

# ANALISIS SENTIMEN COVID 19 BERDASARKAN KOMENTAR PADA MEDIA SOSIAL TWITTER MENGGUNAKAN PHYTON

Sartika Lina Mulani Sitio<sup>1</sup>

Program Studi Teknik Informatika

<sup>1</sup>Universitas Pamulang, Alamat, Jl. Raya Puspitex No.46, Buaran Tangerang Selatan

\*E-mail: dosen00847@unpam.ac.id

## ABSTRAK

Berdasarkan data yang diperoleh (JHU, 2020), per 15 Maret 2020, epidemi COVID-19 telah mencapai 1.986.986 kasus, jumlah orang yang terinfeksi mencapai 1.986.986, dan jumlah kematian mencapai 126.812. Di Indonesia sendiri tercatat 4.839 kasus dengan 459 kematian per 15 April 2020 (BNPB, 2020). Efek COVID-19 yang meluas, menimbulkan kekhawatiran yang sangat tinggi bagi kehidupan masyarakat. Analisis sentimen digunakan untuk menganalisis komentar individu pengguna Twitter terkait Covid19 dengan menggunakan tagar seperti #coronavirus dan #covid19. Tujuan dari penelitian ini adalah untuk menganalisis suasana hati berdasarkan pendapat individu dan mengklasifikasikannya menjadi komentar positif, netral dan negatif. Data diproses menggunakan tweet dari Twitter API dan diproses, dianalisis dan divisualisasikan sebagai tweet menggunakan Python. Tweet yang diterima diproses terlebih dahulu untuk mengambil data tweet, yang digunakan untuk data selanjutnya. Tahapan pra-pemrosesan yang dilakukan meliputi pembersihan, reduksi data, integrasi data, dan proses konversi data. Hasil yang diperoleh menunjukkan bahwa polaritas emosi yang digunakan dapat memberikan informasi dan mengolah data sesuai dengan komentar dari masing-masing komunitas dengan nilai polaritas tertinggi pada status netral 58,94% yaitu polaritas status positif sebesar 33,68% dan negatif 7,36%.

**Kata kunci:** Analysis, Sentimen, Twitter, Python, Preprocessing.

## ABSTRACT

Based on data obtained (JHU, 2020), as of March 15, 2020, the COVID-19 epidemic has reached 1,986,986 cases, the number of infected people has reached 1,986,986, and the death toll has reached 126,812. In Indonesia alone, there were 4,839 cases with 459 deaths as of April 15, 2020 (BNPB, 2020). The widespread effects of COVID-19 have caused very high concern for people's lives. Sentiment analysis is used to analyze individual Twitter users' comments regarding Covid19 using hashtags such as #coronavirus and #covid19. The purpose of this study was to analyze moods based on individual opinions and classify them into positive, neutral and negative comments. The data is processed using tweets from the Twitter API and processed, analyzed and visualized as tweets using Python. Received tweets are processed first to retrieve tweet data, which is used for further data. The pre-processing stages include cleaning, data reduction, data integration, and data conversion processes. The results obtained indicate that the emotional polarity used can provide information and process data according to comments from each community with the highest polarity value at neutral status 58.94%, namely positive status polarity 33.68% and negative 7.36%.

**Keywords:** Analysis, Sentimen, Twitter, Python, Preprocessing.

## 1. PENDAHULUAN

Media sosial adalah salah satu contoh terbesar penghasil data internet. Jutaan orang dan organisasi menggunakan media sosial sebagai wujud eksistensinya di dunia maya (Sivarajah, 2019). Data Twitter adalah salah satu sumber data penelitian yang paling sering digunakan (Cheng, 2013) dan menjelaskan mengapa Twitter digunakan sebagai sumber informasi: data dan

konten dari berbagai sumber, lampiran, komunikasi waktu nyata. Data besar dihasilkan secara real time dengan menciptakan hubungan sosial antara pengguna dan berbagi informasi dan acara. COVID-19, juga dikenal sebagai virus corona, sangat terkenal di seluruh dunia pada awal tahun 2020.

