Balancing Risk and Reward: Unveiling the Credit Conundrum in P2P Lending - A Tale of Default and Profit Scoring

Marko Dávid Vateha[1]

[1] University of Economics in Bratislava, Faculty of Economics and Finance, Department of Banking and International Finance, Dolnozemská cesta 1, Bratislava 852 31 Slovak republic

MarkoDavid.Vateha@euba.sk

https://doi.org/10.53465/EDAMBA.2023.9788022551274.282-291

Abstract. A credit risk assessment is a vital component of the lending process, particularly in the rapidly growing realm of peer-to-peer (P2P) lending. This empirical study delves into the credit risk assessment methods of default and profit scoring, employing machine learning techniques on a publicly available dataset sourced from P2P lending platform – Lending Club. Our investigation yields insightful findings, emphasizing the paramount importance of accurate credit risk evaluation and their implications for loan portfolio returns. The outcomes of our analysis reveal that profit scoring outperforms default scoring in terms of higher annualized returns on loan portfolio. Notably, this superior performance of profit scoring is primarily attributed to its ability to intelligently accept more loans. This is due to the fact that traditional default modelling approaches do not take into account the possibility that certain defaulted loans would generate positive annualized returns as debtors may default at the end of the loan life cycle. By considering not only the risk of default but also the potential profitability of a loan, profit scoring enables lenders to make informed decisions and optimize their portfolio returns effectively. Our findings further reinforce the need for lenders to adopt advanced credit risk modelling techniques, such as profit scoring, to navigate the dynamic P2P lending landscape successfully.

Keywords: P2P Lending, Default, Profit Scoring, Machine Learning

JEL classification: *G21, C55*

1 Introduction

The evaluation of credit risk plays a pivotal role in lending decisions, as financial institutions strive to strike a delicate balance providing a credit to as many customers as possible with the loan default rates [14]. Traditionally, default scoring models have been employed to assess the likelihood of loan repayment, focusing primarily on the borrower's creditworthiness and historical repayment patterns. However, as the lending landscape evolves, alternative approaches, such as profit scoring [21], have gained prominence in capturing the multidimensional aspects of credit risk.

Default scoring methods primarily aim to predict the probability of default. Such models are trained on a sample of loans with binary target variable – Default or Full repayment of all liabilities [3]. While default scoring is effective in identifying high-risk borrowers and minimizing default rates, it often neglects an essential aspect of lending: the potential profitability of the loans.

In contrast, profit scoring takes a more comprehensive approach to credit risk assessment by considering the potential returns associated with lending to a particular borrower [21]. The profit scoring models aim to maximize the profitability of the loan portfolio while managing credit risk implicitly. This approach leads to prediction of future annualized return on a loan according to the parameters of loan application.

In this article, we delve into the key differences between default scoring and profit scoring in the context of loan assessment. We aim to shed light on the advantages and limitations of each approach and explore how they impact lending decisions and loan portfolio performance. Drawing on an empirical study utilizing peer-to-peer (P2P) lending data, we examine the effectiveness of both default scoring and profit scoring in terms of loan portfolio profitability.

By comparing the outcomes of default scoring and profit scoring models, we aim to provide valuable insights for lenders and financial institutions seeking to enhance their credit risk assessment strategies. Understanding the trade-offs between default scoring and profit scoring is crucial for informed decision-making in lending, as it allows lenders to strike a balance between mitigating default risk and maximizing the profitability of their loan portfolios.

2 Literature Review

In a variety of application fields, including language models or image recognition, machine learning (ML) and artificial intelligence (AI) have attained human-level performance. Yet, Munkhdalai et al. [18] perceive expert-based credit risk models as those which rule the financial industry. Financial Stability Board (FSB) [7] finds it challenging to conduct a broad assessment of the effectiveness of ML models since the predictive power of these models has often only been examined in experimental settings of academic research. Fintech companies, such as P2P lending platforms, in particular tend to apply machine learning into their processes [2].

