

ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Herwinsyah¹, Arita Witanti²

Informatika, Universitas Mercu Buana Yogyakarta

Jl. Raya Wates-Jogjakarta, Karanglo, Argomulyo, Kec. Sedayu, Kabupaten Bantul, Daerah
Istimewa Yogyakarta

e-mail: herwinsy@gmail.com, arita@mercubuana-vogya.ac.id

Abstract

This research was conducted to find information about the tendency of Indonesian people regarding the Covid-19 vaccination. The method that the author uses is by collecting data from Twitter social media using the API key provided by Twitter. The process of collecting data using a Python application with several libraries such as tweepy, pandas, numpy and nltk. After the data is crawled, then the data is cleaned with several data cleaning processes such as remove username, remove url, lower case, remove stopwords and lemmatize. Then the results are labeled with the textblob and sklearn libraries. then the data is analyzed using the Support Vector Machine (SVM) algorithm with the best comparison being 20 testing data and 80 training data or as many as 942 testing data and 3766 training data, the prediction results for testing data are f1 score 0.93, accuracy score 0.88, precision score 0.88 and a recall score of 0.99. The results showed that from 4,078 tweet data, there were 2,525 positive sentiments (43.0%), 771 negative sentiments (16.4%), and 1,912 neutral sentiments (40.6%). The results of 80% (3766) of training data and 20% (942) of test data obtained an accuracy score of 73.6%. From this study, it can be concluded that the tendency of Indonesian people when sampling data is taken is more accepting (positive responses) to government policies regarding the Covid-19 vaccination program. In the future, it is hoped that there will be a library that supports text data processing such as regional languages, because researchers found that during the data cleaning process there was a lot of word elimination, because many regional languages were used by the Indonesian people in writing on social media.

Keyword: analysis, covid-19, sentiment, SVM, vaccination,

PENDAHULUAN

Pandemi *corona virus disease* 2019 atau lebih dikenal dengan nama covid-19 diperkirakan menyerang seluruh belahan dunia dimulai pada akhir tahun 2019 tepatnya pada Oktober 2019, dimulai dari negara tirai bambu China tepatnya di kota Wuhan. Penyakit ini disebabkan oleh virus SARS-CoV-2, yang diserang oleh virus ini utamanya pada bagian pernafasan dalam dan paru-paru. Akibat yang bisa dirasakan penderita adalah sesak pernafasan, hilangnya penciuman bahkan sampai pada sindrom pernafasan akut yang dapat menyebabkan kematian (Novina Putri Bestari, 2021).

Cepatnya penyebaran virus ini di seluruh dunia bahkan di Indonesia, mengharuskan pemerintah mengambil banyak langkah-langkah kebijakan. Pemerintah yakin dengan dilakukan vaksinasi covid-19 dapat menjadi langkah awal menuju masyarakat yang lebih sehat. Akan tetapi di masyarakat beredar beberapa masalah terkait kebijakan pemerintah tersebut. Bahkan di media sosial juga ramai memperbincangkan efek positif dan efek negatif jika dilakukan vaksinasi terhadap tubuh manusia bahkan kominfo juga mengkalkulasi banyaknya isu di harian berita mencapai 1387 isu hoax yang beredar di masyarakat (Mutiar Nabila, 2021). Untuk mengetahui hal ini penulis menggunakan analisis sentiment sebagai alat dalam memperoleh hasil respon animo dari masyarakat terkait kebijakan vaksinasi covid-19 tersebut.

Sentimen analisis adalah proses mendeteksi kecenderungan seseorang melalui tulisan atau teks atau mengekstrak data teks untuk mendapatkan informasi tentang sentimen yang biasanya dituliskan didalam media sosial baik bermakna positif, negatif atau netral. Proses analisis ini biasanya dilakukan secara otomatis melalui beberapa aplikasi atau platform yang mendukung untuk dilakukannya analisis teks.

