**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

# Analisis Sentimen Masyarakat Indonesia terhadap Keterlibatan Bill Gates dalam Program Vaksin TBC di Media Sosial X Menggunakan SVM

**Fakhita Fahraini*, Sriani**

FakultasSains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: [1*]fakhita0701212103@uinsu.ac.id, [2]sriani@uinsu.ac.id
Email Penulis Korespondensi: fakhita0701212103@uinsu.ac.id
Submitted **21-12-2025;** Accepted **27-01-2026;** Published **19-02-2026**

**Abstrak**

Keterlibatan Bill Gates dalam program pengembangan vaksin tuberkulosis (TBC) menimbulkan beragam respons di kalangan masyarakat Indonesia, khususnya yang diekspresikan melalui media sosial X, sehingga penting untuk dikaji karena sentimen publik dapat memengaruhi tingkat kepercayaan masyarakat terhadap program kesehatan dan keberhasilan upaya pencegahan TBC. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap keterlibatan Bill Gates dalam program vaksin TBC berbasis data media sosial X menggunakan algoritma Support Vector Machine (SVM). Data penelitian terdiri atas 784 tweet berbahasa Indonesia yang dikumpulkan melalui teknik web scraping berdasarkan kata kunci terkait isu vaksin TBC dan Bill Gates. Data yang diperoleh selanjutnya melalui tahapan pra-pemrosesan teks yang meliputi pembersihan data (cleaning), case folding, tokenisasi, normalisasi kata, penghapusan stopword, dan stemming guna meningkatkan kualitas dan konsistensi teks. Representasi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), sementara pembagian data dilakukan dengan skema 80% sebagai data latih dan 20% sebagai data uji. Model klasifikasi dibangun menggunakan algoritma SVM dengan kernel linear untuk memisahkan kelas sentimen positif dan negatif secara optimal. Hasil pengujian menunjukkan bahwa model SVM berbasis TF-IDF mampu mencapai tingkat akurasi sebesar 96,8%, dengan nilai presisi 98,3%, recall 79,2%, dan F1-score sebesar 86,0%. Analisis distribusi sentimen menunjukkan bahwa mayoritas tweet didominasi oleh sentimen negatif, sementara sentimen positif muncul dalam jumlah yang lebih kecil. Kontribusi utama penelitian ini terletak pada penerapan analisis sentimen berbasis media sosial secara spesifik pada isu vaksin TBC yang dikaitkan dengan tokoh global dalam konteks Indonesia, serta pembuktian bahwa kombinasi SVM dan TF-IDF efektif dalam mengklasifikasikan opini publik secara akurat. Temuan ini diharapkan dapat menjadi dasar bagi pemerintah dan pemangku kepentingan kesehatan dalam merancang strategi komunikasi publik yang lebih tepat dan berbasis data.

**Kata Kunci:** Analisis Sentimen; Vaksin TBC; Bill Gates; Support Vector Machine; TF-IDF

**Abstract**

The involvement of Bill Gates in the development program of the tuberculosis (TB) vaccine has generated diverse responses among the Indonesian public, particularly as expressed through social media platform X, making it an important issue to examine since public sentiment may influence public trust in health programs and the success of TB prevention efforts. This study aims to analyze public sentiment in Indonesia toward Bill Gates' involvement in the TB vaccine program based on social media data from platform X using the Support Vector Machine (SVM) algorithm. The research data consist of 784 Indonesian-language tweets collected through web scraping techniques using keywords related to the TB vaccine issue and Bill Gates. The collected data then underwent several text preprocessing stages, including data cleaning, case folding, tokenization, word normalization, stopword removal, and stemming to improve text quality and consistency. Feature representation was performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, while the dataset was split into 80% training data and 20% testing data. The classification model was built using an SVM algorithm with a linear kernel to optimally separate positive and negative sentiment classes. The experimental results show that the SVM model combined with TF-IDF achieved an accuracy of 96.8%, with a precision of 98.3%, a recall of 79.2%, and an F1-score of 86.0%. Sentiment distribution analysis indicates that the majority of tweets were dominated by negative sentiment, while positive sentiment appeared in a smaller proportion. The main contribution of this study lies in the application of social media–based sentiment analysis specifically to the TB vaccine issue associated with a global public figure in the Indonesian context, as well as in demonstrating that the combination of SVM and TF-IDF is effective in accurately classifying public opinion. These findings are expected to serve as a data-driven reference for government institutions and public health stakeholders in designing more effective public communication strategies.

**Keywords**: Sentiment Analysis; TB Vaccine; Bill Gates; Support Vector Machine; TF-IDF

## 1. PENDAHULUAN

Tuberkulosis (TBC) merupakan salah satu penyakit menular yang hingga saat ini masih menjadi permasalahan kesehatan global yang serius. Penyakit ini disebabkan oleh bakteri Mycobacterium tuberculosis dan menyerang terutama organ paru-paru, meskipun dapat pula memengaruhi organ lainnya. Berdasarkan laporan Organisasi Kesehatan Dunia (World Health Organization), TBC masih menyebabkan jutaan kasus baru setiap tahunnya dengan angka kematian yang signifikan, khususnya di negara-negara berkembang [1], [2]. Indonesia termasuk dalam kelompok negara dengan beban TBC tertinggi di dunia, sehingga membutuhkan penanganan yang berkelanjutan dan berbasis strategi jangka panjang [3], [4]. Berbagai upaya telah dilakukan oleh pemerintah dan lembaga internasional untuk menekan penyebaran TBC, mulai dari peningkatan deteksi dini, pengobatan berkelanjutan, hingga pengembangan vaksin TBC generasi baru sebagai langkah pencegahan yang lebih efektif [5], [6].

Salah satu inisiatif yang mendapat perhatian luas adalah pengembangan vaksin TBC yang didukung oleh berbagai lembaga internasional, termasuk Yayasan Bill & Melinda Gates [7], [8]. Keterlibatan Bill Gates sebagai tokoh global dalam isu kesehatan publik sering kali menjadi sorotan masyarakat karena rekam jejaknya dalam mendukung program vaksinasi dan pengendalian penyakit menular di berbagai negara [9], [10]. Namun, keterlibatan tersebut juga

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

memunculkan pro dan kontra di tengah masyarakat. Di satu sisi, Bill Gates dipandang sebagai filantropis yang berkontribusi besar dalam pengembangan kesehatan global. Di sisi lain, muncul sentimen negatif berupa ketidakpercayaan, teori konspirasi, serta penyebaran misinformasi yang berpotensi memengaruhi persepsi publik terhadap program vaksinasi, termasuk vaksin TBC yang sedang dikembangkan [11], [12].

Perkembangan media sosial turut mempercepat penyebaran opini, informasi, dan disinformasi terkait isu kesehatan. Media sosial X (Twitter) menjadi salah satu platform utama yang digunakan masyarakat untuk menyampaikan pandangan, baik dalam bentuk dukungan maupun penolakan, terhadap isu vaksin TBC dan keterlibatan Bill Gates di dalamnya [13], [14]. Opini yang berkembang di media sosial dapat membentuk persepsi kolektif masyarakat dan berpengaruh terhadap tingkat kepercayaan publik terhadap kebijakan kesehatan. Oleh karena itu, pemahaman terhadap kecenderungan sentimen masyarakat menjadi hal yang penting agar pihak terkait dapat merancang strategi komunikasi kesehatan yang tepat dan berbasis data [15], [16].

Analisis sentimen merupakan pendekatan yang banyak digunakan untuk mengidentifikasi kecenderungan opini publik secara otomatis dengan memanfaatkan teknik text mining dan machine learning. Metode ini mampu mengklasifikasikan opini ke dalam kategori sentimen positif atau negatif berdasarkan pola kata yang muncul dalam teks. Salah satu algoritma yang banyak diterapkan dalam klasifikasi teks adalah Support Vector Machine (SVM) karena kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan tingkat akurasi yang baik. Beberapa penelitian terdahulu menunjukkan bahwa SVM efektif digunakan dalam analisis sentimen media sosial, seperti pada penelitian [17], yang memperoleh akurasi tinggi dalam analisis sentimen politik di Twitter, serta penelitian [18], yang membuktikan kinerja SVM dalam klasifikasi teks pada kasus pencemaran nama baik [19], [20].

Meskipun demikian, berbagai penelitian sebelumnya menunjukkan variasi pendekatan metode, jenis dataset, serta hasil performa yang berbeda-beda. Penelitian oleh Angelina pada tahun menggunakan algoritma Naïve Bayes untuk analisis sentimen isu kesehatan berbasis Twitter dengan dataset berskala kecil, namun memiliki keterbatasan pada akurasi klasifikasi [21]. Penelitian lain oleh Palomino tahun 2022 menerapkan algoritma K-Nearest Neighbor (KNN) dengan hasil akurasi yang cukup baik, tetapi kurang optimal dalam menangani data berdimensi tinggi [22]. Sementara itu, penelitian Purnomo tahun 2023 memanfaatkan Random Forest untuk klasifikasi sentimen vaksin COVID-19, namun memerlukan waktu komputasi yang lebih besar [23]. Penelitian Munandar, Farikhin, dan Widodo tahun 2023 menggunakan Deep Learning berbasis LSTM dengan performa tinggi, tetapi membutuhkan dataset yang sangat besar dan proses pelatihan yang kompleks [24]. Adapun penelitian Emarapenta, Sinulingga, Cesar, dan Sitorus tahun 2024 menunjukkan bahwa SVM memberikan keseimbangan antara akurasi, efisiensi komputasi, dan kestabilan model pada data teks berukuran menengah [25].

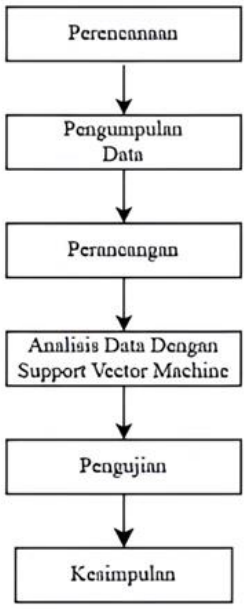
Berdasarkan perbandingan tersebut, SVM dipilih dalam penelitian ini karena memiliki keunggulan dalam mengklasifikasikan data teks berdimensi tinggi, tidak memerlukan dataset yang sangat besar seperti metode deep learning, serta mampu menghasilkan performa yang stabil dan konsisten dibandingkan algoritma klasifikasi lainnya pada kasus analisis sentimen media sosial. Ringkasan perbandingan penelitian terkait, yang mencakup penulis, tahun, metode, jenis dataset, tingkat akurasi, dan keterbatasan masing-masing penelitian, disajikan secara sistematis dalam Tabel 1 untuk memperjelas posisi dan kontribusi penelitian ini dibandingkan penelitian sebelumnya. Meskipun penelitian terkait analisis sentimen menggunakan SVM telah banyak dilakukan, masih terbatas penelitian yang secara khusus menganalisis sentimen masyarakat Indonesia terhadap isu vaksin TBC yang dikaitkan dengan keterlibatan tokoh global seperti Bill Gates berbasis data media sosial X. Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap keterlibatan Bill Gates dalam program vaksin TBC menggunakan algoritma Support Vector Machine (SVM). Hasil penelitian ini diharapkan dapat memberikan gambaran objektif mengenai persepsi publik serta menjadi dasar pertimbangan bagi pemerintah dan organisasi kesehatan dalam merumuskan strategi komunikasi dan edukasi kesehatan yang lebih efektif dan tepat sasaran.