An attempt to incorporate machine learning advancements into the field of credit scoring is not brand new. The initial initiatives started in 2003 when Baesens et al. [1]

started examining the performance of several categorization approaches. According to the literature review, intelligent systems appear to have the necessary components to perform better than conventional methods.

Conventional credit scoring models rely mostly on linear statistical models [9]. However, Khandani et al. [12], similarly as Gambacorta et al. [9], declares the need to address the problem of credit quality assessment with more advanced, non-linear techniques. He relies on the results of an empirical test, in which he showed that such more complex models can outperform traditional linear approaches by 6 to 23% in terms of minimising realised losses, thus also bringing a financial dimension. The literature review shows a bit conflicting evidence about performance of linear models relative to the more advanced ones. The conclusions of Munkhdalai et al. [18] or Finlay [8] put logistic regression level-headed comparing to non-linear predictive models, whereas Chang et al. [4] study reveals that ensemble learning algorithms, such as XGBoost, dominate logistic regression by a large margin. A comprehensive study of Lessmann et al. [15] compares individual classifiers with ensemble classifiers. He concludes that logistic regression does not lag behind other individual classifies and is comparable also with neural network classifier, which is supposed to capture for non-linear relationship, yet it falls short in comparison to ensemble models. Ensemble algorithm combines multiple individual models, such as decision trees or neural networks, to make more accurate and robust predictions by aggregating their outputs.

Serrano-Cinca et al. [21] highlights the importance of incorporating profit-based assessment methods to optimize lending decisions and enhance the overall economic performance of P2P lending platforms. The study demonstrates that by considering profit as the primary criterion for loan approval, P2P lending platforms can achieve higher annualized returns on their loan portfolios. A more recent re-examination of profit scoring as a viable alternative to traditional credit scoring methods in P2P lending is brought by Lyócsa et al. [16]. According to the empirical research of the study, profit scoring performs better than default scoring when it comes to producing greater annualized returns on loan portfolios. This outcome is mostly attributable to taking more loans rather than depending heavily on strict default risk assessment.

In an analysis of credit card underwriting, Krivorotov [13] concludes that profit scoring approach reshuffles a bank’s credit card portfolio substantially and may possibly make the credit card portfolios riskier.

3 Research Methods

This chapter discusses machine learning (ML) techniques used in the modelling section, explains how we arrived at the annualized rate of return of loans using modified internal rate of return (MIRR), and provides a brief introduction to the data sample utilized in the research part.

3.1 Statistical Methods

Statistical learning is a key tool in analysing and modelling various real-world phenomena. In the field of credit risk assessment, statistical models are crucial for predicting the likelihood of default and managing risk [3]. For the aim of credit scoring, we chose three essential statistical learning tools: support vector machine (SVM), Extreme Gradient Boosting (XGBoost), and linear regression (LR) or Logistic Regression (LogReg) and their regularized variants (L1 and L2 penalty).

Linear regression (LR) is a popular statistical modelling technique used to establish a relationship between a dependent variable and one or more independent variables. It aims to fit a linear equation that best represents the data. LR aims to predict actual value of response variable and is solely used for regression tasks [17]. However, in some cases, the basic linear regression model may suffer from issues like overfitting or high sensitivity to outliers. To address these challenges, various regularization techniques can be employed. Ridge regression (L2 penalty regularization) adds a penalty term to the ordinary least squares (OLS) objective function, which helps shrink the coefficient estimates towards zero and hence reduces model complexity [10]. Lasso regression (L1 penalty regularization), on the other hand, not only introduces a penalty term but also performs feature selection by enforcing some coefficients to be exactly zero [6]. This makes lasso regression useful for feature selection and creating more interpretable models. These regularization techniques provide flexible tools to improve the performance and interpretability of linear regression models in various scenarios.

Logistic regression (LogReg) is a statistical method used for classification tasks (mostly binary), where a goal is to predict the probability of an instance belonging to a specific class [19]. Unlike linear regression, logistic regression uses a logistic or sigmoid function to transform the linear combination of predictors into a probability score. This score represents the likelihood of an instance belonging to the positive class.

