

Abstrak

Selama pandemi Covid-19 banyak kegiatan sekolah maupun kantor yang dilakukan secara daring. Diperlukan jaringan internet untuk mendukung kegiatan tersebut, maka banyak orang yang berinisiatif untuk berlangganan layanan internet IndiHome. Banyak pelanggan yang merasa tidak puas dengan pelayanan IndiHome dengan menuliskan *tweet* di Twitter. Diperlukan *sentiment analysis* pada komentar *tweet* untuk mengevaluasi layanan IndiHome berdasarkan opini pelanggan. Dengan menggunakan metode *Multinomial Naive Bayes* diterapkan untuk klasifikasi sehingga menghasilkan akurasi optimal. Pada klasifikasi *Multinomial Naive Bayes* fitur diasumsikan diambil dari distribusi Multinomial sederhana. Proses analisis sentimen terdiri dari proses *labeling* data menggunakan *VADER Lexicon*, lalu *pre-processing*, selanjutnya akan dihitung jumlah *term frequency* dan *documen frequency*. Data akan dibagi menjadi data training dan testing menggunakan *K-fold Cross Validation* serta klasifikasi teks menggunakan *Multinomial Naive Bayes* dengan mencari nilai probabilitas tertinggi. Pengujian terdiri dari perbandingan jumlah data *training* dan data *testing* dengan menggunakan variasi *k-fold* untuk memperoleh nilai rata-rata akurasi terbaik terdapat pada variasi 11 *fold* yaitu sebesar 76,07%.

Kata kunci : *Multinomial Naive Bayes*, Klasifikasi, *Sentiment analysis*, *VADER Lexicon*

Abstract

During COVID-19 pandemic, many activities in schools or offices that use online method. Internet connection with high speed is needed to support these activities, thus, a lot of people initiated to subscribe internet service from Indihome. Many customers are not happy with services from Indihome by writing *tweets* on Twitter. *Sentiment analysis* is needed on *tweet* comments to evaluate Indihome's service, depend on customer's opinion. *Multinomial Naive Bayes* method is applied to classify so the result's accuracy is optimal. On *Multinomial Naive Bayes* classification, features are assumed to get from simple Multinomial distribution. The process of Sentiment Analysis is formed by data labeling process using VADER Lexicon, and then Pre-processing, thus the amount of term frequency and document frequency will be calculated. The data will divided into data training and testing using K-fold Cross Validation along with text classification using Multinomial Naive Bayes by searching the highest probability value. The testing consists of comparison of training data and testing data by using K-fold variation to get the mean of the best accurate value found in 11-fold variation that is 76.7%.

Keywords: *Multinomial Naive Bayes, Classification, Sentiment analysis, VADER Lexicon*

