default, proving that artificial intelligence can increase accuracy and default risk by 70%. They also proposed the separation of small business subsets to increase the performance of default predictions. Ko et al. (2022) proposed a wide range of prediction models to mitigate the risk of default and asymmetric information on P2P lending platforms, stating that LightGBM outperformed the other methodologies, with a model accuracy of 68.57% and a revenue improvement of 23.8 million US dollars. They argue that the Lending Club, despite being the largest P2P lending platform in the USA, still has a high rate

of default, which proves its ineffectiveness when classifying debt. Similarly, Muslim et al. (2022) applied an improved LightGBM based on swarm algorithms to predict default rates on P2P platforms. Their study indicated that the performance increased after feature selection using a swarm algorithm, with LightGBM + ACO achieving the highest level of accuracy (95.64%).

In terms of profitability, the body of work done by Serrano-Cinca and Gutiérrez-Nieto (2016) is worthy of mention, which used a multivariate regression along with a decision tree model (CHAID) to develop a profit scoring system. The internal rate of return (IRR) of each loan was used as a profitability measure. Using exploratory data analysis and multivariate linear regression, they found that the explanatory factors for predicting default and profitability differed. Similar to other studies, the Lending Club subgrade was found to be significant. Although the loan purpose and housing characteristics were generally significant, they seemed to be more significant in predicting profitability. Moreover, a nonlinear relationship (inverted and U-shaped) existed between the internal rate of return and its factors. With respect to profitability, the decision tree has a mean internal rate of return of 5.98%, outperforming the Lending Club's mean of 3.92%. Their study also suggested that nonlinear data mining techniques can be useful for developing profit-scoring systems.

Similarly, Guo et al. (2016) predicted the expected return on a loan using an instance-based model (IOM), by considering the investment decision on a P2P platform as a portfolio optimization problem with boundary constraints. They compared the results of this instance-based model with those of two rating-based models, the restricted Boltzmann machine (RBM) and RBM+, and showed that the IOM outperformed them in terms of prediction accuracy and investment performance. The probability of default was estimated using logistic regression, which was then applied to the IOM to calculate the expected returns. Similarly, Bastani et al. (2019) propose a two-stage system for credit and profit scoring. In Stage 1, they attempted to identify non-default loans, which were then moved to Stage 2, where they estimated profitability using the internal rate of return. Wide and deep learning using Google was used to build predictive models in both stages. This model can achieve both memorization and generalization, avoiding the overgeneralization frequently observed in deep learning modelling techniques. The factors predicting default and the internal rate of return were similar to each other and those studied by Serrano-Cinca et al. (2015) and Serrano-Cinca and Gutiérrez-Nieto (2016), respectively. These two studies used different statistical techniques to address the imbalance problem of the sample, namely, random undersampling, random oversampling, and SMOTE. Nonetheless, the latter method appeared to perform consistently better than the other two. Their results indicated that the proposed scoring approach outperformed existing credit and profit scoring approaches, and the combination of wide and deep learning with the SMOTE method achieved the highest performance.

Elliott and Curet (1999) devised the first framework for invoice discounting, a generic term for financing solutions that use invoices as collateral for loans. They suggested using an inductive algorithm such as case-based reasoning (CBR) and noticed a lack of knowledge regarding invoice-discounting cases. To the best of our knowledge, few similar studies have been published to date. Dorfleitner et al. (2017) estimated default events in the online invoice trading market using logit models. Their study suggests that interest

rates, duration, and advance rates are determinant factors in predicting default, where the higher the gross yield and the longer the time to maturity, the higher the probability of default and the loss rate. However, the advance rate is negatively related to default, reflecting the platform's ability to assess sellers' creditworthiness. This implies that the larger and more creditworthy a company is, the more easily it can obtain greater loans. It has also been concluded that risk can be effectively reduced by diversifying the portfolio of invoices and that investors prefer to deal with a higher risk with higher interest rates rather than lower advance rates. Furthermore, the default and average net returns are lower in a fixed-price regime than in an auction price mechanism. Zhang and Thomas (2015) consider the merits of including economic variables in a logistic regression-based credit scorecard in an invoice-discounting context. By doing so, they wanted to directly estimate a short-term, dynamic version of the probability of default (i.e., Point-in-Time (PIT) probability of default). Perko (2017) also adapted the definition of "probability of default" as a more short-term and fine-grained concept; the outcome should be predicted at a fixed timeframe of 30 days in advance, rather than overdue days.

Other authors have extensively evaluated credit scoring systems and the estimation of default in commercial credit. However, this study focuses on the estimation of the probability of default and pricing optimization on crowdlending platforms, thus we have not referred to the authors in this literature review. A summary of related bodies of research can be found in Table 9 in Appendix, where additional information exists regarding the accuracy of the models and the samples used.

Dataset and descriptive analysis

Institutional background and data

We used publicly available data from the crowdlending platform Kriya,¹ whose headquarters are in the United Kingdom (UK). The company was formerly known as MarketInvoice and was later rebranded as MarketFinance before being given its current name; it specializes in factoring and P2B. According to the information provided on their website, as of November 2023, they had funded around 3.4 billion pounds of sterling (GBP) to small and medium-sized enterprises (SMEs), with an expected net yield of between 4 and 6% and a loan collection rate of 78.01%. After data cleansing, the dataset contained 46,761 observations and 26 variables with loans from March 2011 to December 2019. Some original variables were dropped since considering them for predictive purposes—such as trade payment, trade settlement, and delinquent dates—was not sensible. This was also the case for the currency variables; this study focused on GBP given the low number of observations in other currencies. Meanwhile, the trade expected pay date was subtracted to the advance date, creating the variable "days", which represents the maturity of the transaction. In addition, the following macroeconomic variables were tested for model fitting: GDP, exports, import price index, producer price index, consumer price index, money supply, M1, US/GBP spot exchange rate, and EUR/GBP spot exchange rate. However, after feature selection, only the first three showed significance and were considered for further analysis. The information provided by the platform was rather limited since no variables related existed regarding financial grade, income, or solvency; financial information about the borrower, seller, or protection provider was

¹ More information at: <https://kriya.co>.

Table 1 Summary of numerical variables

| | Advance rate (%) | Annualized gross yield (%) | Total face value (£) | Total crystallized loss (£) | Days | GDP (%) | Exports (%) | IPI (%) |
|--------|------------------|----------------------------|----------------------|-----------------------------|------|---------|-------------|---------|
| Min | 2.52 | 2.71 | 1 | 0 | 1 | -0.82 | -4.58 | -5.58 |
| 1°Q | 70 | 7.94 | 9588 | 0 | 35 | 0.61 | 0 | -0.30 |
| Median | 80 | 10.62 | 22,511 | 0 | 48 | 0.94 | 1.70 | -0.03 |
| Mean | 76.22 | 10.99 | 57,243 | 108 | 55 | 0.89 | 1.86 | -0.05 |
| 3°Q | 85 | 12.87 | 55,415 | 0 | 72 | 1.24 | 4.41 | 0.54 |
| Max | 99.63 | 49.02 | 36,792,829 | 488,899 | 404 | 2.20 | 7.18 | 3.14 |

"IPI" stands for Import Price Index

Table 2 Summary of categorical variables

| Trade type (%) | Trade payment status (%) | Default (%) | Trade band (%) |
|--------------------|---------------------------|-------------|----------------|
| Whole ledger: 12.8 | Debtor paid: 88.4 | No: 98.3 | 1: 13.3 |
| Multi debtor: 5.5 | Seller repurchased: 11.0 | Yes: 1.7 | 2: 10.5 |
| Standard: 80.9 | Repurchase-requested: 0.4 | | 3: 10.1 |
| License fee: 0.8 | Overdue: 0.1 | | 4: 15.3 |
| | Partially paid: 0.04 | | 5: 23.7 |
| | | | 6: 11.6 |
| | | | 7: 9.0 |
| | | | 8: 3.5 |
| | | | 9: 2.0 |
| | | | 10: 1.1 |

missing. As such, it was necessary to study the extent the available information allowed for an analysis of the probability of default and consequently whether investors could accurately assess risk, which can be done in other crowdlending platforms that do not restrict this type of information.

Table 10 lists all the variables used in the present study. In addition to the invoice variables (advance rate, annualized gross yield, total face value, days) sourced from the Kriya database, a set of macroeconomic variables (GDP, exports, and import price index) were considered for the model fitting, while the remaining variables were only used in the descriptive analysis.