Analisis sentimen bertujuan untuk mengekstraksi atribut pada dokumen atau teks yang berisi komentar untuk mengetahui respon didalamnya agar dapat digolongkan menjadi respon positif, negatif dan netral. Berdasarkan tujuan dari analisis sentimen tersebut sangat memungkinkan dalam mengatasi permasalahan yang muncul dari opini publik terhadap kebijakan pemerintah tentang vaksin covid-19, baik opini masyarakat yang menerima kebijakan tersebut, maupun yang menolak.

Penelitian dengan judul Analisis Sentimen Terhadap Opini Publik Covid-19 Di Media Sosial Twitter Menggunakan Naive Bayes Classifier Dan Support Vector Machine ini fokus pada pencarian data tentang pendapat masyarakat terutama di Indonesia tentang Kesehatan mental selama pandemic covid-19. Analisis Sentimen dalam Penanganan Covid-19 di Indonesia Menggunakan Naive Bayes Classifier. Penelitian ini bertujuan mengetahui bagaimana sentimen masyarakat terhadap kebijakan PSBB melalui tanggapan di media sosial twitter.

Berdasarkan beberapa penelitian di atas terdapat persamaan dengan penelitian penulis yang terletak pada analisis sentimen atau sentimen analisis, dan pada media twitter, namun terdapat perbedaan yaitu penelitian sebelumnya menggunakan *Naive Bayes Classifier* sedangkan penulis menggunakan *algoritma support vector machine*. Tujuan penelitian ini adalah melakukan perancangan, Membuat implementasi sistem, dan mengetahui bagaimanakah tingkat akurasi sistem yang didapatkan untuk analisis sentimen masyarakat warga negara Indonesia terhadap vaksinasi covid-19 pada media sosial twitter menggunakan algoritma Support Vector Machine.

Support Vector Machine (SVM) merupakan suatu teknik untuk melakukan prediksi, baik prediksi dalam kasus regresi maupun klasifikasi (M. Fachrurrozi & Novi Yusliani, 2015). Teknik *SVM* digunakan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal untuk memisahkan observasi yang memiliki nilai variabel target yang berbeda (Williams, Wilkins, 2011). *Hyperplane* ini dapat berupa *line* pada *two dimensional* dan dapat berupa *flat plane* pada *multiple dimensions*.

METODE PENELITIAN

Data yang digunakan oleh penulis dalam penelitian ini adalah *tweet* atau cuitan masyarakat di wilayah Indonesia yang terdapat pada aplikasi media sosial *twitter*. *Twitter* adalah sebuah situs *microblogging*, yang populer karena sebuah pesan singkat yang lebih dikenal dengan sebutan *tweet*. *Tweet* merupakan ungkapan yang dari para pengguna yang bertujuan mengungkapkan isi dari dalam pikiran atau isi hati yang biasanya berkaitan dengan kondisi sosial di masyarakat (Bakliwal, A., Arora, P., Madhappan, S., Kapre, N., Singh, M., & Varma, 2012)

Data mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar, meliputi metode -metode yang merupakan irisan dari *artificial intelligence*, *statistics machine learning*, , dan *database systems* (Suyanto, 2017).

Text Mining dapat didefinisikan sebagai “penemuan informasi baru dan tidak diketahui sebelumnya oleh komputer, dengan secara otomatis mengekstrak informasi dari sumber – sumber teks tak terstruktur yang berbeda” (Tan, 2018) Dalam *text mining* terdapat beberapa tahapan untuk memproses data teks tersebut

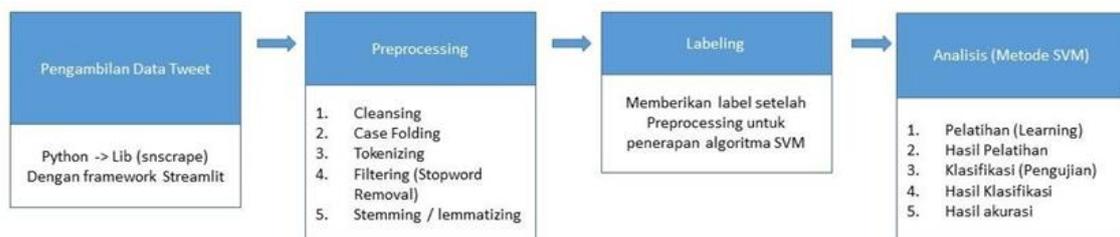
1. *Cleaning* adalah proses menghilangkan semua karakter di dalam *tweet* yang tidak termasuk alfabet, agar dapat mengurangi karakter atau simbol yang tidak dikehendaki atau tidak memiliki makna dalam analisis sentimen.