Support Vector Machine (SVM) is a powerful and versatile supervised machine learning algorithm used for both classification and regression tasks [20]. It aims to find an optimal hyperplane that maximally separates different classes or predicts continuous values with the highest margin of confidence. SVM achieves this by transforming data into a higher-dimensional space using kernel functions, allowing for both linear and nonlinear decision boundaries. It is known for its ability to handle high-dimensional data and effectively deal with outliers.

Extreme Gradient Boosting (XGBoost) is a powerful and widely used machine learning algorithm known for its exceptional predictive performance when tuned properly [5]. It belongs to the gradient boosting family and is based on the concept of ensemble learning. XGBoost combines the predictions of multiple weak decision trees to create a strong predictive model. It utilizes a gradient boosting framework, where each subsequent tree is built to correct the errors made by the previous trees.

In our study we need to employ algorithms capable of performing both classification and regression. The main difference between classification in default scoring and regression in profit scoring lies in the nature of the prediction task and the objective of the analysis. In default scoring, the goal is to classify instances into binary classes (e.g., default or non-default) based on applicants' creditworthiness. On the contrary, in profit scoring, the objective is to estimate the potential profitability of each instance or transaction, typically in terms of expected profit or return on investment. Regression in profit scoring involves predicting a continuous variable and aims to optimize financial outcomes by evaluating the potential monetary gains or losses [16].

3.2 Modified Internal Rate of Return

Lending Club provides data about applicant characteristics, as well as the history of debtor repayments for accepted loans. In order to test superiority of profit scoring, we need to calculate the annualized modified internal rate of returns (MIRR) from the payments history [22]. The derived MIRR will become the target variable of our regression models as opposite to the competing default models trying to classify loans into correct category – Default and Non-default.

We presume initial transfer of funds following the approval of a loan application to be a single payment. Then, we assume the repayments of debtor towards the platform to be reinvested at the median IRR of loans for the last 12 months. With this approach we are able to obtain an annualized rate or return and hence to compare the loans with different maturities.

3.3 Data

Lending Club is a prominent peer-to-peer (P2P) lending platform based in the United States. It operates as an online marketplace connecting borrowers and investors, offering a streamlined alternative to traditional banking channels. Anonymised data about its accepted and rejected applications, as well as the final status of the loan – Repaid or Default – is accessible and free to download for investors on their webpage.

The dataset [11] covers period of 2007-2018 in which 2.2 mil. loans were granted by the platform. There are more than 150 features in the dataset. In the data transformation process, the original set of records was reduced to 1.2 mil loans with more than 530 columns. The features were transformed in the nature that continuous variables are standardized, and categorical data is dummy encoded. Feature standardization involves transforming the values of each feature in the data to have a zero-mean (by subtracting the mean) and unit variance, ensuring that the features are on the same scale. Variables with a high prevalence of missing values were deleted in the process to ensure better data quality.

Table 1. Descriptive statistics of Modified Internal Rate of Return of loans in portfolio

		Annualized rate of return (in %)				
	Default rate	Mean	Median	STD	10th percentile	90th percentile
Full data	19.53%					
Non-Default		13.27	12.27	4.86	8.33	19.45
Default		-44.51	-40.85	35.83	-97.97	1.27

Train data	19.45%					
Non-Default		13.30	12.30	4.88	8.35	19.43
Default		-44.86	-41.12	36.05	-98.05	1.50
Test data	20.02%					
Non-Default		13.35	12.36	4.87	8.33	19.55
Default		-44.39	-39.90	35.99	-98.11	1.11

4 Results of the Research

In Fig 1. we highlight the fact discussed in the introduction of the paper. A default rate analysis (1.2 mil. loans) reveals that in 19.53% of cases liabilities are not fully repaid. If we zoom in into subsample of loans with positive annualized returns, we find that even defaulted loans fall into the profitable category. About 11% of loans when borrowers did not fully meet their obligations (i. e. default) are still relatively lucrative (approx. 2% of entire loan portfolio) as depicted in the right-hand side of Fig. 1.

