



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijrset@gmail.com](mailto:ijrset@gmail.com)

[www.ijrset.com](http://www.ijrset.com)

# Predicting Default Risk on Peer-to-Peer Lending Imbalanced Datasets: A Review

Prof.Tanvi.A.Ghodke, Parth Suryawanshi,Suyash, Nalwade,Sourav Yewale

Department of Computer Engineering, KJ College of Engineering and Management Research Pune, India

**ABSTRACT:** P2P lending's basic tenet is disintermediation, or getting rid of middlemen like banks. P2P lending is a fantastic approach to apply for a loan for small businesses and some people without sufficient credit or credit history. The information asymmetry in this approach, which might not accurately predict the default risk of loan, is the core issue with P2P lending, though. Due to the fact that lenders rely only on the information provided by borrowers in making funding decisions, P2P lending datasets are unbalanced and contain an uneven distribution of completely paid and defaulted loans. Real-world examples of imbalanced datasets include credit card fraud in transactions, subpar products in the factory, and other occurrences. Sadly, the uneven data are hostile to the traditional machine learning techniques. In our example, models devoid of any adaptive techniques would concentrate on understanding the typical payback.

**KEYWORDS**—:Peer-to-Peer lending, imbalanced datasets, re-sampling, machine learning.

## I. INTRODUCTION

P2P lending is a way to obtain credit without a financial institution like a bank being involved in the decision-making process and has the potential to obtain better terms than in the conventional banking system [1]. P2P lending also offers an online platform to connect borrowers and lenders directly. In order to avoid the brick-and-mortar operating costs, P2P lending can offer a lower interest rate for borrowers than that of banks. Therefore, P2P lending is an alternative to traditional banking.

Due to the fact that completely paid and default loans are not equal, peer-to-peer lending datasets are typically unbalanced. In our dataset, the proportion of fully paid to default loans is roughly 3.5:1. There are many unbalanced datasets in the real world, including those used for fraud detection, risk management, medical diagnosis, and other purposes. Because classifiers tend to discover the dominant class rather than the minority class, it is challenging to generate a prediction on such an unbalanced dataset. As a result, the classification's results will be biased. In this instance, it is crucial to solve the issue with the classification of the unbalanced dataset.

## II. RELATED WORK

**Amir F. Atiya**, -Prediction of bankruptcies with a focus on neural network (NN) models. Second, we create a model for predicting bankruptcy using NN.

We provide brand-new indicators for the NN system, which are inspired by one of the conventional credit risk models created by Merton. We demonstrate that using these indicators in addition to conventional financial ratio indicators significantly increases the prediction accuracy (out-of-sample) (from 81.46% to 85.5% for a three-year projection).

**R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu**- This study investigates the P2P loan features, assesses their credit risk, and measures loan performances using data from Lending Club, one of the well-known online P2P lending institutions. We discover that factors such as credit rating, debt-to-income ratio, FICO score, and revolving line usage are significant contributors to loan defaults. Loans with worse credit ratings and longer terms have higher death rates. The outcome is in line with the Cox Proportional Hazard test, which contends that as a borrower's credit risk increases, so does their likelihood of defaulting on a loan. Finally, we discover that higher interest rates offered to high-risk borrowers are insufficient to offset their increased chance of defaulting on their loans.

**Y. Jin and Y. Zhu**, - In this paper builds a data mining (DM) method is suggested to forecast P2P loan performance before it is funded. We examine the traits of the loan and its application using data from Lending Club, and we use

random forest to select the features for the modelling stage. The forecast is divided into three or four categories rather than simply two, the default and not default classes, which makes it different from previous risk prediction models. Then, using two metrics—average percent hit rate and area of the lift cumulative curve—we compare five DM models: two decision trees (DTs), two neural networks (NNs), and one support vector machine (SVM) to assess the accuracy of the predictions.

**A. Byanjankar, M. Heikkila, and J. Mezei**-In this paper due to the absence of knowledge on the creditworthiness of the borrowers, peer-to-peer lending carries a significant risk of investment failure. Peer-to-peer lending's credit risk is also increased by information asymmetry, the unsecured nature of loans, and a lack of strict norms and regulations. This study suggests a credit assessment approach for categorising peer-to-peer loan applications into default and nondefault groups using artificial neural networks. According to the findings, the neural network-based credit scoring model is efficient at identifying default applications.

