Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method

Arya Rafif Muhammad Fikri[1], Jondri[2], Widi Astuti[3]
School of Computing, Informatics Study Program, Telkom University, Bandung, Indonesia
Email: [1]aryarafif@student.telkomuniveristy.ac.id, [2]jondri@telkomuniveristy.ac.id, [3]widiwdu@telkomuniversity.ac.id
Email Penulis Korespondensi: jondri@telkomuniversity.ac.id

Abstract−Customer satisfaction is one of the factors that can be used to measure the success of service in a company. In the era of the 2000s until now, internet service providers have continued to grow throughout the world, including in Indonesia. IndiHome and First Media are companies that provide internet services that make it easy for the public to communicate and obtain information. With many uses of IndiHome and First Media internet services, there are often several obstacles that cause various responses from users. Users usually channel these responses to IndiHome or First Media customer care on Twitter. Sentiment analysis on Twitter needs to be done to see how IndiHome and First Media users respond to the internet services that have been provided. The first process in sentiment analysis is taking data from Twitter, then data cleaning and feature extraction are carried out, after that the data is classified using the ensemble stacking method. The dataset for this study was obtained from Twitter using the Twitter API and the *Tweepy library*. The dataset that has been collected is 6.962 *tweets* for the IndiHome dataset and 8,089 *tweets* for the First Media dataset. This study conducts sentiment analysis using the *Ensemble Stacking* with three base classifiers and a meta classifier. The base classifier used is *Naïve Bayes, K-Nearest Neighbor, and Decision Tree,* while the meta classifier used is *Logistic Regression.* This study uses the *term frequency-inverse document frequency* (TF-IDF) to determine the frequency value of a word in a document. This study uses two test scenarios: testing without *oversampling* and testing with *oversampling* on the dataset. The results show that *Ensemble Stacking* with *term frequency-inverse document frequency* feature extraction produces the highest accuracy, with an accuracy value of 88.27% on the IndiHome dataset and 92.56% on the First Media dataset by *oversampling* on both datasets.

Keywords: IndiHome;First Media;*Oversampling*;*Ensemble Stacking*;TF-IDF