2. *Case Folding* biasa disebut penyeragaman kata dengan cara mengubah seluruh kata menjadi huruf kecil (*lowercase*). Hanya huruf a sampai z yang dapat diterima, karakter selain huruf dihilangkan.
3. *Tokenizing* adalah suatu tahapan pemotongan *string* kata berdasarkan penyusunan kata tersebut. Hasil proses *cleansing* dan *case folding* kemudian dilanjutkan proses *tokenizing* dimana kalimat *tweet* dipenggal kata perkata.
4. *Filtering* adalah pengambilan kata-kata penting dari hasil *Tokenizing* atau biasa disebut pengeliminasi kata-kata sesuai dengan kaidahnya. Algoritma *stopword removal* atau *remove stopwords* adalah salah satu yang digunakan untuk melakukan tahapan *filtering*
5. “*Lemmatizing* adalah proses untuk menemukan bentuk dasar dari sebuah kata” (Anton Karl Ingason; Sigrún Helgadóttir; Hrafn Loftsson; Eiríkur Rögnvaldsson, 2015).
6. *Labeling* merupakan penerapan pada *data mining* dengan metode *SVM* data yang telah didapat harus didefinisikan terlebih dahulu sebagai kalimat yang mengandung nilai positif atau nilai negatif. Masing-masing data *tweet* harus diberi label sebagai *tweet* positif atau *tweet* negatif.

Dalam penelitian ini cuitan yang digunakan hanya yang mengandung opini terhadap pelaksanaan vaksinasi covid-19. Data cuitan yang diambil akan diproses melalui beberapa proses pembersihan selanjutnya oleh penulis akan diklasifikasikan sentimen yang terkandung didalam *tweet* tersebut yang kemudian ditandai apakah *tweet* tersebut masuk kedalam 2 kategori sentimen, yaitu *tweet* yang mengandung sentimen negatif atau mengandung sentimen positif. Kemudian data akan diberi nilai (pembobotan) setelah itu kan dianalisis menggunakan *support vector machine (svm)*.

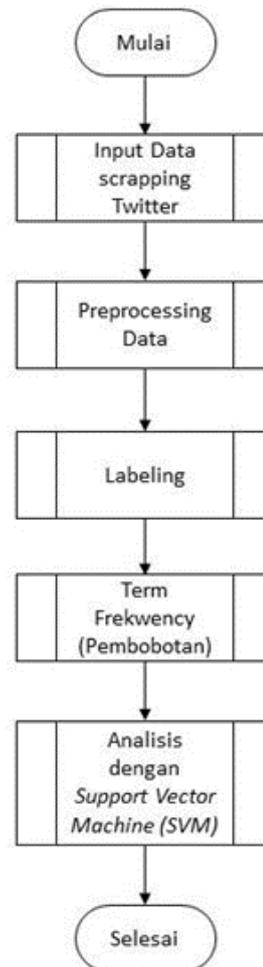
Alat penelitian yang digunakan penulis dalam penelitian ini adalah sebuah *device* perangkat laptop dengan spesifikasi yang cukup untuk menjalankan beberapa *software* dan aplikasi yang digunakan dalam proses penelitian. Perangkat lunak atau *software* yang digunakan penulis dalam penelitian antara lain *Operating System Windows 10 Pro*, *web browser Google Chrome*, *web browser mozilla firefox*, bahasa pemrograman *Python ver. 3.7*, *anaconda console 3*, *library python tweepy*, *library python snsrape*, *microsoft excel* dan aplikasi media sosial *twitter*.

