

**SENTIMEN ANALISIS OPINI MASYARAKAT PADA MEDIA SOSIAL
TWITTER TERHADAP VAKSIN BERBAYAR MENGGUNAKAN METODE
NAÏVE BAYES CLASSIFIER (NBC)**

*SENTIMENT ANALYSIS OF PUBLIC OPINION ON SOCIAL MEDIA TWITTER
ON PAID VACCINE USING NAÏVE BAYES CLASSIFIER (NBC) METHOD*

**¹Angga Aditya Permana, ²Muhammad Fany Fahrezi, ³Dyas Yudi Priyanggodo, ⁴Dinar
Ajeng Kristiyanti, ⁵Manorang Sihotang**

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang Jalan
Perintis Kemerdekaan 1/33 Cikokol Kota Tangerang

⁴ Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika,
Jl. Kramat Raya No. 98, Senen, Jakarta Pusat, Indonesia

⁵ Universitas Pramita Indonesia, Jl. Kampus Pramita, Tangerang 15810, Indonesia
Co Responden Email: anggaumt@gmail.com

Receive: 27 Desember 2021

Accepted: 30 Desember 2021

Abstract

The existence of a paid vaccine plan by the government has generated various responses from the public. The number of notifications that come from many sources can affect people's perceptions. The negative perception of paid vaccines experienced by the community can be a trigger for anger. The purpose of this study is to conduct sentiment analysis using R programming, to find out public opinions on the news about paid vaccines in Indonesia. This study uses the naive Bayes classifier method which is classified into positive, negative, and neutral opinions. The results of this sentiment analysis will be described in the form of barplot visualization, word cloud, and in the form of UML (Unified Modeling Language) diagrams then the process will be continued by making a system design. This sentiment analysis is web-based and can be accessed using a browser. That way this research will produce a sentiment analysis information using rstudio in retrieving and classifying data.

Keywords : analysis, sentiment, vaccine, paid, rstudio

Abstrak

Adanya rencana vaksin berbayar oleh pemerintah menimbulkan berbagai macam tanggapan dari masyarakat. Banyaknya pemberitahuan yang berasal dari banyak sumber dapat mempengaruhi persepsi masyarakat. Persepsi negatif terhadap vaksin berbayar yang dialami masyarakat dapat menjadi pemicu terjadinya kemarahan. Tujuan penelitian ini adalah untuk melakukan analisis sentimen menggunakan pemrograman R, untuk mengetahui opini-opini masyarakat terhadap kabar mengenai vaksin berbayar yang ada di Indonesia. Penelitian ini menggunakan metode naïve bayes classifier yang diklasifikasikan kedalam opini positif, negatif, dan netral. Hasil analisis sentimen ini akan dijabarkan dalam bentuk visualisasi barplot, wordcloud, dan dalam bentuk diagram UML (*Unified Modelling Language*) kemudian proses akan dilanjutkan dengan membuat rancangan sistem. Analisis sentiment ini adalah berbasis web dan dapat diakses menggunakan browser. Dengan begitu penelitian ini akan menghasilkan sebuah informasi analisis sentimen dengan menggunakan rstudio dalam mengambil dan mengklasifikasi data.

Kata Kunci : analisis, sentimen, vaksin, berbayar, rstudio

PENDAHULUAN

Pada dunia teknologi informasi di zaman *modern* saat ini, *internet* menjadi hal yang sangat wajar bagi masyarakat dunia terutama sosial media. Sosial media merupakan salah satu pengaruh dari *internet* yang banyak disukai orang, sosial media digunakan sebagai media komunikasi terbuka dan masyarakat dapat mengemukakan pendapatnya dengan bebas pada setiap harinya. *Twitter, Facebook, Instagram* merupakan salah satu contoh dari sosial media tersebut.

Dalam kehidupan bermasyarakat, mengungkapkan sentimen atau pendapat kepada orang lain telah menjadi suatu aktivitas yang dilakukan setiap harinya. Menurut KBBI, sentimen adalah pendapat atau 3 pandangan yang didasarkan pada perasaan yang berlebihan terhadap sesuatu masalah. Secara umum terdapat dua tipe informasi *tekstual* yaitu fakta dan opini. Fakta adalah pernyataan objektif mengenai entitas dan kejadian di dunia sedangkan opini adalah pernyataan subjektif yang merefleksikan sentimen atau persepsi orang mengenai suatu permasalahan yang di dunia.

