

Credit Risk Prediction for Peer-To-Peer Lending Platforms: An Explainable Machine Learning Approach

^{1*}Chong Pei Swee, ²Farid Meziane, and ³Jane Labadin

^{1,3}Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

²School of Computing and Engineering, University of Derby, UK

email: ^{1*}chongpeiswee@gmail.com, ²F.Meziane@derby.ac.uk, ³ljane@unimas.my

*Corresponding author

Received: 23 Jun 2022 | Accepted: 30 August 2022 | Early access: 19 September 2022

Abstract - Small and medium enterprises face the challenge of obtaining start-up fund due to the strict rules and conditions set by banks and financial institutions. The plight yields to the growth in popularity of online peer-to-peer lending platforms which are an easier way to obtain loan as they have fewer rigid rules. However, high flexibility of loan funding in peer-to-peer lending comes with high default probability of loan funded to high-risk start-ups. An efficient model for evaluating credit risk of borrowers in peer-to-peer lending platforms is important to encourage investors to fund loans and justify the rejection of unsuccessful applications to satisfy financial regulators and increase transparency. This paper presents a supervised machine learning model with logistic regression to address this issue and predicts the probability of default of a loan funded to borrowers through peer-to-peer lending platforms. In addition, factors that affect the credit levels of borrowers are identified and discussed. The research shows that the most important features that affect probability of default are debt-to-income ratio, number of mortgage account, and Fair, Isaac and Company Score.

Keywords: Credit Risk Evaluation, Peer-to-Peer Lending, Logistic Regression; Explainable Machine Learning; Explainable AI.

1 Introduction

Peer-To-Peer (P2P) lending platforms are online services provided by financial institutions as an intermediary to initiate loans for private individuals (Bachmann et al., 2011). Loans for borrowers are funded by multiple investors, bound with agreed-upon terms and conditions, with profits generated from the interest made on the loans as the borrowers are given a certain duration to pay back the loan and interest. The higher the investment risk, the higher is the interest rate. Due to a reduction in loans to small businesses from banks, P2P lending has gained popularity for personal, small business start-ups and SMEs loans as these tend to have high failing rate to pay back their loans and with low credit scores. Indeed, P2P lending allows individuals and businesses to loan money directly from investors or lenders without going through the strict requirements and criteria of traditional banks and financial institutions. Although these platforms provide several instruments to assess and limit credit risks, they do not guarantee the repayment of loans (Meyer, 2007).

The most common credit score for risks assessment is the “Fair, Isaac and Company” (FICO) score. The FICO score is not suitable for P2P lending since these platforms are meant for relatively high-risk start-ups, and for those that failed to secure loans from banks due to their low credit scores. Small and medium-sized enterprises (SMEs) which are categorized as high-risk client by financial institution play an important role in many economies, and to encourage their growth, a reliable and accurate clients’ credit risk evaluation is critical to build confidence among investors so that more funds are available on P2P lending platforms. This paper presents a supervised machine learning model that predicts the probability of default by considering more information related to the clients rather than just evaluating their credit score using FICO. The focus will be on solving the credit