Descriptive analysis

A summary of the variables is presented in Tables 1 and 2. Considering the mean values, a typical invoice was funded at a 76% rate, with an annualized gross yield of 11% and a maturity of 55 days. The platform advance rate ranged from 2.5% to the total value of the invoice, with yields reaching as high as 49% and maturities of over a year in some cases. However, the invoice amount varied from a few thousand to several million GBP, indicating that the platform financed a wide range of businesses in terms of income. The total crystallized losses were very low, with a mean absolute value of 108 GBP, mainly because of the low default rate of 1.7%. Regarding macroeconomics, the UK has a steady mean GDP growth rate of 0.89%. Exports experienced a slightly higher growth rate of

Table 3 Correlation matrix

| | AR | AGY | TFV | Days | GDP | Exports | IPI |
|---------|----------|----------|----------|----------|----------|----------|----------|
| AR | 1 | -0.17*** | 0.01 | -0.17*** | 0 | 0 | -0.13*** |
| AGY | -0.17*** | 1 | -0.08*** | -0.22*** | 0.06*** | -0.11*** | -0.01 |
| TFV | 0.01 | -0.08*** | 1 | 0.03*** | -0.01 | -0.03*** | 0 |
| Days | -0.17*** | -0.22*** | 0.03*** | 1 | -0.02*** | 0.03*** | 0.06*** |
| GDP | 0 | 0.06*** | -0.01 | -0.02*** | 1 | 0.09*** | 0.08*** |
| Exports | 0 | -0.11*** | -0.03*** | 0.03*** | 0.09*** | 1 | -0.36*** |
| IPI | -0.13*** | -0.01 | 0 | 0.06*** | 0.08*** | -0.36*** | 1 |

This table reports Spearman’s correlation coefficients with the significance of the correlation test. Advance rate (AR) is the percentage of an invoice’s face value that a factor pays upon its purchase. Annualized gross yield (AGY) is the actual yearly rate of return accrued from factoring, considering the effect of compounding interest. Total face value (TFV) is the total amount of the invoice (in GBP). Days, or maturity, is the number of days elapsed from the advance date to the expected pay date. GDP, Exports and IPI are the gross domestic product, level of exports, and import price index in the UK, respectively. Trade type and band are categorical variables, and we exclude them from the correlation matrix. *Significant at 10%; ** significant at 5%; *** significant at 1%

1.86%, and the import price index did not change significantly, considering the mean values. Regarding the categorical variables, the most common trade operation was the standard discount (80.9%), wherein the platform discounted a single invoice with negotiated conditions for the seller, followed by the entire ledger (12.8%), wherein the sellers had a special agreement with the platform to discount their entire portfolio of invoices. This is probably why some of the funded invoices had very low face values and maturity. Most loans were traded in Bands 5 (23.7%), 4 (15.3%), and 1 (13.3%), which may be due to the industry in which the borrower is positioned. In addition, 88% of these loans were in “repaid” status and only 11.4% were repurchased, while other statuses were virtually non-existent.

Table 3 presents the correlation matrices between variables. In general, the data did not exhibit multicollinearity. Exports and the import price index were the most noticeable pairs of correlated variables. When the import price index increased, the production costs of companies increased, which in turn affected their competitiveness and reduced their export levels, thus creating a negative correlation. Maturity was also negatively correlated with advance rate and annualized gross yield. In the first case, the platform may have been trying to reduce the risk involved in a higher fund rate by providing a lower maturity period, limiting investors’ exposure to it. In the second case, because high interest rates with larger periods of maturity result in greater costs for the seller, many of these loans were repaid as soon as possible, sometimes before the expected payment date. This allows for invoices with long periods of maturity and low interest rates to be obtained, mainly because they are repaid earlier than expected. However, those repaid over longer timeframes usually have lower interest rates to compensate for the costs incurred. Financially healthy companies can do this if they are not overcharged interest when assuming a longer period of maturity, thus creating this negative relationship. The platform’s correct credit risk assessment is also supported by the fact that the advance rate and annualized gross yield are negatively correlated, such that higher fund rates are granted with lower interest rates to financially healthy companies, and vice versa.

For outliers, the dataset was primarily affected by the invoice variables, as shown in Fig. 1. It contains several extreme outliers that must be removed from the training

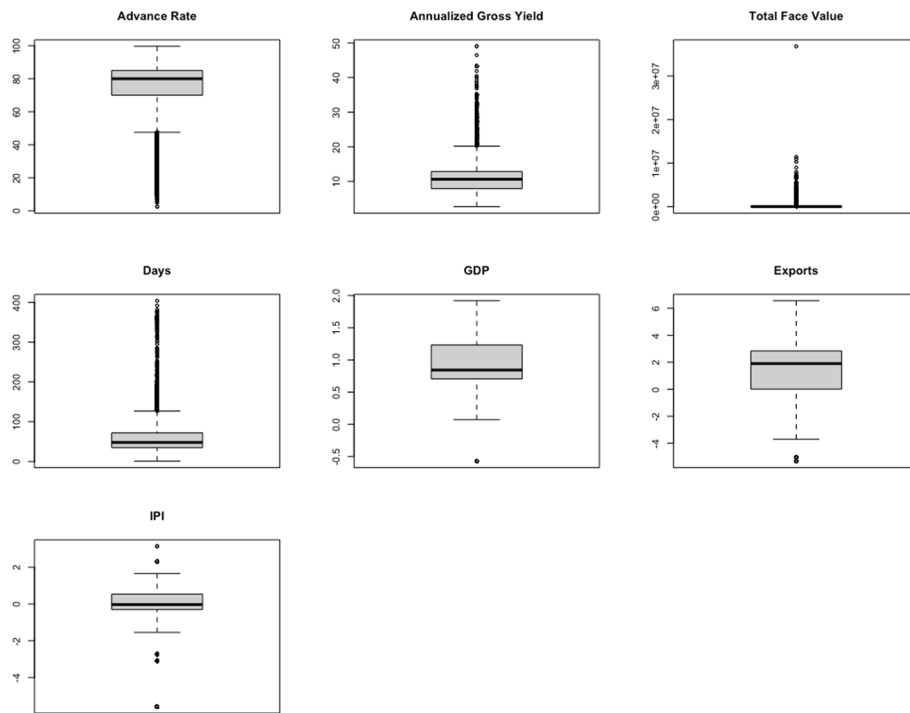


Fig. 1 Boxplots of the main variables considered in the analysis. *Note:* Trade type and trade band are categorical variables; therefore, we exclude them from the boxplot analysis. The definitions of the variables are listed in Table 3

dataset because logistic regression might be very sensitive to these outliers. Therefore, no bias was generated during model fitting. Nonetheless, the test dataset remained untouched; therefore, realistic predictions could be made using any type of data point. Consequently, 10% top and bottom winsorization of the training data was performed for model fitting.

Pre-processing and sampling

First, the data were cleaned by rejecting missing or irrelevant observations. Loans pending repayment were discarded. Therefore, only those that had been completed were considered because they had been fully paid, partially recovered, or completely lost. Two levels² from the categorical variable trade type were excluded because they had very low representation in the dataset. Given the divergence in variable scales, we standardized the variables by subtracting their mean and dividing by their standard deviation.

Following these adjustments, feature selection was performed using the information gain ratio and chi-squared filters. Given a training set S that is partitioned into V subsets S_1, \dots, S_V according to V different values of feature X , the mutual information of features X and class Y was defined based on Kotsiantis and Pintelas (2004, p. 50) as

$$Gain(X) = info(S) + \sum_{v=1}^V \frac{|S_v|}{|S|} info(S_v) \tag{1}$$

² Jr. Supply Chain Finance and Purchase Order.

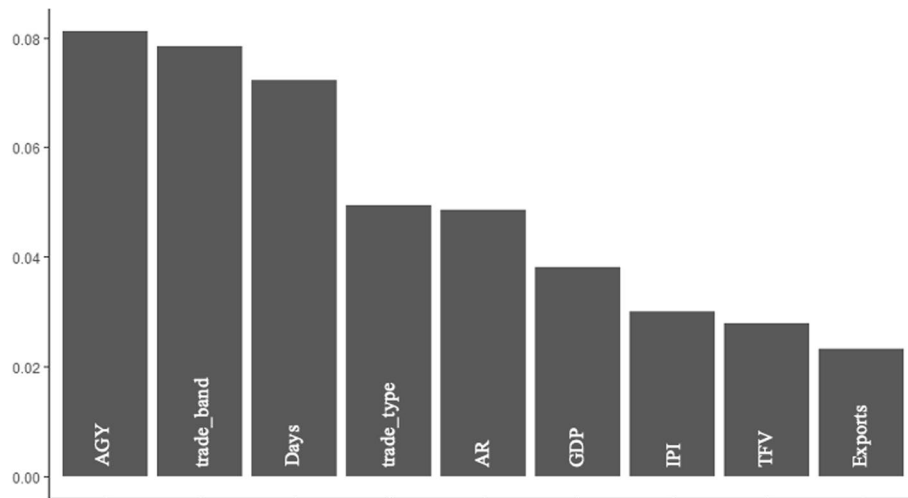


Fig. 2 Feature selection. *Note.* The results were obtained using the information gain ratio and chi-square filters. The definitions of the variables are listed in Table 3

However, the information-gain filter is strongly biased toward features with different values. This can be corrected using the following calculation, which represents the potential information gain by partitioning S into V subsets.

$$\text{splitinfo}(S) = - \sum_{v=1}^V \left(\frac{|S_v|}{|S|} \right) \log_2 \left(\frac{|S_v|}{|S|} \right) \quad (2)$$

The information gain ratio expresses the proportion of information gained by the partition.

$$\text{Gain ratio}(X) = \frac{\text{Gain}(X)}{\text{split info}(X)} \quad (3)$$

The chi-square filter X_c^2 , measures the divergence of the feature distribution by comparing the observed and expected values.

$$X_c^2 = \sum_{i=0}^N \frac{(O_i - E_i)^2}{E_i} \quad (4)$$

where c is the number of degrees-of-freedom, O_i is the observed value, and E_i is the expected value. Statistics were compared using a chi-square table.

