

SKRIPSI

ANALISIS SENTIMEN VAKSINASI COVID-19 DI INDONESIA



Michael Alvino Wijaya Darmawan

NPM: 6181801041

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2023**

UNDERGRADUATE THESIS

**SENTIMENT ANALYSIS OF COVID-19 VACCINATION IN
INDONESIA**



Michael Alvino Wijaya Darmawan

NPM: 6181801041

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2023**

LEMBAR PENGESAHAN

ANALISIS SENTIMEN VAKSINASI COVID-19 DI INDONESIA

Michael Alvino Wijaya Darmawan

NPM: 6181801041

Bandung, 12 Januari 2023

Menyetujui,

Pembimbing

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

**Digitally signed
by Maria V.
Claudia M.**

Maria Veronica, M.T.

Anggota Tim Penguji

**Digitally signed
by Vania Natali**

Vania Natali, M.T.

Mengetahui,

Ketua Program Studi

**Digitally signed
by Mariskha Tri
Adithia**

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS SENTIMEN VAKSINASI COVID-19 DI INDONESIA

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 12 Januari 2023

A handwritten signature in black ink is written over a 5000 Rupiah postage stamp. The stamp features the Garuda Pancasila emblem and the text '5000', 'WELFARE', and 'TEMPERATURE'. The signature is stylized and covers most of the stamp's details.

Michael Alvino Wijaya Darmawan
NPM: 6181801041

ABSTRAK

Wabah *Covid-19* adalah sebuah pandemi global yang terjadi hampir di seluruh dunia yang mengharuskan masyarakat untuk melakukan vaksinasi. Penyelenggaraan vaksinasi menuai pro dan kontra dalam masyarakat khususnya pada media sosial, tidak semua masyarakat setuju tentang keharusan untuk melakukan vaksinasi. Maka dari itu, penelitian ini dibuat untuk menganalisis sentimen tentang vaksinasi *Covid-19* di Indonesia.

Analisis sentimen yaitu analisis teks untuk mengidentifikasi opini subjektif dalam sumber materi seperti kalimat atau dokumen. Metode dari penelitian untuk melakukan analisis sentimen diawali dengan penggunaan API *Twitter* untuk pengambilan data *tweet*. Setelah data dikumpulkan, proses pembersihan dan normalisasi dilakukan agar data dapat diproses lebih lanjut. Data yang telah melalui normalisasi kemudian dikategorikan ke dalam kategori positif, negatif, atau netral. Analisis sentimen dilakukan setelah data dikategorikan dengan menggunakan visualisasi data untuk mempermudah melihat pola-pola yang ada. Selain melakukan analisis, pembuatan model untuk melakukan prediksi sentimen dibangun dalam penelitian ini. Model dibuat dengan teknik *Support Vector Machine* untuk klasifikasi dan hasil dari model dievaluasi untuk melihat apakah model yang dihasilkan baik. Model yang telah dibuat dimasukkan ke dalam perangkat lunak agar dapat digunakan.

Dalam penelitian yang telah dilakukan, data yang diproses berjumlah 4519 buah yang diambil dari bulan Januari 2022 hingga Maret 2022 dengan kata kunci 'vaksin covid' dan 'vaksinasi covid'. Didapatkan hasil analisis sentimen pada data vaksinasi *Covid-19* di Indonesia dengan sentimen positif sebesar 39.1%, dan yang paling sedikit adalah sentimen negatif yaitu sebesar 24.5%. Pembangunan model dengan *Support Vector Machine* berhasil dibuat dengan mendapatkan parameter yang dipilih oleh metode *Grid Search* dengan melakukan kombinasi parameter yang dimasukkan. Model yang dihasilkan menggunakan parameter *kernel sigmoid*, *cost* sebesar 100 dan *gamma* sebesar 0.5 memiliki akurasi sebesar 81.19%.

Kesimpulan yang diperoleh dalam penelitian ini adalah pengumpulan data Twitter dengan kata kunci 'vaksin covid' dan 'vaksinasi covid' menggunakan *library tweepy* dengan *Python* berhasil dilakukan. Selain itu, analisis sentimen terkait vaksinasi *Covid-19* di Indonesia dapat dilakukan dengan menerapkan teknik pra-pengolahan teks. Pengembangan model dengan menggunakan teknik klasifikasi *Support Vector Machine* dan pengembangan perangkat lunak untuk memprediksi sentimen telah berhasil dibangun.

