

ANALISIS SENTIMEN OPINI TERHADAP VAKSIN COVID-19 PADA MEDIA SOSIAL TWITTER MENGUNAKAN *NAÏVE BAYES* DAN *DECISION TREE*

Ahmad Zamsuri¹, Elvira Asril², Muhammad Sadar³, Walhidayat⁴, Fajrizal⁵
Syahtriatna⁶, Yogo Turnandes⁷

(^{1,3,5}Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning)

(^{2,6}Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Lancang Kuning)

(^{4,7}Program Studi Bisnis Digital Fakultas Ilmu Komputer Universitas Lancang Kuning)

(Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015)

e-mail: ¹ahmadzamsuri@unilak.ac.id, ²elviraasil@unilak.ac.id, ³sadarzen@unilak.ac.id,

⁴walhidayat@unilak.ac.id, ⁵fajrizal@unilak.ac.id, ⁶eet@unilak.ac.id, ⁷turnandes@unilak.ac.id

Abstrak

Wabah virus korona yang biasa disebut dengan COVID-19 ditetapkan secara resmi sebagai pandemic global oleh World Health Organization (WHO). Untuk meminimalisir dampak yang disebabkan oleh virus salah satu langkah yang tepat adalah dengan mengembangkan vaksin. Akan tetapi dengan adanya vaksinasi untuk masyarakat Indonesia tersebut menimbulkan kontroversial sehingga mengundang banyak kalangan untuk memberikan penilaian opininya. Keterbatasan tempat membuat masyarakat sulit dalam menyampaikan opininya. Oleh karena itu masyarakat memilih sosial media sebagai tempat untuk menyalurkan opini masyarakat. Dalam penelitian ini dilakukan perbandingan algoritma *Naïve Bayes*, dan *Decision Tree* untuk melakukan analisa sentiment tentang vaksin Covid – 19. hasil yang di dapat dari analisis menunjukkan bahwa pada umumnya masyarakat memeberikan tanggapan positif terhadap kebijakan vaksinasi yang dilakukan pemerintah indonesia.

Kata Kunci: Covid-19, Vaksin, Twitter, *Naïve Bayes* dan *Decision Tree*, sentiment analisis.

Abstract

The corona virus outbreak, commonly referred to as COVID-19, has been officially declared a global pandemic by the World Health Organization (WHO). To minimize the impact caused by the virus, one of the right steps is to develop a vaccine. However, the existence of vaccination for the people of Indonesia has caused controversy so that it invites many people to give an opinion assessment. The limited space makes it difficult for people to express their opinions. Therefore, people choose social media as a place to channel public opinion. In this study, a comparison of the *Naïve Bayes* algorithm and *Decision Tree* was carried out to conduct a sentiment analysis about the Covid-19 vaccine. The results obtained from the analysis showed that in general the public gave a positive response to the vaccination policy carried out by the Indonesian government.

Keywords: Covid-19, Vaksin, Twitter, *Naïve Bayes* dan *Decision Tree*, sentiment analisis.

1. PENDAHULUAN

Wabah virus korona yang biasa disebut dengan COVID -19 ditetapkan secara resmi sebagai pandemi global oleh World Health Organization (WHO) pada tanggal 11 Maret 2020. Pusat penyebaran virus tersebut dimulai pada akhir tahun 2019 lalu, dan sampai saat ini sudah menyebar ke berbagai negara-negara dengan tingkat penularan mencapai nilai 40 juta lebih kasus dan jumlah kematian mencapai 1 juta lebih hingga Oktober 2020. Efek lanjutan dari COVID -19 ini berpotensi membawa tantangan besar bagi sistem kesehatan dunia dan memiliki

konsekuensi yang luas pada ekonomi global jika penyebaran virus tidak dikendalikan secara efektif [1].

Melihat pesatnya penyebaran COVID -19 dan dampak yang akan ditimbulkan apabila tidak secepatnya diselesaikan, salah satu solusi untuk mengurangi tingkat penyebaran virus yaitu dengan pembuatan vaksin. Vaksin akan melindungi masyarakat terhadap penyebaran virus dan juga dapat menurunkan tingkat penyebaran virus tersebut. Dalam menyikapi hal tersebut, Pemerintah Indonesia juga turut aktif dalam rencana kegiatan vaksinasi yang akan diberikan kepada masyarakatnya. Untuk mempercepat penanganan COVID -19, Pemerintah Republik Indonesia telah mengeluarkan vaksinasi untuk masyarakat Indonesia pada akhir tahun 2020 [2].

Dengan adanya vaksinasi untuk masyarakat Indonesia tersebut dinilai kontroversial sehingga mengundang banyak kalangan untuk memberikan pendapat dan opininya. Masyarakat bebas memberikan respon dan opininya di berbagai media. Sosial media yang paling banyak digunakan masyarakat untuk menyampaikan pendapat dan opininya. Berdasarkan data dari Global Digital Statistic “Digital, Social & Mobile in 2019” di We Are Social (2019), 2019 merupakan tahun dengan jumlah pengguna media sosial dengan jumlah

150 juta pengguna. Twitter menjadi media sosial yang memiliki pengguna aktif paling banyak di Indonesia dengan jumlah mencapai 52% dari seluruh pengguna sosial media di Indonesia[3].