Saat ini *twitter* adalah media sosial yang disukai banyak orang dari seluruh penjuru dunia dan setiap orang dapat mengungkapkan pendapat mereka pada setiap harinya. Data yang dihasilkan dari *twitter* bisa sangat berguna jika dianalisis karena data tersebut dapat diekstrak menjadi informasi penting melalui data *mining*.

Adapun sentimen yang menjadi objek penelitian ini adalah melihat opini pengguna *twitter* yang sedang membahas “Vaksin Berbayar”. dalam penelitian ini, penulis melihat salah satu media sosial populer yang disebut *twitter* dan membuat model untuk mengklasifikasikan “*tweet*” menjadi sentimen yang positif, negatif, dan netral.

Text Mining adalah salah satu teknik yang dapat digunakan untuk melakukan klasifikasi dokumen dimana *Text Mining* merupakan variasi dari *Data Mining* yang berusaha menemukan pola menarik dari sekumpulan data *tekstual* yang berjumlah besar (Syakuro, 2017).

Analisis sentimen sangat berguna untuk menganalisis opini pengguna *twitter* tentang berbagai isu atau topik yang sedang ramai dibicarakan. Topik terkait dengan Vaksin Berbayar saat ini menjadi salah satu topik yang banyak dibicarakan oleh pengguna *Twitter* di Indonesia. Berdasarkan latar belakang diatas, maka masalah yang dapat diambil yaitu cara mengetahui hasil persentase terhadap opini pengguna *Twitter* berdasarkan kata kunci “Vaksin Berbayar”. Penelitian ini dilakukan untuk mengetahui persentase opini positif, negatif, dan netral dengan proses klasifikasinya yaitu dengan metode *Naïve Bayes Classifier (NBC)*. (Sutanty, n.d.)

METODE PENELITIAN

A. Registrasi twitter API

Proses pengumpulan data yang telah dilakukan berasal dari tweet yang dikirimkan oleh masyarakat di twitter. Data yang diambil berupa tweet yang menggunakan kata “Vaksin Berbayar”. Data diperoleh dengan memanfaatkan tools Rstudio dan twitter API. Sebelum melakukan pengumpulan atau *crawling* data kita diharuskan memiliki API *key* untuk autentifikasi dengan Twitter.

Proses mendapatkan API *key* :

1. Masuk ke dalam akun Twitter.

<http://jurnal.umat.ac.id/index.php/jt/index>

2. Masuk ke App Twitter atau buka <https://apps.twitter.com/app/new>
3. Lalu *Sign In API Development for Application on Twitter*
4. Kemudian mengisi formulir dengan informasi yang dibutuhkan.
5. Mendapatkan *api key, api secret, access token, dan access token secret*.

Setelah berhasil registrasi *twitter API* dan mendapatkan token, Kemudian baru dapat melakukan proses pengambilan atau *crawling* data, tetapi pertama harus dilakukan autentifikasi antara *twitter* dengan *Rstudio* menggunakan *package* dari *Rstudio*, setelah proses autentifikasi sukses maka dapat mencari data dengan *keyword* yang diinginkan.

B. *Pre-processing* Data

Data Tweet yang sudah diambil masih berupa data mentah yang belum siap diolah, oleh karena itu dilakukan tahap *pre-processing* untuk mendapatkan data yang siap untuk diolah pada proses selanjutnya. Tahapan *pre-processing* yang dilakukan adalah :

1. *Case Folding*

Dalam sebuah tweet sering kali memiliki banyak perbedaan penggunaan pada bentuk huruf, pada bagian ini dilakukan perubahan seluruh huruf kapital (*uppercase*) dikembalikan menjadi huruf kecil (*lowercase*) agar seragam.

2. *Cleansing*

Cleansing adalah proses yang dilakukan untuk membersihkan fitur-fitur yang tidak diperlukan dalam pengambilan data yang ada pada Twitter, seperti URL, Username, dan lain-lain.

3. *Tokenizing*

Tokenizing adalah proses yang dilakukan untuk memenggal kalimat menjadi beberapa bagian atau kata berdasarkan tanda bacanya seperti koma, titik, dan tanda baca lainnya.

4. *Stopword Removal*

Merupakan proses yang dilakukan dengan cara menghilangkan kata yang tidak diperlukan. Jika kata tersebut dibuang maka tidak akan mengubah atau menghilangkan informasi yang berada dalam kalimat tersebut. Seperti kata hubung yang, akan, di, pada dan lain-lain.