# 2. METODOLOGI PENELITIAN

## 2.1 Kerangka Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif dengan memanfaatkan data sekunder yang diperoleh dari media sosial X (Twitter). Data yang digunakan berupa tweet yang membahas isu vaksin tuberkulosis (TBC) oleh Bill Gates. Pengambilan data dilakukan menggunakan teknik web scraping dengan bantuan Google Colab dan bahasa pemrograman Python. Penggunaan data media sosial dalam penelitian kuantitatif dinilai efektif untuk merepresentasikan opini publik secara luas dan actual. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada pengukuran dan analisis numerik terhadap data teks dalam jumlah besar untuk mengidentifikasi pola sentimen masyarakat secara objektif. Kerangka penelitian disusun untuk memberikan gambaran sistematis mengenai tahapan yang dilakukan selama proses penelitian. Tahapan tersebut meliputi perencanaan penelitian, pengumpulan data, perancangan sistem, analisis data menggunakan algoritma Support Vector Machine (SVM), hingga penarikan kesimpulan. Alur kerangka penelitian ini disajikan dalam Gambar 1 yang menunjukkan hubungan antar tahapan penelitian secara berurutan dan terstruktur [26].

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom




**Gambar 1.** Kerangka Penelitian

### 2.2 Rencana Pembahasan

#### 2.2.1 Perencanaan

Pada tahap perencanaan, peneliti merancang penelitian secara menyeluruh dengan menentukan topik, metode, serta sumber data yang digunakan. Metode yang diterapkan dalam penelitian ini adalah Support Vector Machine (SVM) untuk melakukan analisis sentimen terhadap tweet di media sosial X. Fokus penelitian diarahkan pada isu vaksin TBC yang dikaitkan dengan Bill Gates, mengingat isu tersebut menimbulkan beragam respons publik. Pada tahap ini juga ditentukan tahapan analisis, teknik pemrosesan data, serta metode evaluasi yang akan digunakan untuk mengukur kinerja model klasifikasi [27].

#### 2.2.2 Pengumpulan Data

Pengumpulan data dilakukan melalui dua pendekatan, yaitu field research dan library research.

a. Field Research

Penelitian lapangan dilakukan dengan mengamati secara langsung objek penelitian, yaitu tweet yang membahas vaksin TBC oleh Bill Gates pada media sosial X. Data diperoleh melalui proses crawling menggunakan bahasa pemrograman Python dengan kata kunci “tbc_billgates”. Dari proses tersebut diperoleh sebanyak 784 tweet berbahasa Indonesia. Dataset yang dihasilkan kemudian disimpan dalam format CSV untuk memudahkan proses pengolahan data pada tahap selanjutnya. Contoh dataset hasil crawling disajikan pada Tabel 1.

**Tabel 1.** Contoh *Dataset*

| No | Tweet |
|---|---|
| 1. | Indonesia Jadi Lokasi Uji Coba Vaksin TBC Buatan Bill Gates https://t.co/FayMU4GYVH |
| 2. | Anggota DPRD Jatim Ma mulah Harun meminta pemerintah tidak gegabah dalam mengadopsi vaksin TBC dari Bill Gates. Ia menekankan pentingnya data dan kajian yang transparan agar masyarakat tidak menjadi objek uji coba. #FraksiPKBJatim #PKBJatimPeduli #JatimLebihBaik #BersamaPKB https://t.co/Fhjh39mLH8 |
| 3. | Vaksin tbc dari BIll Gates kalo disuntik bukan menyembuhkan akan kejang2 dan menambah penyakit..lihat aja dari jarumnya..Bisnis atau Pekerjaan tdk benar pasti akan ketahuan dahulu sama seprrti vaksin covid kita memperingati bahwa vaksin covid dan wabahnya itu direncanakan |
| 4. | salut bgt sama langkah visioner dari Presiden Prabowo dan Bill Gates ini. Kalo serius berantas TBC dan jamin makan bergizi masa depan anak bangsa pasti lebih cerah https://t.co/2vh34b2Shr |
| 5. | Vaksin TBC dari Bill Gates bukan mencegah tapi menambah penyakit..lihat aja vaksin covid setelah orang habis divaksin covid banyak org pada bertambah penyakit seperti jantung gagal ginjal kaki dll...Bill gates pengen berbuat demikian agar manusia berkurang sdgkan pejabat |

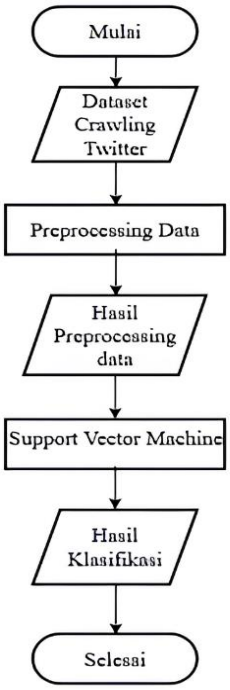
b. Library Research

Penelitian kepustakaan dilakukan dengan mengumpulkan referensi dari berbagai sumber literatur seperti buku, jurnal ilmiah, makalah, dan penelitian terdahulu yang relevan dengan topik analisis sentimen, text mining, dan algoritma Support Vector Machine. Literatur ini digunakan sebagai dasar teori dan pembanding hasil penelitian.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

### 2.2.3 Perancangan

Pada tahap perancangan, peneliti merancang sistem analisis sentimen setelah dataset diperoleh. Sistem yang dirancang bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap isu vaksin TBC oleh Bill Gates menggunakan algoritma SVM. Perancangan sistem digambarkan dalam bentuk flowchart yang menunjukkan alur proses mulai dari input data tweet, pra-pemrosesan teks, pembobotan kata menggunakan TF-IDF, proses klasifikasi dengan SVM, hingga keluaran berupa hasil klasifikasi sentimen. Flowchart sistem ditunjukkan pada Gambar 2.




**Gambar 2.** *Flowchart* SVM

### 2.2.4 Analisis Data dengan Support Vector Machine

Dalam melakukan analisis data menggunakan metode Support Vector Machine, penelitian ini memerlukan perangkat lunak. Perangkat lunak yang digunakan meliputi Google Chrome, Operation Windows 11, Google Colab serta bahasa pemrograman Python. Pembobotan kata akan dilakukan menggunakan TF-IDF, dan analisis sentimen akan diterapkan menggunakan Support Vector Machine. Sebelum melakukan analisis sentimen dengan Support Vector Machine, data akan diproses melalui tahapan pemrosesan dan pembobotan kata menggunakan TF-IDF [28].

### 2.2.5 Pengujian

a. *Text Preprocessing*

1. *Cleaning*, yaitu menghapus karakter khusus, simbol, URL, dan tanda baca yang tidak diperlukan.
2. *Case folding*, yaitu mengubah seluruh huruf menjadi huruf kecil (*lowercase*).
3. *Tokenizing*, yaitu memecah teks menjadi unit kata.
4. *Normalization*, yaitu mengubah kata tidak baku menjadi kata baku sesuai KBBI.
5. *Stopword removal*, yaitu menghapus kata-kata yang tidak memiliki makna penting.
6. *Stemming*, yaitu mengubah kata berimbuhan menjadi kata dasar.

b. Pelabelan Data

Data teks kemudian diberi label sentimen, yaitu positif dan negatif, untuk memudahkan proses pembelajaran mesin dan evaluasi hasil klasifikasi.

c. Pembobotan Kata

Pembobotan kata dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk merepresentasikan teks ke dalam bentuk numerik yang dapat diproses oleh algoritma SVM [29].

### 2.2.6 Kesimpulan

Setelah data diklasifikasi dan dianalisis menggunakan metode SVM, kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Nilai evaluasi ini digunakan untuk menilai efektivitas algoritma SVM dalam menganalisis sentimen masyarakat terhadap isu vaksin TBC oleh Bill Gates di media sosial X [30].

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

# 3. HASIL DAN PEMBAHASAN

Pada bagian ini akan diuraikan hasil implementasi dan pengujian sistem klasifikasi sentimen menggunakan metode Support Vector Machine (SVM) untuk menganalisis opini publik terhadap vaksin tuberkulosis Bill Gates. Penelitian ini melibatkan pengolahan 784 data tweet yang dibagi menjadi 80% data latih (627 dokumen) dan 20% data uji (157 dokumen). Implementasi sistem dilakukan menggunakan bahasa pemrograman Python pada platform Google Colaboratory dengan memanfaatkan berbagai library seperti Pandas, NumPy, Sastrawi, NLTK, dan Scikit-learn.

## 3.1 Analisis Data

Secara spesifik, bab ini menguraikan prosedur sistematis yang dilakukan untuk mempersiapkan dan memproses data guna klasifikasi sentimen. Rangkaian langkah-langkah dalam proses, yakni Representasi Data, Preprocessing Data, Pelabelan Data, dan Pembobotan TF-IDF

### 3.1.1 Representasi Data

Data set tersebut diperoleh dengan menggunakan pendekatan crawling yang memanfaatkan tweet harvest dan auth tokens untuk mengakses akun Twitter guna mencari frasa "tbc_billgates." Sebanyak 784 tweet berhasil dikumpulkan. Setelah proses pengumpulan selesai, data set tersebut diunduh dalam format ".csv" untuk analisis lebih lanjut. Setelah itu, data disusun dalam tabel untuk memudahkan analisis.

### 3.1.2 Preprocessing Data

Setelah memahami pentingnya prapemrosesan (preprocessing) sebagai langkah pertama dalam klasifikasi teks, di mana data harus dipersiapkan dan diubah ke format yang lebih ideal untuk mendapatkan informasi berkualitas tinggi. Contoh data tweet mentah sebelum serangkaian proses tersebut diaplikasikan dapat dilihat pada tabel 2 berikut.

**Tabel 2.** Contoh Data Tweet

| Data Tweet |
|---|
| Indonesia Jadi Lokasi Uji Coba Vaksin TBC Buat Bill Gates |
| Mau ketawa bacanya sama Bill Gates aja tunduk |
| Tersiratnya. TNI vs Geng Solo yg dibackup Bill Gates |
| Ini sejalan dengan Bill Gates sebelumnya |

Berikut adalah langkah-langkah preprocessing data:

a. Cleaning

Data Tweet dibersihkan untuk menghilangkan karakter yang tidak diperlukan seperti tanda baca, angka, URL, dan nama pengguna. Karakter tersebut dapat dihilangkan atau diganti dengan karakter tertentu. Berikut contoh dari proses sebelum dan sesudah cleaning dapat dilihat pada tabel 3 berikut.