All the metrics above reported similar results. The variables that were not significant were excluded from the analysis. Those that were finally included in the model development are shown in Fig. 2.

The data were split into training and testing sets with a 75–25% ratio, respectively. The test dataset collected only loans from 2019, whereas the training dataset gathered a wider selection of older data to ensure intertemporal validation of the models (Lau 1987; Serrano-Cinca et al. 2015). Outliers were handled in the training dataset only by winsorizing the top and bottom 10% of the invoice variables, as mentioned previously. The target variable of this study represented loan status and had two classes: “Yes” if

the loan was default and “No” if the loan was not in default. The positive class, also known as the minority class, represented 1.7% of the total loans, while the negative class, also known as the majority class, accounted for 98.3%.

The dataset suffered from a severe class imbalance problem that would have biased the predictive models towards the majority class (Bastani et al. 2019). “A classifier derived from an imbalanced dataset typically has a low error rate for the majority class and an unacceptable error rate for the minority class” (Kotsiantis and Pintelas 2004, p.48). However, in this case, the misclassification cost for the minority class was higher than for the majority class. That is, the outcome of accepting a defaulting loan was higher than that of rejecting a non-defaulting loan because the former assumes an actual loss, whereas the latter implies an opportunity cost. To resolve the class distribution issue, three well-known sampling techniques—random undersampling, random oversampling, and SMOTE— were applied and compared.

The random under-sampling technique balances classes by randomly eliminating observations from the majority class. This technique can eliminate potentially useful information from the analysis, whereas in random oversampling, the replication of observations in the minority class increases the likelihood of overfitting (Kotsiantis and Pintelas 2004). In both techniques, a parameter controls the final ratio of the classes. Finally, in the synthetic minority oversampling technique, new observations are not replicated but are synthetically derived from the original observations in the minority class. By means of an imaginary segment, these are created by connecting each minority class observation with K nearest neighbors (K -NN). K -NN calculates the distance between a current observation and all other observations by selecting the K^{th} nearest observation with the least distance. This study involved both categorical and numerical data, thus the Gower distance (Gower 1971) was used and estimated by considering two individuals— i and j —compared to a character, k . An S_{ijk} score of zero was assigned when they differed, and a positive fraction or unity when they had some degree of similarity. The possibility of making comparisons can be represented by a quantity, δ_{ijk} , which is 1 when the character k can be compared for i and j or 0 otherwise. The similarity between i and j , S_{ij} , can be expressed as the average score obtained for all possible comparisons.

$$S_{ij} = \frac{\sum_{k=1}^p S_{ijk}}{\sum_{k=1}^p \delta_{ijk}} \quad (5)$$

After obtaining K nearest neighbors, the distance between the observation under consideration and its nearest neighbors is multiplied by a random number between 0 and 1, which is added to the original observation (Chawla et al. 2002) and which “causes the selection of a random point along the line segment between two specific features,” (Chawla et al. 2002, p. 328).

The equation below (Zhu et al. 2019) represents the SMOTE calculation, where x_i is the original observation, x_n is the K th-nearest neighbor, $R \in \{0, 1\}$ is a random number, and x_{new} is the resulting artificial observation:

$$x_{new} = x_i + R(x_n - x_i) \quad (6)$$

To select the perfect ratio between the minority and majority classes, multiple scenarios were tested using logistic regression as a reference and the F1-score as the maximization target. Specifically, logistic regression with SMOTE was evaluated using the five nearest neighbors at a rate of 29 with different balancing rates (ranging from 1 to 50%). The F1-score was higher at a balancing rate of approximately 10% and 36%, although the latter reported a considerably higher sensitivity.

Typically, in an imbalanced dataset, accuracy is not a desirable metric for model comparisons (Bastani et al. 2019). Therefore, this study focused on precision³ and sensitivity⁴; the former penalizes the misclassification of negatives, whereas sensitivity is only to the detriment of the misclassification of positives. The F1-score is the harmonic mean of both metrics and gives them equal importance, thus influencing our decision to use it for model parameter optimization and final selection between models.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1 - score} = \frac{2TP}{2TP + FP + FN} \quad (9)$$

where TP is the actual number of positives, FP the number of false positives, and FN the number of false negatives.

For easier comparison with other studies, other commonly used metrics were also provided (specificity,⁵ accuracy, and McFadden's R^2).

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (10)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$R_{McFadden}^2 = 1 - \frac{\log(L_c)}{\log(L_{null})} \quad (12)$$

where TN is the number of true negatives, L_c is the (maximized) likelihood value from the current fitted model, and L_{null} is the corresponding value for the null model (with only an intercept and no covariates).

As previously mentioned, different balancing rates were considered using logistic regression and SMOTE (using the five nearest neighbors at a rate of 29). The best results for F1-score and sensitivity were obtained at a balancing rate of 36%. To provide a fair

³ *Precision or Positive predictive value* (PPV) is the number of true positive results divided by all positive results (including misclassifications).

⁴ *Sensitivity or recall* is the proportion of positives that are correctly classified (true positive rate).

⁵ *Specificity* is the proportion of negatives that are correctly classified (true negative rate).

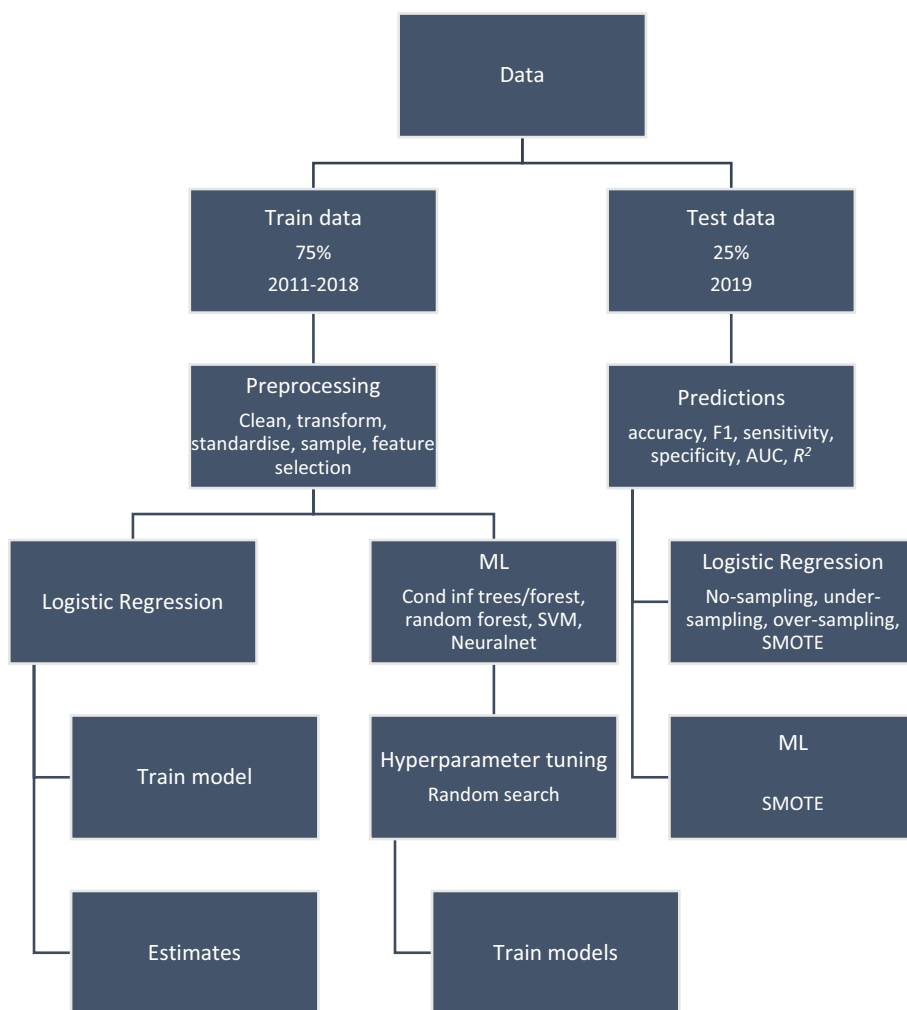


Fig. 3 Summary of the research procedure

comparison between the undersampling and oversampling techniques, we used rates of 0.345 and 29 for the former and latter, respectively. At these rates, the three sampling techniques yielded three datasets, all with a minority class ratio of 36%. These datasets were used to evaluate the model performance of the sampling techniques with the logistic regression (Table 5). After the entire preprocessing, the final training dataset contained 1,830 observations with the undersampling technique and 53,064 observations with the oversampling techniques. The experimental design is illustrated in Fig. 3.

Predicting invoice-lending default

Logistic regression

This section analyzes the determinants of default using a binary logistic regression model, which compares the relationship between predictor variables, $X = \{x_1 \cdots x_i\}$, and the categorical response variable, $Y = \{0, 1\}$. Predicted scores ($\hat{\pi}$) and observed probabilities (Y) can also be compared using the “maximization of a log-likelihood” (LL) function.