Tahapan proses penelitian menyesuaikan dengan alur data yang diterima yang merupakan data yang di *scraping* dari media sosial *Twitter* kemudian melalui proses *preprocessing* data dan pembobotan selanjutnya ke tahap analisis data dengan metode *support vector machine (SVM)*.



Gambar 1. Tahapan Proses Penelitian

Hasil penelitian yang dilakukan penulis diimplementasikan dengan bahasa pemrograman dan *framework* streamlit. Tampilan interface dari aplikasi sentimen analisis berbasis web ini terdiri dari 4 bagian besar yaitu *Crawling Data*, *Preprocessing Data*, *Labeling*, dan *Support Vector Machine*.

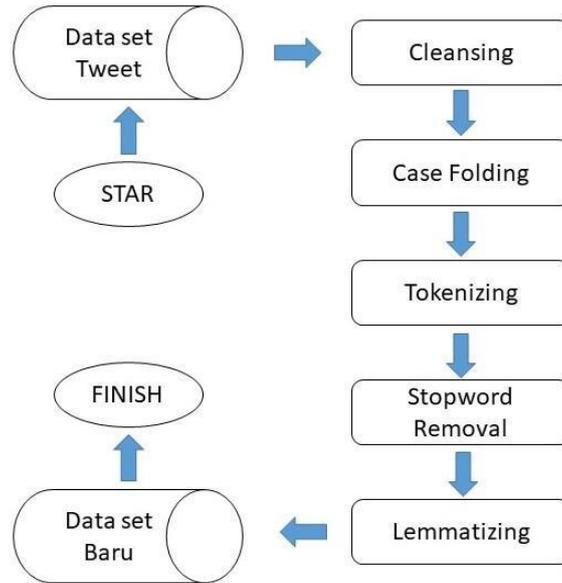


Gambar 2. Flowchart system

Tahap implementasi ini menjelaskan proses secara keseluruhan dari mulai proses pengumpulan data sampai dengan analisis. Proses pengumpulan data dilakukan dengan *Crawling / scraping* data pada aplikasi *twitter* dengan menggunakan bahasa pemrograman *python*. Kemudian pembersihan data pada tahap *preprocessing* data, selanjutnya pembobotan dan terakhir analisis data dengan metode *support vector machine (svm)*.

Data yang dikumpulkan adalah *tweet* yang mengandung kata-kata dari unsur kata positif atau negatif. Kalimat yang positif mengandung kata-kata seperti siap, menerima, mau, dan kata negatif yang mengandung kata berlawanan dari kata positif. Data yang dikumpulkan akan dimasukkan ke dalam satu *database* untuk kemudian di cluster berdasarkan cluster positif atau cluster negatif. Dari hasil *crawling* yang kemudian dibuka dengan aplikasi *microsoft excel*, telah didapatkan sebanyak 14.099 baris data, dimana data tersebut masih perlu dilakukan proses pembersihan data atau *preprocessing* data.

Preprocessing data merupakan salah satu tahapan proses yang harus dilakukan untuk memproses data pada proses *mining*. Kondisi data yang dihasilkan tidak selalu dalam kondisi yang siap untuk diproses. Data tersebut seringkali terdapat berbagai permasalahan yang bisa mengganggu hasil dari proses *mining* itu sendiri, seperti data yang *missing value*, *outliers*, data *redundant*, ataupun format data yang tidak sesuai dengan sistem. Oleh sebab itu tahap *Preprocessing* ini dibutuhkan untuk mengatasi permasalahan tersebut. Beberapa proses yang dibutuhkan untuk mengklasifikasi dokumen yang bertipe teks yaitu *cleansing (remove punctuation)*, *case folding*, *tokenizing*, *filtering (stopword removal)*, dan *lemmatizing*.