Berdasarkan (JHU, 2020), sebaran global kasus COVID-19 per 15 Maret 2020 mencapai 1.986.986 kasus terkonfirmasi dan 126.812 meninggal. Sementara itu, Indonesia mencatat 4.839 kasus per 15 April 2020 dengan 459

kematian (BNPB, 2020). Efek COVID-19 meluas dan dapat memengaruhi perilaku masyarakat selama pandemi. Sementara aspek kemanusiaan dan keamanan dari pandemi menjadi perhatian utama di seluruh dunia, pembatasan jarak sosial, karantina, dan perintah tinggal di rumah dapat berdampak signifikan pada konsumsi media dan dapat meningkat hingga 60%. Sebuah survei baru oleh tim media AS (Nielsen, AS, 2020). Data yang dikumpulkan dalam penelitian ini menggunakan tweet dari API Twitter untuk memproses ulang, menganalisis, dan memvisualisasikan tweet tersebut dengan Python. Aliran data Twitter kemudian diproses dan disanitasi untuk menganalisis data yang dapat diklasifikasikan berdasarkan opini menggunakan algoritma penambangan teks menggunakan gumpalan teks Python. Ekstraksi fitur dilakukan pada hubungan antar kata menggunakan metode bigram dan n-gram.

Data yang dihasilkan oleh Twitter penting dan berguna bagi komunitas atau organisasi ketika diproses dan dianalisis. Salah satunya adalah analisis sentimen Twitter. Ini adalah metode yang dapat Anda gunakan untuk menyelidiki apakah konten yang dikumpulkan positif, negatif, atau netral. Analisis sentimen diterapkan pada penelitian ini untuk menganalisis opini individu tentang peristiwa pandemi COVID-19 dalam data Twitter yang membahas topik COVID-19. Penyebaran pandemi ini ke seluruh dunia dan pemberlakuan pembatasan interaksi sosial telah mempengaruhi situasi sosial masyarakat, merebaknya masalah positif dan negatif, bahkan kepanikan media (Muhardi, 2020).

## 2. TINJAUAN PUSTAKA

Analisis sentimen, atau yang biasa dikenal dengan opinion mining, adalah bidang penelitian text mining yang bertujuan untuk menentukan suatu masalah, peristiwa, atau persepsi umum (audiens) atau subjektivitas terhadap suatu masalah. Analisis sentimen adalah tugas klasifikasi yang menempatkan teks dalam bias positif atau negatif (18). Secara teknis, analisis sentimen dapat dibagi menjadi empat pendekatan: pembelajaran mesin, pendekatan berbasis kamus, pendekatan berbasis aturan, dan pendekatan pemodelan statistik (19). Perolehan data adalah proses mengumpulkan dan mengukur informasi secara sistematis tentang variabel sasaran. Pengumpulan data yang akurat penting untuk menjaga integritas penelitian. Pembersihan data adalah proses mengidentifikasi informasi yang tidak lengkap atau tidak relevan dan memodifikasi atau menghapus data yang kotor (Saini, 2019). TextBlob TextBlob adalah pustaka Python (2 dan 3) untuk memproses data teks. Ini menyediakan API sederhana untuk melompat ke tugas NLP

(Natural Language Processing) umum seperti: B. Bagian dari penandaan ucapan, ekstraksi frase kata benda, analisis sentimen, klasifikasi, terjemahan, dll. Pustaka Python untuk memproses data tekstual. Ini menyediakan API yang konsisten untuk menganalisis tugas NLP (Natural Language Processing) umum seperti penandaan bagian-ucapan, ekstraksi kata benda, dan analisis sentimen (Praneeth, 2020).

