

Machine Learning Based Ex-Ante Loan Default Prediction in Peer-to-Peer Lending using Logistic Regression

Shrinidhi T.P.V.* and Nigama K**

*School of Management,
SASTRA Deemed University, Thanjavur, India*

Abstract: Peer-to-peer (P2P) lending services are establishing themselves as a major alternative source of funding, although a lack of traditional underwriting processes increases credit risk. This paper provides an interpretable analytical structure for forecasting p2p loan default by using exclusively the credit-related and financial data available at the time of loan origination. To guarantee ex-ante prediction validity and stop information leakage, a sizable consumer loan dataset is handled using stratified sampling, feature selection and preprocessing. To predict default probabilities and investigate the impact of important borrower and loan parameters, a logistic regression model is used. Despite class imbalance, the experimental findings show good predictive performance with a ROC–AUC score of 0.702. Feature importance analysis indicates that interest rate, loan duration, and debt to income ratio are the most significant indicators, while stronger credit profiles are linked with lower default risk. Overall, the findings confirm that transparent and data-driven statistical models remain important for early-stage credit risk assessment in P2P lending environments and offer useful information for platform operators and lenders.

Keywords: P2P lending, Fintech, Loan Default Prediction, Logistic Regression, Credit Risk Assessment

JEL Classification Number: G21, G23, C25, C53

*Email: shrinidhitpv@gmail.com; **Corresponding author. Email: nigama@mba.sastra.edu