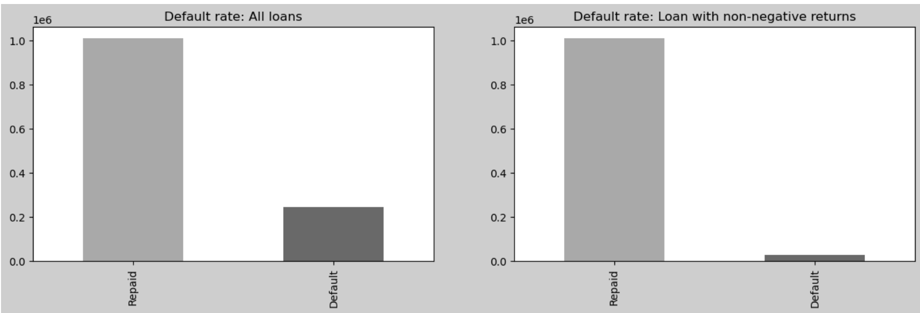

Fig. 1. Default rate in the portfolio of analysed Lending Club loans.

A deep dive into distribution of annualized rate of returns in the loan portfolio is presented in Fig 2. The left bottom panel exhibits drill-down on non-performing loans, which can nudge some investors to target even defaulted one as some bring profits.

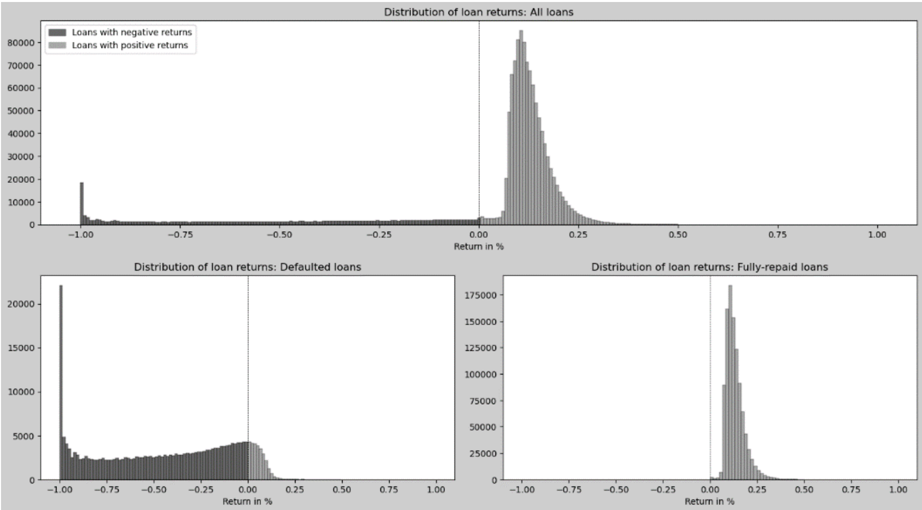

Fig. 2. Profit distribution of P2P lending loan portfolio of Lending Club. Notes: Loans that led to negative internal rate of return are denoted with dark grey. The upper panel depicts distribution of entire portfolio of accepted credit applications, the lower panel provides split view on segment of defaulted (left side) and fully repaid loan (right side).

In the modelling part we employed individual linear algorithm learning (i.e. LR and LogReg), regularized forms of such a linear model (L1 or L2 penalty, respectively), individual non-linear model (namely RBF kernel SVM) and ensemble method which aggregates results of multiple base models (such as XGBoost). Table 2. presents out-of-sample performance of all the models and the results provide compelling evidence about potential financial benefits of proposed paradigm shift from credit to profit scoring. As we are using diverse set of machine learning algorithms, we believe the research delivers adequately robust findings. If we investigate nominal yields of lending platform, an average gross profit of default classification models is \$9,8 mil., whereas algorithms projecting profits average at \$12,9 mil. The difference between two families of modelling paradigms is more than 30%. Even though the total gross profit does not account for differences in term lengths (loan maturities), it is quite substantial margin. Except of Gaussian RBF kernel algorithm, which is a negative outlier among profit scoring models, all others profit-driven approaches to credit approval are superior to their default-oriented counterparts by more than 10% in terms of arithmetic average returns or median returns.