## 3. METODE

### 3.1 Metode Pengumpulan Data

Data tanggapan dan opini masyarakat Indonesia mengenai vaksin COVID-19 dikumpulkan menggunakan teknologi web scraping berupa tweet di media sosial Twitter. Aktivitas scraping web dilakukan menggunakan API Twitter dengan paket "rtweet"[3]. "Rtweet" adalah paket yang dikembangkan untuk mengumpulkan dan mengelola data Twitter menggunakan API Twitter R-Application[2][4]. Dua kata kunci yang digunakan untuk mengumpulkan tanggapan dan opini umum tentang vaksin COVID-19 dalam proses web scraping adalah "Covid 19" dan "Corona 19". Kata kunci yang digunakan diyakini mampu menjangkau opini seluruh masyarakat Indonesia tentang vaksin COVID-19 di media sosial Twitter. Tweet yang dikumpulkan adalah tweet yang diposting di media sosial Twitter antara tanggal 25 Oktober hingga 3 November 2020 karena pembatasan pendataan.

### 3.2 Metode Persiapan Data

Dari tweets yang terambil, dilakukan penyaringan data/filter data dengan cara menghapus tweets yang bersumber dari akun selain akun masyarakat. Kegiatan ini mengacu kepada penelitian yang dilakukan oleh Luong & Houston, 2018, yang mengklasifikasikan akun tweets ke dalam akun pemerintah, layanan, sekolah, perusahaan, agen transportasi umum, dan masyarakat. Pada penelitian ini, peneliti menghapus tweets yang bersumber dari akun selain akun masyarakat, dengan cut off jumlah tweets sebanyak 20 tweets. Selanjutnya, dilakukan kegiatan persiapan data/preprocessing untuk mempersiapkan data agar siap untuk dianalisis. Kegiatan preprocessing tersebut meliputi tahapan pembersihan tweets terhadap unsur-unsur yang tidak dibutuhkan dalam analisis, yaitu delete duplicate atau menghapus tweets respon & opini masyarakat yang sama persis, delete URL atau menghapus link yang terdapat pada tweets, menghapus mentions & hastags, menghapus emoji, menghapus punctuation, melakukan normalisasi kata, menghapus kata yang tidak penting (stopwords removal), dan

mengubah format tulisan menjadi huruf kecil. Tahapan normalisasi kata berguna untuk mentransformasi kata singkatan, typo, dan kata berlebih menjadi sebuah kata yang formal. Kamus yang digunakan tersebut menggunakan 'kamus alay' yang dibuat oleh Salsabila, 2018. Penghapusan kata yang tidak penting (stopwords) berguna untuk mengurangi waktu sistem dalam merunning data. Kamus stopwords yang digunakan pada penelitian ini menggunakan kamus yang dibuat oleh Tala, 2013. Tahapan preprocessing merupakan tahapan yang paling penting dalam penelitian yang menggunakan data hasil text mining, karena pada tahapan ini sangat menentukan hasil analisis yang akan didapatkan.

### 3.3 Analisis Data

Adapun metode analisis data yang dilakukan adalah sebagai berikut[1]:

#### a. Analisis Sentimen

Kegiatan klasifikasi tweet atau analisis sentimen dalam penelitian ini dilakukan dengan menggunakan metode berbasis leksikon atau kamus positif-negatif. Kamus positif dan negatif yang digunakan dalam penelitian ini adalah kamus positif dan negatif yang dibuat oleh Liu, Hu & Cheng. Telah diterjemahkan ke dalam bahasa Indonesia dan telah disesuaikan secara bahasa. Kamus ini sebelumnya diterapkan pada beberapa penelitian sebelumnya