If:

$$\pi_i = Pr(Y = 1|X = x_i) \quad (13)$$

then:

$$odds(Y = 1) = \frac{\hat{\pi}(Y = 1)}{1 - \hat{\pi}(Y = 1)} \quad (14)$$

$$= \exp(\hat{b}_0 + \hat{b}_1 X_1 + \dots + \hat{b}_i X_i) \quad (15)$$

By logit transformation:

$$logit(\hat{\pi}) = \ln\left(\frac{\hat{\pi}(Y = 1)}{1 - \hat{\pi}(Y = 1)}\right) \quad (16)$$

$$= \hat{b}_0 + \hat{b}_1 X_1 + \dots + \hat{b}_i X_i \quad (17)$$

By algebraic manipulation:

$$\hat{\pi} = \frac{e^{B_0 + B_1 X_1 + \dots + B_i X_i}}{1 + e^{B_0 + B_1 X_1 + \dots + B_i X_i}} = \frac{1}{1 + e^{-(B_0 + B_1 X_1 + \dots + B_i X_i)}} \quad (18)$$

Maximum log-likelihood estimation:

$$LL = \sum_{i=1}^N [Y_i \ln(\hat{\pi}_i) + (1 - Y_i) \ln(1 - \hat{\pi}_i)] \quad (19)$$

Table 4 presents the coefficient estimates, standard errors, and significance of the logistic model. Unstandardized coefficients were used to study the direct effects of the variables on default status and standardized coefficients to rank the predictors. These results show that annualized gross yield, advance rate, and maturity were significant in determining default, in line with Turiel and Aste (2020) for the P2P market and Dorfleitner et al. (2017) for the P2B market. Gross yield has a positive effect on the probability of default, which is consistent with how borrowers accepting higher interest rates are able to recognize lower creditworthiness (Stiglitz and Weiss 1981; Dorfleitner et al. 2017). Furthermore, a negative relationship exists between the advance rate and default, which was also discovered by Dorfleitner et al. (2017), showing that the advance rate can be an effective mechanism to dissuade debtors the platform considers riskier to default. Furthermore, the nominal value was negatively related to default; thus, lower face-value invoices were more likely to be unpaid. As in the previous case, the platform correctly assesses risk by granting higher capital to companies that are less likely to default, thereby creating a negative relationship. However, maturity is positively linked, increasing the probability of default in long-term transactions, similar to Dorfleitner et al. (2017). Trade type is exceedingly significant in determining default when it is a standard or multi-debtor transaction. A strong positive relationship exists between these types of transactions and defaults. Furthermore, GDP and exports have positive and significant relationships with default. This might have been counterintuitive initially

Table 4 Logistic regression estimations for probability of default

| Coef. | Unstandardized | Standardized |
|------------------------|----------------|--------------|
| Intercept | −6.907 *** | −6.944 *** |
| Advance rate | −0.023 *** | −0.297 *** |
| Annualized gross yield | 0.129 *** | 0.578 *** |
| Total face value | −0.077 ** | −0.661 *** |
| Days | 0.014*** | 0.409 *** |
| GDP | 0.239 ** | 0.110 ** |
| Exports | 0.046 *** | 0.127 *** |
| IPI | 0.003 | 0.005 |
| Trade_typeMultiDebtor | 3.231 *** | 3.241 *** |
| Trade_typeStandard | 2.960 *** | 2.932 *** |
| Trade_typeLicenceFee | −0.191 | −0.238 |
| Trade_band2 | 0.210 | 0.218 |
| Trade_band3 | −0.176 | −0.190 |
| Trade_band4 | −0.507 * | −0.518 * |
| Trade_band5 | 0.006 | −0.012 |
| Trade_band6 | 0.134 | 0.111 |
| Trade_band7 | −0.314 | −0.333 |
| Trade_band8 | −0.227 | −0.254 |
| Trade_band9 | −0.937 | −0.967 * |
| Trade_band10 | −2.116** | −2.125 ** |

This table reports the results of the logistic regression estimates for the model presented at the beginning of this section. The advance rate is the percentage of an invoice's face value that a factor pays for a purchase. Annualized gross yield is the actual yearly rate of return earned on factoring, considering the effects of compounding interest. Total face value (with log transformation) is the total number of invoices (in GBP). Days, or maturity, is the number of days elapsed from the advance date to the expected pay date. GDP, Exports and IPI are the gross domestic product, level of exports, and import price index in the UK, respectively. Trade type reflects the type of operation (standard, multi-debtor, whole ledger, or license fee). The trade band reflects the borrower's industry. The intercept was constant. The total number of observations was 46,761. The McFadden's R^2 is 0.072 and 0.073, respectively. *Significant at 10%; ** significant at 5%; *** significant at 1%

Table 5 Logistic regression prediction on a balanced minority class ratio of 36%

| | Accuracy (%) | F1-score | Sensitivity (%) | Specificity (%) | AUC (%) | McFadden's- R^2 |
|--|--------------|----------|-----------------|-----------------|---------|-------------------|
| Logistic regression (without sampling) | 97.4 | 0.043 | 5.1 | 98.5 | 72.9 | 0.073 |
| Logistic regression (under-sampling) | 73.3 | 0.047 | 56.1 | 73.5 | 73.0 | 0.132 |
| Logistic regression (over-sampling) | 75.5 | 0.047 | 53.2 | 75.7 | 73.1 | 0.128 |
| Logistic regression (SMOTE) | 76.0 | 0.051 | 55.9 | 76.2 | 73.2 | 0.137 |

The mean results comprise over 200 iterations for the random sampling procedure. McFadden's pseudo- R^2

because positive economic cycles are associated with a higher rate of default. Nonetheless, because lending standards usually fluctuate with the economic cycle, in stable cycles, lending standards may decrease, making it easier to access credit even for borrowers with higher probabilities of default. As for the trade bands, 10 was highly significant, while 4 and 9 were only significant at the 10% level. Considering the standardized coefficients, the following variables were ranked as the most important based on the size of their effect: trade type, multi-debtor and standard, trade band 10, total face value, annualized gross yield, and maturity.

For the model predictions, Table 5 shows the results in terms of performance metrics with a 36% balanced minority class ratio. Better models were achieved with greater balance at a class distribution of less than 50%. Moreover, increasing these to higher ratios can bias the results towards very high sensitivity rates to the detriment of poor specificity. The data were randomly sampled, thus the mean of 200 iterations was calculated.

In general, a high level of overall accuracy was obtained when no sampling techniques were used; however, the resulting sensitivity rate was not acceptable. A clear difference existed between the results using sampling techniques and those that did not, especially in terms of sensitivity, where the former correctly classified 50% more loan defaults. All sampling techniques reported similar results, although SMOTE slightly outperformed the others for almost all metrics, which is in accordance with Bastani et al. (2019). It correctly classified 56% of the loans in default, while preserving its capacity to do so for 76% of the NPLs. A relatively good area under the curve (AUC⁶) score of 73.2% was achieved using SMOTE: higher than that of any of the alternative models used, as shown in Table 5 and Fig. 5.

Predicting default with machine learning alternatives

This section analyzes the performance of alternative models in predicting the default rate in the invoice-lending segment. Our objective was to evaluate the precision of alternative logistic regression techniques, such as conditional inference trees, random forests, support vector machines, and neural networks. Logistic regression reported a higher F1-score on the SMOTE dataset, thus it was used for alternative model fitting. All models were validated using the same data split as in the logistic regression (75% training—25% testing); however, the package used for modelling⁷ (Bischl et al. 2016) allowed for tenfold cross-validation within the training dataset, which we used to validate the models. This procedure splits the training dataset into ten folds where each fold is excluded iteratively and used to validate the performance of the models against them. Hyperparameter optimization was performed using a random search, with the objective of maximizing the F1-score (see Table for the hyperparameter selection).

Conditional inference tree

Decision trees are nonparametric supervised learning algorithms that do not require special assumptions, making them versatile. Their main characteristic is a feature space, which is recursively partitioned by grouping observations that have similar response values (Strobl et al. 2009). In other words, they attempt to describe the conditional distribution of the response variable Y , given a set of m covariates X , restricting the feature space X of the covariates to r disjoint nodes B_1, \dots, B_r , where $X = \bigcup_{k=1}^r B_k$ (Hothorn et al. 2006):

$$D(Y|X) = D(Y|f(X_1, \dots, X_m)) \quad (20)$$

⁶ The AUC is a measure of a model's ability to distinguish between classes, where a higher score indicates better predictive performance in classification.

⁷ R's mlr package.

The algorithm selects the variable from the covariate vector \mathbf{X} with the strongest association with \mathbf{Y} , searches for the best split point, and splits the variable into r disjointed nodes. As the splitting process continues, the level of purity⁸ is observed. “In each node, the variable that is most strongly associated with the response variable (i.e., that produces the highest impurity reduction or the lowest p value) is selected for the next split (Strobl et al. 2009, p. 327). For conditional inference trees, p values were used instead of entropy measures. In this study, the association between \mathbf{Y} and X_j , where $j = 1, \dots, m$, was measured using the Bonferroni test, which formulated a global hypothesis of independence in terms of m partial hypothesis, $H_0^j : D(Y|X_j) = D(Y)$. The variable with the lowest p value was selected. When insufficient evidence existed to reject H_0 , at a pre-specified level α , the recursion was halted (Hothorn et al. 2006). Permutation tests were performed because the distribution $D(Y|X_j)$ is usually unknown. The split-point A^* was obtained with a test statistic maximized over all possible A subsets (Hothorn et al. 2006), which measured the discrepancy between $\{Y_i|X_{ji} \in A\}$ and $\{Y_i|X_{ji} \notin A\}$.