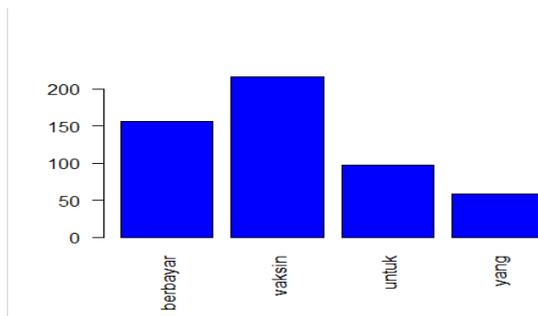
METODE ANALISIS SISTEM

1. Visualisasi *barplot* dan *wordcloud*

Hasil dari pembobotan term akan menampilkan visualisasi berupa *barplot* dan *wordcloud*. pembobotan kata pada *barplot* menunjukkan kata paling banyak muncul dari kata-kata yang lain. Pada *wordcloud* ada kata yang berukuran besar dan kecil menunjukkan hasil kata dari *wordcloud* yang paling banyak di gunakan dalam *tweet* sedangkan kata yang berukuran kecil mengartikan sedikitnya kemunculan kata yang di gunakan dalam *tweet* mengenai topik vaksin berbayar. Pada samping *barplot* juga ada frekuensi kemunculan. Hasil visualisasi *barplot* dan *wordcloud* ada pada gambar dibawah ini :



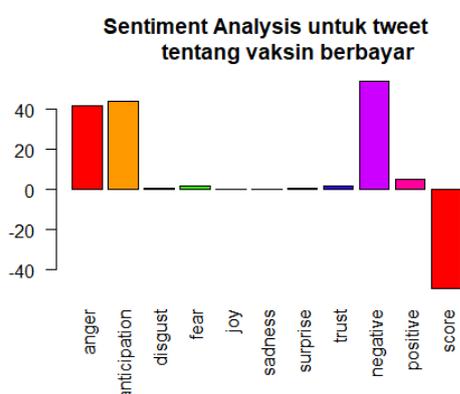
Gambar 1 tampilan workcloud



Gambar 2 tampilan barplot

2. Analisis sentimen berdasarkan emosi

Tahap selanjutnya adalah analisis sentimen berdasarkan emosi yang dihasilkan dari kata yang telah melakukan tahap *preprocessing* dan akan masuk ke dalam beberapa kategori seperti marah, senang, sedih, positif, negatif, dan lain-lain. Hasil analisis sentimen berdasarkan emosi ini akan menampilkan sebuah visualisasi *barplot* yang memiliki beberapa warna dan keterangan angka yang menyatakan berapa banyak jumlah setiap sentimen seperti gambar dibawah ini:



Gambar 3 hasil analisis sentimen berdasarkan emosi

3. Klasifikasi dengan metode naïve bayes

Dalam mengklasifikasi dengan metode *naïve bayes classification* proses klasifikasi menentukan kata tersebut masuk kedalam klasifikasi model kata positif, negatif, dan netral, Jika dalam klasifikasi data mendapatkan *score* 0 itu tandanya hasil klasifikasi termasuk netral, artinya kata – kata tersebut lebih banyak netral dari pada kata positif atau negatif. Jika dalam klasifikasi mendapatkan *score* 1 itu artinya termasuk klasifikasi positif, dan jika *score* -1 maka artinya negatif. Setelah itu menentukan data *training* dan data *testing* yang sudah di lakukan tahap *text preprocessing*. Data *training* berjumlah 208 kata dan data *testing* berjumlah 75 kata, lalu akan di lanjutkan dengan menghitung *prior probabilitas* nya dalam data *training* dan *testing*.