Kata-kata kunci: Analisis Sentimen, Vaksinasi Covid-19 di Indonesia, Support Vector Machine

ABSTRACT

The Covid-19 outbreak is a global pandemic that has occurred almost all over the world which requires people to vaccinate. The implementation of vaccination reaps pros and cons in society, especially on social media, not all people agree on the need to vaccinate. Therefore, this research was created to analyze sentiment about Covid-19 vaccination in Indonesia.

Sentiment analysis is text analysis to identify subjective opinions in source material such as sentences or documents. The research method for conducting sentiment analysis begins with the use of the Twitter API to retrieve tweet data. After the data is collected, the process of cleaning and normalization is carried out so that the data can be processed further. Data that has gone through normalization are then categorized into positive, negative, or neutral categories. Sentiment analysis is performed after the data is categorized by using data visualization to make it easier to see existing patterns. In addition to conducting analysis, modeling to predict sentiment was built in this study. The model is created using the Support Vector Machine technique for classification and the results of the model are evaluated to see if the resulting model is good. The model that has been made is entered into the software so that it can be used.

In the research that was conducted, 4519 data were processed taken from January 2022 to March 2022 with the keywords 'vaksin covid' and 'vaksinasi covid'. The results of sentiment analysis were obtained on the Covid-19 vaccination data in Indonesia with a positive sentiment of 39.1%, and the least negative sentiment was 24.5%. Model development with the Support Vector Machine was successfully created by obtaining the parameters selected by the Grid Search method by carrying out a combination of the parameters entered. The resulting model uses sigmoid kernel parameters, a cost of 100 and a gamma of 0.5 has an accuracy of 81.19%.

The conclusion obtained in this study is that Twitter data collection with the keywords 'vaksin covid' and 'vaksinasi covid' using the `tweepy` library with Python was successfully carried out. In addition, sentiment analysis related to the Covid-19 vaccination in Indonesia can be carried out by applying text pre-processing techniques. Model development using the Support Vector Machine classification technique and software development to predict sentiment has been successfully constructed.

Keywords: Sentiment Analysis, Covid-19 Vaccination in Indonesia, Support Vector Machine

Skripsi ini ditujukan untuk diri sendiri, keluarga, dan semua yang telah membantu saya dalam mengerjakan penelitian ini

KATA PENGANTAR

Puji dan Syukur yang penulis panjatkan kepada Tuhan Yang Maha Esa atas semua kelancaran yang diberikan dalam menyelesaikan penelitian yang berjudul "Analisis Sentimen Vaksinasi *Covid-19* Di Indonesia". Penulis menyadari banyak pihak yang telah memberikan dukungan dan bantuan selama menyelesaikan penelitian ini. Oleh karena itu, penulis dengan hormat mengucapkan terima kasih dan mendoakan semoga Tuhan memberikan balasan terbaik kepada:

- Ibu Mariskha Tri Adithia, P.D.Eng yang telah memberikan bimbingan dan arahan selama proses bimbingan skripsi berlangsung.
- Ibu Maria Veronica, M.T. dan Ibu Vania Natali, M.T. yang telah melakukan pengujian dan memberikan masukan yang membangun terkait penelitian yang sudah dilaksanakan.
- Keluarga khususnya mami yang selalu mendukung penulis dalam penelitian ini.
- Sahabat terdekat semasa SMA.
- Sahabat terdekat semasa kuliah Gian, Yalvi, Mario yang telah mendukung penulis dalam penelitian ini.
- Seluruh teman di angkatan 18 jurusan Informatika.

Penulis menyadari bahwa di dalam penelitian ini masih banyak kekurangannya. Oleh karena itu, penulis sangat mengharapkan kritik dan juga saran yang sifatnya membangun untuk karya yang lebih baik lagi kedepannya. Harapannya, penelitian ini akan memberikan manfaat bagi para pembaca.