**Tabel 3.** Contoh Data *Cleaning*

| Data Tweet | Hasil Cleaning |
|---|---|
| Indonesia Jadi Lokasi Uji Coba Vaksin TBC Buat Bill Gates | Indonesia Jadi Lokasi Uji Coba Vaksin TBC Buat Bill Gates |
| Mau ketawa bacanya sama Bill Gates aja tunduk | Mau ketawa bacanya sama Bill Gates aja tunduk |
| Tersiratnya. TNI vs Geng Solo yg dibackup Bill Gates | Tersiratnya. TNI vs Geng Solo yg dibackup Bill Gates |
| Ini sejalan dengan Bill Gates sebelumnya | Ini sejalan dengan Bill Gates sebelumnya |

b. Case Folding

Tujuan dari tahap ini adalah mengubah seluruh huruf besar pada data menjadi huruf kecil sehingga tidak ada perbedaan antara kata yang ditulis dengan huruf besar dan huruf kecil. Contoh dari proses sebelum dan sesudah case folding dapat dilihat pada tabel 4 sebagai berikut.

**Tabel 4.** Contoh Data Case Folding

| Data Tweet | Hasil Case Folding |
|---|---|
| Indonesia Jadi Lokasi Uji Coba Vaksin TBC Buat Bill Gates | indonesia jadi lokasi uji coba vaksin tbc buat bill gates |
| Mau ketawa bacanya sama Bill Gates aja tunduk... | mau ketawa bacanya sama bill gates aja tunduk... |
| Tersiratnya. TNI vs Geng Solo yg dibackup Bill Gates | tersiratnya. tni vs geng solo yg dibackup bill gates |
| Ini sejalan dengan Bill Gates sebelumnya... | ini sejalan dengan bill gates sebelumnya... |

c. Tokenizing

Tokenisasi melibatkan pemecahan teks menjadi unit yang lebih kecil seperti kata, frasa, atau kalimat. Hal ini memungkinkan untuk melakukan analisis lebih lanjut pada tingkat token. Tokenisasi penting untuk menyusun teks

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

sebelum menerapkan algoritma analisis, terutama dalam analisis sentimen media sosial. Berikut adalah contoh dari proses sebelum dan sesudah proses tokenizing dapat dilihat pada table 5.

**Tabel 5.** Contoh Data Tokenizing

| Data Tweet | Hasil Tokenizing |
|---|---|
| indonesia jadi lokasi uji coba vaksin tbc buat bill gates | ['indonesia', 'jadi', 'lokasi', 'uji', 'coba', 'vaksin', 'tbc', 'buat', 'bill', 'gates'] |
| mau ketawa bacanya sama bill gates aja tunduk... | ['mau', 'ketawa', 'bacanya', 'sama', 'bill', 'gates', 'aja', 'tunduk'] |
| tersiratnya. tni vs geng solo yg dibackup bill gates | ['tersiratnya.', 'tni', 'vs', 'geng', 'solo', 'yg', 'dibackup', 'bill', 'gates'] |
| ini sejalan dengan bill gates sebelumnya... | ['ini', 'sejalan', 'dengan', 'bill', 'gates', 'sebelumnya'] |

d. Normalization

Proses normalisasi dilakukan untuk mengubah kata yang dihilangkan menjadi kata yang mempunyai arti yang sama berdasarkan Kamus Besar Bahasa Indonesia (KBBI), sehingga data lebih mudah diolah. Sebagai contoh, "utk" diubah menjadi "untuk" dan "yg" diubah menjadi "yang". Berikut adalah contoh dari proses sebelum dan sesudah proses normalisasi dapat dilihat pada tabel 6.

**Tabel 6.** Contoh Data Normalization

| Data Tweet | Hasil Normalization |
|---|---|
| indonesia jadi lokasi uji coba vaksin tbc buat bill gates | indonesia jadi lokasi uji coba vaksin tbc buat bill gates |
| mau ketawa bacanya sama bill gates aja tunduk | mau ketawa bacanya sama bill gates saja tunduk |
| tersiratnya. tni vs geng solo yg dibackup bill gates | tersiratnya. tni vs geng solo yang dibackup bill gates |
| ini sejalan dengan bill gates sebelumnya | ini sejalan dengan bill gates sebelumnya |

e. Filtering

Proses penyaringan dilakukan untuk menghilangkan kata-kata yang tidak bermakna pada data. Fase ini menggunakan algoritma stoplist atau stopword yang berisi daftar kata-kata umum yang dianggap tidak relevan dengan analisis. Berikut adalah contoh dari proses sebelum dan sesudah proses dari filtering dapat dilihat pada tabel 7:

**Tabel 7.** Contoh Data Filtering

| Data Tweet | Hasil Filtering |
|---|---|
| ['indonesia', 'jadi', 'lokasi', 'uji', 'coba', 'vaksin', 'tbc', 'buat', 'bill', 'gates'] | ['indonesia', 'lokasi', 'uji', 'coba', 'vaksin', 'tbc', 'bill', 'gates'] |
| ['mau', 'ketawa', 'bacanya', 'sama', 'bill', 'gates', 'aja', 'tunduk'] | ['ketawa', 'bacanya', 'bill', 'gates', 'tunduk'] |
| ['tersiratnya.', 'tni', 'vs', 'geng', 'solo', 'yang', 'dibackup', 'bill', 'gates'] | ['tersiratnya.', 'tni', 'geng', 'solo', 'dibackup', 'bill', 'gates'] |
| ['ini', 'sejalan', 'dengan', 'bill', 'gates', 'sebelumnya'] | ['sejalan', 'bill', 'gates', 'sebelumnya'] |

f. Stemming

Pada tingkat akar, data dengan imbuhan termasuk prefiks, sufiks, dan sisipan dihilangkan. Tujuannya adalah mengembalikan kata tersebut ke bentuk dasarnya. Berikut adalah contoh dari proses sebelum dan sesudah proses stemming dapat dilihat pada tabel 8:

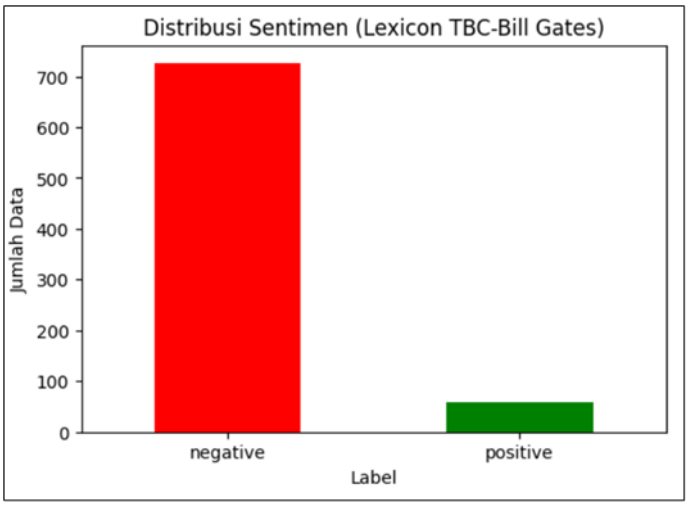
**Tabel 8**. Contoh Data Stemming

| Data Tweet | Hasil Stemming |
|---|---|
| ['indonesia', 'lokasi', 'uji', 'coba', 'vaksin', 'tbc', 'bill', 'gates'] | indonesia lokasi uji coba vaksin tbc bill gates |
| ['ketawa', 'bacanya', 'bill', 'gates', 'tunduk'] | ketawa baca bill gates tunduk |
| ['tersiratnya.', 'tni', 'vs', 'geng', 'solo', 'dibackup', 'bill', 'gates'] | tersirat tni vs geng solo backup bill gates |
| ['sejalan', 'bill', 'gates', 'sebelumnya'] | jalan bill gates belum |

### 3.1.3 Pelabelan Data

Pelabelan data mengacu pada proses penambahan huruf atau identifikasi ke data dalam bentuk angka, teks, atau gambar, untuk memungkinkan mesin memahami atau memproses data secara lebih efisien dan akurat. Pelabelan ini memungkinkan data untuk dianalisis dan dikelompokkan berdasarkan informasi tentang data. Dalam penelitian ini, misalnya, teks klasifikasi ditandai dengan label "positif" atau "negatif" berbasis sentiment dapat dilihat pada gambar 3 berikut.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom




**Gambar 3.** Label Data Positif dan Negatif

### 3.1.4 Pembobotan TF-IDF

Proses pembobotan TF-IDF terdiri dari dua bagian: TF (frekuensi term) dan IDF (frekuensi dokumen invers). TF menunjukkan seberapa sering sebuah kata muncul dalam dokumen. Dengan kata lain, semakin banyak kata yang terkandung dalam setiap dokumen, maka semakin besar nilai TFnya. IDF adalah jumlah dokumen yang mengandung setiap kata. Berikut adalah contoh perhitungan TF-IDF dari data latih dan data uji dapat dilihat pada tabel 9 berikut.

**Tabel 9.** Sampel Data Latih

| Data Latih |
|---|
| 'beda', 'uji', 'vaksin', 'bill, 'gates', 'klinis', takut' |
| 'indonesia', 'ujicoba, 'vaksin', 'tbc', 'bill, 'gates', 'menkesnya' |
| 'ujicoba, 'vaksin', 'tbc', 'bill, 'gates' |
| 'bill, 'gates', 'tbc', 'ujicoba', 'vaksin', 'indonesia' |
| 'yakin', 'vaksin', 'bill, 'gates', 'tbc' |

Pertama, tentukan nilai DF pada data pelatihan pada tahap penimbangan TF-IDF. Nilai DF dapat ditemukan di Tabel 10 berikut.