Random forest

“A random forest is a classifier consisting of a collection of tree-structured classifiers $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$, where $\{\Theta_k\}$ are independent identically-distributed random vectors and each tree casts a unit vote for the most popular class at input \mathbf{x} ,” (Breiman 1996, p. 6). Random forests try to predict the response variable with a predictor $\varphi(x, \mathcal{L})$ where x is the input vector and \mathcal{L} is a learning dataset. To obtain a better predictor, repeated bootstrap samples are taken from the learning dataset \mathcal{L}^B , each of them consisting of N cases drawn at random with replacement, forming a sequence of predictors $\varphi(x, \mathcal{L}^B)$ (Breiman 1996, p. 123). Each predictor forms a decision tree that votes for the most popular class (mode) at input \mathbf{x} . These procedures are known as bootstrap aggregation or bagging. In addition, random forests use feature bagging, in which a random subset of features x is selected for each built tree. Decision trees are usually correlated with strong predictors, thus feature bagging can reduce this correlation, making random forests more accurate. Conditional random forests are a special type of random forest wherein the trees used for bagging are conditional inference trees.

Support vector machine

In support vector machines, the input vectors are nonlinearly mapped to a high-dimensional feature space, where a hyperplane or a set of hyperplanes is constructed to provide a linear decision function with a maximal margin between the vectors of the two classes (Cortes and Vapnik 1995).

Using a set of labelled training patterns:

$$(y_1, \mathbf{x}_1), \dots, (y_k, \mathbf{x}_k) \quad y_i \in \{-1, 1\} \tag{21}$$

are linearly separable if there exists a vector \mathbf{w} and scalar b such that:

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \quad \text{if} \quad y_i = 1 \tag{22}$$

⁸ The number of observations with a majority for a response class that becomes isolated.

$$w \cdot x_i + b \leq -1 \quad \text{if } y_i = -1 \tag{23}$$

The optimal hyperplane that separates the data with maximal margin is given by:

$$w_0 \cdot x + b_0 = 0 \tag{24}$$

$$w_0 = \sum_{i=1}^k y_i a_i^0 x_i \tag{25}$$

where $a_i^0 > 0$ and $\Lambda_0^T = (a_1^0 \dots a_t^0)$ form a vector of parameters.

and the Lagrangian problem that needs to be minimized is:

$$\mathcal{L} = \sum_{i=1}^k \alpha_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j y_i y_j x_i \cdot x_j \tag{26}$$

A kernel function is often used to transform the feature space when the data cannot be split by a hyperplane without error, which occurs when points overlap.

Artificial neural networks

The basic elements of an artificial neural network are the nodes that represent the neurons of the biological brain. These nodes were connected and received stimulation signals from the input variables. This is not performed directly but with weights and activation functions. The neuron output signal O is expressed by the following relationship (Abraham 2005):

$$O = f(net) = f\left(\sum_{j=i}^n w_j x_j\right) \tag{27}$$

where $w = \{w_1 \dots w_n\}$ is the weight vector, $x = \{x_1 \dots x_n\}$ and the function $f(net)$ is referred to as an activation (transfer) function. The variable net is defined as the scalar product of the weight and input vectors.

$$net = w^T x = w_1 x_1 + \dots + w_n x_n \tag{28}$$

where T is the transpose of the matrix and the output value O is computed as

$$O = f(net) = \begin{cases} 1 & \text{if } w^T x \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

where θ is the threshold level. If the activation function reaches the threshold level, the signal is transmitted to the connected node. Figure 4 illustrates the general structure of an artificial neural network.

Our network architecture was similar to that of the extreme-learning machine used by Pang et al. (2021), differing only in the learning method. Specifically, nine input variables existed (Fig. 2) with their respective weights, one hidden layer, and one output node (default). A logistic activation function was used: resilient backpropagation with weight backtracking as the learning algorithm and cross-entropy as the error function.

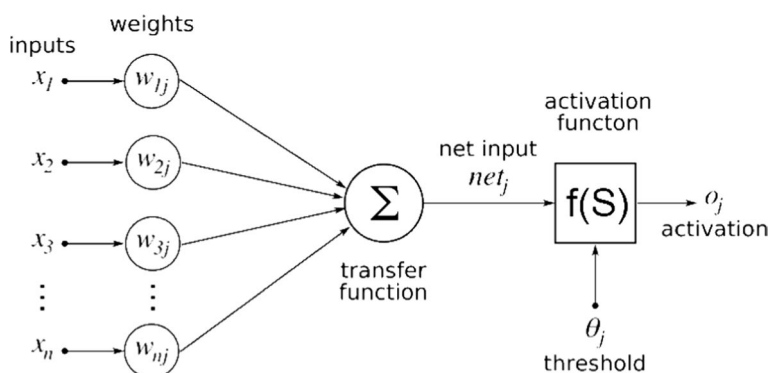


Fig. 4 Illustration of an artificial neural network. Note. Image by Geetika Saini, Artificial neural network, CC BY-SA 4.0

Table 6 Performance measures for a balanced minority class ratio of 36%

| | Accuracy (%) | F1-score | Sensitivity (%) | Specificity (%) | AUC (%) |
|-----------------------------|--------------|----------|-----------------|-----------------|---------|
| Cond. inference tree | 63.6 | 0.026 | 41.6 | 63.9 | 52.8 |
| Cond. inference tree forest | 75.4 | 0.023 | 25.5 | 76.0 | 47.4 |
| Random forest | 50.6 | 0.020 | 43.1 | 50.7 | 45.6 |
| SVM | 88.4 | 0.045 | 24.1 | 89.1 | 55.0 |
| Neuralnet | 85.5 | 0.053 | 35.8 | 86.0 | 71.7 |

Hyperparameters were searched only for the conditional inference trees and applied to the forests. Results on conditional inference trees may have varied as they have high variance by default. The forests were built with 200 trees each

Benchmark predictions

Table 6 presents the results for the performance measures. Apart from the neural network, the alternative models did not show any skill, with an AUC score below or slightly above 50% (Fig. 5). Using the DeLong test to compare the significance between the AUC scores, we developed the following two-sided contrast based on DeLong et al. (1988):

$$\begin{cases} H_0 : AUC_p = AUC_k \\ H_1 : AUC_p \neq AUC_k \end{cases}$$

where p and k represent the AUC of the two models. Table 7 presents the estimates and significance levels for the scores. Our results showed that the AUC score of the logistic regression was statistically significant in all models. The neural network also showed significance at the 5% and 10% levels. However, in contrast to the logistic regression, greater significance was observed in favor of the latter at the 5% level. The other methods did not show statistical significance.

The conditional inference tree forest and SVM lacked sensitivity to the detriment of greater specificity, whereas the conditional inference tree and random forest had acceptable sensitivity rates but poor specificity. The neural network is a more balanced model and has a greater F1-score and AUC percentage. Indeed, it achieved the greatest F1-score of all the models assessed, including logistic regression. A greater specificity rate and overall accuracy existed than in logistic regression. However, the AUC was slightly lower and predicted 20% fewer defaulted loans. These results experienced some misclassifications, one of particular importance being that some of the