Bandung, Januari 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
DAFTAR KODE PROGRAM	xxiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Metodologi	3
1.6 Sistematika Pembahasan	4
2 LANDASAN TEORI	5
2.1 Twitter [1]	5
2.2 Analisis Sentimen [2]	6
2.3 Data Mining [3] [4]	6
2.4 Machine Learning [5]	8
2.5 Natural Language Processing (NLP) [6]	9
2.6 Ekstraksi Fitur [7] [8]	12
2.7 Teknik Klasifikasi [9]	14
2.7.1 Teknik Klasifikasi Support Vector Machine (SVM) [10] [11]	14
2.8 Evaluasi Teknik Klasifikasi [9] [12]	20
2.8.1 Teknik Pembagian Data dan Optimasi Parameter	21
3 STUDI DAN EKSPLORASI TOOLS	23
3.1 Tools dan Library yang Digunakan	23
3.2 Eksplorasi	23
3.2.1 Eksplorasi Pengumpulan Data	23
3.2.2 Eksplorasi Pembersihan Data	26
3.2.3 Eksplorasi Natural Language Processing	27
3.2.4 Pelabelan Data Manual	29
3.2.5 Pelabelan Data dengan Lexicon Based [13]	29
3.2.6 Eksplorasi Visualisasi Data	31
3.2.7 Eksplorasi Ekstraksi Fitur	32
3.2.8 Eksplorasi Support Vector Machine	33
3.2.9 Eksplorasi Pembuatan Perangkat Lunak	35
4 PEMBANGUNAN MODEL DAN EKSPERIMEN	39

4.1	Analisis Masalah	39
4.2	Deskripsi Dataset	40
4.3	Proses Pembersihan Data	40
4.4	Proses Normalisasi Data	41
4.5	Proses Pelabelan Data	43
4.6	Visualisasi Data	44
4.7	Visualisasi Data Lanjutan	50
4.8	Proses Ekstraksi Fitur	54
4.9	Pemodelan Support Vector Machine (SVM) dan Evaluasi Teknik Klasifikasi	54
4.9.1	SVM dengan Train-Test Split	54
4.9.2	SVM dengan Cross-Validation	57
4.9.3	SVM dengan Grid Search	59
4.10	Kesimpulan Analisis Sentimen Data Twitter	61
5	IMPLEMENTASI PERANGKAT LUNAK DAN PENGUJIAN	63
5.1	Pembuatan Mockup Perangkat Lunak	63
5.1.1	Mockup Halaman Prediksi Sentimen	63
5.1.2	Mockup Halaman Deskripsi <i>Dataset</i>	65
5.2	Hasil Implementasi Perangkat Lunak	66
5.2.1	Halaman Prediksi Sentimen	66
5.2.2	Halaman Deskripsi Dataset	68
5.3	Pengujian Perangkat Lunak	69
5.3.1	Kalimat yang Berhasil Diprediksi	69
5.3.2	Kalimat yang Tidak Berhasil Diprediksi	70
6	KESIMPULAN DAN SARAN	73
6.1	Kesimpulan	73
6.2	Saran	73
	DAFTAR REFERENSI	75
	A KODE PROGRAM	77