**Tabel 10.** Nilai DF Dari Data Latih

| Term | D1 | D2 | D3 | D4 | D5 | DF |
|---|---|---|---|---|---|---|
| Beda | 1 | 0 | 0 | 0 | 0 | 1 |
| Uji | 3 | 0 | 0 | 0 | 0 | 1 |
| Klinis | 3 | 0 | 0 | 0 | 0 | 1 |
| Vaksin | 4 | 2 | 1 | 1 | 1 | 5 |
| Tbc | 1 | 2 | 1 | 1 | 1 | 5 |
| Bill | 1 | 1 | 1 | 1 | 1 | 5 |
| Gates | 1 | 1 | 1 | 1 | 1 | 5 |
| Indonesia | 0 | 2 | 0 | 1 | 0 | 2 |
| Menkesnya | 0 | 2 | 0 | 0 | 0 | 1 |
| Ujicoba | 0 | 0 | 2 | 1 | 0 | 2 |
| Takut | 1 | 0 | 0 | 0 | 0 | 1 |
| Yakin | 0 | 0 | 0 | 0 | 1 | 1 |

Menentukan nilai IDF (Inverse Document Frequency) untuk setiap kata dilakukan setelah mengumpulkan nilai frekuensi istilah. Perhitungan IDF dari setiap kata dan hasilnya dapat dilihat pada tabel 11 sebagai berikut.

$$IDF = \ln\left(\frac{D+1}{df(t)+1}\right) + 1 \qquad (1)$$

Contoh perhitungan nilai IDF :

$$IDF = \ln\left(\frac{D+1}{df(t)+1}\right) + 1$$

$$IDF = \ln\left(\frac{5+1}{1+1}\right) + 1$$

$$= \ln\left(\frac{6}{2}\right) + 1$$

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$= ln(3) + 1$
$= 1.098 + 1$
$= 2.099$

**Tabel 11.** Nilai IDF Dari Data Latih

| Term | DF | IDF |
|---|---|---|
| Beda | 1 | 2.099 |
| Uji | 1 | 2.099 |
| Klinis | 1 | 2.099 |
| Vaksin | 5 | 1.000 |
| Tbc | 5 | 1.000 |
| Bill | 5 | 1.000 |
| Gates | 5 | 1.000 |
| Indonesia | 2 | 1.693 |
| Menkesnya | 1 | 2.099 |
| Ujicoba | 2 | 1.693 |
| Takut | 1 | 2.099 |
| Yakin | 1 | 2.099 |

Menentukan nilai TF-IDF dilakukan setelah menentukan nilai TF dan IDF. Untuk menentukan nilai TF-IDF dan hasilnya dapat dilihat pada tabel 12 sebagai berikut.

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \quad (2)$$

Contoh perhitungan TF-IDF sebagai berikut :

$TF - IDF = 1 \times 1.693 = 1.693$

**Tabel 12**. Nilai TF-IDF Dari Data Latih

| Term | TF | | | | |
|---|---|---|---|---|---|
| | D1 | D2 | D3 | D4 | D5 |
| Beda | 2.099 | 0 | 0 | 0 | 0 |
| Uji | 6.297 | 0 | 0 | 0 | 0 |
| Klinis | 6.297 | 0 | 0 | 0 | 0 |
| Vaksin | 4.000 | 2.000 | 1.000 | 1.000 | 1.000 |
| Tbc | 1.000 | 2.000 | 1.000 | 1.000 | 1.000 |
| Bill | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Gates | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Indonesia | 0 | 3.386 | 0 | 1.693 | 0 |
| Menkesnya | 0 | 4.198 | 0 | 0 | 0 |
| Ujicoba | 0 | 0 | 3.386 | 1.693 | 0 |
| Takut | 2.099 | 0 | 0 | 0 | 0 |
| Yakin | 0 | 0 | 0 | 0 | 2.099 |

Nilai TF-IDF harus dinormalisasi agar mendapatkan interval yang konsisten untuk setiap titik data.

$$\text{Wnorm}(t,d) = \frac{\text{TF-IDF}(t,d)}{\sqrt{\sum i(TF-IDF(t,d)^2}} \quad (3)$$

Contoh perhitungan data ternormalisasi pertama

$$\text{Wnorm}(\text{beda}) = \frac{\text{TF} - \text{IDF}(beda)}{\sqrt{\sum i(TF - IDF(t,d)^2}}$$
$= 6.297^2 + 6.297^2 + 4.000^2 + 2.099^2 + 2.099^2 + 1.000^2 + 1.000^2 + 1.000^2$
$= 39.652 + 39.652 + 16.000 + 4.406 + 4.406 + 1.000 + 1.000 + 1{,}000$
$= \sqrt{107.116}$
$= 10.350$
$= \frac{2.099}{10.350}$
$= 0.203$

Berikut adalah tabel dari hasil proses normalisasi data latih yang dapat dilihat pada tabel 13:

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

**Tabel 13.** Normalisasi Data Latih

| NO | D1 | D2 | D3 | D4 | D5 |
|---|---|---|---|---|---|
| 1 | 0.203 | 0 | 0 | 0 | 0 |
| 2 | 0.608 | 0 | 0 | 0 | 0 |
| 3 | 0.608 | 0 | 0.415 | 0 | 0 |
| 4 | 0.386 | 0.320 | 0.254 | 0.321 | 0.345 |
| 5 | 0.097 | 0.320 | 0.254 | 0.321 | 0.345 |
| 6 | 0.097 | 0.160 | 0.254 | 0.321 | 0.345 |
| 7 | 0.097 | 0.160 | 0.254 | 0.321 | 0.345 |
| 8 | 0 | 0.542 | 0 | 0.543 | 0 |
| 9 | 0 | 0.671 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0.861 | 0.543 | 0 |
| 11 | 0.203 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | 0.742 |

### 3.2 Klasifikasi SVM

Dengan mengoptimalkan margin (jarak) antara hiperplane dan titik data terdekat (yang disebut vektor pendukung), teknik klasifikasi Support Vector Machine (SVM) berusaha untuk mengidentifikasi hiperplane terbaik untuk memisahkan kelas data. Berikut adalah penjelasan tentang nilai Kernel Linier dan Matriks Hessian dalam kaitannya dengan klasifikasi SVM.

a. Linier kernel

SVM menggunakan teknik matematis yang disebut kernel (fungsi kernel) untuk menangani data yang tidak dapat dipisahkan secara linier di ruang fitur aslinya. Dikenal sebagai "Trik Kernel," kernel memungkinkan SVM untuk secara implisit mentransformasi data ke ruang dimensi yang lebih tinggi di mana pemisahan linier menjadi mungkin tanpa memerlukan perhitungan eksplisit koordinat data di ruang dimensi tersebut. Berikut contoh representase data untuk 5 data yang dapat dilihat pada tabel 14.

**Tabel 14.** Representase Data

| | X1 | X2 | X3 | X4 | X5 |
|---|---|---|---|---|---|
| Y1 | K(X1, Y1) | K(X2,Y1) | K(X3,Y1) | K(X4,Y1) | K(X5,Y1) |
| Y2 | K(X1,Y2) | K(X2,Y2) | K(X3,Y2) | K(X4,Y2) | K(X5,Y2) |
| Y3 | K(X1,Y3) | K(X2,Y3) | K(X3,Y3) | K(X4,Y3) | K(X5,Y3) |
| Y4 | K(X1,Y4) | K(X2,Y4) | K(X3,Y4) | K(X4,Y4) | K(X5,Y4) |
| Y5 | K(X1,Y5) | K(X2,Y5) | K(X3,Y5) | K(X4,Y5) | K(X5,Y5) |

Contoh perhitungan kolom 1 baris 1.

$$K(x,y) = x * y$$
$$K(x,y) = \big((x1y1 * x1y1) + (x1y2 * x1y2) + (x1y3 * x1y3) + (x1y4 * x1y4) + (x1y5 * x1y5)\big)$$
$$K(x,y) = (0.203 * 0.203 + 0 * 0 + 0 * 0 + 0 * 0 + 0 * 0)$$
$$= 0.041$$

Berikut adalah tabel 15 hasil perhitungan linier kernel.

**Tabel 15.** Normalisasi Wnorm^2

| 0Term | D1 | D2 | D3 | D4 | D5 |
|---|---|---|---|---|---|
| Beda | 0.041 | 0 | 0 | 0 | 0 |
| Uji | 0.369 | 0 | 0 | 0 | 0 |
| Klinis | 0.369 | 0 | 0 | 0 | 0 |
| Vaksin | 0.149 | 0.102 | 0.064 | 0.103 | 0.119 |
| Tbc | 0.009 | 0.102 | 0.064 | 0.103 | 0.119 |
| Bill | 0.009 | 0.025 | 0.064 | 0.103 | 0.119 |
| Gates | 0.009 | 0.025 | 0.064 | 0.103 | 0.119 |
| Indonesia | 0 | 0.293 | 0 | 0.294 | 0 |
| Menkesnya | 0 | 0.450 | 0 | 0 | 0 |
| Ujicoba | 0 | 0 | 0.743 | 0.294 | 0 |
| Takut | 0.041 | 0 | 0 | 0 | 0 |
| Yakin | 0 | 0 | 0 | 0 | 0.524 |

b. Matrik Hessian

Struktur kuadratik dari masalah optimisasi dual didefinisikan oleh matriks Hessian dalam SVM, dan sifat positif definit dan semi-definitnya sangat penting untuk memastikan bahwa pelatihan SVM dapat dilakukan secara efektif guna mengidentifikasi solusi optimal global. Parameter yang diterapkan adalah sebagai berikut yang terdapat pada tabel 16.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

**Tabel 16.** Nilai Parameter

| Parameter | Nilai |
|---|---|
| α (Ai) | 0 (untuk D1,D2,D5); 0.578 (untuk D3,D4) |
| C | 1 |
| Λ | 0.1 |
| T | 0.01 |
| Iterasi | 2 |

Berikut tabel 17 hasil perhitungan matriks hessian.

**Tabel 17.** Matriks Hessian

| Term | D1 | D2 | D3 | D4 | D5 |
|---|---|---|---|---|---|
| Beda | 0.051 | 0.010 | 0.010 | 0.010 | 0.010 |
| Uji | 0.379 | 0.010 | 0.010 | 0.010 | 0.010 |
| Klinis | 0.379 | 0.010 | 0.010 | 0.010 | 0.010 |
| Vaksin | 0.159 | 0.133 | 0.108 | -0.113 | -0.123 |
| Tbc | 0.029 | 0.041 | 0.034 | -0.021 | -0.023 |
| Bill | 0.029 | 0.035 | 0.034 | -0.021 | -0.023 |
| Gates | 0.029 | 0.035 | 0.034 | -0.021 | -0.023 |
| Indonesia | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| Menkesnya | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| Ujicoba | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |
| Takut | 0.051 | 0.010 | 0.010 | 0.010 | 0.010 |
| Yakin | 0.010 | 0.010 | 0.010 | 0.010 | 0.010 |

Iterasi 0 pada sequntial training dapat dilihat pada tabel 18 berikut.