Analisis sentimen merupakan suatu metode untuk mengklasifikasikan tingkat emosional seseorang baik netral, positif ataupun negatif [4]. Salah satu penerapan analisis sentimen dilakukan pada Twitter, dimana sistem akan secara otomatis mengambil data lewat postingan atau tweet dari pengguna dan sistem akan melakukan klasifikasi untuk menilai apakah postingan tersebut mengandung kalimat netral, positif atau negatif [5]. Berdasarkan klasifikasi tersebut, pengguna dapat mengetahui dan menilai tentang suatu topik tertentu serta dapat membuat keputusan [6]. Metode Naïve Bayes dan Support Vector Machine (SVM) digunakan untuk melakukan analisis sentimen melalui komentar pada postingan sosial media instagram terkait pada kasus COVID-19 pada postingan World Health Organization (WHO) mengenai Covid-19. Hasil penelitian diperoleh bahwa dengan menggunakan algoritma Naïve Bayes mendapatkan akurasi 78,02% dan algoritma SVM mendapatkan nilai akurasi 80,23%. Melalui optimasi Particle Swarm Optimization (PSO), akurasi metode SVM meningkat menjadi 81,16% dan akurasi Naïve Bayes menjadi 79,07% [7].

Untuk Mengetahui opini positif atau negatif suatu peristiwa atau isu dapat dilakukan analisis sentimen data Twitter dengan kata kunci ‘Vaccine Covid-19’ menggunakanmeto Naïve Bayes. Hasil pengukuran sentimen dengan lebih dari 3,4 ribu tweet negatif, lebih dari 2,4 ribu tweet positif, dan sisa 301 tweet dianggap netral selama periode seminggu pada minggu kedua dan ketiga Januari 2021. Analisis pada periode tersebut menunjukkan 39% sentimen positif, 56% sentimen negatif, dan 1% sentimen menguntungkan. Pendapat negatif pada periode tersebut muncul karena masyarakat tidak percaya bahwa vaksin itu aman [8]. Penelitian analisis sentimen masyarakat mengenai pendapat tentang COVID-19 melalui media sosial Twitter juga diteliti menggunakan KNN serta Naïve Bayes. Dengan menggunakan KNN diperoleh akurasi sebesar 85% sedangkan untuk nilai akurasi Naïve Bayes mendapatkan nilai 82%. Melalui hasil tersebut algoritma KNN terlihat lebih baik dalam melakukan analisis sentimen untuk mengklasifikasi komentar pada media sosial Twitter jika dibandingkan dengan algoritma Naïve Bayes [9]. Selain itu algoritma Naïve Bayes juga digunakan untuk analisis sentimen menggunakan twitter API yang terhubung dengan RStudio. Dan hasil yang didapat dari penelitian analisis sentimen ini yaitu pada data tweet Maret 2020, sentimen positif 51% dan sentimen negatif 49% dengan tingkat akurasi 0,7586, spesifisitas 0,6667, prevalensi 0,5862. Sedangkan data tweet Juni 2020 menunjukkan, 59% sentimen negatif dan 41% sentimen positif. Dengan tingkat akurasi 0,6486, spesifisitas 0,6111, prevalensi 0,5135 [10].

Algoritma Naïve Bayes dikenal mampu melakukan klasifikasi dengan baik berdasarkan penalaran probabilitas. Algoritma SVM juga merupakan salah algoritma machine learning yang dikenal cukup baik dalam melakukan klasifikasi berdasarkan pembobotan yang diproses dalam algoritma tersebut. Dalam penelitian ini dilakukan perbandingan model algoritma SVM serta

Naïve Bayes untuk melakukan analisis sentimen pada postingan Twitter berdasarkan kata kunci vaksin Covid -19.

Berdasarkan gambaran umum yang di sajikan tersebut diatas, maka permasalahan yang diangkat pada kajian penelitian adalah bagaimana melakukan analisis sentiment pada Twitter berdasarkan kata kunci vaksin Covid -19 dengan menggunakan algoritma *Naïve Bayes*, dan *Decision Tree*

Penelitian terkait yang pernah dilakukan adalah sebagai berikut:

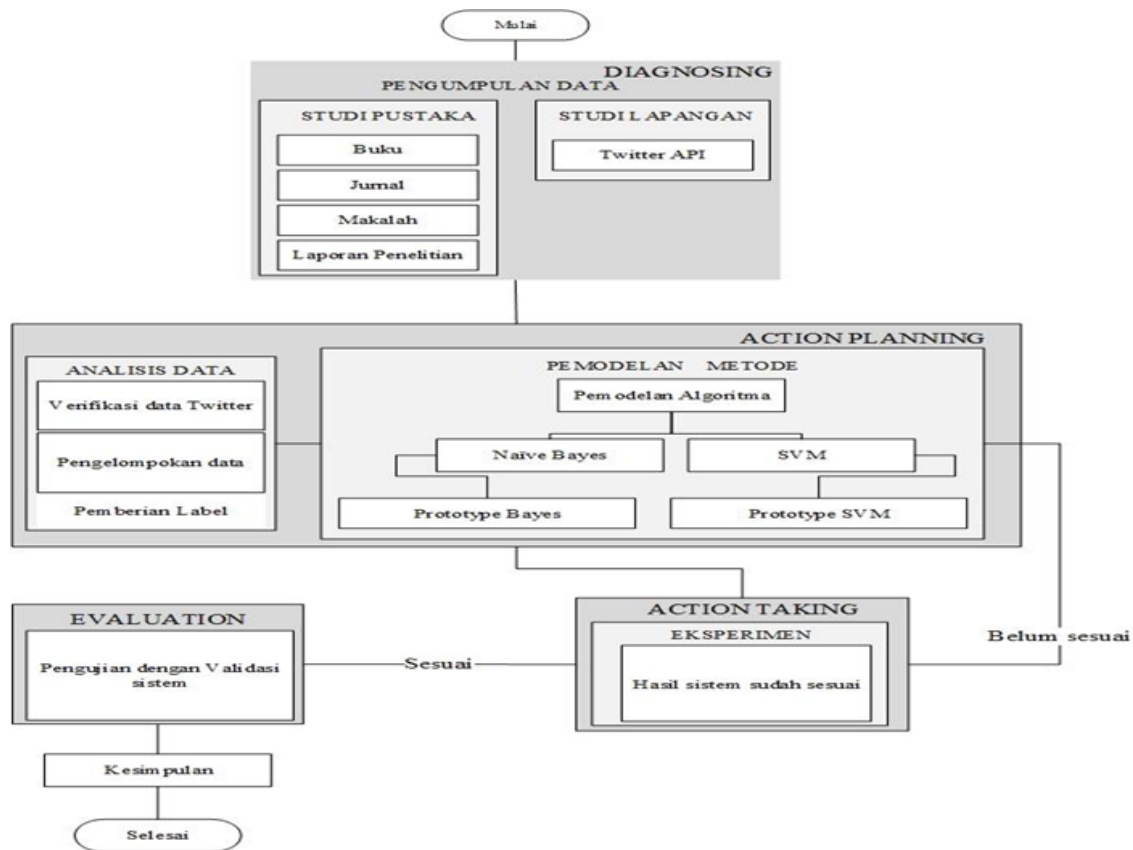
Organisasi kesehatan Dunia menyatakan virus baru terus bermunculan yang merupakan masalah bagi kesehatan masyarakat. Pada Desember 2019 sebuah epidemi kasus pernafasan rendah terdeteksi. World Health Organization (WHO) menyatakan penyakit tersebut disebabkan oleh coronavirus (CoV) baru. CoV baru ini adalah coronavirus 2019 (COVID-19) [1]. Virus COVID-19 ini telah menyebar dengan cepat hampir diseluruh dunia. Meskipun penyebarannya cepat diseluruh dunia, namun karakteristik penyakit pernafasan akut 2019-n CoV sebagian masih belum jelas [2]. Informasi dan cara pencegahan virus ini pun sudah tersebar diberbagai media sosial. Media sosial yang saat ini banyak digunakan salah satunya adalah Instagram. Instagram merupakan media yang kerap digunakan sebagai sarana berbagi informasi tentang berbagai hal. Hal ini karena Instagram menjadi media yang populer saat ini [3]. Dengan berbagai pengguna dari berbagai dunia dari mulai anak-anak hingga orang dewasa turut serta dalam popularitas Instagram.

Analisis setimen adalah proses mengklasifikasi dokumen teks yang berupa opini berdasarkan sentimen untuk menentukan apakah suatu tanggapan tertentu bersifat positif, netral atau negatif [4]. Analisis sentimen juga merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat [5]. Algoritma yang digunakan dalam melakukan klasifikasi pada penelitian ini yaitu Support Vector Machine dan Naïve Bayes. Kelebihan dari Algoritma Support Vector Machine yaitu mampu mengidentifikasi hyperplane terpisah yang memaksimalkan margin antara dua kelas yang berbeda [6]. Namun algoritma Support Vector Machine memiliki kekurangan yaitu pada masalah pemilihan fitur yang sesuai. Particle Swarm Optimization (PSO) ditambahkan untuk meningkatkan kinerjanya.