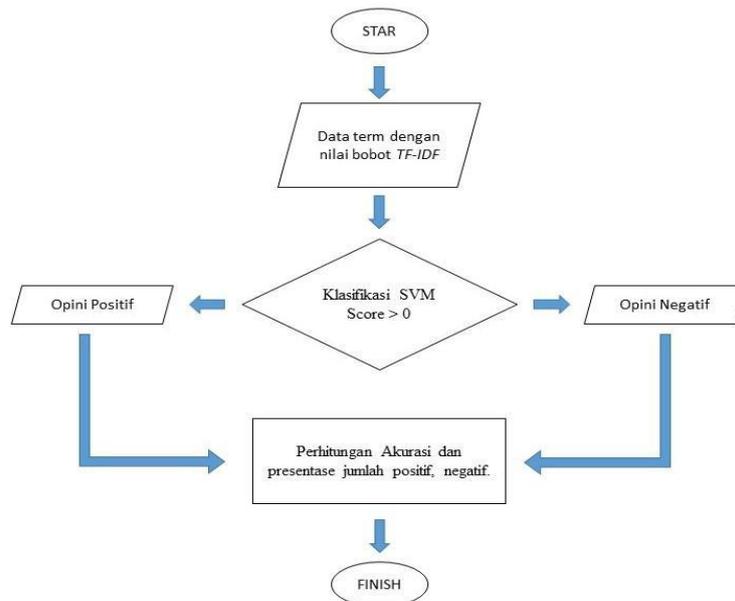


Gambar 3. Tahapan *Preprocessing* Data

Selanjutnya proses yang akan dilakukan adalah melabel data yang telah diproses sebelumnya. Dalam memproses data pada *metode support vector machine (SVM)* data *tweet* yang telah didapatkan harus diberikan label, apakah *tweet* tersebut termasuk kedalam sentiment tweet positif atau sentimen negatif.

Setelah proses *preprocessing* maka langkah selanjutnya adalah proses *term weighting* (pembobotan). Pada setiap kata yang dihasilkan akan dilakukan pembobotan berdasarkan metode TF (*Term Frequency*) – IDF (*Inverse Document Frequency*).

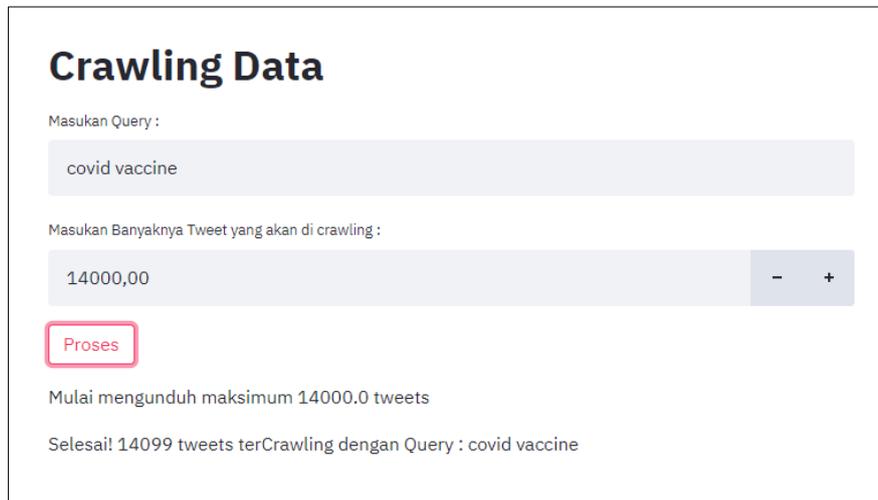
Proses selanjutnya dengan menerapkan hasil pembobotan pada proses sebelumnya dengan metode SVM. Proses ini akan menghasilkan nilai atau pola yang akan digunakan pada proses *testing*, kemudian penilaian akan dibuat dengan memberi *score* pada dokumen. Keputusan dan analisis yang digunakan mengacu kepada tingkat akurasi dan jumlah dokumen pada *class* positif dan negative.



Gambar 4. Tahap Proses Metode SVM

HASIL DAN PEMBAHASAN

Pada halaman depan user akan langsung diarahkan ke bagian pertama yaitu *Crawling Data*. Pada bagian ini user dapat melakukan *crawling* data sesuai dengan *query* yang diinginkan.