**Tabel 18.** Sequential Training Iterasi 0

| Term | Iterasi 0 |
|---|---|
| Beda | 0 |
| Uji | 0 |
| Klinis | 0 |
| Vaksin | 0 |
| Tbc | 0 |
| Bill | 0 |
| Gates | 0 |
| Indonesia | 0 |
| Menkesnya | 0 |
| Ujicoba | 0 |
| Takut | 0 |
| Yakin | 0 |

Pada Iterasi 0, seluruh parameter Lagrange Multiplier (αi) diinisialisasi ke nilai nol (α(0)=[0,0,0,0,0]T). Model kemudian menghitung output prediksi (ui) dan error (Ei) menggunakan rumus ui=∑jαjyjK(xj,xi). Karena semua αi bernilai nol, output model juga nol, sehingga semua dokumen (D1 hingga D5) melanggar kondisi Karush-Kuhn-Tucker (KKT). Algoritma Sequential Minimal Optimization (SMO) kemudian memilih pasangan αi dan αj yang paling melanggar kondisi KKT untuk diperbarui. Berdasarkan analisis kontras pada Matriks Hessian Total (HReg), Dokumen 3 (D3, y3=+1) dan Dokumen 4 (D4, y4=−1) dipilih sebagai pasangan awal yang paling signifikan untuk dioptimalkan (i=3,j=4). Perhitungan pembaruan α memerlukan nilai ukuran langkah (η), yang didapatkan dari elemen Matriks Hessian Total (Hij). Nilai η dihitung menggunakan elemen diagonal H33 dan H44, serta elemen off-diagonal H34 dari Matriks Hessian Total yang sudah diregularisasi (λ2=0.01).

$\eta = H_{33} + H_{44} - 2H_{34}$
$\eta = 1.0019 + 0.8988 - 2(-0.5635)$
$\eta = 1.9007 + 1.1270 = 3.0277$

Perubahan pada α4 (Δα4) dihitung menggunakan *error* awal (E3 dan E4) dan η. Perubahan parameter α4 dihitung :

$$\Delta\alpha 4 = \frac{y4(E3 - E4)}{\eta}$$

$$\Delta\alpha 4 = \frac{(-1) \times ((-1) - (+1))}{3.0277}$$

$$= 0.6606$$

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Nilai Δα4 digunakan untuk memperbarui α4, dan kemudian α3 dihitung berdasarkan kendala keseimbangan. Parameter α4 yang baru (sebelum pemotongan) adalah
$\alpha 4^{new,unclipped} = \alpha 4^{old} + \Delta\alpha 4 = 0 + 0.6606 = 0.6606$.
Karena nilai 0.6606 berada dalam batas $0 \le \alpha \le 1$ (C = 1), maka:
$\alpha 4^{new}=0.6606$
Perhitungan utama di balik Δα4 = +0.6606 adalah:
1. Error Awal: E3=−1, E4=+1.
2. Ukuran Langkah (η): η = H33 + H44 − 2H34 = 1.0019 + 0.8988 – 2 (−0.5635) = 3.0277.
3. Δα4 : ηy4(E3−E4) = 3.0277(−1)×(−2) = 0.6606.

Pembaruan α3 harus menjaga kendala keseimbangan ∑αiyi=0. Karena semua α lainnya nol pada awal iterasi 1, maka:
α3newy3 + α4newy4 = 0
Nilai α3 dihitung untuk memenuhi kendala keseimbangan (α3y3 + α4y4 = 0):
$\alpha 3^{new}(+1) + 0.6606(-1) = 0$
$\alpha 3^{new} = 0.6606$
Setelah pembaruan yang signifikan pada Iterasi 1 (α3=α4=0.6606), algoritma *Sequential Training* melanjutkan ke Iterasi 2. Pada tahap ini, perhitungan Δα berfokus pada penyesuaian halus pada *Support Vectors* (D3 dan D4) dan memverifikasi kondisi KKT pada *Non-Support Vectors* (D1,D2,D5).

a. Data Awal (Akhir iterasi 1)
   1. αold (diperbarui): α3=0.6606, α4=0.6606.
   2. αold (tetap nol): α1,α2,α5=0.
   3. y3=+1, y4=−1.
   4. η=3.0277 (tidak berubah karena elemen Hessian tetap).
b. Hitung output baru

$$u_i \sum_{k=3}^{4} \alpha_k \, y_k \, H_{ki} \qquad (4)$$

   1. $u_3$: $\alpha_3 y_3 H_{33} + \alpha_4 y_4 H_{43}$
   2. $u_3 = (0.6606)(+1)(1.0019) + (0.6606)(-1)(-0.5635)$
   3. $u_3 = 0.6619 + 0.3722 = 1.0341$
   4. $u_4$: $\alpha_3 y_3 H_{34} + \alpha_4 y_4 H_{44}$
   5. $u_4 = (0.6606)(+1)(-0.5635) + (0.6606)(-1)(0.8988)$
   6. $u_4 = -0.3722 - 0.5937 = -0.9659$
c. Hitung error baru
   1. $E_3 = u_3 - y_3 = 1.0341 - (+1) = +0.0341$
   2. $E_4 = u_4 - y_4 = -0.9659 - (-1) = +0.0341$
d. Hitung perubahan $\Delta\alpha_4$

$\Delta\alpha_4 = y_4(E_3 - E_4)/\eta$
$\Delta\alpha_4 = (-1) \times (0.0341 - 0.0341)/\, 3.0277 = 0.0000$

Perhitungan menunjukkan bahwa E3 dan E4 hampir sama, yang berarti gradien di titik tersebut sangat mendekati nol. Ini mengindikasikan bahwa model berada sangat dekat dengan solusi optimal (0.578). Perubahan Δα yang dihitung di sini mendekati nol. Perhitungan Δα pada Iterasi 2 menunjukkan bahwa gradien fungsi objektif di sekitar $\alpha_3 = 0.6606$ dan $\alpha_4 = 0.6606$ mendekati nol, yang secara matematis mengindikasikan bahwa model telah mencapai konvergensi dan solusi optimal telah ditemukan. Nilai α akhir ditetapkan berdasarkan solusi optimal dari masalah optimasi Kuadratik (QP) dengan Regularisasi $\lambda = 0.01$. Hasilnya menunjukkan bahwa hanya dua dokumen yang memiliki nilai α lebih besar dari nol (αi>0), yang menjadikannya *Support Vector* yang dapat dilihat pada tabel 19 berikut :

**Tabel 19.** Nilai Alpha Akhir

| Dokumen | Nilai Alpha Akhir | Status |
|---|---|---|
| D1 | 0 | Non-Support Vector |
| D2 | 0 | Non-Support Vector |
| D3 | 0.578 | Support Vector |
| D4 | 0.578 | Support Vector |
| D5 | 0 | Non-Support Vector |

Langkah berikutnya adalah menghitung vector bobot (W) dan bias (b). Setelah $\alpha_{akhir}$ ditetapkan, vektor bobot (W) dan *bias* (b) dihitung untuk mendefinisikan *hyperplane* pemisah (W·x+b=0).
Vector bobot (W) dihitung sebagai kombinasi linier dari support vector :

$$W = \sum_{i \in SV} \alpha_i y_i x_i$$
$$W = \alpha_3 y_3 x_3$$
$$W = (0.578)(+1)x_3 + (0.578)(-1)x_4$$
$$W = 0.578(x_3 - x_4)$$

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Nilai W adalah vektor multidimensi yang menentukan orientasi *hyperplane*. Setelah menyelesaikan nilai bobot (W), selanjutnya adalah menghitung bias (b). *Bias* (b) dihitung menggunakan salah satu *Support Vector* (misalnya D3) untuk memenuhi kondisi *margin*:

$(y_i(W \cdot x_i + b) = 1)$

$b = y_3 - W \cdot x_3$

Dengan mensubstitusi nilai $W \cdot x_3 = 0.8994$ yang dihitung dari elemen Matriks Hessian:

$b = (+1) - 0.8994 = 0.1006$

Model klasifikasi akhir didefinisikan oleh Bobot $W=0.578(x_3-x_4)$ dan Bias $b=0.1006$. *Hyperplane* yang dihasilkan ini dapat digunakan untuk mengklasifikasikan dokumen baru sebagai kelas +1 atau −1. Berdasarkan hasil analisis dan pemrosesan yang telah dilakukan, dapat dilihat bahwa kelas atau kategori yang relevan untuk data pelatihan dpat dilihat pada tabel 20 sebagai berikut :

**Tabel 20.** Kelas Data Latih

| Data Latih | Kelas |
|---|---|
| 'beda', 'uji', 'vaksin', 'bill, 'gates', 'klinis', takut' | Positif |
| 'indonesia', 'ujicoba, 'vaksin', 'tbc', 'bill, 'gates', 'menkesnya' | Positif |
| 'ujicoba, 'vaksin', 'tbc', 'bill, 'gates' | Positif |
| 'bill, 'gates', 'tbc', 'ujicoba', 'vaksin', 'indonesia' | Negative |
| 'yakin', 'vaksin', 'bill, 'gates', 'tbc' | Negative |

Berdasarkan parameter model W dan b yang telah ditetapkan, berikut disajikan sampel data uji pada tabel 21.

**Tabel 21.** Sampel Data Uji

| Data Uji | Kelas |
|---|---|
| 'langkah', 'keren', 'bill', 'gates', 'tim', 'ayo, 'indonesia', 'dukung', 'penuh', 'sehat', 'tbc', 'malaria' | ? |
| 'sabar', 'kamu', 'umum', vaksin', 'tbc', 'sudah', 'buzzer', 'saja' 'gercep', 'banget', 'duit', 'bill', 'gates' | ? |

Fungsi klasifikasi hiperplane SVM, yang telah ditentukan oleh Vektor Bobot dan Bias, digunakan untuk menentukan kelas data uji baru dengan memasukkan nilai fitur (nilai TF-IDF yang dinormalisasi) dari data tersebut. Berikut perhitungan tf-idf pada uji pertama:

$$\text{IDF(langkah)} = \log_{10}\left(\frac{784}{1}\right)$$

$$= \log_{10}(784)$$

$$= 2{,}894$$

Berikut perhitungan untuk normalisasi data pada uji pertama.

$$\text{Wnorm(langkah)} = \frac{\text{TF} - \text{IDF}(langkah)}{\sqrt{\sum \text{i}(TF - IDF(t,d)^2}}$$

$$= \frac{2.894}{10.139}$$

$$= 0.285$$

Berikut adalah tabel 24 yang berisi Data Uji pertama TF-IDF dan Normalisasi.

