

ANALISIS SENTIMEN BERDASARKAN TOPIK TERKAIT WABAH COVID-19 DI TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA) DAN NAIVE BAYES CLASSIFIER (NBC)

Pangky Putra Aziztiya, Muhammad Habibi, Puji Winar Cahyo

INTISARI

Latar Belakang: Pada tanggal 11 Maret 2020, WHO menetapkan bahwa Virus Corona (COVID-19) sebagai pandemi. Selama berlangsungnya masa pandemi, wabah ini mengalami berbagai macam perubahan atau biasa disebut mutasi. Proses perubahan atau mutasi tersebut menghasilkan berbagai varian-varian jenis baru dengan tipe dan pola transmisi serta virulensa berbeda-beda. Meluasnya wabah COVID-19 secara global menjadikan media sosial Twitter menjadi salah satu alat yang paling banyak digunakan untuk mempublikasikan dan mencari informasi.

Tujuan: Penelitian ini bertujuan untuk membentuk pemodelan topik terkait wabah COVID-19 di platform media sosial Twitter dan menganalisis sentiment positif dan negatif di setiap topik yang sudah didapatkan dengan mengkombinasikan antara kedua metode Latent Dirichlet Allocation (LDA) dan Naïve Bayes Classification (NBC).

Metode Penelitian: Metode yang digunakan pada penelitian ini adalah Latent Dirichlet Allocation (LDA) yang dikombinasikan dengan Naïve Bayes Classifier (NBC)

Hasil: Dari hasil penelitian yang sudah dilakukan maka diperoleh hasil dari topic coherence dengan jumlah topik ideal adalah 5 topik. Dari kelima topik tersebut terdapat pembahasan yang berbeda-beda pada topik pertama membahas mengenai varian pada Covid-19, topik kedua membahas mengenai vaksinasi, topik ketiga membahas mengenai varian Covid delta dan omicron, topik keempat membahas mengenai vaksinasi booster sedangkan untuk topik kelima membahas mengenai vaksinasi booster tingkat kabupaten. Setelah melakukan proses pemodelan topik maka dapat dilakukan proses testing NBC dan mendapatkan nilai akurasi yang cukup baik yaitu 89%. Dan dapat diketahui topik yang paling banyak diperbincangkan terkait data covid-19 adalah mengenai vaksinasi booster dengan komposisi jumlah sentiment negatif lebih banyak dibandingkan dengan sentiment positif.

Kesimpulan: pada penelitian ini berhasil membuat lima topic modelling menggunakan metode Latent Dirichlet Allocation (LDA) dan mendapatkan nilai akurasi training dan testing yang baik melalui metode Naïve Bayes Classifier (NBC) sehingga pada penelitian ini berhasil mengkombinasikan antara kedua metode tersebut.

Kata-kunci: Corona Virus Disease, COVID-19, Vaksinasi, Omicron, Delta, Topic Modelling, Text Mining, Latent Dirichlet Allocation, NBC, Naïve Bayes Classifier.

SENTIMENT ANALYSIS BASED ON TOPICS RELATED TO THE COVID-19 OUTBREAK ON TWITTER USING LATENT DIRICHLET ALLOCATION (LDA) AND NAIVE BAYES CLASSIFIER (NBC)

Pangky Putra Aziztiya, Muhammad Habibi, Puji Winar Cahyo

ABSTRACT

Background: On March 11, 2020, WHO determined that the Corona Virus (COVID-19) was a pandemic. During the pandemic, this outbreak underwent various changes or commonly called mutations. The process of change or mutation produces various new types of variants with different types and patterns of transmission and virulence. The global spread of the COVID-19 outbreak has made Twitter one of the most widely used tools to publish and find information.

Objective: This study aims to form a modeling of topics related to the COVID-19 outbreak on the Twitter social media platform and analyze positive and negative sentiments in each topic that has been obtained by combining the two Latent Dirichlet Allocation (LDA) and Naïve Bayes Classification (NBC) methods. Beginning with modeling the topic using the Latent Dirichlet Allocation so that the topics that have been obtained will be searched for the sentiment value of each topic using the Naïve Bayes Classifier method.

Method: The method used in this study is Latent Dirichlet Allocation (LDA) combined with Naïve Bayes Classifier (NBC).

Result: From the research that has been done, the results obtained from topic coherence with the ideal number of topics are 5 topics. Of the five topics there are different discussions, the first topic discusses variants in Covid-19, the second topic discusses vaccination, the third topic discusses the delta and omicron variants of Covid, the fourth topic discusses booster vaccination, while the fifth topic discusses vaccination. district level boosters. After carrying out the topic modeling process, the testing and get a fairly good accuracy value of 89%. And it can be seen that the most discussed topic related to Covid-19 data is the booster vaccination with a composition of more negative sentiments than positive sentiments.

Conclusion: this study succeeded in making five topic modeling using the Latent Dirichlet Allocation accuracy values training and testing through the Naïve Bayes Classifier (NBC) method so that in this study succeeded in combining the two methods.

Keywords: Corona Virus Disease, COVID-19, Vaksinasi, Omicron, Delta, Topic Modelling, Text Mining, Latent Dirichlet Allocation, NBC, Naïve Bayes Classifier