#### b. Mengelompokkan komentar masyarakat dari twitter

Gambar 1 menunjukkan model survei yang diusulkan dalam survei ini. Tahap awal adalah proses streaming data API Twitter, yang membutuhkan empat kunci untuk mengakses: kunci konsumen, kunci rahasia konsumen, kunci token akses, dan kunci rahasia token akses. Variabel # coronavirus # COVID-19 digunakan sebagai istilah pencarian. Dalam penelitian ini, kami melakukan pengumpulan data 4000 tweet dari data Twitter untuk menganalisis sentimen pengguna Twitter menggunakan variabel COVID-19 yang merupakan pandemi virus corona menggunakan Python. Kumpulan data tersebut kemudian disimpan dalam database mongodbnosql. Pengumpulan data yang disimpan dalam database diproses dengan pembersihan data sebelum data dianalisis. Proses ekstraksi fitur dimulai dengan menganalisis 100 kata teratas yang biasa ditemukan dalam tweet dan mencari kumpulan kata dalam tweet yang telah melalui proses pembersihan. collection.Counter: Menampilkan kata yang paling sering digunakan dan seberapa sering kata tersebut digunakan. Juga, nilai hubungan antara 100 kata terbaik dihitung menggunakan bigram. Perhitungan hubungan antara dua kata. Klasifikasi analisis

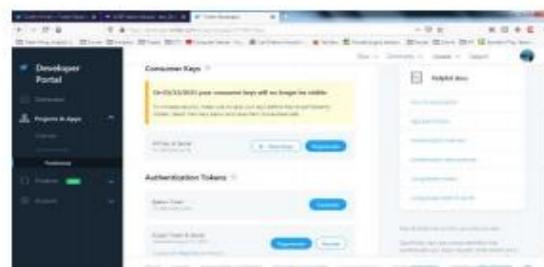
sentimen menggunakan sifat-sifat emosi dari textblob mengidentifikasi sikap data teks pada topik yang diminati dalam hal emosi (polaritas, subjektivitas). Skor polaritas mengambang dalam kisaran -1.0 hingga 1.0. Nilai yang lebih dekat ke 1 menunjukkan lebih banyak positif dan nilai yang lebih dekat ke -1 menunjukkan lebih banyak negatif.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Implementasi

#### 4.1.1 Halaman Informasi Keys dan Token

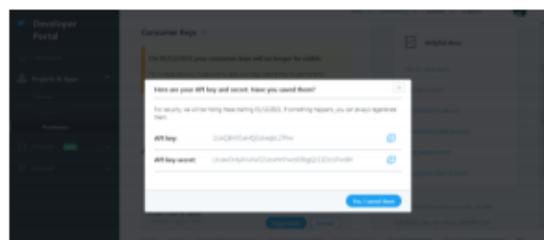
Halaman ini memberikan informasi yang diperlukan mengenai kunci konsumen dan token akses. Kunci API unik untuk setiap aplikasi yang Anda buat. Token akses dibuat kapan pun Anda perlu melihatnya.



Gambar 1. 1 Halaman Informasi Keys dan Tokens

#### 4.1.2 Halaman Melihat Api Keys dan Token

Setelah divalidasi oleh Twitter, aplikasi dapat menjalankan proses crawl data dengan parameter tertentu. Parameter yang perlu Anda masukkan termasuk kata kunci dan jumlah tweet. Data Twitter yang dapat diperoleh dalam waktu maksimal satu bulan.

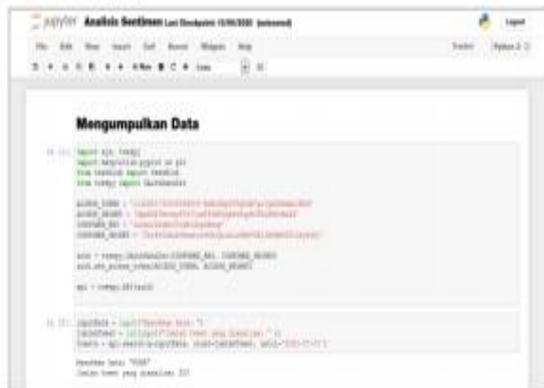


Gambar 1. 2 Halaman untuk melihat API Keys dan Tokens

#### 4.1.3 Proses Pengumpulan Data Twitter

Pada tahap pertama, data diolah terlebih dahulu. Prosesnya dimulai dengan pemenggalan kepala setiap kata. Proses penyaringan kemudian terdiri dari menghilangkan kata-kata yang tidak berarti, simbol, dan berbagai karakter lainnya.

