JIUBJ

Jurnal Ilmiah Universitas Batanghari Jambi, 22(2), Juli 2022, 1205-1208
Lembaga Penelitian dan Pengabdian kepada Masyarakat universitas Batanghari Jambi
ISSN 1411-8939 (Online), ISSN 2549-4236 (Print)
DOI 10.33087/jiubj.v22i2.2397

Analisis Sentimen Distribusi Vaksin COVID-19 di Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*

Kevin Manurip*, Debi Irawan
Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI
*Correspondence email: manuripkevin@gmail.com, debi.irawan@email.unikom.ac.id

Abstrak. Sejak pemerintah Indonesia secara resmi mengumumkan kasus pertama COVID-19, konten media tradisional dan media sosial terkait COVID-19 meningkat drastis. Di satu sisi, ada pembicaraan media yang luas tentang pencegahan, pengenalan dan pencegahan gejala, pengenalan gejala dan pengobatan. Analisis sentimen, atau yang biasa disebut dengan opinion mining, adalah area penelitian yang menganalisis opini, perasaan, peringkat, penilaian, sikap, dan perasaan tentang suatu entitas dan diimplementasikan dalam konten media sosial. Hal ini menarik dan penting bagi pemangku kepentingan tertentu yang ingin mengetahui perasaan dan pendapat baik dan buruk masyarakat Indonesia terkait kegiatan pendistribusian vaksin untuk menghadapi COVID 19. Dari penelitian ini didapatkan tingkat performansi sistem yang dibangun untuk mengetahui keakuratan antara informasi yang diminta pengguna mengenai vaksin Sinovac terungkap dari total 1524 tweet, 819 tweet positif, 452 tweet netral, 253. Terdapat tweet negatif dari. Hasil klasifikasi vaksin AstraZeneca mengungkapkan total 211 tweet dengan emosi positif, 80 tweet dengan emosi netral, dan 31 tweet dengan emosi negatif. Hasil klasifikasi emosi berdasarkan data scraping menggunakan kata kunci vaksin Astrazeneca sebanyak 1266 tweet, dengan skor emosi positif 712 tweet, skor emosi netral 344 tweet, dan skor emosi negatif 210 tweet.

Kata Kunci: Klasifikasi; *Naïve Bayes Classifier;* Analisis Sentimen; *TextBlob;* Covid-19; Distribusi Vaksin

Abstract. *Since the Indonesian government officially announced the first case of COVID-19, traditional media and social media content related to COVID-19 has increased dramatically. On the one hand, the media talks about prevention, symptom recognition and about prevention, symptom recognition and treatment are massive. Sentiment Analysis or commonly called opinion mining, is a field of study that analyzes opinions, sentiments, evaluations, judgments, attitudes, and emotions towards entities and is implemented on social media content. This becomes interesting and important for certain parties who want to know the good and bad sentiments or opinions given by the Indonesian people towards the distribution of vaccines for the handling of COVID-19. From this research, the level of capability of the system that has been built to find the accuracy between the information requested by the user on the Sinovac vaccine results from a total of 1524 tweets, there are 819 positive tweets, 452 neutral tweets, and 253 negative tweets. The results of the AstraZeneca vaccine classification resulted in 211 tweets with a total of 100 positive sentiments, 80 tweets of neutral sentiment, and 31 tweets of negative sentiment. Sentiment classification results based on scraping data with the keyword Astrazeneca vaccine, resulted in 1266 tweets with a positive sentiment value of 712 tweets, neutral sentiment as many as 344 tweets, and negative sentiment as many as 210 tweets.*

Keywords: *Classification; Naïve Bayes Classifier; Sentiment Analysis; TextBlob, Covid-19; Vaccine Distribution*

PENDAHULUAN

Perkembangan kasus COVID-19 di Indonesia terus mengalami kenaikan, dengan terus adanya kenaikan kasus ini menyebabkan kepanikan di tengah masyarakat dengan ada dan tidaknya virus Corona. WHO melaporkan lebih dari 52 juta orang terkonfirmasi positif COVID-19 dan 1,2 juta orang meninggal dunia pada minggu kedua bulan November 2020 (Samsir et al., 2021). Wabah virus Corona telah membawa langkah-langkah yang belum pernah terjadi sebelumnya, yang memaksa pihak berwenang untuk membuat keputusan terkait dengan penerapan lockdown di beberapa daerah yang dilanda pandemi (Yulita et al., 2021). Sejak pemerintah Indonesia secara resmi mengumumkan kasus pertama COVID-19, konten media tradisional dan media sosial terkait COVID-19 meningkat drastis. Kepanikan masyarakat disebabkan sedikitnya oleh dua hal: Pertama, pandemi COVID-19 merupakan hal baru. Kedua, pandemi COVID-19 hadir ketika lanskap media berubah. Hal ini menyebabkan arus informasi melimpah serta informasi terkait pandemi COVID-19 yang kurang akurat meningkat menjadikan fenomena ini dengan istilah infodemik (Widyawati et al., 2021). Pemerintah Republik Indonesia telah mengeluarkan wacana vaksinasi untuk masyarakat Indonesia pada akhir tahun 2020 mendatang. Hal ini menjadi menarik dan penting bagi pihak-pihak tertentu yang ingin mengetahui baik buruknya sentimen atau opini yang diberikan masyarakat Indonesia terhadap kegiatan distribusian vaksin untuk penanganan COVID-19 ini.

METODE

Objek penelitian ini adalah trend di Indonesia saat ini yaitu distribusi vaksin COVID-19. Adapun subjek penelitian ini adalah perpsepsi masyarakat terhadap vaksin COVID-19 melaui review pada data tweet aplikasi Twitter. Populasi dalam penelitian ini adalah semua data review atau ulasan terhadap tweet mengenai vaksin COVID-19 dari tweet pada aplikasi Twitter. Sedangkan sampel diambil dari suatu populasi yang

mana pada penelitian ini digunakan sampel ulasan pengguna Twitter dengan rentang waktu dari bulan Januari 2021 sampai dengan bulan Februari 2021. Jumlah sampel sebanyak 3131 ulasan. Metode yang digunakan untuk mengumpulkan data dari penelitian ini antara lain sebagai berikut:

1. *Scraping Data Twitter*. Data pada penelitian ini adalah data sekunder yang diambil secara *online* dengan bantuan *advance search* Twitter.
2. Studi Pustaka. Studi pustaka dilakukan dengan menggunakan beberapa kajian literatur, buku, maupun referensi jurnal yang sekiranya berkaitan dengan tujuan, rumusan, batasan, dan metode penelitian.

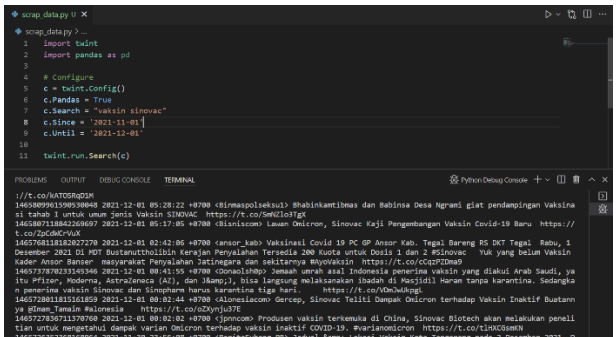
Jenis data yang digunakan pada penelitian ini adalah data sekunder, hal ini dikarenakan peneliti mendapatkan data melalui media perantara. Pada penelitian ini data sekunder berperan sebagai main data dalam proses pengolahan untuk mengetahui persepsi masyarakat terhadap vaksin COVID-19. Adapun sumber data yang digunakan dalam penelitian ini adalah dataset reviews terhadap topik tersebut yang diambil dari website aplikasi Twitter. Metode *Naïve Bayes* adalah metode klasifikasi untuk penambangan teks yang digunakan dalam analisis sentimen. Pendekatan ini secara teoritis baik dalam hal konsistensi data dan klasifikasi perhitungan. Algoritma *Naïve Bayes Classifier* umumnya digunakan dalam teknik klasifikasi, khususnya Twitter, menggunakan berbagai metode seperti *Unigram Naïve Bayes, Multinomial Nave Bayes*, dan *Maximum Entropy Classification*. Fitur utama dari klasifikasi *Naïve Bayes* adalah untuk mendapatkan hipotesis yang kuat dari setiap kondisi atau peristiwa (Pristiyono et al., 2021). Dicatat bahwa *Naïve Bayes* tampil lebih akurat dari algoritma *Support Vector Machine*, K-*Neares Neighbor* dalam menentukan polaritas sebenarnya dari kalimat yang disajikan, bahkan dalam kumpulan data yang tidak seimbang.

Metode analisis data yang digunakan dalam penelitian ini, sebagai berikut:

1. Metode Machine Learning yang digunakan adalah algoritma *Naïve Bayes Classifi*er yang diimplementasikan untuk melakukan klasifikasi ulasan yang berbentuk kelas positif maupun negatif.
2. Bahasa Pemrograman Python untuk melakukan *scraping* data Twitter berdasarkan *hashtag*, kata, dan waktu tertentu.
3. Aplikasi WEKA, digunakan untuk memvisualisasikan data-data hasil dari analisis untuk melihat secara detail sentimen publik berdasarkan label yang sudah dibuat, yaitu: positif dan negatif.

HASIL DAN PEMBAHASAN

Penelitian Analisis Sentimen ini diimplementasikan pada platform Jupyter Notebook sebagai Text Editor untuk melakukan compile code dalam menjalakan proses sentimen analisis. Implementasi metode ini dimulai dengan tahap mengumpulkan list library yang akan digunakan dalam proses analisis sentimen, seperti: pandas, nltk, string, dan textblob. Pengumpulan data training dilakukan dengan metode Scraping Data Twitter dengan menggunakan TWINT (Twitter Inteligent Tool). Proses ini dilakukan dengan menggunakan project tersebut karena dapat menggamil dari Twitter API dengan batas atas 3200 tweets, dapat melakukan inisialisasi dengan cepat, dapat digunakan secara anonim tanpa perlu mendaftar ke Twitter. *Library Pandas* untuk memvisualisasikan data tahap pertama serta menetapkan tanggal pengambilan data pada tanggal 1 November 2021 sampai 1 December 2021. Setelah data-data berhasil ditampilkan, maka selanjutnya simpan data tersebut ke dalam *Microsoft Excel* dengan menggunakan ektensi .xlsx ataupun .csv bila diperlukan. Data yang dikumpulkan adalah data mentah dan belum melewati *preprocessing stage* (tahap pra-pemrosesan). Langkah selanjutnya adalah membersihkan *noise* dari setiap kalimat tweet agar memudahkan berikutnya, menerjemahkan ke dalam bahasa Inggris agar memudahkan *library TextBlob* melakukan proses mendeteksi kata-kata ke dalam 1206entiment menjadi lebih mudah



Sumber: data olahan

Gambar 1. Scraping Data Twitter dengan TWINT Project

Setelah data tweet melewati prapemrosesan maka akan dilanjutkan dengan melakukan klasifikasi masing-masing tweet. Langkah pertama dengan membuat algoritma sederhana berdasarkan fungsi yang diambil dari *package TextBlob* dan *NaiveBayesClassifier*. Selanjutnya proses penyaringan data tweet ke dalam dua bagian fungsi, yaitu: mencari subjektifitas dan menentukan polaritas dengan mengunakan library *TextBlob NaiveBayesClassifier*. *Package NaiveBayesClassifier* bertujuan untuk menyediakan akses ke aktivitas pemrosesan teks umum melalui antarmuka yang sudah memiliki database kata-kata bersentimen. Skor polaritas memiliki tipe data float [-1.0, 1.0] berikut dengan subjektifitas memiliki tipe data float dengan nilai [0.0, 1.0] di mana nilai 0.0 sangat

objektif dan 1.0 sangat subjektif. Setelah mengetahui nilai subjektifitas, maka polaritas akan memiliki nilai yang menjadi acuan bagi sentiment (*textblob*) untuk menentukan positif, netral, ataupun negatif. Pada dasarnya analisis sentimen di-implementasikan dengan metode *machine learning*. Data dalam bentuk *text*, contohnya tweet yang masuk akan dipilah-pilah terlebih dahulu. Proses tersebut disebut juga sebagai tokenisasi. Tokenisasi dilakukan untuk mempermudah proses analisis dari sebuah kalimat text. Setelah itu, sentimen dari input dapat ditentukan dengan cara mengklasifikasikan kata-kata yang sudah dipisahkan sebelumnya dengan *sentiment lexicon*, sehingga memunculkan polaritas dan subjektifitas pada *tweet* yang ada.

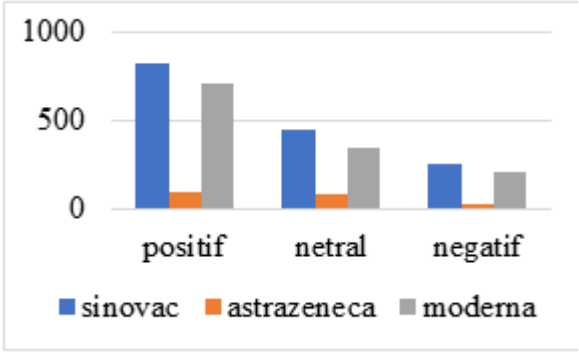
Klasifikasi sentimen berdasarkan data-data yang diambil dari tiga kali pengambilan data (scraping data twitter) dilihat dari tiga jenis vaksin yang telah beredar dan didistribusikan oleh pemerinta pada rentang waktu 1 November 2021 – 1 Desember 2021 yaitu Coronavac-Sinovac, Oxford-AztraZeneca, dan Booster Moderna dengan masing-masing berjumlah: Sinovac (1524 tweet), AstraZeneca (211), dan Moderna (1266) menghasilkan nilai akhir pengolahan data sebagai berikut:

Tabel 1. Jumlah data sentimen vaksin 1 November – 1 Desember, 2021

Vaksin	Positif	Netral	Negatif
Sinovac	819	452	253
AstraZeneca	100	80	31
Moderna	712	344	210

Sumber: data olahan

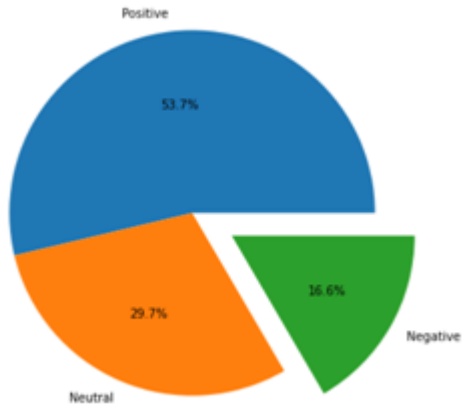
Melihat statistik tabel di atas dapat menyimpulkan secara pragmatis dan kasat mata bahwa distribusi vaksin COVID-19 dengan mengusung tiga buah vaksin, yaitu: Sinovac, AstraZeneca, dan Moderna (*booster*) di Indonesia mayoritas menuai kesan yang baik (dilihat dari diagram batang sentimen positif). Tabel 2 diagram batang menyimpulkan secara pragmatis dan kasat mata bahwa distribusi vaksin COVID-19 dengan mengusung tiga buah vaksin, yaitu: Sinovac, AstraZeneca, dan Moderna (booster) di Indonesia mayoritas menuai kesan yang baik (dilihat dari diagram batang sentimen positif).



Sumber: data olahan

Gambar 1. Grafik Masing-Masing Jenis Vaksin

Berikutnya dilanjutkan dengan menyimpulkan *DataFrame* ke dalam bentuk diagram lingkaran mengandalkan library *matplotlib*. Hasil diagram lingkaran dari masing-masing vaksin memiliki tingkat presentase yang cukup unik.



Sumber: data olahan

Gambar 2. Diagram Lingkaran Vaksin Sinovac

SIMPULAN

Setelah melakukan berbagai macam pengujian pada sistem, dapat ditarik kesimpulan sebagai berikut:

1. Metode Naïve Bayes Classification pada TextBlob dapat memvisualisasikan dan memberikan hasil yang cukup akurat berdasarkan variabel yang sudah ditetapkan yaitu positif, netral, dan negatif.
2. Akurasi tingkat analisis algoritma Naïve Bayes Classifier cukup baik terkait data-data yang telah dikumpulkan dengan jumlah total data 3001 tweets yang diambil pada rentan waktu 1 November 2021 - 1 Desember 2021, walaupun tidak banyak dikarenakan vaktor semakin banyak data akan lebih mempengaruhi tingkat akurasi.
3. Metode scraping dan dilakukan analisis dengan bantuan library TextBlob—Naïve Bayes Classifier untuk masing-masing data vaksin dengan kata scraping: ‘vaksin sinovac’, ‘vaksin astrazeneca’, dan ‘vaksin moderna’ menghasilkan data yang cukup sesuai.

DAFTAR PUSTAKA

Pristiyono, Ritonga, M., Ihsan, M. A. Al, Anjar, A., & Rambe, F. H. 2021. Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. IOP Conference Series: Materials Science and Engineering, 1088(1), 012045. https://doi.org/10.1088/1757-899x/1088/1/012045

Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. 2021. Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5, 157–163. https://doi.org/10.30865/mib.v5i1.2604

Widyawati, N., Santoso, W. M., Maulana, I., Daraini, I. N., Windarsih, A., & Alamsyah, P. 2021. COVID-19 in Twitter: Issues and Actors. *Jurnal Masyarakat dan Budaya*, 23(1), 105–125. https://doi.org/10.14203/jmb.v23i1.1287

Yulita, W., Nugroho, E. D., & Algifari, M. H. 2021. Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. 2(2), 1–9.