Crawling Data

Masukan Query :

covid vaccine

Masukan Banyaknya Tweet yang akan di crawling :

14000,00 - +

Proses

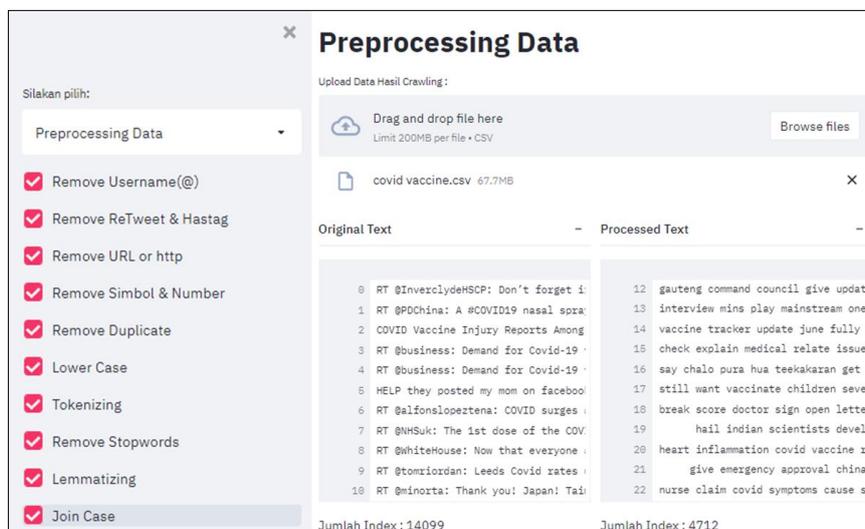
Mulai mengunduh maksimum 14000.0 tweets

Selesai! 14099 tweets terCrawling dengan Query : covid vaccine

Gambar 5. Halaman *Crawling Data*

Pada proses ini data hasil *crawling* akan banyak mengalami perubahan, proses preprocessing yang dilakukan antara lain:

- Remove Username (@)
- Remove ReTweet & Hashtag
- Remove URL or http
- Remove Symbol & Number
- Remove Duplicate
- Lower Case
- Tokenizing
- Remove Stopwords
- Lemmatizing
- Join Case



Preprocessing Data

Silakan pilih:

Preprocessing Data

- Remove Username(@)
- Remove ReTweet & Hastag
- Remove URL or http
- Remove Simbol & Number
- Remove Duplicate
- Lower Case
- Tokenizing
- Remove Stopwords
- Lemmatizing
- Join Case

Upload Data Hasil Crawling:

Drag and drop file here
Limit 200MB per file • CSV

Browse files

covid vaccine.csv 67.7MB

Original Text - Processed Text

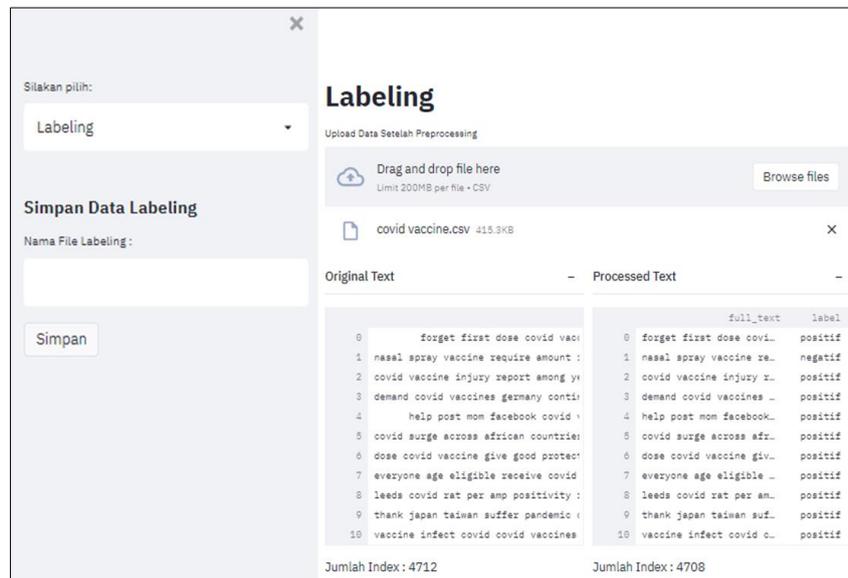
0 RT @InverclydeHSCP: Don't forget f	12 gauteng command council give updat
1 RT @PDCChina: A #COVID19 nasal sprax	13 interview mins play mainstream one
2 COVID Vaccine Injuz Reports Among	14 vaccine tracker update june fully
3 RT @business: Demand for Covid-19	15 check explain medical relate issue
4 RT @business: Demand for Covid-19	16 say chalo puxa hua teekakazan get
5 HELP they posted my mom on facebook	17 still want vaccinate children seve
6 RT @alfonslopezrena: COVID surges	18 break score doctor sign open lette
7 RT @NHSuk: The 1st dose of the COV	19 hail indian scientists devel
8 RT @WhiteHouse: Now that everyone	20 heart inflammation covid vaccine r
9 RT @tomziordan: Leeds Covid rates	21 give emergency approval china
10 RT @minorta: Thank you! Japan! Tai	22 nurse claim covid symptoms cause s