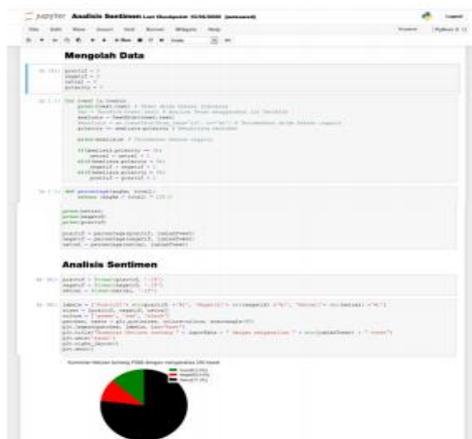
Python juga menyediakan data stopwords khusus untuk bahasa Indonesia. Stopwords adalah kumpulan kata-kata yang tidak berarti dan dihapus atau dihilangkan oleh proses.



Gambar 1.3 Pengumpulan Data Twitter

#### 4.1.4 Klasifikasi Data Twitter

Proses klasifikasi ini dilakukan dengan menghasilkan model yang dapat digunakan oleh kelas emosi untuk mengklasifikasikan data baru yang belum tersedia. Untuk visualisasi output yang lebih baik, hasil klasifikasi ditampilkan dalam bentuk diagram lingkaran.

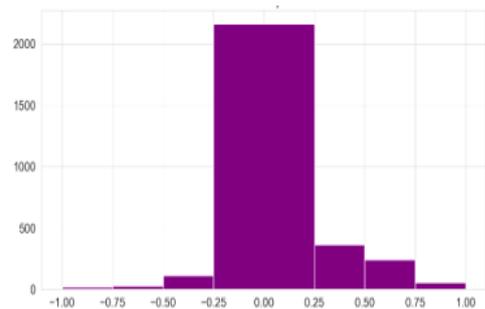


Gambar 1.4 Klasifikasi Sentimen Data Twitter

#### 4.1.5 Analisis Sentimen Covid 19

Nilai polaritas pada analisa sentiment ini antara 1 sampai -1. Opini dengan nilai polaritas mengarah ke nilai 1 menunjukkan opini dengan status positif, sedangkan polaritas mengarah ke nilai -1 menunjukkan negative, untuk opini netral nilai polaritas berkisar ke 0. Pada Gambar dibawah ini ditampilkan grafik histogram dari nilai polaritas tweet COVID-19, nilai polaritas tertinggi pada status netral 58,94% yaitu polaritas status positif sebesar 33,68% dan

negatif 7,36%.



Gambar 1.5 Analisis Sentimen Covid 19

## 5. KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Kesimpulan yang dapat diambil dari survei ini adalah media sosial sebagai sarana berkomentar oleh masyarakat dapat menjadi sumber data yang bermanfaat jika diolah dengan baik dan benar. Temuan survei menunjukkan bahwa opini publik tentang kebijakan pemerintah di media sosial cenderung positif.

### 5.2 Saran

Usulan penelitian selanjutnya yang akan dikembangkan adalah Penelitian lebih lanjut dapat dilakukan pada pengembangan algoritma atau pemilihan tugas lain dalam data mining. Misalnya, sehubungan dengan analisis estimasi atau prediksi kebijakan yang mungkin dihasilkan dari kebijakan lain.

## DAFTAR PUSTAKA

- [1] Ge, Z., Song, Z., Ding, S.X. and Huang, B., (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, 5, pp.20590-20616.
- [2] Han, J. dan M. Kamber. (2006). *Data Mining: Concepts and Techniques*, Second Edition. Morgan Kaufmann Publishers. San Francisco
- [3] Hikmawan, S., Pardamean, A., & Khasanah, S. N. (2020). Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, 20(2), 167-176.
- [4] Lisbach, V. M. (2013). *Linguistic Identity Matching*. [www.springer.com](http://www.springer.com).
- [5] Maharani, E. M. (2016). Perancangan Sistem Analisis Sentimen Komentar Pelanggan Menggunakan Metode Naive Bayes Classifier.
- [6] Muhandi, A. B. (2020). Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain.