

Enhancing P2P Credit Scoring through Social Media Sentiment Analysis: A Hybrid Clustering and Logistic Regression Approach

Dr. Salma Arabi

University Hassan 1st, ENCG Settat, Morocco

Dr. Tarik Rachid

ESCA Business School, Casablanca, Morocco

Abstract

The rapid growth of peer-to-peer (P2P) lending platforms has significantly transformed access to credit, particularly for borrowers with limited or no formal banking history. Traditional credit scoring models, primarily based on historical financial data, often fail to capture the behavioral and emotional dimensions that influence repayment behavior. This study investigates the extent to which sentiment analysis derived from social media can complement conventional financial data to improve the prediction of repayment outcomes in the P2P lending context.

We utilize a dataset combining borrower information from LendingClub with publicly available social media content. Textual data are processed using advanced Natural Language Processing (NLP) techniques, including data cleaning, tokenization, lemmatization, and sentiment extraction. A hybrid methodological framework is proposed, integrating unsupervised clustering with an enhanced logistic regression model. The performance of this approach is compared with standard machine learning models, including Random Forest, XGBoost, and artificial neural networks.

The results demonstrate that models incorporating sentiment analysis outperform traditional approaches, achieving higher predictive accuracy and significantly reducing classification errors. Emotional and behavioral signals extracted from social media emerge as statistically significant explanatory variables. These findings have important practical implications for credit risk assessment, decision-making optimization, and financial inclusion. The study also outlines recommendations for responsible implementation and highlights future research directions, including multi-platform data integration and real-time sentiment monitoring.

Keywords

Peer-to-Peer Lending (P2P), Credit Scoring, Social Media Sentiment Analysis, Behavioral Finance, Natural Language Processing (NLP), Machine Learning, Credit Risk Prediction.