2. METODE PENELITIAN

Pada penelitian ini menggunakan metode *Naïve Bayes*, dan *Decision Tree* untuk menghasilkan akurasi terbaik dalam menganalisa komentar Instragram terhadap informasi berkaitan dengan Covid-19. Adapun kerangka berfikir pada penelitian ini seperti pada gambar berikut:

Alur analisis sentimen vaksin Covid-19 menggunakan algoritma *Naïve Bayes* dan SVM disajikan dalam Gambar 2. Proses pengumpulan data didapatkan melalui API Twitter untuk mengambil data postingan melalui kata vaksin Covid-19. Setelah data diperoleh maka data tersebut menjadi *dataset* awal untuk dilakukan proses *over sampling* yang membentuk data ke dalam pola yang telah ditentukan dan format yang sesuai untuk proses *preprocessing*.



Gambar 1. Metode Penelitian

Selanjutnya dilakukan *preprocessing* yaitu dengan melakukan sortir ke dalam proses *case folding*, *tokenize*, *stop removal*, *stemming*, *translation* dan *normalization*. Data yang telah dilakukan *preprocessing* selanjutnya siap untuk diimplementasikan ke dalam algoritma *Naïve Bayes*, *SVM* dan *Decision Tree*. Data yang menjadi *input* dibagi menjadi data *training* dan data *testing*. Dalam tahap evaluasi *K-Fold Cross Validation* dilakukan validasi untuk pengujian algoritma, dimana *dataset* diatur melalui skenario 5 *subset* dan di poses sebanyak 5 kali perulangan untuk mendapatkan validasi pengujian. Pada bagian akhir dilakukan validasi pengujian terhadap model yang dibentuk untuk mengukur performa algoritma melalui parameter akurasi, presisi, *recall* dan waktu training sehingga dapat dibuat kesimpulan performa algoritma yang sesuai untuk analisis sentimen vaksin Covid – 19.

3. HASIL DAN PEMBAHASAN

Hasil yang didapat dari berbagai tahapan yang dilakukan sesuai dengan metode yang menjadi rujukan, maka dapat di jelaskan sebagai berikut

3.1 Hasil Analisa Penelitian

1. Data Collection

Crawling adalah semacam pengambilan data dari media sosial kemudian di kumpulkan menjadi satu untuk di evakuasi dan di bentuk agar menjadi sebuah penelitian. Data diambil dengan cara crawling dari media social twitter dengan keywords: vaksin, vaksinasi, dst. Dari tanggal (2021/Juli/15.) sampai dengan tanggal (2022/Juni/30.). Data yang diambil adalah teks tweet dari pengguna yang terelasi dengan kwyword yang dicari. Data yang terkumpul adalah 2.180 data tweet berbentuk teks. Tools yang digunakan untuk crawling adalah: Twint, Python, dan Jupyter Notebook. Data yang dikumpulkan disimpan ke dalam file .csv.

2. Data Pre-processing

Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Salah satu fungsi utama data preparation adalah memastikan keakuratan dan konsistensi data mentah yang disiapkan untuk pemrosesan dan analisis. Dengan demikian hasil analitik akan valid. Tahap data pre-processing yang dilakukan pada penelitian ini adalah: regex text cleaning, text case folding, tokenizing, removing stop words, dan text stemming.

2.1 Regex Text Cleaning

Regex adalah singkatan dari Regular Expression. Regex merupakan sebuah teks (string) yang mendefinisikan sebuah pola pencarian sehingga dapat membantu kita untuk melakukan matching (pencocokan), locate (pencarian), dan manipulasi teks.

Pada tahap ini, dilakukan pembersihan teks dari regex seperti link URL, hashtag, nomor, symbol, dan mention yang ada pada teks tweet. Berikut adalah contoh hasil dari regex cleaning:

2.1 Text Case Folding

Case folding adalah salah satu bentuk text preprocessing yang paling sederhana dan efektif meskipun sering diabaikan. Tujuan dari case folding untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima.

Tabel 1. Analisa Sampel Data *Text Case Folding*

Before	After case_folding
Kajian saintifik mendapati vaksin itu selamat dan berkesan AWANInews AWANI DisiplinMalaysia	kajian saintifik mendapati vaksin itu selamat dan berkesan awaninews awani disiplinmalaysia
Railfriends Pelanggan KA jarak jauh yg belum mendapatkan vaksin ketiga booster wajib menunjukkan hasil negatif tes PCR atau Rapid Test Antigen yang masih berlaku pada saat boarding Kebijakan berlaku mulai keberangkatan Juli Selengkapnya	railfriends pelanggan ka jarak jauh yg belum mendapatkan vaksin ketiga booster wajib menunjukkan hasil negatif tes pcr atau rapid test antigen yang masih berlaku pada saat boarding kebijakan berlaku mulai keberangkatan juli selengkapnya
Berikan Edukasi kepada Anakanak Tentang Bahaya Vaksin Covid supaya nanti Jika Sekolah Melakukan Vaksinasi Untuk Ana...	berikan edukasi kepada anakanak tentang bahaya vaksin covid supaya nanti jika sekolah melakukan vaksinasi untuk ana...
Badan cekat cekit habis vaksin booster tadi pagiii jam berapa ya aku vaksin	badan cekat cekit habis vaksin booster tadi pagiii jam berapa ya aku vaksin

2.2 Slang Word Adjustment

Slang words merupakan bahasa gaul yang sering digunakan oleh anak-anak muda di Amerika atau di Inggris. Remaja baik di Amerika Serikat, Inggris, bahkan di Indonesia menggunakan bahasa gaul dalam percakapan atau tulisan non-formal. Karena di twitter penulisan tidak diatur sehingga user bebas menulis sesuai keinginannya. Hal ini menyebabkan harusnya penyesuaian slang word yang digunakan oleh user ke dalam Bahasa Indonesia yang sesuai dengan kaidah.