DAFTAR GAMBAR

1.1	(a) Kalimat yang mengutarakan opini positif terhadap vaksinasi <i>Covid-19</i> . (b) Kalimat yang mengutarakan opini negatif terhadap vaksinasi <i>Covid-19</i>	2
2.1	Contoh dari <i>Tweet</i>	5
2.2	Tahapan dalam <i>Knowledge Discovery in Databases</i> (KDD) [14]	7
2.3	Contoh penentuan <i>hyperplane</i> dalam <i>Support Vector Machine</i> ¹	15
2.4	Contoh variabel dalam <i>Support Vector Machine</i> [15]	15
2.5	Contoh pembagian 3 kelas (kelas merah, biru, dan hijau) dalam SVM ²	17
2.6	Contoh Kasus Non-Linier SVM. Gambar bagian kanan adalah data dalam bidang 2 dimensi. Gambar bagian kiri adalah setelah proyeksi ke dalam bidang 3 dimensi. ³	17
2.7	Contoh bentuk pemisahan data dalam <i>Linear Kernel</i> ⁴	18
2.8	Contoh bentuk pemisahan data dalam <i>Polynomial Kernel</i> ⁴	19
2.9	Contoh bentuk pemisahan data dalam <i>RBF Kernel</i> ⁴	19
2.10	Contoh bentuk pemisahan data dalam <i>Sigmoid Kernel</i> ⁴	20
2.11	Contoh <i>Confusion Matrix</i> [12]	21
2.12	Contoh <i>Cross Validation k = 5</i> [16]	22
3.1	Contoh tampilan awal <i>Twitter Developer</i>	24
3.2	Visualisasi <i>Word Cloud</i> Persebaran Kata pada Data <i>Twitter</i> tentang Vaksinasi <i>Covid-19</i> di Indonesia	31
3.3	Visualisasi <i>Pie Chart</i> Persebaran Sentimen pada Data <i>Twitter</i> tentang Vaksinasi <i>Covid-19</i> di Indonesia	32
3.4	Tampilan awal <i>PAGE</i>	36
3.5	Tampilan <i>Main Application Toolbar</i>	36
3.6	Tampilan <i>Widget Toolbar</i>	36
3.7	Tampilan <i>Widget Tree</i>	37
3.8	Tampilan <i>Attribute Editor</i>	37
3.9	Tampilan <i>Preview Page</i>	38
3.10	Tampilan <i>Test Preview Page</i>	38
4.1	Diagram Alir Proses yang Dilakukan dalam Penelitian	39
4.2	Visualisasi <i>Pie Chart</i> Persebaran Sentimen pada Data <i>Twitter</i>	44
4.3	Visualisasi <i>Bar Chart</i> Persebaran Sentimen pada Data <i>Twitter</i>	45
4.4	Visualisasi <i>Bar Chart</i> Persebaran Sentimen pada Data <i>Twitter</i>	45
4.5	Visualisasi <i>Word Cloud</i> Persebaran Kata pada Data <i>Twitter</i>	46
4.6	Visualisasi <i>Bar Chart</i> Persebaran Kata pada Data <i>Twitter</i>	47
4.7	Visualisasi <i>Bar Chart</i> Persebaran Kata pada Sentimen Positif	47
4.8	Visualisasi <i>Bar Chart</i> Persebaran Kata Positif	48
4.9	Visualisasi <i>Bar Chart</i> Persebaran Kata pada Sentimen Negatif	48
4.10	Visualisasi <i>Bar Chart</i> Persebaran Kata Negatif	49
4.11	Visualisasi <i>Bar Chart</i> Persebaran Kata pada Sentimen Netral	50
4.12	Visualisasi <i>Bar Chart</i> Persebaran Kata Netral	50
4.13	Visualisasi <i>Pie Chart</i> Persebaran Sentimen Bulan Januari	51

4.14	Visualisasi <i>Word Cloud</i> Persebaran Kata Bulan Januari	51
4.15	Visualisasi <i>Pie Chart</i> Persebaran Sentimen Bulan Februari	52
4.16	Visualisasi <i>Word Cloud</i> Persebaran Kata Bulan Februari	52
4.17	Visualisasi <i>Pie Chart</i> Persebaran Sentimen Bulan Maret	53
4.18	Visualisasi <i>Word Cloud</i> Persebaran Kata Bulan Maret	53
4.19	Visualisasi Nilai Evaluasi dengan <i>Cost</i> 1 dan <i>Gamma</i> 1	55
4.20	Visualisasi Nilai Evaluasi dengan <i>Cost</i> 10 dan <i>Gamma</i> 1	56
4.21	Visualisasi Nilai Evaluasi dengan <i>Cost</i> 100 dan <i>Gamma</i> 1	57
4.22	Visualisasi Nilai Evaluasi setiap Iterasi <i>K-Fold</i>	58
5.1	Contoh <i>Mockup</i> Halaman Prediksi Sentimen	63
5.2	Contoh Hasil Prediksi	64
5.3	Contoh Hasil Prediksi tanpa Masukan	64
5.4	Contoh <i>Pop-up</i> Keluar dari Perangkat Lunak	65
5.5	Contoh <i>Mockup</i> Halaman Deskripsi <i>Dataset</i>	65
5.6	Contoh Halaman Prediksi Sentimen	66
5.7	Contoh Hasil Prediksi Sentimen	67
5.8	Contoh Hasil Prediksi Sentimen tanpa Masukan	67
5.9	Contoh <i>Pop-up</i> Keluar dari Perangkat Lunak	68
5.10	Contoh Halaman Deskripsi <i>Dataset</i>	68