**Tabel 22.** TF-IDF dan Normalisasi Data Uji Pertama

| Term | TF | DF | IDF | TF-IDF | NORM |
|---|---|---|---|---|---|
| langkah | 1 | 1 | 2,894 | 2,894 | 0.285 |
| keren | 1 | 1 | 2.894 | 2.894 | 0.285 |
| Bill | 1 | 3 | 2.418 | 2.418 | 0.238 |
| gates | 1 | 3 | 2.418 | 2.418 | 0.238 |
| Tim | 1 | 1 | 2.894 | 2.894 | 0.285 |
| Ayo | 1 | 1 | 2.894 | 2.894 | 0.285 |
| indonesia | 1 | 5 | 2.293 | 2.293 | 0.226 |
| dukung | 1 | 1 | 2.894 | 2.894 | 0.285 |
| penuh | 1 | 1 | 2.894 | 2.894 | 0.285 |
| sehat | 1 | 1 | 2.894 | 2.894 | 0.285 |
| Tbc | 1 | 5 | 2.293 | 2.293 | 0.226 |
| malaria | 1 | 1 | 2.894 | 2.894 | 0.285 |

Langkah selanjutnya dilakukan perhitungan nilai linier kernel untuk data uji pertama.

$$x_{uji} = W.\left(K(x_3, x_{uji}) - K(x_4, x_{uji})\right) + b$$

$$= 0.578 \text{ x } (0.698) + 0.100$$

$$= 0.403 + 0.100$$

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$= 0.504$

Hasil klasifikasi data uji pertama > 0, diklasifikasikan sebagai Positif (+1). Setelah mengetahui data uji positif langkah selanjutnya adalah menentukan data uji kedua apakah hasil klasifikasinya sebagai < 0 .
Berikut adalah perhitungan untuk TF-IDF data uji kedua.

$$\text{IDF(duit)} = \log_{10}\left(\frac{784}{1}\right)$$
$$= \log_{10}(784)$$
$$= 2.894$$

Berikut adalah perhitungan untuk normalisasi data pada uji kedua dapat dilihat pada tabel 23:

$$\text{Wnorm(duit)} = \frac{\text{TF} - \text{IDF}(duit)}{\sqrt{\sum \text{i}(TF - IDF(t,d)^2}}$$
$$= 2.894/9.876$$
$$= 0.293$$

**Tabel 23.** TF-IDF dan Normalisasi Data Uji Kedua

| Term | TF | DF | IDF | TF-IDF | NORM |
|---|---|---|---|---|---|
| sabar | 1 | 1 | 2.894 | 2.894 | 0.293 |
| kamu | 1 | 1 | 2.894 | 2.894 | 0.293 |
| umum | 1 | 1 | 2.894 | 2.894 | 0.293 |
| vaksin | 1 | 5 | 2.293 | 2.293 | 0.232 |
| Tbc | 1 | 5 | 2.293 | 2.293 | 0.232 |
| sudah | 1 | 1 | 2.894 | 2.894 | 0.293 |
| buzzer | 1 | 1 | 2.894 | 2.894 | 0.293 |
| Saja | 1 | 1 | 2.894 | 2.894 | 0.293 |
| gercep | 1 | 1 | 2.894 | 2.894 | 0.293 |
| banget | 1 | 1 | 2.894 | 2.894 | 0.293 |
| Duit | 1 | 1 | 2.894 | 2.894 | 0.293 |
| Bill | 1 | 3 | 2.418 | 2.418 | 0.244 |
| Gates | 1 | 3 | 2.418 | 2.418 | 0.244 |

Selanjutnya adalah menghitung linier kernel untuk data uji kedua.

$$\text{x}_{uji2} = \text{W}.\left(K\left(\text{x}_3, \text{x}_{uji}\right) - \text{K}\left(\text{x}_4, \text{x}_{uji}\right)\right) + b$$
$$= 0.578\ x\ (-0.778) + 0.100$$
$$= -0.349$$

Hasil klasifikasi data uji kedua < 0, maka diklasifikasikan sebagai Negatif (-1). Berikut adalah tabel 24 untuk kelas data uji.

**Tabel 24.** Kelas Data Uji

| Data Uji | Kelas |
|---|---|
| 'langkah', 'keren', 'bill', 'gates', 'tim', 'ayo, 'indonesia', 'dukung', 'penuh', 'sehat', 'tbc', 'malaria' | Positif |
| 'sabar', 'kamu', 'umum', vaksin', 'tbc', 'sudah', 'buzzer', 'saja' 'gercep', 'banget', 'duit', 'bill', 'gates' | Negatif |

### 3.3 Pembahasan

Support Vector Machine (SVM) digunakan dalam studi ini sebagai metode klasifikasi untuk menganalisis opini publik dan sentimen seputar vaksin tuberkulosis Bill Gates. Google Colaboratory (Colab) dan bahasa pemrograman Python digunakan untuk pemrosesan dan analisis 784 data dalam dataset. Untuk mengevaluasi efektivitas model klasifikasi, data dibagi menjadi 80% data pelatihan (627 dokumen) dan 20% data uji (157 dokumen). Pandas, NumPy, Sastrawi, dan Lexicon (penandaan/pelabelan analisis sentimen) adalah perpustakaan utama yang digunakan.Diharapkan metode ini dapat mengklasifikasikan opini publik secara lebih akurat ke dalam kategori emosi positif, atau negatif sehingga memberikan gambaran yang jelas tentang bagaimana masyarakat memandang masalah ini. Karena dapat menangani masukan teks berdimensi tinggi dan menghasilkan pemisahan kelas optimal menggunakan hiperplane terbaik, metode SVM digunakan dalam penelitian ini. Langkah pertama dalam pendekatan SVM adalah pengumpulan dan pembersihan data (prapemrosesan), yang meliputi tokenisasi, penghapusan kata, penyesuaian huruf besar/kecil, dan stemming untuk memastikan kualitas teks yang akan dianalisis. Agar algoritma SVM dapat memproses data yang telah diolah, data tersebut terlebih dahulu diubah menjadi representasi numerik menggunakan teknik seperti TF-IDF.

a. *Library*

Pada penelitian ini menggunakan *library Lexicon* untuk pelabelan data, dan library NumPy (np), Pandas (pd) untuk pengelolaan data dan operasi numerik dasar. *Regular Expression* (re) digunakan untuk membersihkan teks dari pola yang tidak diinginkan seperti URL dan mention. Sastrawi menangani operasi linguistik spesifik bahasa Indonesia selama tahap Prapemrosesan Teks, seperti stemming untuk mencari kata dasar, sementara NLTK digunakan untuk mengakses daftar

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

*stopword* yang perlu dihilangkan. Scikit-learn, yang menyediakan TfidfVectorizer untuk penimbangan fitur dan SVC sebagai model klasifikasi Mesin Vektor Dukungan (SVM), merupakan dasar dari pembelajaran mesin. Evaluasi model (classification_report, confusion_matrix) dan pembagian data (train_test_split).

```
import pandas as pd
import re
import string
import matplotlib.pyplot as plt # Dipertahankan untuk Visualisasi Distribusi Label
import seaborn as sns # Dipertahankan untuk Confusion Matrix (jika diaktifkan)
import os
import numpy as np

!pip install Sastrawi

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import SVC
from sklearn.metrics import confusion_matrix, classification_report # <-- Dipakai
import nltk
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

**Gambar 4.** Import Library

b. Preprocessing Data

*Pre-processing* adalah langkah pertama dalam pemrosesan data teks yang mengubah teks mentah menjadi format yang lebih terstruktur dan relevan. Berikut rincian tahapan preprocessing teks :

1. *Case Folding*

Pada tahap ini, semua huruf dalam teks diubah menjadi huruf kecil (*lower case*). Berikut adalah tampilan kode dan hasil *case folding*.

```
data['case_folding'] = data['full_text'].astype(str).str.lower()
data['tokenized'] = data['case_folding'].apply(lambda x: x.split())
data['stopword_removed'] = data['tokenized'].apply(
    lambda tokens: [w for w in tokens if w not in stop_words]
)
data['stemmed'] = data['stopword_removed'].apply(
    lambda tokens: [stemmer.stem(w) for w in tokens]
)
```

**Gambar 5.** Tampilan Kode Case folding

Pada gambar 5 untuk bagian teks dari kolom 'full_text' diubah menjadi *string*. Metode '.str.lower()' diterapkan untuk mengubah semua karakter menjadi huruf kecil, dan hasilnya disimpan dalam kolom baru 'case_folding' dapat dilihat pada gambar 6.

| | case_folding |
|---|---|
| 0 | apa kabar ujicoba vaksin tbc?? sibuk soal pa... |
| 1 | indonesia jadi lokasi uji coba vaksin tbc buat... |
| 2 | kak  mau nanya dong vaksin tbc dari pemerinta... |
| 3 | mau ketawa bacanya sama bill gates aja tunduk... |
| 4 | tersiratnya. tni vs geng solo yg dibackup bil... |

**Gambar 6.** Hasil Case Folding

2. Tokenization

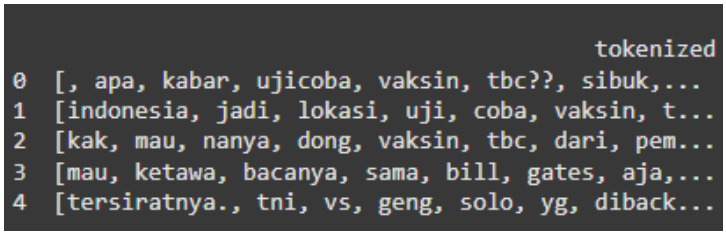
Pada tahap ini teks dipecah menjadi daftar kata (token). Berikut adalah tampilan kode dan hasil tokenization.

```
data['case_folding'] = data['full_text'].astype(str).str.lower()
data['tokenized'] = data['case_folding'].apply(lambda x: x.split())
data['stopword_removed'] = data['tokenized'].apply(
    lambda tokens: [w for w in tokens if w not in stop_words]
)
data['stemmed'] = data['stopword_removed'].apply(
    lambda tokens: [stemmer.stem(w) for w in tokens]
)
```

**Gambar 7.** Kode Tokenization

Pada gambar 7 untuk data di kolom 'case_folding' dipecah menjadi daftar kata (token) menggunakan fungsi '.split()', dan hasilnya disimpan dalam kolom 'tokenized' dapat dilihat pada gambar 8.

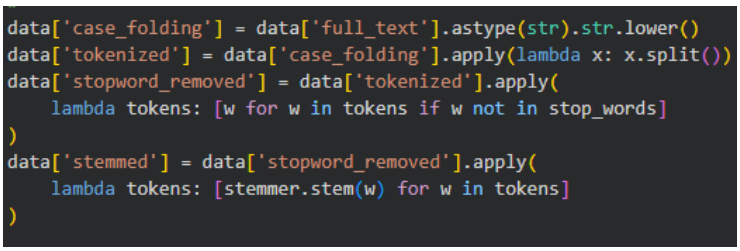
**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



| | tokenized |
|---|---|
| 0 | [, apa, kabar, ujicoba, vaksin, tbc??, sibuk,... |
| 1 | [indonesia, jadi, lokasi, uji, coba, vaksin, t... |
| 2 | [kak, mau, nanya, dong, vaksin, tbc, dari, pem... |
| 3 | [mau, ketawa, bacanya, sama, bill, gates, aja,... |
| 4 | [tersiratnya., tni, vs, geng, solo, yg, diback... |

**Gambar 8.** Hasil Tokenization

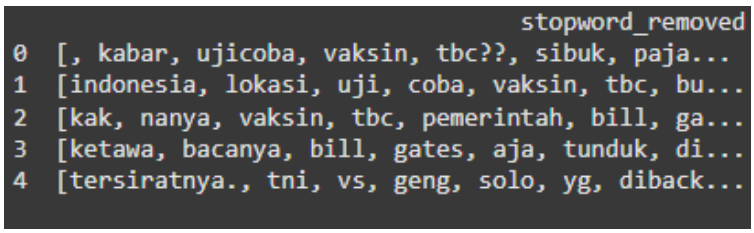
3. *Stopword Removal*
   Pada bagian ini menghapus kata-kata umum seperti dan, di, ke. Berikut adalah tampilan kode dan hasil *stopword removal*.