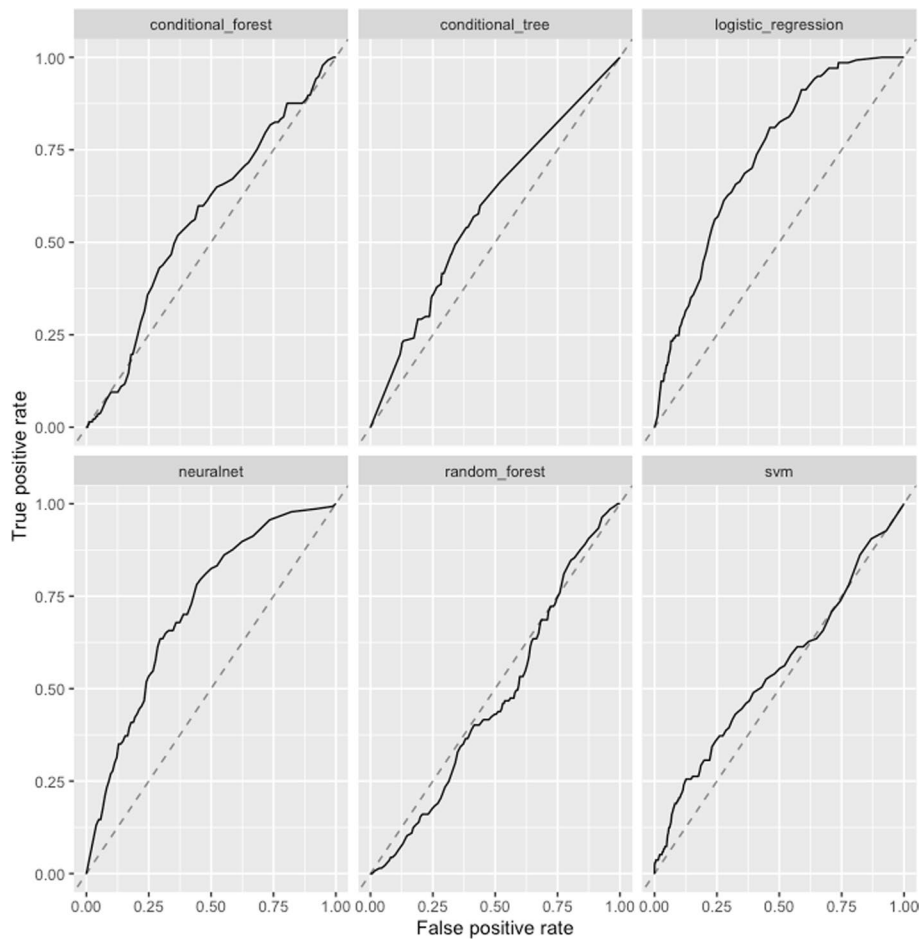


Fig. 5 ROC curve comparison

Table 7 DeLong test estimates and significance of AUC scores

| | Log. Reg | CT | CF | R. Forest | SVM | NN |
|-----------|-----------|----------|----------|-----------|----------|----------|
| Log. Reg | – | 3.609*** | 3.159*** | 3.488*** | 3.473*** | 2.311** |
| CT | –3.609*** | – | –0.271 | 0.463 | –0.223 | –2.222** |
| CF | –3.159*** | 0.271 | – | 0.576 | 0.014 | –1.731* |
| R. Forest | –3.488*** | –0.463 | –0.576 | – | –0.725 | –2.114** |
| SVM | –3.473*** | 0.223 | –0.014 | 0.725 | – | –1.792* |
| NN | –2.311** | 2.222** | 1.731* | 2.114** | 1.792* | – |

This table states the estimates and significance levels of the De Long test between paired AUC scores. Log. Reg: logistic regression. CT: conditional inference tree. CF: conditional inference tree forest. R. Forest: random forest. SVM: support vector machine. NN: neuralnet. *Significant at 10%; ** significant at 5%; *** significant at 1%

non-default loans were classified as default, which negatively impacted specificity rates. This was due to a trade-off between sensitivity and specificity, which affected all the models. However, the greater detection of default loans is more than compensated for by the decrease in specificity rates.

Table 8 Annualized gross yield and implicit and observed probabilities of default mean evolution

| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|------|
| AGY | 21.80 | 16.95 | 16.60 | 15.13 | 11.56 | 11.44 | 11.18 | 10.52 | 8.90 |
| IPD | 7.45 | 6.39 | 6.66 | 6.51 | 6.66 | 7.13 | 7.43 | 7.57 | 6.37 |
| OPD | 4.50 | 4.06 | 1.79 | 3.29 | 1.80 | 2.16 | 1.83 | 1.23 | 1.15 |

AGY denotes annualized gross yield. IPD is implicit probability of default. OPD is the observed probability of default

Pricing mechanism with implicit probability of default

This section evaluates the pricing of the platform by comparing the implicit and observed probabilities of default. Many factors can affect invoice prices. However, only the platform knows the extent each is considered. The Discussion section revisits this debate. The price paid by the seller is assumed to exclusively represent the borrower's probability of default, thus the valuation of an invoice is based on the expected payoff and can be expressed as

$$P(1 + i_t) = rC \times PD_t + C \times (1 - PD_t) \quad (30)$$

where, P is the price, i is the interest rate, r is the recovery rate, C is the total advance amount of the invoice, and PD is the probability of default.

By algebraic manipulation, the implicit probability of default in the price is estimated as

$$PD_t = \frac{C - P(1 + i)}{C(1 - r)} \quad (31)$$

This is the borrower's probability of default, intrinsic to the price charged to the seller.

Table 8 compares the implicit and observed probabilities of default. Implicit default experienced a slight uptrend and is consistently higher than the observed default for each year. This trend remains even with a lowering of the observed probability of default in later years. Notably, the implicit probability of default does not decrease like the observed default. Explanations for this lowering observed probability of default may be higher experience in loan selection, which may have led to a higher proportion of non-performing loans gradually being rejected; the increase in business volume, because the higher the number of observations, the higher the convergence of the default rate towards a lower level (law of large numbers); and the positive economic cycle experienced during this period (2011–2019). These results may indicate a low correlation between the implicit probability of default and observed default, with a Pearson correlation coefficient of 0.31. The Mann–Whitney–Wilcoxon test was used to compare the two medians, given that they did not follow a normal distribution and were not paired samples. Furthermore, the medians of the two samples were compared as follows:

$$\begin{cases} H_0 : \text{implicit}PD = \text{observed}PD \\ H_1 : \text{implicit}PD \neq \text{observed}PD \end{cases}$$

The statistics were given by:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 \quad (32)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 \quad (33)$$

$$U = \min\{U_1, U_2\} \quad (34)$$

where n_1 and n_2 are the sample sizes; R_1 and R_2 are the rank sums and U_1 and U_2 are the statistics of samples 1 and 2, respectively.

The Mann–Whitney–Wilcoxon test reported a p -value of $4.1e-05$, which led us to reject the null hypothesis of equal medians between samples at the 1% significance level. By computing the mean difference between the two probabilities of default, implicit default was four times higher than observed default. There is evidence to consider this difference significant, especially if the platform did not consider its own probability of default, the probability of default of the seller, or if some of these transactions were guaranteed by insurance. Given that the borrower's probability of default is the most relevant factor to consider when pricing an invoice and implicit default did not decrease in the same fashion as the observed default, these results may indicate that sellers pay a premium not realistically adjusted to the real probability of default. An explanation for paying such a premium may be that many of these borrowers must deal with credit rationing (Cowling 2010; Lee et al. 2015) in traditional credit markets, either by not having access to credit or having it at an even higher interest rate. In addition, the high flexibility of this lending method, which can remotely finance an invoice within a few hours, could explain why a higher price is charged. Hence, the platform may take advantage of this situation by charging a premium based on market conditions rather than on operational risk, leading to an imbalance between price and actual risk.

Discussion of results

Research on the prediction of default for invoice trading platforms is a novel topic (Dorfleitner et al. 2017) which is investigated in this study. However, unlike previous papers, the current study employed artificial intelligence techniques and sampling combined to enhance the accuracy and effectiveness of predictions. Current results show that these techniques can effectively improve the detection of defaults by up to 56% while maintaining levels of specificity above 70%. Not addressing the problem of an imbalanced dataset results in models with very high overall accuracy and specificity rates but unacceptable sensitivity rates. Therefore, by applying sampling techniques, the robustness of our models increased since they helped the learning process with artificial data. This in turn improved the reaction to unseen data. Once intertemporal validation was ensured, another layer of robustness was introduced by using predictions made with a 1-year lag. Feature selection in the preprocessing stage allowed for the construction of models with relevant factors determining the default,

which creates less complex models that are easier to interpret and less prone to overfitting, as well as being more robust. Of the models evaluated, the neural network was the best in terms of the F1-score, whereas the logistic regression had a similar score with 20% more sensitivity, which could be more appropriate from an investor's perspective. Bearing this in mind, limited disclosure on the Kriya platform prevents the advantages of more advanced techniques from being employed, and asymmetric information from being reduced. This negatively affects the performance of any model trying to predict default probability and limits the possibility of investors correctly assessing credit risk. Disclosing the borrower's and seller's financial information on the invoice, as well as information on the protection provider, is essential to permit a complete assessment of transaction risk.

Furthermore, the current study of the platform's credit pricing with an implicit probability of default indicates a consistent discrepancy between the observed and implicit probabilities of default. The price charged to the sellers represented a probability of default four times higher than the observed probability of default, and the differences between the two were statistically significant. Consequently, this study assessed whether the invoice price was realistically adjusted to the borrower's probability of default, or whether the platform was overcharging sellers for other reasons. Therefore, standard valuation models for invoice lending usually consider a borrower's probability of default as the only relevant event (Nava et al. 2019). This is a simplification of reality since the price that the seller pays reflects not only that but also the probability of default of the seller (Nava et al. 2019), the value of a flexible method (where an invoice can be sold in just a few hours [Dziuba 2018]), and situations wherein sellers with lower creditworthiness may encounter credit rationing or higher interest rates in traditional commercial credit (Li 2016). The price also reflects the operational risk of this form of financing (Liang et al. 2022). Factors contributing to the apparent discrepancy may include risks associated with invoice factoring such as fraud, the authenticity of sold invoices (fraudulent activities from debtors), and/or operational issues impacting the platform's reliability and investors' confidence. This uncertainty may prompt investors to demand higher interest rates to offset the perceived risks. Other factors, such as the economic cycle or fierce competition between platforms, may also increase a platform's probability of default (Yoon et al. 2019), thereby influencing the pricing mechanism.

This study had two practical implications. First, although the neural network model exhibited a high F1-score, logistic regression demonstrated comparable overall performance with a 20% higher sensitivity. From an investor's standpoint, prioritizing sensitivity, which entails identifying true-positive cases, may be more appropriate because it mitigates the risk of overlooking potential defaults. When selecting a predictive model, investors should carefully evaluate the tradeoff between specificity and sensitivity. Second, the invoice-trading P2B market may appeal to small businesses that have limited collateral or lack a credit history and that have trouble being granted traditional bank loans. However, during the examined period, the discrepancy between the observed and implicit probabilities may indicate an inefficiency in the market, likely due to

limited competition, resulting in small and medium enterprises (SMEs) being financed at higher interest rates than what their actual risk profile suggests. Investors should make informed decisions regarding their investments in online invoice trading platforms to avoid being aware of the disparity between the observed and implicit probabilities of default in the pricing mechanism of these platforms.