Jumlah Index: 14099

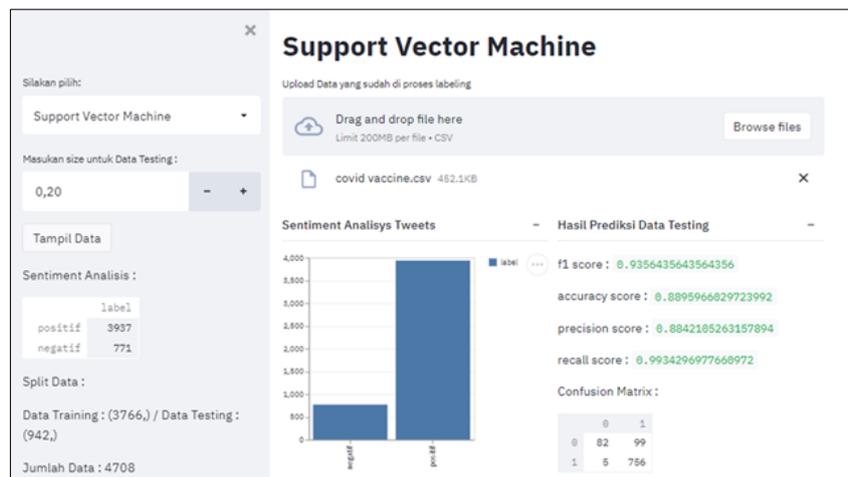
Jumlah Index: 4712

Gambar 6. Halaman *Preprocessing Data*

Labeling data merupakan salah satu syarat yang harus ada untuk dapat diaplikasikan ke dalam metode *support vector machine*. Terlihat hasil akhir *preprocessing* terdapat 4712 indeks data dan setelah melalui proses *labeling* dimana pada proses ini juga terdapat proses menghapus data indeks yang kosong, sehingga hasil akhir menjadi 4708 data *index*, ini menunjukkan terdapat data *index* kosong sebanyak 4 *row index* data. Data tweet yang telah melalui proses *labeling* data, selanjutnya di simpan ke dalam file berekstensi csv, untuk diproses selanjutnya.



Gambar 7. Halaman *Labeling* Data



Gambar 8. Halaman Implementasi SVM

Hasil penelitian analisis sentimen masyarakat terhadap vaksinasi covid 19 di media sosial *twitter* dengan algoritma *support vector machine* adalah:

1. Analisis sentimen dengan Algoritma svm bisa digunakan untuk mengukur sentimen masyarakat terhadap kebijakan pemerintah tentang himbuan vaksinasi covid.