DAFTAR TABEL

2.1	Contoh Tokenisasi	9
2.2	Contoh <i>Stemming</i>	10
2.3	Contoh awalan dan akhiran yang tidak diizinkan	11
2.4	Contoh <i>Case Folding</i>	12
2.5	Contoh Filtering	12
2.6	Contoh <i>Kasus</i>	13
3.1	Contoh <i>Cleaning Data</i>	27
3.2	Contoh <i>Case Folding Data</i>	27
3.3	Contoh <i>Stemming Data</i>	28
3.4	Contoh <i>Filtering Data</i>	28
3.5	Contoh Tokenisasi Data	29
3.6	Contoh Pelabelan Data	29
3.7	Contoh kamus kata	30
3.8	Skor sentimen	30
3.9	Contoh perhitungan skor	30
3.10	Contoh Hasil Ekstraksi Fitur dengan TF-IDF	33
3.11	Hasil Akurasi <i>Linear Kernel</i>	34
3.12	Hasil Akurasi <i>Polynomial Kernel</i>	34
3.13	Hasil Akurasi <i>RBF Kernel</i>	35
3.14	Hasil Akurasi <i>Sigmoid Kernel</i>	35
4.1	Contoh data yang diambil	41
4.2	Contoh penerapan proses pembersihan pada data	41
4.3	Contoh penerapan proses <i>case folding</i> pada data	42
4.4	Contoh penerapan proses perubahan kata tidak baku pada data	42
4.5	Contoh penerapan proses <i>filtering</i> pada data	42
4.6	Contoh hasil penerapan algoritma Nazief-Adriani pada data	43
4.7	Contoh penerapan proses tokenisasi pada data	43
4.8	Contoh hasil dari proses penghapusan kalimat yang kurang dari 3 kata	43
4.9	Contoh penerapan skor sentimen pada data	44
4.10	Contoh pemberian label sentimen pada data	44
4.11	Contoh Hasil Ekstraksi Fitur dengan TF-IDF	54
4.12	Contoh Nilai Evaluasi dengan $Cost = 1$	55
4.13	Contoh Nilai Evaluasi dengan $Cost = 10$	56
4.14	Contoh Nilai Evaluasi dengan $Cost = 100$	57
4.15	Contoh Nilai Evaluasi untuk setiap Iterasi Fold	58
4.16	Contoh Parameter	59
4.17	Contoh Hasil Parameter <i>Grid Search</i>	59
4.18	Hasil Nilai Evaluasi dengan $Kernel = Sigmoid$, $Cost = 100$ dan $Gamma = 0.5$	59
4.19	<i>Confusion Matrix</i>	59
4.20	Hasil Nilai <i>Confusion Matrix</i> Kategori Negatif	60
4.21	Hasil Nilai <i>Confusion Matrix</i> Kategori Netral	60

4.22 Hasil Nilai <i>Confusion Matrix</i> Kategori Positif	60
5.1 Contoh Kalimat Normal	69
5.2 Contoh Kalimat <i>Slang</i>	70
5.3 Contoh Kalimat Normal	70
5.4 Contoh Kalimat <i>Slang</i>	71

DAFTAR KODE PROGRAM

3.1	Contoh kode untuk otentikasi <i>Key</i> dan <i>Token</i>	24
3.2	Contoh kode untuk melakukan penarikan data dengan <i>Cursor</i> setelah otentikasi <i>key</i> dan <i>token</i> dilakukan	25
3.3	Contoh kode untuk melakukan penarikan data	25
3.4	Contoh kode untuk memasukkan data ke <i>dataframe</i> & pembuatan <i>CSV</i>	26
3.5	Contoh kode untuk penggabungan kumpulan data	26
3.6	Contoh kode untuk <i>cleaning</i> data	26
3.7	Contoh kode untuk <i>case folding</i> data	27
3.8	Contoh kode untuk <i>stemming</i> data	28
3.9	Contoh kode untuk <i>filtering</i> data	28
3.10	Contoh kode untuk tokenisasi data	29
3.11	Contoh kode untuk visualisasi <i>word cloud</i>	31
3.12	Contoh kode untuk visualisasi <i>pie chart</i>	31
3.13	Contoh kode untuk ekstraksi fitur	32
3.14	Contoh kode SVM <i>Linear</i>	33
3.15	Contoh kode SVM <i>Polynomial</i>	34
3.16	Contoh kode SVM <i>RBF</i>	34
3.17	Contoh kode SVM <i>Sigmoid</i>	35
A.1	Tweepy.py	77
A.2	Preprocessing.py	77
A.3	Lexicon.py	80
A.4	Analysis.py	81
A.5	SVMFinal.py	87
A.6	FinalGuiSupport.py	90
A.7	FinalGui.py	91
A.8	FunctionGui.py	95