Conclusions

This study developed several machine-learning models to predict default in the invoice-trading P2B market. We used publicly available data from the crowdlending platform Kriya and estimated several techniques such as logistic regression, conditional inference trees, random forest, support vector machines, and neural networks. The current findings demonstrate that implementing these techniques leads to a substantial enhancement in the detection of defaults, achieving an improvement of up to 56%. Remarkably, these improvements were achieved while maintaining specificity levels above 70%. Furthermore, these results were obtained despite limited information provided by the platform. The inclusion of neither the financial information of the borrower nor that of the invoice seller was possible and it is a limitation of this study. Solvency ratios, insurance, and collateral information are variables that can significantly increase model performance, hence greater information disclosure in line with what peer-to-peer lending platforms offer is desirable to increase transparency and correctly assess risk.

In addition, this study examined the platform's credit pricing using the implicit probability of default. We discovered a consistent disparity between the observed and implicit default probabilities. Notably, the price charged to sellers reflects a significantly higher probability of default than the observed probability. Several factors could explain this difference, in addition to the borrowers' risk of default, such as inclusion in the pricing mechanism of the platform's and seller's probabilities of default, the flexibility of this form of financing, or credit rationing. However, if the borrower's probability of default is considered the main factor when pricing an invoice, this may indicate that sellers pay a premium that does not accurately reflect the actual risk involved.

The current study focused on the problem of estimating defaults in the P2B invoice market, which is considered short-term lending. This study can be applied to medium- and long-term lending in P2B. Other methodologies proven effective in related studies include LightGBM with and without swarming techniques (Ko et al. 2022; Muslim et al. 2022) and learning-to-rank methodologies. Future studies in this area should change their focus from the borrower's probability of default to the platform's risk of default, as in Ahelegbey et al. (2019), Chen et al. (2022), and Liang et al. (2022). In addition, niche segments in well-studied P2P markets, such as equity and real estate, can be explored.

Appendix

See Tables 9, 10, 11.

Table 9 Summary of related studies

| Year | Author | Data | | | Method | Accuracy |
|------|-----------------------------------|---------|--|------------------|----------------------|-------------------------------|
| | | Period | Platform | Size | | |
| 2014 | Carmichael | 2007/13 | L.Club | 2,227,987 | Log. Reg | n.d |
| 2016 | Li et al | 2007/15 | L.Club | 160,956 | Log. Reg | 86.16% |
| 2015 | Serrano-Cinca et al | 2008/14 | L.Club | 24,449 | Log. Reg | 80.60% |
| 2017 | Möllenkamp | 2012/13 | L.Club | 143,654 | Log. Reg | .019 Cox-Snell R ² |
| 2019 | Zhu et al | 2019 | L.Club | 115,000 | R. Forest | 98% |
| | | | | | D. Trees | 95% |
| | | | | | SVM | 75% |
| | | | | | Log. Reg | 73% |
| 2015 | Malekipirbazari and Aksakalli | 2012/14 | L.Club | 350,000 | R. Forest | 78% |
| | | | | | KNN | 70.1% |
| | | | | | SVM | 63.3% |
| | | | | | Log. Reg | 54.5% |
| 2017 | Fu | n.d | L.Club | 1,320,000 | ANN + R. Forest | 73.5% |
| 2018 | Kim and Cho | 2015/16 | L.Club | 855,502 | DenseNet | 79.6% |
| 2016 | Serrano-Cinca and Gutiérrez-Nieto | 2007/12 | L.Club | 40,901 | MLR | .5 R ² |
| | | | | | D. Trees | 5.98% IRR |
| 2015 | Guo et al | 2016 | L.Club | n.d | RBM | 4.18%—.25 R ² |
| | | | | | RBM + | 5.33%—.25 R ² |
| | | | | | IOM | 6.17%—.26 R ² |
| | | n.d | Prosper | 4,128 | RBM | 5.41%—.16 R ² |
| | | | | | RBM + | 6.06%—.19 R ² |
| | | | | | IOM | 6.42%—.21 R ² |
| 2019 | Bastani et al | 2007/13 | L.Club | 80,000 | Wide & deep learning | 16.19% IRR |
| 2017 | Dorfleitner et al | 2011/17 | Kriya | 19,566 | Logit/Tobit | .07-.08 Pseudo-R ² |
| 2022 | Ko et al | 2018/19 | L. Club | 60,993 | LightGBM | 68.57% |
| | | | | | CNN | 67.27% |
| | | | | | Log. Reg | 66.87% |
| | | | | | LDA | 66.81% |
| | | | | | ANN | 66.85% |
| | | | | | Bay. Class | 64.27% |
| | | | | | R. Forest | 63.89% |
| | | | | | D. Trees | 63.63% |
| 2022 | Muslim et al | 2019 | L. Club | 20,875,146 | LightGBM | 94.38% |
| 2022 | Ahelegbey et al | 2015 | ECAI | 813 (SMEs) | Log. Reg | 85.6% (AUROC) |
| | | | | | | 94.9% (AUROC) |
| 2022 | Liang et al | 2013/18 | Wangdaizhijia & National Bureau Statistics | 1262 (platforms) | L.STACK | 95.90% (AUC) |

Table 10 List of variables

| Advance rate (AR) | Percentage of an invoice’s face value which a factor pays upon its purchase |
|------------------------------|---|
| Annualized gross yield (AGY) | Actual yearly rate of return earned on factoring considering the effect of compounding interest |
| Total face value (TFV) | Total amount of an invoice (in pounds sterling) |
| Days/Maturity | Number of days elapsed from the advance date to the expected pay date, also known as maturity |
| GDP | Gross domestic product (seasonally-adjusted, quarterly, percentage of change) |
| Exports | Exports (seasonally-adjusted, quarterly, percentage of change) |
| IPI | Import price index (not seasonally-adjusted, quarterly, percentage of change) |
| Trade type | Factor that reflects the type of trade. May be standard, whole ledger, multi debtor or license fee |
| Trade band | Factor that reflects the industry/sector of the borrower |
| Default | Factor that reflects if the loan was default or not. Target variable of the study |
| Total crystallized loss | Real loss perceived after discounting recovered amount (in pounds sterling) |
| Trade payment state | Factor that reflects the status of the loan. May be paid, seller-repurchased, repurchase-requested, overdue or partially-paid |

Table 11 Hyperparameter tuning

| | teststat | testtype | mtry | minrow | minsplit | minbucket | maxdepth | nree |
|------------------------|----------|------------|--------|---------|--------------|-----------|----------|----------|
| Cond. inf. tree | max | Bonferroni | 8 | – | 43 | 4 | 17 | – |
| Cond. inf. tree forest | max | Bonferroni | 8 | – | 43 | 4 | 17 | 200 |
| Random forest | – | – | 8 | 4 | – | – | 17 | 200 |
| | C | epsilon | hidden | stepmax | learningrate | algorithm | err.fct | act.fct |
| SVM | 1 | 0.1 | – | – | – | – | – | – |
| Neuralnet | – | – | 1 | 100,000 | 0.1 | rprop+ | ce | Logistic |

Any hyperparameter not specified in this table was considered with its default value by the R mlr package. Hyperparameters were searched only for the conditional inference trees and applied over the forests for comparison purposes. Results on conditional inference trees may have varied as they have high variance by default. Teststat is the type of test statistic applied for variable selection. Testtype specifies the distribution of the test statistic. Mtry is the number of input variables randomly sampled at each node. Minrow is the minimum number of observations in a node. Minsplit is the minimum sum of weights in a node considered for splitting. Minbucket is the minimum sum of weights in a terminal node. Maxdepth is the maximum depth of the tree. Ntree is the number of trees. C is the cost of constraints violation. Epsilon is the insensitive-loss function used. Hidden is the number of hidden layers. Stepmax is the maximum number of steps for training the net. Learningrate is the rate at which the net learns. Algorithm specifies the type of algorithm used to calculate the net. Err.fct is the error function. Act.fct is the activation function

Abbreviations

- ACO Ant colony optimization
- AGY Annualized gross yield
- AI Artificial intelligence
- AR Advance rate
- AUC Area under the curve
- CBR Case-based reasoning
- ECAI European External Credit Assessment Institution
- FICO Fair Isaac Corporation
- FN False negative
- FP False positive
- GDP Gross domestic product
- IOM Instance-based model
- IPD Implicit probability of default
- IPI Import price index
- IRR Internal rate of return

| | |
|-------|---|
| K-NN | K nearest neighbors |
| ML | Machine learning |
| NPLs | Non-performing loans |
| OPD | Observed probability of default |
| P2B | Peer-to-business |
| P2P | Peer-to-peer |
| PD | Probability of default |
| PIT | Point-in-time |
| RBM | Restricted Boltzmann machine |
| SMEs | Small and medium-sized enterprises |
| SMOTE | Synthetic minority oversampling technique |
| TFV | Total face value |
| TN | True negative |
| TP | True positive |
| UK | United Kingdom |
| US | United States of America |

Acknowledgements

Not applicable.

