# ABSTRACT

Today's social media can represent the behavior of its users, so it plays an important role in assessing one's credibility. The increasing number of social media users today makes the data generated more and more (big data) and there are also more opportunities to use this data for useful insights in various fields, such as in the financial sector, which is used by the Peer-to-Peer Lending platform in assessing creditworthiness. There needs to be the right method and approach in processing social media data because there is quite a lot of data that is less relevant on social media for creditworthiness. For this reason, this study aims to create a creditworthiness model using social media data from the LinkedIn platform in the form of account profile data which is represented as demographic attributes and textual upload data which represented as personality attributes based on The Big Five Personality theory.

To assess the creditor's characteristics in terms of creditworthiness, the 5C principles are used, which are Condition, Character, Capacity, Collateral, and Capital. This study proposes two of the total five credit principles, which are Condition and Character. Demographic attributes represents the Condition principle and personality attributes represents the Character principle.

The data collection in this study used the web scraping method on LinkedIn social media and the sample data used was 85 data. Each data is given a value according to the category classified into two groups, namely Eligible and Not Eligible. The model used is a prediction analysis method as classification method that exists in data mining with decision tree and random forest algorithms followed by performance evaluation using cross validation and confusion matrix methods.

This study shows that the creditworthiness assessment model with a combination of demographic attributes and personality attributes using random forest algorithm shows the highest performance evaluation value. The resulting accuracy value is 90.14%, the recall value is 82.93%, the precision value is 89.47%, and the f-measure value is 86.08%. Personality attributes can be considered as complementary data because they can increase the value of performance evaluation.

The data sources used in this study are still limited, so it is hoped that further research can use data sources from other social media as comparisons. In addition, the other social media features that represent other creditworthiness principles can be used. The result of this study can be a recommendation for financial institutions, especially Peer-to-Peer Lending, to be able to utilize data from social media in making innovative credit scoring as a determinant of creditworthiness.

Keywords: big data, big five personality, classification, creditworthiness, social media.