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Wabah *Covid-19* adalah sebuah pandemi global yang terjadi hampir di seluruh dunia. Pandemi *Covid-19* disebabkan oleh virus bernama *SARS-CoV-2* yang menyerang saluran pernapasan. Jika manusia terkena penyakit ini, ada beberapa gejala umum yang akan terjadi seperti demam, batuk, kehilangan rasa atau bau, dan juga gejala serius seperti kesulitan bernapas, kesulitan berbicara atau bergerak, maupun nyeri dada. *Covid-19* ditemukan pertama kali di kota Wuhan yang berlokasi di negara Cina pada bulan Desember 2019. Setelah *Covid-19* menyebar luas di berbagai negara, *World Health Organization* (WHO) atau Badan Kesehatan Dunia mendeklarasikan secara resmi bahwa virus *Corona* (*Covid-19*) merupakan pandemi pada tanggal 12 Maret 2020 [17]. Pada 6 Maret 2020, pemerintah Indonesia mengumumkan kasus pertama *Covid-19* di Indonesia [18].

Semakin luasnya penyebaran wabah *Covid-19*, pemerintahan Indonesia mengadakan program vaksinasi untuk setiap warganya. Vaksinasi adalah kegiatan yang dilaksanakan tenaga medis dengan memberikan vaksin melalui alat suntik maupun dengan cara diteteskan ke dalam mulut untuk meningkatkan produksi antibodi dengan tujuan untuk menjaga sistem imun dan kekebalan tubuh terhadap penyakit tertentu [19].

Vaksinasi *Covid-19* dilakukan di seluruh dunia yang terkena dampak pandemi, termasuk Indonesia. Pada tanggal 13 Januari 2021, Indonesia menerima vaksin *Covid-19* untuk pertama kalinya, dan Presiden Joko Widodo mengadakan program vaksinasi *Covid-19* di Indonesia sampai saat ini. Masuknya vaksin dan penyelenggaraan program vaksinasi ini menjadi pembicaraan dalam media sosial [20].

Dalam era digital, banyak media sosial yang dipakai oleh orang-orang untuk saling terhubung antara satu dengan yang lain, salah satunya *Twitter*. *Twitter* adalah layanan jejaring sosial, dapat diartikan sebagai *microblog* yang memungkinkan pengguna untuk mengirim dan melakukan interaksi melalui pesan yang dikenal sebagai "*tweets*". *Tweets* adalah sebuah pesan yang dapat berbentuk teks yang memiliki panjang maksimum 280 karakter, selain itu pengguna dapat mengirimkan gambar dan video berdurasi pendek. *Tweet* tersebut digunakan sebagai sarana utama pengguna untuk berkirim pesan secara global yang banyak dipakai orang-orang untuk berkomentar, bercerita, maupun berpendapat [1].

Data yang tersebar dalam media sosial seperti *Twitter* dapat dimanfaatkan untuk melihat pandangan masyarakat terhadap suatu hal. Tanggapan yang masyarakat kirimkan lewat media sosial memiliki sentimen yang bermacam-macam. Dalam Kamus Besar Bahasa Indonesia (KBBI), sentimen yaitu sebuah pendapat atau pandangan yang didasarkan pada perasaan yang berlebihan terhadap sesuatu sedangkan kata sentimen dalam konteks analisis sentimen yaitu pendapat subjektif seseorang yang berhubungan dengan topik tertentu. Pengertian dari analisis sentimen adalah analisis teks untuk mengidentifikasi dan mengklasifikasikan pendapat subjektif dalam materi sumber seperti kalimat atau dokumen. Analisis sentimen bertujuan untuk mengidentifikasi sikap seorang penulis yang berhubungan dengan topik tertentu atau polaritas kontekstual yang berkategori positif, negatif, dan netral dari sumber teks tersebut [2].