Author contributions

C-MC contributes to the development of the models, the statistical programming, the analysis of the data, and the writing of the paper. L-OG and P-DS contributes with the design of the work and the analysis of the data. All authors read and approved the final manuscript.

Funding

We acknowledge the funding provided by the Galician Regional Government [ED431C 2020/18] co-funded by the European Regional Development Fund (ERDF/FEDER) within the period 2020–2023.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration

Competing interests

The authors declare that they have no competing interests.

Received: 20 January 2023 Accepted: 7 February 2024

Published online: 08 June 2024

References

- Ahelegbey D, Giudici P, Hadji-Misheva B (2019) Latent factor models for credit scoring in P2P systems. *Physica A* 522:112–121. <https://doi.org/10.1016/j.physa.2019.01.130>
- Avgeri E, Psillaki M (2023) Factors determining default in P2P lending. *J Econ Stud*. <https://doi.org/10.1108/JES-07-2023-0376>
- Bastani K, Asgari E, Namavari H (2019) Wide and deep learning for peer-to-peer lending. *Expert Syst Appl* 134:209–224. <https://doi.org/10.1016/j.eswa.2019.05.042>
- Bischi B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G, Jones Z (2016) mlr: Machine Learning in R. *J Mach Learn Res* 17(170):1–5
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140. <https://doi.org/10.1007/bf00058655>
- Carmichael D (2014) Modeling default for peer-to-peer loans. *Soc Sci Res Netw*. <https://doi.org/10.2139/ssrn.2529240>
- Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16:321–357. <https://doi.org/10.1613/jair.953>
- Chen X, Chong Z, Giudici P, Huang B (2022) Network centrality effects in peer to peer lending. *Physica A Stat Mech Appl*. <https://doi.org/10.1016/j.physa.2022.127546>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1023/a:1022627411411>
- Cowling M (2010) The role of loan guarantee schemes in alleviating credit rationing in the UK. *J Financ Stab* 6(1):36–44. <https://doi.org/10.1016/j.jfs.2009.05.007>
- Cumming D, Zhang Y (2016) Are crowdfunding platforms active and effective intermediaries? *Soc Sci Res Netw*. <https://doi.org/10.2139/ssrn.2882026>
- DeLong E, DeLong D, Clarke-Pearson D (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. <https://doi.org/10.2307/2531595>
- Dorfleitner G, Rad J, Weber M (2017) Pricing in the online invoice trading market: first empirical evidence. *Econ Lett* 161:56–61. <https://doi.org/10.1016/j.econlet.2017.09.020>

- Dziuba D (2018) Crowdfunding platforms in invoice trading as alternative financial markets. *Roczniki Kolegium Analiz Ekonomicznych/Szkoła Główna Handlowa* 49:455–464
- Elliott J, Curet O (1999) Invoice discounting - a strategic analysis using case-based reasoning. *Appl Innov Expert Syst* 1. https://doi.org/10.1007/978-1-4471-0575-6_15
- Faraway J (2016) Extending the linear model with R. In Chapman and Hall/CRC eBooks. <https://doi.org/10.1201/9781315382722>
- Fu Y (2017) Combination of random forests and neural networks in social lending. *J Financ Risk Manag* 06(04):418–426. <https://doi.org/10.4236/jfrm.2017.64030>
- Gower J (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27(4):857. <https://doi.org/10.2307/2528823>
- Grosan C, Abraham A (2005) Artificial neural networks. *Intell Syst Ref Lib* 17:281–323. https://doi.org/10.1007/978-3-642-21004-4_12
- Guo Y, Zhou W, Luo C, Liu C, Xiong H (2016) Instance-based credit risk assessment for investment decisions in P2P lending. *Eur J Oper Res* 249(2):417–426. <https://doi.org/10.1016/j.ejor.2015.05.050>
- Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 15(3):651–674. <https://doi.org/10.1198/106186006x133933>
- Kim J, Cho S (2018) Deep dense convolutional networks for repayment prediction in peer-to-peer lending. *Adv Intell Syst Comput* 771:134–144. https://doi.org/10.1007/978-3-319-94120-2_13
- Ko P, Lin P, Do H, Huang Y (2022) P2P lending default prediction based on AI and statistical models. *Entropy*. <https://doi.org/10.3390/e24060801>
- Kotsiantis S, Pintelas P (2004) Mixture of expert agents for handling imbalanced data sets. *Ann Math Comput Teleinform* 1:46–55
- Lau A (1987) A five-state financial distress prediction model. *J Account Res* 25(1):127. <https://doi.org/10.2307/2491262>
- Lee N, Sameen H, Cowling M (2015) Access to finance for innovative SMEs since the financial crisis. *Res Policy* 44(2):370–380. <https://doi.org/10.1016/j.respol.2014.09.008>
- Li J (2016) Going online? The motive of firms to borrow from the crowd. <https://www.federalreserve.gov/conferences/files/going-online-the-motive-of-firms-to-borrow-from-the-crowd.pdf>
- Li J, Mirza N, Rahat B, Xiong D (2020) Machine learning and credit ratings prediction in the age of fourth industrial revolution. *Technol Forecast Soc Chang* 161:120309. <https://doi.org/10.1016/j.techfore.2020.120309>
- Li Z, Yao X, Wen Q, Yang W (2016) Prepayment and default of consumer loans in online lending. *Soc Sci Res Netw*. <https://doi.org/10.2139/ssrn.2740858>
- Liang K, Zhang C, Jiang C (2022) Analyzing default risk among P2P platforms based on the LAS-STACK method by considering multidimensional signals under specific economic contexts. *Electron Commer Res* 22(1):77–111. <https://doi.org/10.1007/s10660-021-09505-9>
- Liu J, Li X, Wang S (2020) What have we learnt from 10 years of fintech research? A scientometric analysis. *Technol Forecast Soc Chang* 155:120022. <https://doi.org/10.1016/j.techfore.2020.120022>
- Malekipirbazari M, Aksakalli V (2015) Risk assessment in social lending via random forests. *Expert Syst Appl* 42(10):4621–4631. <https://doi.org/10.1016/j.eswa.2015.02.001>
- Möllenkamp N (2017) Determinants of loan performance in P2P lending. In: 9th IBA Bachelor Thesis conference, 1–4. <https://journals.open.tudelft.nl/sure/article/view/2551/2808>
- Muslim M, Dasril Y, Saman M, Ifriza Y (2022) An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending. *Indonesian J Electr Eng Comput Sci* 28(2):1002–1011
- Nava I, Cuccio D, Giada L, Nordio C (2019) A simple factoring pricing model. *Soc Sci Res Netw*. <https://doi.org/10.2139/ssrn.3428749>
- Nigmonov A, Shams S, Alam K (2022) Macroeconomic determinants of loan defaults: evidence from the US peer-to-peer lending market. *Res Int Bus Finance*. <https://doi.org/10.1016/j.ribaf.2021.101516>
- Osborne J (2008) Best practices in quantitative methods. SAGE Publications, Inc. eBooks. <https://doi.org/10.4135/9781412995627>
- Pang S, Hou X, Xia L (2021) Borrowers' credit quality scoring model and applications, with default discriminant analysis based on the extreme learning machine. *Technol Forecast Soc Chang* 165:120462. <https://doi.org/10.1016/j.techfore.2020.120462>
- Perko I (2017) Behaviour-based short-term invoice probability of default evaluation. *Eur J Oper Res* 257(3):1045–1054. <https://doi.org/10.1016/j.ejor.2016.08.039>
- Serrano-Cinca C, Gutiérrez-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. <https://doi.org/10.1016/j.dss.2016.06.014>
- Serrano-Cinca C, Gutiérrez-Nieto B, López-Palacios L (2015) Determinants of default in P2P lending. *PLoS ONE* 10(10):e0139427. <https://doi.org/10.1371/journal.pone.0139427>
- Stiglitz J, Weiss A (1981) Credit rationing in markets with rationing credit information imperfect. *Am Econ Rev* 71(3):393–410
- Strobl C, Malley J, Tutz G (2009) An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14(4):323–348. <https://doi.org/10.1037/a0016973>
- Turiel J, Aste T (2020) Peer-to-peer loan acceptance and default prediction with artificial intelligence. *R Soc Open Sci* 7(6):191649. <https://doi.org/10.1098/rsos.191649>
- Yoon Y, Li Y, Feng Y (2019) Factors affecting platform default risk in online peer-to-peer (P2P) lending business: an empirical study using Chinese online P2P platform data. *Electron Commer Res* 19:131–158. <https://doi.org/10.1007/s10660-018-9291-1>
- Zhang J, Thomas L (2015) The effect of introducing economic variables into credit scorecards: an example from invoice discounting. *J Risk Model Validation* 9(1):57–78
- Zhu L, Qiu D, Ergu D, Ying C, Liu K (2019) A study on predicting loan default based on the random forest algorithm. *Procedia Comput Sci* 162:503–513. <https://doi.org/10.1016/j.procs.2019.12.017>

Ziegler T, Shneor R, Garvey K, Wenzlaff K, Yerolemou N, Hao R, Zhang B (2017) Expanding horizons: the 3rd European alternative finance industry report. Soc Sci Res Netw. <https://doi.org/10.2139/ssrn.3106911>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.