Tabel 2. Hasil Analisa Sampel Data *Slang Word Adjusment*

Before	After slang_word
kajian saintifik mendapati vaksin itu selamat dan berkesan awaninews awani disiplinmalaysia	kajian saintifik mendapati vaksin itu selamat dan berkesan awaninews awani disiplinmalaysia

railfriends pelanggan ka jarak jauh yg belum mendapatkan vaksin ketiga booster wajib menunjukkan hasil negatif tes pcr atau rapid test antigen yang masih berlaku pada saat boarding kebijakan berlaku mulai keberangkatan juli selengkapnya	railfriends pelanggan kak jarak jauh yang belum mendapatkan vaksin ketiga booster wajib menunjukkan hasil negatif tes pcr atau rapid test antigen yang masih berlaku pada saat boarding kebijakan berlaku mulai keberangkatan juli selengkapnya
berikan edukasi kepada anakanak tentang bahaya vaksin covid supaya nanti jika sekolah melakukan vaksinasi untuk ana...	berikan edukasi kepada anakanak tentang bahaya vaksin covid supaya nanti jika sekolah melakukan vaksinasi untuk anak...
badan cekat cekit habis vaksin booster tadi pagiii	badan cekat cekit habis vaksin booster tadi pagiii
jam berapa ya aku vaksin	jam berapa ya aku vaksin

2.3 Tokenizing

Tokenizing atau disebut juga tahap Lexical Analysis adalah proses pemotongan teks menjadi bagian-bagian yang lebih kecil, yang disebut token. Pada proses ini juga dilakukan penghilangan angka, tanda baca dan karakter lain yang dianggap tidak memiliki pengaruh terhadap pemrosesan teks.

Tabel 3. Hasil Analisa Sampel Tokenizing Data

Before	After Tokenizing
kajian saintifik mendapati vaksin itu selamat dan berkesan awaninews awani disiplinmalaysia	['kajian', 'saintifik', 'mendapati', 'vaksin', 'itu', 'selamat', 'dan', 'berkesan', 'awaninews', 'awani', 'disiplinmalaysia']
railfriends pelanggan kak jarak jauh yang belum mendapatkan vaksin ketiga booster wajib menunjukkan hasil negatif tes pcr atau rapid test antigen yang masih berlaku pada saat boarding kebijakan berlaku mulai keberangkatan juli selengkapnya	['railfriends', 'pelanggan', 'kak', 'jarak', 'jauh', 'yang', 'belum', 'mendapatkan', 'vaksin', 'ketiga', 'booster', 'wajib', 'menunjukkan', 'hasil', 'negatif', 'tes', 'pcr', 'atau', 'rapid', 'test', 'antigen', 'yang', 'masih', 'berlaku', 'pada', 'saat', 'boarding', 'kebijakan', 'berlaku', 'mulai', 'keberangkatan', 'juli', 'selengkapnya']
berikan edukasi kepada anakanak tentang bahaya vaksin covid supaya nanti jika sekolah melakukan vaksinasi untuk ana...	['berikan', 'edukasi', 'kepada', 'anakanak', 'tentang', 'bahaya', 'vaksin', 'covid', 'supaya', 'nanti', 'jika', 'sekolah', 'melakukan', 'vaksinasi', 'untuk', 'ana...']
badan cekat cekit habis vaksin booster tadi pagiii	['badan', 'cekat', 'cekit', 'habis', 'vaksin', 'booster', 'tadi', 'pagiii']
jam berapa ya aku vaksin	['jam', 'berapa', 'ya', 'aku', 'vaksin']

2.4 Stop Words Removal

Dalam dunia pemrograman seperti NLP (Natural Language Processing), stopwords merupakan kata yang diabaikan dalam pemrosesan dan biasanya disimpan di dalam stop lists. Stop list ini berisi daftar kata umum yang mempunyai fungsi tapi tidak mempunyai arti. Stopword Removal adalah proses filtering, pemilihan kata-kata penting dari hasil token yaitu kata-kata apa saja yang di gunakan untuk mewakili dokumen.

