

ABSTRACT

Credit scoring is very helpful for lenders in evaluating the credit worthiness. Insufficient data availability, presents a challenge to P2P Lending companies to use alternative data as a substitute or complement to credit scoring modeling indicators. P2P Lending uses social media data to compensate for the lack of data. However, heterogeneous social media data is very difficult to identify the credit worthiness. With the least relevance between social media data with content related to credit scoring, it is necessary to adjust the theoretical approach and appropriate data processing. This study aims to determine the credit worthiness of someone using LinkedIn social media data. This study use data mining methods to perform prediction analysis with classification method using decision tree and random forest algorithm. The results of this study indicate that the combination of user demographic attributes and user generated content data produces the best credit scoring model with an accuracy value of 87.12%, with these results, P2P Lending has the opportunity to open new segments of consumers who do not have a complete credit history. Other findings also explain the data user generated content is more suitable to be used as a complementary data credit scoring model.

Keywords: Credit Scoring, Classification, P2P Lending, Social Media