

INTISARI

ANALISIS SENTIMEN BERDASARKAN TOPIK MENGGUNAKAN *LATENT DIRICHLET ALLOCATION (LDA) DAN NAÏVE BAYES CLASSIFIER (NBC)*

Oleh:

Sinta Fitrianingrum

14/364009/PA/15875

Perkembangan internet semakin pesat menyebabkan semakin meningkatnya pengguna media sosial. Salah satu media sosial yang populer adalah Twitter. Sebanyak 40% pengguna media sosial Twitter menggunakan media sosial mereka untuk mengakses berita. Tidak hanya membaca berita namun mereka juga bisa menanggapi berita tersebut sesuai opini mereka masing-masing. Dari opini-opini tersebut dapat dicari topik utama atau tren yang sedang dibicarakan serta bagaimana pandangan masyarakat terhadap tren tersebut. Oleh karena itu, dibutuhkan analisis sentimen yang dapat menilai sentimen dari opini yang disampaikan masyarakat. Dibutuhkan juga pemodelan topik untuk melihat tren topik yang sedang dibicarakan.

Proses analisis sentimen menggunakan algoritma *Naïve Bayes*, dan proses pemodelan topik akan menggunakan *Latent Dirichlet Allocation*. Data yang digunakan adalah data *tweet* yang di-*post* dan *me-mention* akun berita @bbcbrk, @cnbreaking, dan @nytimes. Data diambil sebanyak 5505 dan diambil mulai 2 April 2018 Hingga 8 April 2018.

Pengujian klasifikasi sentimen dilakukan dengan membandingkan Multinomial *Naïve Bayes* dan Bernoulli *Naïve Bayes*. Pengujian klasifikasi sentimen menggunakan *10-Fold Cross Validation* dengan iterasi sebanyak 10 kali. Dari hasil pengujian klasifikasi sentimen didapatkan akurasi terbesar dihasilkan oleh algoritma *Bernoulli Naïve Bayes* dengan rata-rata akurasi sebesar 85,29%. Pengujian pemodelan topik pada parameter *passes* dilakukan dengan melihat tren kestabilan *perplexity*. Pengujian pemodelan topik pada parameter jumlah topik dilakukan dengan melihat rata-rata *perplexity* terendah. Hasil pengujian didapatkan tren *perplexity* mulai stabil pada *passes* ke 10 dan rata-rata *perplexity* terendah pada saat jumlah topik sama dengan 35.

Kata Kunci: analisis sentimen, *naive bayes*, pemodelan topik, *latent dirichlet allocation*

ABSTRACT

SENTIMENT ANALYSIS BASED ON TOPICS USING *LATENT DIRICHLET ALLOCATION (LDA) AND NAÏVE BAYES CLASSIFIER (NBC)*

By:

Sinta Fitrianingrum
14/364009/PA/15875

As the development of the Internet increasingly rapidly led to the increasing social media users. One of the popular social media is Twitter. 40% of Twitter users use their social media to access news. Not only read the news but they can also respond to the news according to their respective opinions. From these opinions can be searched the main topics or trends being discussed and how people view the trend. Therefore, it takes a sentiment analysis that can assess the sentiments of the opinions conveyed by the community. It also takes topic modeling to see the trending topics being discussed.

The process of sentiment analysis uses the Naïve Bayes algorithm, and the topic modeling will process using Latent Dirichlet Allocation. The data used are tweet data which is post and mention to news account @bbcbrk, @cnnbreaking, and @nytimes. Data taken as many as 5505 and taken from 2 April 2018 Until 8 April 2018.

Sentimental classification testing is done by multiplying the Multinomial Naïve Bayes and Bernoulli Naïve Bayes. Testing sentiment classification using 10-Fold Cross Validation with iteration 10 times. From the result of classification of sentiments obtained the greatest accuracy produced by Bernoulli Naïve Bayes algorithm with average accuracy of 85.29%. Testing of topic modeling on parameter passes is done by looking at stability trends of perplexity. The topic modeling test on the topic number parameter is performed by looking at the lowest perplexity average. The test results showed a trend of perplexity starting to stabilize at the 10th passes and the lowest perplexity average when the number of topics equals 35.

Keyword: *sentiment analysis, naïve bayes, topic modeling, latent dirichlet allocation*