2. Hasil penelitian sebanyak 4708 *index* data *tweet* yang diproses, sentimen analisis yang diperoleh adalah respon positif sebanyak 3937 (83,6%) dan respon negatif sebanyak 771 (16,4%).
3. Hasil akurasi yang diperoleh dengan perbandingan 80% (3766) data *training* dan 20% (942) data *testing* diperoleh hasil akurasi *score* sebesar 89%, *f1 score* sebesar 93%, *precision score* sebesar 88% dan *recall score* sebesar 99%. Dari hasil *confusion matrix* 942 *index* data *testing* dihasilkan *true* positif sebanyak 756 data, *true* negatif sebanyak 82 data, *false* positif sebanyak 99 data dan *false* negatif sebanyak 5 data.

KESIMPULAN

Sentimen analisis yang peneliti lakukan dengan menggunakan algoritma *Support Vector Machine (SVM)* jelas dapat mengukur tingkat respon dari masyarakat terhadap kebijakan pemerintah tentang vaksinasi covid-19.

SARAN

Dari penelitian yang dilakukan penulis melihat masih adanya celah yang kurang dalam beberapa proses analisis, misalnya belum adanya *library* yang belum mendukung untuk menghitung polaritas dengan data text yang berbahasa Indonesia. Sehingga beberapa kali dalam penelitian ini penulis menemukan kendala dalam memproses *text processing*. Berharap kedepannya dapat tersedia *library* yang mensupport untuk polaritas dalam proses *labeling* data.

Berdasarkan penelitian dan uraian yang telah disampaikan bahwasanya penelitian ini masih jauh dari kata sempurna, semoga kedepannya akan ada penelitian-penelitian selanjutnya yang juga mengangkat kasus yang sama ataupun berbeda dengan metode yang berbeda juga agar dapat mendapat perbandingan yang lebih variatif dan hasil yang lebih akurat.

DAFTAR PUSTAKA

- Anton Karl Ingason; Sigrún Helgadóttir; Hrafn Loftsson; Eiríkur Rögnvaldsson. (2015). *Advances in natural language processing*. In *Science* (Vol. 349, Issue 6245). <https://doi.org/10.1126/science.aaa8685>
- Bakliwal, A., Arora, P., Madhappan, S., Kapre, N., Singh, M., & Varma, V. (2012). *Mining Sentiment from Tweets*. *3rd Workshop on Sentiment and Subjectivity Analysis*, 11–18.
- Ihsanuddin. (2020). Fakta Lengkap Kasus Pertama Virus Corona di Indonesia. <https://nasional.kompas.com/read/2020/03/03/06314981/fakta-lengkap-kasus-pertama-virus-corona-di-indonesia?page=all>
- Kementerian Dalam Negeri. (2020). *Pedoman Umum Menghadapi Pandemi Covid-19 bagi Pemerintah Daerah Pencegahan, Pengendalian, Diagnosisi dan Manajemen*. Kementerian Dalam Negeri.
- Komite Peninjau untuk Pedoman Vaksinasi. (2020). *Vaksinasi dan Kesehatan Anak*. Yayasan Umum Pusat Penelitian Vaksinasi.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher.
- M. Fachrurrozi, M. T., & Novi Yusliani, M. T. (2015). *Analisis Sentimen Pengguna Jejaring Sosial*. *June 2016*, 1–6. <https://www.researchgate.net/publication>

- Mutiara Nabila. (2021). Kominfo: 1.387 Hoaks Beredar, Isu Vaksin Covid-19 Paling Dominan. <https://teknologi.bisnis.com/read/20210127/101/1348349/kominfo-1387-hoaks-beredar-isu-vaksin-covid-19-paling-dominan>
- Novina Putri Bestari, C. I. (2021). *Ahli AS: Covid Berasal dari China, Nyebar Mulai Oktober 2019*. <https://www.cnbcindonesia.com/tech/20210204161414-37-221089/ahli-as-covid-berasal-dari-china-nyebar-mulai-oktober-2019>
- Soleman, B. T. (1984). *Struktur dan Proses Sosial*. Rajawali.
- Suyanto, S. M. (2017). *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Informatika.
- Tan, A.-H. (2018). *Text Mining : The state of the art and the challenges*. Singapore Mangement University.
- Williams, Wilkins. (2011). *Nursing: Manfsirkan Tanda-tanda dan Gejala Penyakit*. PT. Indeks.