```
data['case_folding'] = data['full_text'].astype(str).str.lower()
data['tokenized'] = data['case_folding'].apply(lambda x: x.split())
data['stopword_removed'] = data['tokenized'].apply(
    lambda tokens: [w for w in tokens if w not in stop_words]
)
data['stemmed'] = data['stopword_removed'].apply(
    lambda tokens: [stemmer.stem(w) for w in tokens]
)
```

**Gambar 9.** Kode Stopword Removal

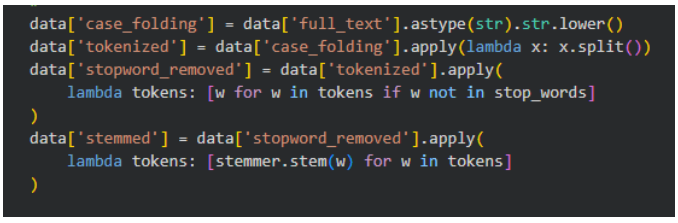
Pada gambar 9 iterasi dilakukan pada setiap token dalam kolom 'tokenized'. Kata-kata yang tidak ada dalam '*set* stop_words' (yang berisi *stopword* Bahasa Indonesia dari NLTK) dipertahankan, dan hasilnya disimpan di kolom 'stopword_removed' dapat dilihat pada gambar 10.



| | stopword_removed |
|---|---|
| 0 | [, kabar, ujicoba, vaksin, tbc??, sibuk, paja... |
| 1 | [indonesia, lokasi, uji, coba, vaksin, tbc, bu... |
| 2 | [kak, nanya, vaksin, tbc, pemerintah, bill, ga... |
| 3 | [ketawa, bacanya, bill, gates, aja, tunduk, di... |
| 4 | [tersiratnya., tni, vs, geng, solo, yg, diback... |

**Gambar 10.** Hasil Stopword Removal

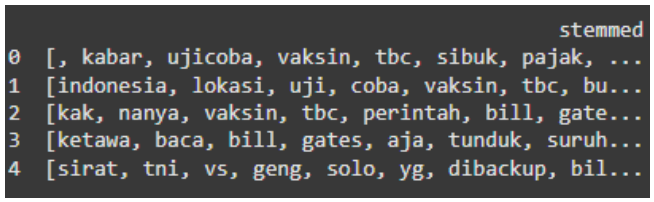
4. *Stemming*
   Pada bagian ini Kata diubah ke bentuk dasar menggunakan *Sastrawi Stemmer*. Berikut adalah tampilan kode dan hasil *stemming*.



```
data['case_folding'] = data['full_text'].astype(str).str.lower()
data['tokenized'] = data['case_folding'].apply(lambda x: x.split())
data['stopword_removed'] = data['tokenized'].apply(
    lambda tokens: [w for w in tokens if w not in stop_words]
)
data['stemmed'] = data['stopword_removed'].apply(
    lambda tokens: [stemmer.stem(w) for w in tokens]
)
```

**Gambar 11.** Kode Steming

Pada gambar 11 fungsi 'stemmer.stem(w)' dari pustaka Sastrawi diterapkan pada setiap kata dalam kolom 'stopword_removed'. Hasilnya (kata-kata dasar) disimpan dalam kolom 'stemmed' dapat dilihat pada gambar 12.



| | stemmed |
|---|---|
| 0 | [, kabar, ujicoba, vaksin, tbc, sibuk, pajak, ... |
| 1 | [indonesia, lokasi, uji, coba, vaksin, tbc, bu... |
| 2 | [kak, nanya, vaksin, tbc, perintah, bill, gate... |
| 3 | [ketawa, baca, bill, gates, aja, tunduk, suruh... |
| 4 | [sirat, tni, vs, geng, solo, yg, dibackup, bil... |

**Gambar 12.** Hasil Steming

5. Normalisasi Kata
   Pada tahap ini kata-kata tidak baku diubah ke bentuk baku. Berikut tampilan kode dan hasil normalisasi kata.

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

```
def normalize_words(tokens):
    return [normalization_dict.get(w, w) for w in tokens]

data['normalized'] = data['stemmed'].apply(normalize_words)
data['clean_text'] = data['normalized'].apply(lambda tokens: " ".join(tokens))

data[['full_text', 'case_folding', 'tokenized',
      'stopword_removed', 'stemmed', 'normalized',
      'clean_text']].to_csv("preprocessing_result.csv", index=False)
```

**Gambar 13.** Normalisasi Kata

Pada gambar 13 fungsi ‘normalization_dict’ mendefinisikan pemetaan kata-kata non-baku ke kata-kata baku. Fungsi ‘normalize_words’ mencari setiap token dari kolom 'stemmed' di ‘normalization_dict;’ jika ditemukan, token diganti dengan nilai baku; jika tidak ditemukan, token dipertahankan. Hasilnya disimpan dalam kolom 'normalized' dapat dilihat pada gambar 14.

```
                                          normalized
0  [, kabar, ujicoba, vaksin, tbc, sibuk, pajak, ...
1  [indonesia, lokasi, uji, coba, vaksin, tbc, bu...
2  [kak, nanya, vaksin, tbc, perintah, bill, gate...
3  [ketawa, baca, bill, gates, saja, tunduk, suru...
4  [sirat, tni, vs, geng, solo, yang, dibackup, b...
```

**Gambar 14.** Hasil Normalisasi Kata

*c. Splitting Data*

Pada tahap ini data dibagi menjadi dua yakni data latih dan data uji. Berikut adalah tampilan kode dan hasil *splitting data*.

```
# ======================================
# 8. Split Data
# ======================================
X = data['clean_text']
y = data['label']

test_size_ratio = 0.2
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=test_size_ratio, random_state=42, stratify=y if len(y.unique()) > 1 else None
)

print("\nTotal data:", len(data))
print("Jumlah data latih:", len(X_train))
print("Jumlah data uji:", len(X_test))
```

**Gambar 15.** Splitting Data

Pada gambar 15 fitur (X) diambil dari kolom 'clean_text' (teks yang sudah diproses). Label (y) diambil dari kolom 'label' (hasil *labeling* otomatis). Fungsi train_test_split dari ‘sklearn.model_selection’ digunakan untuk membagi data, test_size=0.2 (20%) menentukan proporsi data uji, ‘random_state=42’ memastikan pembagian data konsisten (dapat direproduksi), dan ‘stratify=y’ memastikan bahwa proporsi setiap kelas label ('positive', 'negative') dipertahankan baik di data latih maupun data uji dapat dilihat pada gambar 16 dan 17.

```
Total data: 784
Jumlah data latih: 627
Jumlah data uji: 157
```

**Gambar 16.** Hasil Splitting Data

| Clean Text (Data Uji) | Aktual | Prediksi |
|---|---|---|
| semenjak bill gates indonesia jual vaksin tbc berita tbc tuju takut masyarakat vaksin | negative | negative |
| ribut vaksin tbc indonesia yang usaha bill gates gak dimaksut tbc computer | negative | negative |
| bpom lampu hijau vaksin tbc bill gates uji klinik ri 2 ribu orang rekrut | negative | negative |
| ahlan wasahlan bill gates bawa rahmah vaksin tbc | negative | negative |
| pro kontra vaksin tbc bill gates uji klinis ahli epidemiologi ugm | negative | negative |
| silah coba buzzer jokowi coba | negative | negative |
| taruna ikrar lampu hijau rencana laksana uji klinis vaksin tbc tuberkulosis tahap tiga kembang konglomerat bill gates indonesia | negative | negative |
| indonesia september 2024 presiden prabowo umum kamis pekan lalu uji klinis sorot serta dana hibah gates indonesia rp 2 5 triliun | negative | negative |
| syarat buzzer 400p 700p wajib sertifikat vaksin tbc bill gates | negative | negative |
| ›ill gates kesini cm dagang vaksin kesini niat ngurusin mbg eh skrng nyambi dagang vaksin tbc sudah obat bodoh aje yang jd rawan | negative | negative |
| kalau dengan menu harga segitu pasti udah vaksin tbc mbg mister bill gates ya | negative | negative |
| ; anak obat rumah sakit elit mah tidak kasih vaksin bill gates kelinci coba orang2 miskin sesuai desain elit global hidup jokowi | negative | negative |
| telat suntik vaksin tbc bill gates | negative | negative |
| vaksin tbc advisory jalan as level 2 sebut risiko terorisme bencana alam sebut vaksin uji vaksin tbc m72 as01e indonesia dukung | negative | negative |
| bill gates dtg 62 tbtb muncul statement jkt status siaga tbc moment nya pas bingit | negative | negative |
| great news indonesia uji coba vaksin tbc amp malaria bareng bill gates moga hasil bangga | negative | negative |
| ı laksana bpom udah ijin rakyat indo dijadiin kelinci coba karena yang nolak muncul berita 274 rw jakarta status siaga tbc tidak | negative | negative |
| ıah bikin virus covid nyamuk wolbachia tbc dan lain lain us bg gugat dokter amp scientis karena simpang obat an amp tangan medis | negative | negative |
| keras menkes era sby vaksin tbc bill gates uji coba via | negative | negative |
| iya cuman point nya datang bill gates ksni negara uji coba vaksin tbc skrg muncul kalo negara tbc sudah sblm sblmnya adem saja | negative | negative |

**Gambar 17.** Data Uji

**JURIKOM (Jurnal Riset Komputer)**, Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9431
Hal 29-47
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

d. Pembobotan Kata (TF-IDF)

Pada tahap ini seluruh bobot dari seluruh kata (term) dihitung berdasarkan intensitas kemunculannya pada *dataset*. Berikut adalah gambar 18 tampilan kode dan hasil dari proses TF-IDF.

```
# ========================================
# 9. TF-IDF Vectorization
# ========================================
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

**Gambar 18.** Hasil TF-IDF

Pada gambar 17 fungsi 'TfidfVectorizer()' diinisialisasi. Fungsi dari 'vectorizer.fit_transform(X_train)' menghitung bobot TF-IDF untuk data latih (X_train) dan sekaligus membangun *vocabulary* dari data tersebut 'vectorizer.transform(X_test)' hanya menerapkan bobot yang sudah dipelajari dari data latih ke data uji (X_test).

### 3.4 Evaluasi Hasil

Analisis model SVM dilakukan untuk mengidentifikasi hiperplane yang dapat mengoptimalkan analisis sentimen berdasarkan fitur-fitur yang ditentukan. Model yang disebutkan di atas kemudian dievaluasi menggunakan data uji untuk menentukan kinerjanya menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Proses ini dilakukan secara iteratif untuk mendapatkan model parameter optimal sehingga hasil klasifikasi lebih akurat dan dapat dianalisis untuk memvalidasi keakuratan model dalam mengklasifikasikan sentimen negatif dan positif. *Confusion Matrix* mencatat hasil klasifikasi 157 dokumen uji dengan membandingkan prediksi model terhadap label aktual.