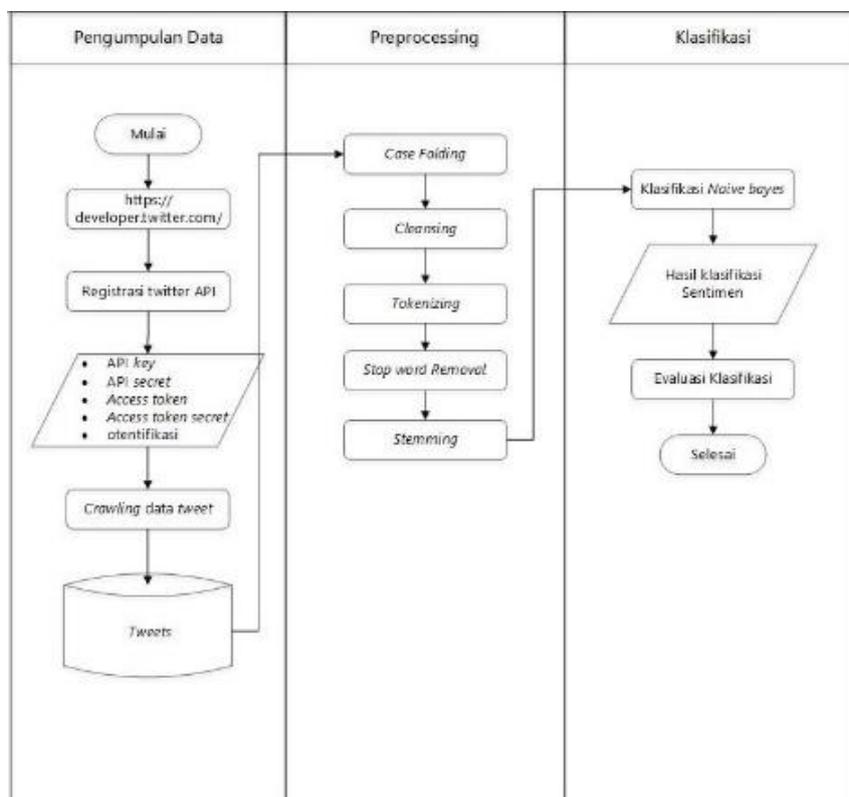
Dari data *training* dengan teks “samsat tidak setuju dengan vaksin” telah di klasifikasikan termasuk model “netral” dengan *prior probabilitas* yang dihasilkan 0.006024096.

Dari data *testing* dengan teks “vaksin berbayar mayoritas setuju” telah di klasifikasi termasuk model “netral” dengan *prior probabilitas* yang dihasilkan 0.02380952.

HASIL DAN PEMBAHASAN

A. Analisis sistem berjalan

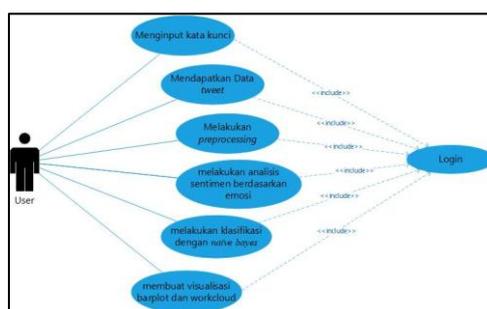
Analisis sistem berjalan dideskripsikan dalam bentuk *flowchart* berikut:



Gambar 4. *flowchart* analisis sentimen

B. UML(Unified Modeling Language)

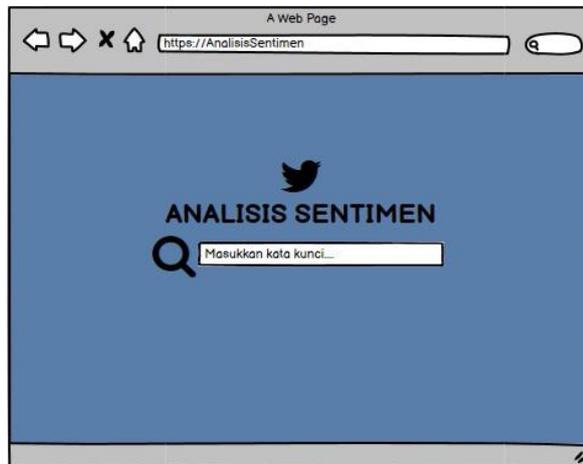
Berikut adalah *usecase* diagram analisis sentimen:



Gambar 5. *usecase* diagram analisis sentimen

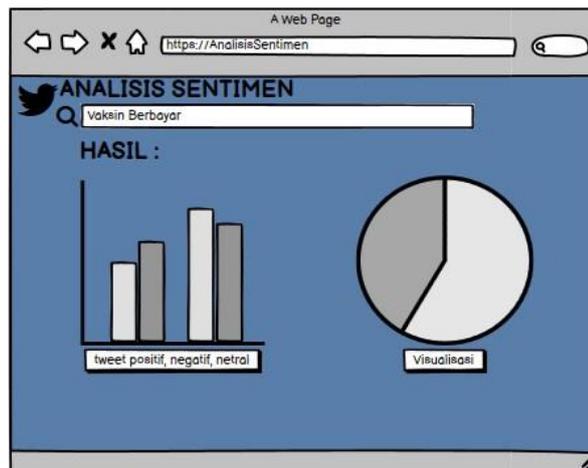
C. Perancangan antar muka

1. Tampilan menu utama



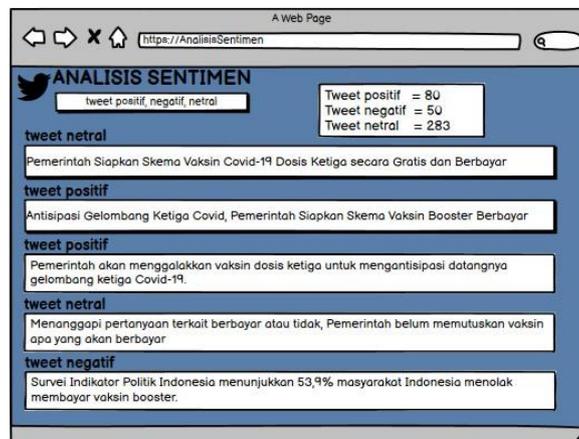
Gambar 6. tampilan menu utama

2. Tampilan hasil masukkan kata kunci



Gambar 7. Tampilan hasil masukkan kata kunci

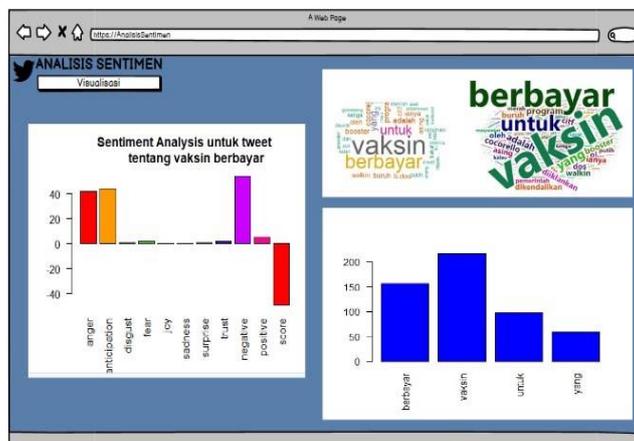
3. Tampilan menu *tweet* positif, negatif, netral



<http://jurnal.umt.ac.id/index.php/jt/index>

Gambar 8. Tampilan menu *tweet* positif, negatif, netral

4. Tampilan menu visualisasi



Gambar 5. Tampilan menu visualisasi

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian, Sentimen analisis opini masyarakat pada media sosial *twitter* nterhadap vaksin berbayar menggunakan metode *naïve baiyes classifier(NBC)*. Dapat diperoleh kesimpulan sebagai berikut:

1. Metode *naïve baiyes* dapat mengklasifikasikan sentimen pengguna *twitter* terhadap vaksin berbayar kedalam sentiment positif, netral, dan negative dengan baik.
2. Berdasarkan emosi, kata “vaksin berbayar” banyak mendapatkan hasil negatif dibanding positif ataupun netral.

DAFTAR PUSTAKA

- Damayanti, R. P. (2018). *Pengaruh Kalimat “ Twitter Please Do Your Magic ” Terhadap Sikap Pengguna Twitter*.
- Dharma Suriyanto, A., & Endah Pangesti, W. (2021). *ANALISA SENTIMEN TERHADAP TAGAR #dirumahaja MELALUI TWITTER DI INDONESIA*. 2(09), 9–17.
- Eka Sembodo, J., Budi Setiawan, E., & Abdurahman Baizal, Z. (2016). *Data Crawling Otomatis pada Twitter*. *August*, 11–16. <https://doi.org/10.21108/indosc.2016.111>
- Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016(Sentika)*, 57–64. <https://fti.uajy.ac.id/sentika/publikasi/makalah/2016/95.pdf>
- Muzaki, A., & Witanti, A. (2021). Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier. *Jurnal Teknik Informatika (Jutif)*, 2(2), 101–107. <https://doi.org/10.20884/1.jutif.2021.2.2.51>

- Rusdaman, D., & Rosiyadi, D. (2019). Metode Naïve Bayes Classifier Dan Support Vector Machine. *Journal of Computer Engineering System and Science*, 4(2), 230–235.
- 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>
- Zuhdi, A. M., Utami, E., & Raharjo, S. (2019). Analisis sentiment twitter terhadap capres Indonesia 2019 dengan metode K-NN. *Jurnal Informa Politeknik Indonusa Surakarta*, 5, 1–7.