Table 2. Results of different modelling approaches applied to test dataset (out-of-sample)

	(in %) Invested loans	MIRR (in %) for entire portfolio			(in mil. \$)
		Median	Mean	STD	Total profit
Credit scoring					
LogReg	58.43	7.23	3.29	15.75	10.53

LogReg-L2	59.76	7.49	3.18	16.27	10.10
LogReg-L1	61.04	7.61	2.80	17.17	8.63
RBF-SVM	66.85	8.22	2.76	19.22	9.55
XGBoost	60.38	7.55	3.12	16.59	10.28
Profit scoring					
LR	66.93	8.43	3.57	17.57	14.27
LR-L2	66.87	8.40	3.53	17.53	13.69
LR-L1	67.92	8.48	3.41	18.30	12.97
RBF-SVM	67.25	8.47	1.88	22.68	9.61
XGBoost	73.10	9.07	3.39	19.89	13.99

The credit approval process is very complex and requires balancing inherent risks and potential rewards. The credit scoring models noticeably overestimate risk of an applicant. Contrarily, the set of algorithms focused on profit projection grants loans substantially more often.

As profit scoring approach outperforms the default classification in average returns across accepted loans by only small margin, if we account for opportunity cost of rejected loan, the difference in performance on entire portfolio in significant (Table 2). Taking this into account, we presume that the profit is generated by the increased acceptance rate that the profit scoring technique associates itself with.

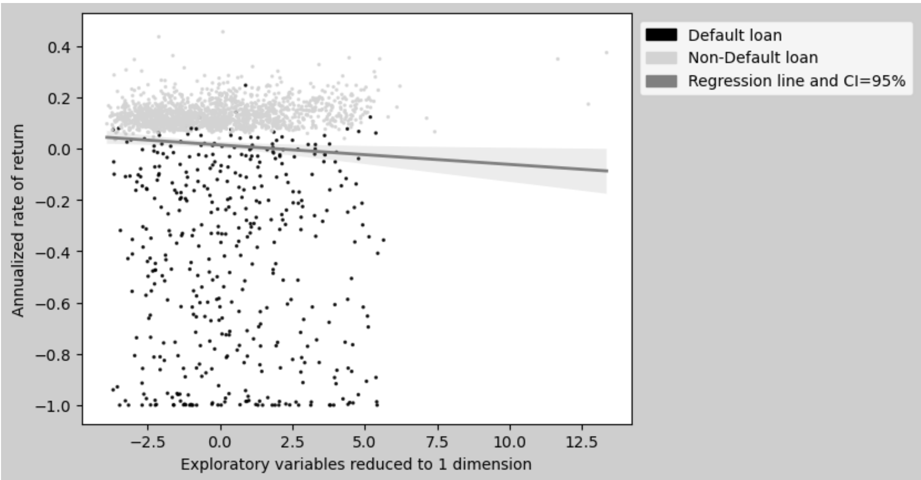

Fig. 3. The relationship between annualized rate of return and independent variables reduced to one dimension. Note: Our dataset includes 522 exploratory variables, and they were linearly reduced via Principal Component Analysis (PCA) to a single variable plotted on X-axis for visualization purposes. This dimensionality reduction explains only 24.8% of variance in original dataset with 522 features. Therefore, the findings should be regarded with some scepticism.

Fig. 3. depicts the most prominent challenge that profit scoring models face. The higher the inherent riskiness of applicant (X-axis), the higher the volatility of returns. The riskier the credit is, the higher the interest rates assigned to the liability and hence higher the potential return. On the other hand, the riskier the customer is, the more likely the default occurs. It is extremely difficult for the algorithm to predict the possible profit on a deal of customer with low creditworthiness. This conclusion is indicated by growing confidence intervals of regression line (at confidence interval of 95%).