Berikut adalah tabel 25 confusion matrix :

**Tabel 25.** Confusion Matrix

| Aktual \ Prediksi | Negatif | Positif |
|---|---|---|
| Negatif | 145 | 0 |
| Positif | 5 | 7 |

Berdasarkan tabel *consufion matrix* dapat dihitung nilai akurasi, presisi, F1-score.

$$\text{Accuracy} = \frac{145 + 7}{157}$$
$$= \frac{152}{157}$$
$$= 0.968$$
$$\text{Precision} = \frac{0.967 + 1.000}{2}$$
$$= 0.9833$$
$$\text{Recall} = \frac{1.000 + 0.583}{2}$$
$$= 0.791$$
$$F1\ Score = \frac{0.983 + 0.736}{2}$$
$$= 0.860$$

Berdasarkan hasil yang diperoleh ditemukan bahwa tingkat akurasi mencapai 96.82%, presisi mencapai 98,34%, recal 79,17%, F1-Score mencapai 86%. Berikut ini adalah gambar 19 hasil confusion matrix.

```
                precision    recall  f1-score   support

      negative       0.97      1.00      0.98       145
      positive       1.00      0.58      0.74        12

      accuracy                           0.97       157
     macro avg       0.97      0.97      0.97         0
  weighted avg       0.98      0.79      0.86       157
```

**Gambar 19**. Hasil Confusion Matrik

## 4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap keterlibatan Bill Gates dalam program vaksin TBC di media sosial X menggunakan metode Support Vector Machine (SVM), dan hasil penelitian menunjukkan bahwa tujuan tersebut berhasil dicapai. Model SVM dengan kernel linear yang dikombinasikan dengan

pembobotan fitur TF-IDF mampu mengklasifikasikan sentimen publik secara efektif dengan tingkat akurasi sebesar 96,82% dan nilai F1-score macro sebesar 86,00%, yang menandakan kinerja model yang sangat baik terutama dalam mengidentifikasi sentimen negatif. Temuan substantif penelitian ini menunjukkan bahwa opini publik di media sosial X terhadap isu vaksin TBC Bill Gates cenderung didominasi oleh sentimen negatif, mencerminkan adanya skeptisisme dan kekhawatiran masyarakat terhadap isu kesehatan global yang melibatkan tokoh internasional. Keberhasilan model ini mengindikasikan bahwa SVM berbasis TF-IDF merupakan pendekatan yang andal dan efisien untuk analisis sentimen pada data teks berbahasa Indonesia yang bersumber dari media sosial. Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh pembuat kebijakan dan pemangku kepentingan di bidang kesehatan sebagai bahan pertimbangan dalam merumuskan strategi komunikasi publik yang lebih tepat sasaran terkait isu vaksinasi. Untuk penelitian selanjutnya, disarankan agar cakupan dataset diperluas dengan melibatkan platform media sosial lain seperti Facebook, Instagram, dan TikTok serta periode waktu yang lebih panjang, menggunakan teknik representasi fitur berbasis word embedding atau model berbasis transformer, serta melibatkan anotator manusia dalam proses pelabelan data guna meningkatkan akurasi dan objektivitas hasil analisis sentimen..

# REFERENCES

[1] S. P. Yuanita *et al.*, “Identifikasi Teknik Data Mining Metode Asosiasi : Systematic Literature Review,” vol. 4, no. 2, pp. 57–61, 2024.
[2] A. Handayani and I. Zufria, “Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM,” *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.
[3] M. Furqan, M. Ikhsan, and R. Aini, “Algoritma Support Vector Machine Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Pandemi Virus Corona Di Media Sosial,” *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 4, pp. 908–915, 2023, [Online]. Available: https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/241%0Afiles/5265/Furqan
[4] D. Sartika, A. Harokan, and L. Suryani, “Analysis of risk factors for the incidence of pulmonary tuberculosis in Tebing Tinggi Community Health Center: A cross-sectional study,” vol. 6, no. 3, pp. 427–435, 2025.
[5] M. Riziq sirfatullah Alfarizi, M. Zidan Al-farish, M. Taufiqurrahman, G. Ardiansah, and M. Elgar, “PENGGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING,” 2023.
[6] H. P. Doloksaribu and Y. T. Samuel, “KOMPARASI ALGORITMA DATA MINING UNTUK ANALISIS SENTIMEN APLIKASI PEDULILINDUNGI,” vol. 16, no. 1, pp. 1–11, 2022.
[7] F. A. Nur, “Analisa Sentimen Twitter Vaksin Covid-19 di Indonesia dengan Metode Support Vector Machine,” vol. 5, no. 3, pp. 1244–1252, 2024.
[8] D. Saputra *et al.*, “Jurnal Wicara Desa , Volume 3 Nomor 3 , Juni 2025 TANJUNG SOCIALIZATION OF TUBERCULOSIS DISEASE PREVENTION TO INCREASE COMMUNITY AWARENESS IN REDUCING TUBERCULOSIS CASES IN TANJUNG VILLAGE 1 Program Studi Matematika , Universitas Mataram , 2 Program Studi Teknik Elektro , Universitas Mataram , 3 Program Studi Arsitektur , Universitas Mataram , 4 Program Pendidikan Bahasa dan Sastra Indonesia , 5 * Program Studi Kehutanan , Universitas Mataram Jurnal Wicara Desa , Volume 3 Nomor 3 , Juni 2025,” vol. 3, pp. 580–594, 2025.
[9] D. A. Punkastyo, F. Septian, and A. Syaripudin, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa,” vol. 5, no. 1, pp. 24–35, 2024.
[10] A. H. Lubis, L. Putri, and A. Lubis, “Sentiment analysis on twitter about the death penalty using the support vector machine method,” vol. 11, no. 2, pp. 312–321, 2024, doi: 10.37373/tekno.v11i2.1096.
[11] R. G. Muhamad Luthfi Bangun Permadi, “PENERAPAN ALGORITMA CNN (CONVOLUTIONAL NEURAL NETWORK) UNTUK DETEKSI DAN KLASIFIKASI TARGET MILITER BERDASARKAN CITRA SATELIT,” vol. 4, no. 2, pp. 134–143, 2024.
[12] R. Merdiansah and A. A. Ridha, “Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT,” vol. 7, pp. 221–228, 2024.
[13] N. Wahyudi, R. Albar, and Y. Rizki, “ANALISIS EFEKTIVITAS METODE SUPPORT VECTOR MACHINE ( SVM ) DALAM MENGURANGI BERITA HOAX BENCANA ALAM DARI DATA TWITTER ANALYSIS OF THE EFFECTIVENESS OF THE SUPPORT VECTOR MACHINE ( SVM ) METHOD IN REDUCING HOAX NEWS ON NATURAL,” vol. 10, no. 1, 2024.
[14] A. Aufa, A. Putri, S. A. Rahmah, T. Informasi, F. Teknik, and U. Dharmawangsa, “IMPLEMENTASI DATA MINING DENGAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS BISNIS PADA PERUSAHAAN ASURANSI,” vol. 5, no. 1, pp. 139–152, 2024, doi: 10.46576/djtechno.
[15] F. Herlando, A. R. Dzikrillah, F. Nufairi, E. Sinduningrum, and M. Sholeh, “ANALISIS PERBANDINGAN SENTIMENT DAN PERBINCANGAN NETIZEN TERHADAP TWITTER PASCA PERGANTIAN NAMA,” vol. 9, no. 1, pp. 360–367, 2024.
[16] R. Ramlan, N. Satyahadewi, and W. Andani, “Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga,” vol. 5, no. 2, pp. 431–445, 2023.
[17] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, and B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM,” vol. 7, no. 1, pp. 11–18, 2026.
[18] F. Abdusyukur, “Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34010/komputa.v12i1.9418.
[19] O. I. Gifari, M. Adha, I. R. Hendrawan, F. Freddy, S. Durrand, and A. S. Literature, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” vol. 2, no. 1, pp. 36–40, 2022.
[20] T. A. M and Q. N. Azizah, “Klasifikasi Tumor Otak Menggunakan Ekstraksi Fitur HOG dan Support Vector Machine,” vol. 4, no. 1, pp. 45–50, 2022.
[21] A. M. T. I. S. Ua *et al.*, “Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru,” vol. 2, no. 2, 2023.

[22] M. A. Palomino, "applied sciences Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis," 2022.
[23] Y. Sergio, V. Putranta, B. Rahayudi, and W. Purnomo, "Analisis Sentimen Masyarakat terhadap Kebijakan Penghapusan Subsidi BBM pada Media Sosial Twitter menggunakan Algoritma Naive Bayes Classifier dengan Ekstraksi Fitur N-Gram TF-IDF," vol. 7, no. 3, pp. 1501–1510, 2023.
[24] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 2, p. 77, 2023, doi: 10.31328/jointecs.v8i2.4747.
[25] J. Emarapenta, B. Sinulingga, H. Cesar, and K. Sitorus, "Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF Sentiment Analysis of Public towards Indonesian Horror Films Using SVM and TF-IDF Methods," vol. 14, no. April, pp. 42–53, 2024.
[26] N. A. Maulana and D. Darwis, "Perbandingan Metode SVM dan Naïve Bayes untuk Analisis Sentimen pada Twitter tentang Obesitas di Kalangan Gen Z Sistem Informasi , Fakultas Teknik dan Ilmu Komputer , Universitas Teknokrat Indonesia , Indonesia Comparison of SVM and Naive Bayes Methods for Sentiment Analysis on Twitter About Obesity Among Gen Z," vol. 5, no. 3, pp. 655–666, 2025.
[27] S. Anadas *et al.*, "Analisis Sentimen Publik Terhadap Program Lapor Mas Wapres Periode 2024 di Media Sosial X Menggunakan Metode Naive Bayes dan Support Vector Machine ( SVM ) Public Sentiment Analysis of the Vice President ' s Lapor Mas Program for the 2024 Period on Social Media X Using the Naive Bayes Method and Support Vector Machine ( SVM )," vol. 5, no. 8, pp. 2183–2192, 2025.
[28] J. Sains, F. A. Ryandi, D. Pratiwi, S. Sari, and J. Sains, "Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM," vol. 7, no. 1, pp. 1–6, 2025.
[29] T. Widyanto, I. Ristiana, and A. Wibowo, "Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter," vol. 6, no. 3, pp. 147–161, 2023.
[30] A. Nurkholis, S. Alim, H. Saputra, and A. Ferriyan, "Sentiment Analysis of COVID-19 Booster Vaccines on Twitter Using Multi-Class Support Vector Machine," vol. 8, no. 1, pp. 29–36, 2025, doi: 10.15408/aism.v8i1.42911.