**C. Serrano-Cinca and B. Gutiérrez-Nieto**-In this paper, Credit scoring programmes calculate the likelihood of loan default. Even though defaulted borrowers do not pay back the entire loan, some sums might be. The riskiest loans also have a high likelihood of default, but they also have high interest rates that can make up for late payments. We concentrate on forecasting the expected profitability of investing in P2P loans, measured by the internal rate of return, in contrast to past studies, which often seek to quantify the chance of default. This study examines 40,901 P2P loans in total. Analysis of the variables affecting loan profitability reveals that they are distinct from those affecting default risk. The findings indicate that the market for peer-to-peer lending is not yet fully efficient.

**Neelam Rout, Debahuti Mishra and Manas Kumar Mallick**-The definition of imbalanced data, examples of imbalanced data, various obstacles associated with handling imbalanced data, imbalance class issues, and performance analysis metrics for imbalanced data are covered in this work. The advantages and disadvantages of the various approaches are then summarised. Finally, examples of imbalanced data sets with IR values ranging from low to high are presented.

**H. Han, W. Y. Wang, and B. H. Mao**-This paper first discusses the significance of imbalanced data sets and the wide range of data mining applications they have, after which it provides an overview of the evaluation metrics and approaches currently used to assess and address the imbalance issue. One oversampling technique addressing this issue is synthetic minority oversampling technique (SMOTE). This study proposes two new minority over-sampling approaches, borderline-SMOTE1 and borderline-SMOTE2, which are based on the SMOTE method and which only oversample minority samples that are close to the borderline. Experiments demonstrate that our approaches outperform SMOTE and random oversampling strategies in terms of TP rate and F-value for the minority class.

### III. EXISTING SYSTEM/OPEN ISSUES

The underlying issue with peer-to-peer lending is the knowledge asymmetry in this paradigm, which may not accurately evaluate the default risk of lending. Due to the fact that lenders base their decision to fund a loan only on the information provided by borrowers, P2P lending data is unbalanced and contains an uneven distribution of fully paid and defaulted loans.

### IV. CONCLUSION

We apply a variety of machine learning algorithms to forecast the default risk of P2P lending and cost-sensitive processing techniques to handle imbalanced datasets. In order to validate our suggested approach, we obtain the dataset from Lending Club. According to the experiment's findings, random undersampling performs the best across all classifiers. The proposed scheme can therefore effectively increase the prediction accuracy for default risk after preprocessing and feature selection.

### REFERENCES

1. F. Atiya, "Bankruptcy prediction for credit risk using neural networks: A survey and new results," IEEE Trans. Neural Netw., vol. 12, no. 4, pp. 929–935, Jul. 2001



**|| Volume 12, Issue 5, May 2023 ||**

**| DOI:10.15680/IJIRSET.2023.1205110 |**

2. R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, “Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending,” *Appl. Econ.*, vol. 47, no. 1, pp. 54–70, Jan. 2015.
3. Y. Jin and Y. Zhu, “A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending,” in *Proc. 5th Int. Conf. Commun. Syst. Netw. Technol.*, Apr. 2015, pp. 609–613.
4. Byanjankar, M. Heikkila, and J. Mezei, “Predicting credit risk in peerto-peer lending: A neural network approach,” in *Proc. IEEE Symp. Ser. Comput. Intell.*, Dec. 2015, pp. 719–725.
5. Serrano-Cinca and B. Gutiérrez-Nieto, “The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending,” *Decis. Support Syst.*, vol. 89, pp. 113–122, Sep. 2016
6. N. Rout, D. Mishra, and M. K. Mallick, “Handling imbalanced data: A survey,” in *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications*. Singapore: Springer, Jan. 2018, pp. 431–443
7. H. Han, W. Y. Wang, and B. H. Mao, “Borderline-SMOTE: A new oversampling method in imbalanced data sets learning,” in *Advances in Intelligent Computing (Lecture Notes in Computer Science)*, vol. 3644. Berlin, Germany: Springer, Sep. 2005, pp. 878–887.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.423**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**9940 572 462** **6381 907 438** **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details