Tabel 4. Hasil Analisa Sampel *Stop Words Removal*

Before	After Stopword Removal
['kajian', 'saintifik', 'mendapati', 'vaksin', 'itu', 'selamat', 'dan', 'berkesan', 'awaninews', 'awani', 'disiplinmalaysia']	['kajian', 'saintifik', 'mendapati', 'vaksin', 'selamat', 'berkesan', 'awaninews', 'awani', 'disiplinmalaysia']
['railfriends', 'pelanggan', 'kak', 'jarak', 'jauh', 'yang', 'belum', 'mendapatkan', 'vaksin', 'ketiga', 'booster', 'wajib', 'menunjukkan', 'hasil', 'negatif', 'tes', 'pcr', 'atau', 'rapid', 'test', 'antigen', 'yang', 'masih', 'berlaku', 'pada', 'saat', 'boarding', 'kebijakan', 'berlaku', 'mulai', 'keberangkatan', 'juli', 'selengkapnya']	['railfriends', 'pelanggan', 'kak', 'jarak', 'vaksin', 'ketiga', 'booster', 'wajib', 'hasil', 'negatif', 'tes', 'pcr', 'rapid', 'test', 'antigen', 'berlaku', 'boarding', 'kebijakan', 'berlaku', 'keberangkatan', 'juli', 'selengkapnya']
['berikan', 'edukasi', 'kepada', 'anakanak', 'tentang', 'bahaya', 'vaksin', 'covid', 'supaya', 'nanti', 'jika', 'sekolah', 'melakukan', 'vaksinasi', 'untuk', 'ana...']	['edukasi', 'anakanak', 'bahaya', 'vaksin', 'covid', 'sekolah', 'vaksinasi', 'ana...']
['badan', 'cekat', 'cekit', 'habis', 'vaksin', 'booster', 'tadi', 'pagiii']	['badan', 'cekat', 'cekit', 'habis', 'vaksin', 'booster', 'pagiii']
['jam', 'berapa', 'ya', 'aku', 'vaksin']	['jam', 'ya', 'vaksin']

2.5 Text Stemming

Stemming adalah metode untuk mencari kata dasar dari sebuah kata. Stemming merupakan salah satu tahapan dalam pre processing. Proses stemming memiliki pengaruh dalam tingkat akurasi temu kembali informasi. Stemming dilakukan dengan cara menghilangkan imbuhan yang terdapat pada kata.

Tabel 5. Analisa Sampel *Stemming Text*

Before	After Stemming
['kajian', 'saintifik', 'mendapati', 'vaksin', 'selamat', 'berkesan', 'awaninews', 'awani', 'disiplinmalaysia']	kaji saintifik dapat vaksin selamat kesan awaninews awan disiplinmalaysia
['railfriends', 'pelanggan', 'kak', 'jarak', 'vaksin', 'ketiga', 'booster', 'wajib', 'hasil', 'negatif', 'tes', 'pcr', 'rapid', 'test', 'antigen', 'berlaku', 'boarding', 'kebijakan', 'berlaku', 'keberangkatan', 'juli', 'selengkapnya']	railfriends langgan kak jarak vaksin tiga booster wajib hasil negatif tes pcr rapid test antigen laku boarding bijak laku berangkat juli lengkap
['edukasi', 'anakanak', 'bahaya', 'vaksin', 'covid', 'sekolah', 'vaksinasi', 'ana...']	edukasi anakanak bahaya vaksin covid sekolah vaksinasi ana
['badan', 'cekat', 'cekit', 'habis', 'vaksin', 'booster', 'pagiii']	badan cekat cekit habis vaksin booster pagiii
['jam', 'ya', 'vaksin']	jam ya vaksin

4. Naïve Bayes Model and Decision Tree

3.1 Data Preparation

Datayang akan diolah adalah text yang sudah melalui tahap pre-processsing dan juga label dari sentiment,. Jumlah data yang akan diolah adalah 2180 data. Data tersebut dibersihkan adari duplikasi data dan juga dari data kolom yang kasong atau Null. Sehingga data menjadi 2175 setelah dibersihkan.

3.2 TF-IDF

TFIDF (Term Frequency Inverse Document Frequency) merupakan metode pembobotan dalam bentuk integrasi antar term frequency dengan inverse document frequency. Metode TFIDF digunakan pada penelitian ini untuk memilih fitur sebagai hasil ringkasan, dengan penerapannya pada seleksi fitur bobot kata.