5 Conclusion

The comparison of the same statistical models in two families of algorithms – classification and regression – shows supremacy of profit scoring approach to credit approval in P2P lending. The dominance is confirmed not only with relative values of average and median profit in out-of-sample portfolio of loans, but also with nominal gross profit in monetary units. Despite that the profit-driven approach is riskier. The variability of annualized returns grows as we change the task from default detection to profit projection. This premise holds across all modelling techniques.

Our findings about growing variability of profits with decreasing credit quality of applicant creates a room for further development in the area of credit risk modelling. The plausible direction of future research could lead to the blend modelling. The best results might be achieved when more prudent credit scoring approach is applied to the class of riskier customers, whereas profit scoring can help to identify profitable deals on the borderline between default and non-default. Alternatively, a quantile regression can offer additional predictive power in profit scoring as quantile regression is an extension of linear regression used when the conditions of linear regression are not met.

Acknowledgement

This paper is an output of the science project VEGA 1/0221/21 Interest rates in the environment with central bank digital currency.

References

1. Baesens, B., et al.: Benchmarking state-of-the-art classification algorithms for credit scoring. In: *Journal of the operational research society* 54 (2003): 627-635.
2. Berg, T., et al.: On the rise of fintechs: Credit scoring using digital footprints. In: *The Review of Financial Studies* 33.7 (2020): 2845-2897.
3. Bluhm, Ch., Overbeck L., Wagner Ch.: Introduction to credit risk modeling. CRC Press, Boca Raton (2016).
4. Chang, Y-Ch., et al.: Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. In: *Applied Soft Computing* 73 (2018): 914-920.

5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.
6. Emmert-Streib, F., Dehmer, M.: High-dimensional LASSO-based computational regression models: regularization, shrinkage, and selection. In: *Machine Learning and Knowledge Extraction* 1.1 (2019): 359-383.
7. Financial Stability Board.: Artificial intelligence and machine learning in financial services. (2017) https://www.fsb.org/wp-content/uploads/P011117.pdf, last accessed 2023/05/08.
8. Finlay, S.: Multiple classifier architectures and their application to credit risk assessment. In: *European Journal of Operational Research* 210.2 (2011): 368-378.
9. Gambacorta, L., et al.: How do machine learning and non-traditional data affect credit scoring? New evidence from a Chinese fintech firm. In: BIS Working Paper No. 834 (2019).
10. Hastie, T.: Ridge regularization: An essential concept in data science. In: *Technometrics* 62.4 (2020): 426-433.
11. Kaggle, https://www.kaggle.com/datasets/wordsforthewise/lending-club, last accessed 2023/05/09.
12. Khandani, A. E., et al.: Consumer credit-risk models via machine-learning algorithms. In: *Journal of Banking & Finance* 34.11 (2010): 2767-2787.
13. Krivorotov, G.: Machine learning-based profit modeling for credit card underwriting-implications for credit risk. In: *Journal of Banking & Finance* 149 (2023): 106785.
14. Lando, D.: Credit risk modeling. Princeton University Press, Princeton (2009).
15. Lessmann, S., et al.: Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. In: *European Journal of Operational Research* 247.1 (2015): 124-136.
16. Lyócsa, Š. et al.: Default or profit scoring credit systems? Evidence from European and US peer-to-peer lending markets. In: *Financial Innovation* 8.1 (2022): 1-21
17. Maulud, D., Abdulazeez, A. M.: A review on linear regression comprehensive in machine learning. In: *Journal of Applied Science and Technology Trends* 1.4 (2020): 140-147.
18. Munkhdalai, L., et al.: An empirical comparison of machine-learning methods on bank client credit assessments. In: *Sustainability* 11.3 (2019): 699.
19. Schein, A. I., Ungar, L. H.: Active learning for logistic regression: an evaluation. In: Penn Engineering Departmental Papers (2007).
20. Steinwart, I., and Christmann, A.: Support vector machines. Springer Science & Business Media, Berlin (2008).
21. Serrano-Cinca, C., Gutiérrez-Nieto, B.: The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending. In: *Decision Support Systems* 89 (2016): 113-122.
22. Xie, M.: Research on the modified internal rate of return. In: *Turkish Journal of Computer and Mathematics Education* 12.11 (2021): 4087-4090.