

ABSTRACT

Twitter is a microblogging social media where the users can search for topics and discuss issues that occur at this time. Twitter allows users to publish short messages (tweets) easily and fast. Most short messages or tweets may contain opinions on products and services related to the community. It can be used as a source of data to assess the sentiment on Twitter.

Naïve Bayes classification is an approach which refers to the Bayes theorem, by combining prior knowledge with new knowledge. Data training used in the process of categorization obtained using KNIME Analytics Platform applications, where the tweet-making process is determined by twitter account services in the city of Bandung and sorted by sentiment category that is positive, negative, and neutral. The applications will focus on Indonesian language tweet, where the Indonesian has a specific morphology in the categorization process. This applications made through several stages in performing a tweet categorizing and one of them is preprocessing which is case folding, parsing, tranformation that is deleting stopwords and stemming, frequency calculations and Naïve Bayes probability.

Keywords: Twitter, KNIME, Naïve Bayes, Indonesian

ABSTRAK

Twitter merupakan salah satu media sosial *microblogging* dimana penggunaannya dapat mencari topik dan membahas isu-isu yang terjadi pada saat ini. *Twitter* membuat pengguna dapat mem-*publish* pesan pendek (*tweet*) dengan mudah dan cepat. Sebagian pesan pendek atau *tweet* dapat berisi opini terhadap produk maupun layanan yang berhubungan dengan masyarakat. Hal tersebut dapat digunakan sebagai sumber data untuk menilai sentimen pada *Twitter*.

Naïve bayes classification adalah pendekatan yang mengacu pada teorema *Bayes*, dengan mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Data latih yang digunakan pada proses pengkategorian didapat menggunakan aplikasi *KNIME Analytics Platform*, dimana pada proses pengambilan *tweet* ditentukan berdasarkan *acount twitter* dinas yang ada di kota Bandung dan dipilah berdasarkan kategori sentimennya yaitu positif, negatif, dan netral. Aplikasi akan fokus pada *tweet* berbahasa Indonesia, dimana Bahasa Indonesia mempunyai morfologi tertentu dalam proses pengkategorian. Pembuatan aplikasi ini melewati beberapa tahapan dalam melakukan pengkategorian *tweet* diantaranya *preprocessing* berupa *case folding*, *parsing*, *tranformation* berupa penghapusan *stopwords* dan *stemming*, perhitungan *frekuensi* dan probabilitas *naïve bayes*.

Kata kunci : *Twitter*, *KNIME*, *Naïve Bayes*, Bahasa Indonesia