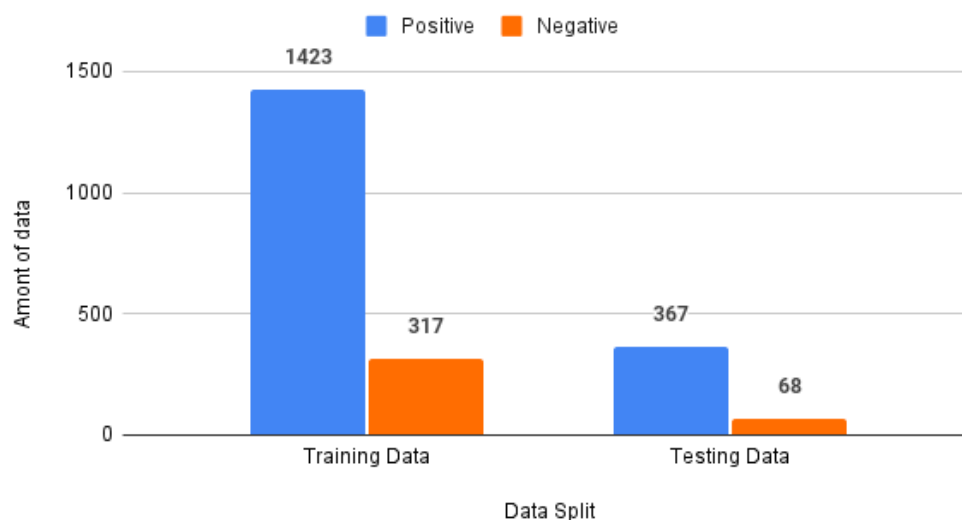
Berikut adalah hasil dari sentiment analysis oleh Naïve Bayes dan Decision Tree:

Tabel 6. Data hasil sentimen

Method	Sentiment	precision	recall	F1-score	support	accuracy
Naïve	N	44%	34%	38%	68	83%
Bayes	P	88%	92%	90%	367	
Decision	N	36%	38%	37%	68	79%
Tree	P	88%	87%	88%	367	

Berdasarkan table hasil sentiment analysis, performa prediksi analisis sentiment dengan menggunakan Naïve Bayes lebih baik dibandingkan dengan Decision Tree. Naïve Bayes memiliki akurasi 83% sedangkan Decision Tree memiliki akurasi 78%. Keduanya memiliki akurasi yang lebih dari 75% bahkan mencapai 80%. Meskipun demikian, jika kita melihat secara spesifik kepada pembagian sentiment Positif (P) dan Negatif (N), hasil performa prediksi untuk sentiment yang Negatif (N) jauh lebih rendah dibanding performa prediksi untuk sentiment yang Positif (P). Hal ini disebabkan oleh sedikitnya data Sentimen Negatif (N) pada data, terutama data training sehingga learner perlu mempelajari lebih banyak pattern dari kalimat sentiment Negatif (N) lebih banyak lagi.

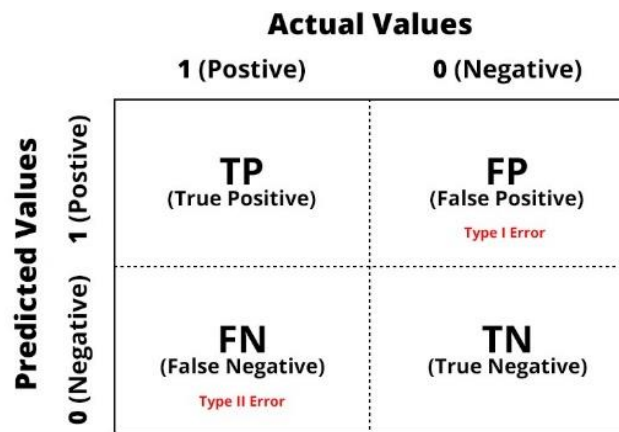
Training Data and Testing Data



Gambar 3. Grafik Perbandingan Data Training dan Data Testing

3.2 Pembahasan

Confusion Matrix yang digunakan adalah pengukuran performa untuk masalah klasifikasi machine learning dimana keluaran dapat berupa dua kelas atau lebih. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual.



Gambar 3. Nilai prediksi

True Positive (TP):

Interpretasi: Anda memprediksi positif dan itu benar.

Sistem memprediksikan bahwa suatu Tweet merupakan Positif (P) dan Tweet tsb memang benar memiliki sentiment Positif (P).

True Negative (TN):

Interpretasi: Anda memprediksi negatif dan itu benar.

Sistem memprediksikan bahwa suatu Tweet merupakan Negatif (N) dan Tweet tsb memang benar memiliki sentiment Negatif (N).

False Positive (FP): (Kesalahan Tipe 1)

Interpretasi: Anda memprediksi positif dan itu salah.

Sistem memprediksikan bahwa suatu Tweet merupakan Positif (P) akan tetapi Tweet tsb ternyata memiliki sentiment Negatif (N).

False Negative (FN): (Kesalahan Tipe 2, kesalahan tipe 2 ini sangat berbahaya)

Interpretasi: Anda memprediksi negatif dan itu salah.

Sistem memprediksikan bahwa suatu Tweet merupakan Negatif (N) akan tetapi Tweet tsb ternyata memiliki sentiment Positif (F).

