

Abstract

Social media is a forum where users can interact with other users and share information through communities and social networks. The large number of uploads from billions of social media users is a source of data to extract and create new information. The study begins with the distribution of a willingness form and a questionnaire to obtain approval from respondents who use Indonesian in their tweets to participate in this study. Agglomerative hierarchical clustering was chosen to enrich the methods for a person's personality prediction based on content on social media. The model in this study has an accuracy of 20.1% with an average silhouette score of -0.23. The high word uniqueness of each processed tweet is a challenge for this model to produce optimal performance. This model can handle large amounts of data in a short time but has not yet provided a more optimal performance than similar cases that were solved by supervised learning.

Keywords: social media, personality, prediction, methods, tweet