Tabel 6. Confusion Matrix dari system Naïve Bayes dan Decision Tree

Naïve Bayes		Actual Values	
		P	N
Predicted	P	338	45
	N	29	23

Decision Tree		Actual Values	
		P	N
Predicted	P	321	42
	N	46	26

Method	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
Naïve Bayes	338	23	45	29
Decision Tree	321	26	42	46

Meskipun tingkat akurasi Naïve Bayes lebih tinggi dibandingkan Decision tree, Decision Tree memiliki performa lebih baik dalam memprediksi sentiment negative, terbukti dengan nilai TN lebih besar meskipun FN nya jauh lebih banyak.

Penyebab miskalsifikasi (kesalahan prediksi):

- o Teks data twitter yang tidak terstruktur dan banyak mengandung kesalahan dalam konteks penulisan oleh pengguna twitter sehingga lolos dalam tahap cleaning.

- Penggunaan Bahasa daerah pada teks yang belum dapat disesuaikan ke Bahasa Indonesia.
- Slang word yang belum bisa menanggulangi penggunaan kata informal yang terus berkembang.
- Tidak seimbang data sentiment positif dan negative sehingga system kurang baik dalam memprediksi terutama sentiment negative.
- Data mengandung konten kmapanye dan politik yang tidak relevan dengan topik

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil kesimpulan bahwa analisis sentimen dari media sosial masyarakat memberikan tanggapan positif terhadap vaksinasi Covid-19 yang dilakukan oleh pemerintah dari dua metode analisa yang digunakan terdapat gap penilaian hasil akurasi perhitungan. Dimana metode Naïve bayes lebih rendah hasil akurasiya berbanding metode decision tree. Diharapkan studi sentimen analisis pengguna media sosial ini dapat menyertakan jumlah tweet yang lebih banyak, agar sebaran nilai toleransi lebih tinggi.

Daftar Pustaka

- [1] World Health Organization (WHO). 2021. Coronavirus disease (COVID-19) pandemic <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, diakses 25 Januari 2021
- [2] Nareza, M. (2021). Jangan Termakan Isu, Ini Fakta Penting Vaksin COVID-19. <https://www.alodokter.com/>, diakses 25 Januari 2021
- [3] Our World In Data. 2021. Coronavirus (COVID-19) Vaccinations. <https://ourworldindata.org/covid-vaccinations>, diakses 25 Januari 2021
- [4] Mubarak, M.S., Adiwijaya, Aldhi, M.D., 2016. Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes. 2nd International Conference on Mathematics: Pure, Applied and Computation (ICOMPAC)
- [5] Tripathy, A., Agrawal, A. and Rath, S.K., 2016. Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, pp.117-126
- [6] M. Adriani, J. Asian, B. Nazief, and H. E. Williams, Stemming Indonesian : A Confix-Stripping Approach, ACM Transactions on Asian Language Information Processing (TALIP), vol. 6, no. 4. pp. 1–33, 2007.
- [7] Ratino, Noor, H., Sita A., Windu, G. 2020. "Sentimen Analisis Informasi Covid-19 menggunakan Support Vector Machine dan Naïve Bayes". Jurnal JUPITER Vol. 12 No. 2 Bulan 10 Tahun 2020, Hal 1 - 11
- [8] Muhammad, A., Agus, A, Fauziah, H. 2020. " Analisis sentimen vaksin COVID-19 di Indonesia menggunakan Algoritma Naïve Bayes". Konferensi Tahunan tentang Ilmu Komputer dan Teknologi Rekayasa (AC2SET). doi: 10.1088 / 1757-899X / 1088/1/012045.
- [9] Ardianne, L., F., Rima, D., R., Nia, A., F. 2021. "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter". Jurnal Dinda. Vol.1x No.1 (2021) Feb 2021.
- [10] Cahyo, P., Nisa, H. 2020. "Analisis Sentimen Covid-19 Sebagai Pandemi Media Sosial". Jurnal Internasional Sistem Informasi & Teknologi . Vol. 4, Tidak. 1, (2020), hlm.509-517.

- [11] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” *KILAT*, vol. 8, no. 1, pp. 81–92, may 2019.
- [12] M. Adriani, J. Asian, B. Nazief, S. M. Tahaghoghi, and H. E. Williams, “Stemming Indonesian: A confix-stripping approach,” *CM Transactions on Asian Language Information Processing*, vol. 6, no. 4, pp. 1–33, dec 2007.
- [13] A. M. Raya, F. Nurbaiti, and D. Sofia, “Klasifikasi Sentimen Masyarakat Terhadap Kenaikan Harga Tiket Pesawat Pada Twitter Menggunakan Naive Bayes,” *Instek*, vol. 4, no. 2, pp. 239–246, 2019.
- [14] Murphy SC. A hands-on guide to conducting psychological research on Twitter. *Social Psychological and Personality Science*. 2017;8(4):396–412.
- [15] Wang W, Hernandez I, Newman DA, He J, Bian J. Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology*. 2016;65(2):355–378.
- [16] Ordenes FV, Theodoulidis B, Burton J, Gruber T, Zaki M. Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*. 2014;17(3):278–295.



ZONAsi: Jurnal Sistem Informasi

is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)