Gambar 1.1: (a) Kalimat yang mengutarakan opini positif terhadap vaksinasi *Covid-19*. (b) Kalimat yang mengutarakan opini negatif terhadap vaksinasi *Covid-19*.

Pada *tweet* yang tersebar dalam platform *Twitter*, banyak kalimat dari pengguna yang berisikan opini di mana opini tersebut dapat mengandung sentimen yang bersifat negatif, positif, maupun netral. Dapat dilihat contoh kalimat pada Gambar 1.1, gambar 1.1a di bagian kiri menunjukkan sentimen positif berisi dukungan terhadap vaksinasi *Covid-19* dan gambar 1.1b di bagian kanan menunjukkan sentimen negatif yang berisi ketidakpercayaan terhadap vaksin.

Dari berbagai kalimat yang berisikan opini, manusia dapat dengan mudah memperkirakan apakah opini tersebut positif, negatif, atau netral. Namun, manusia tidak mungkin untuk menentukan sentimen dari kalimat-kalimat yang sangat banyak. Maka dari itu, untuk tetap bisa melihat sentimen dalam banyak data, dibutuhkan algoritma yang dapat menyelesaikan masalah tersebut.

Kalimat yang diambil dari platform seperti *Twitter* tidak bisa diproses secara langsung untuk diambil maknanya. Kalimat-kalimat yang ada di media sosial tidak semuanya bersifat baku dan juga tidak semuanya kata bermakna, maka dari itu digunakan teknik *Natural Language Processing* (NLP) untuk melakukan pengolahan data agar dapat dimanfaatkan dengan lebih maksimal. *Natural Language Processing* adalah salah satu bidang ilmu yang mempelajari komunikasi antara komputer dan manusia melalui bahasa alami [6].

Setelah melalui tahap NLP, fitur-fitur yang terdapat dalam kalimat akan diambil dengan teknik ekstraksi fitur menggunakan *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk mengukur seberapa penting suatu kata dalam suatu dokumen tertentu [7]. Hasil dari pengambilan fitur data akan digunakan untuk pembuatan model klasifikasi. Ada beberapa teknik klasifikasi yang dapat digunakan seperti *Decision Tree*, *Random Forest*, *Naïve Bayes*, *Support Vector Machine* (SVM), dan lainnya. Teknik klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM). SVM adalah metode untuk melakukan klasifikasi data dengan cara mencari pembatas antar kelompok dengan memaksimalkan nilai margin terhadap kelas data tersebut [10].

Pada penelitian ini, ada beberapa *library* yang akan digunakan seperti *tweepy* yang digunakan untuk mengambil data dari platform *Twitter*, *pandas* yang menyediakan struktur data dan melakukan perubahan format data agar dapat terbaca, *sklearn* untuk proses ekstraksi fitur, teknik klasifikasi SVM, perhitungan evaluasi model, dan *library* lainnya.

Proses yang akan dilakukan untuk melakukan analisis sentimen dimulai dengan pengumpulan data *tweet* dari platform *Twitter* yang berbentuk teks. Dikarenakan platform *Twitter* digunakan oleh masyarakat di berbagai negara dengan bahasa yang beragam, oleh karena itu data *tweet* yang diambil hanya yang berbahasa Indonesia. Data yang telah diambil akan dilakukan pembersihan dan normalisasi data dengan teknik NLP seperti tokenisasi, *case folding*, *stemming* dan *filtering* agar data dapat lebih mudah untuk dilakukan analisis lebih lanjut. Setelah itu, fitur yang ada dalam data diambil dengan proses ekstraksi fitur dengan TF-IDF, fitur yang didapat dimasukkan ke dalam proses teknik klasifikasi SVM untuk membuat model yang akan digunakan untuk melakukan prediksi data baru apakah memiliki sentimen positif, negatif, atau netral. Setelah model dibuat, model diuji dengan menggunakan perhitungan evaluasi model untuk melihat seberapa baik model yang dihasilkan. Hasil dari data yang telah diproses tersebut diambil hal-hal yang penting untuk ditarik kesimpulan agar mendapatkan wawasan baru terhadap data yang dilakukan pengujian.

Penelitian ini menghasilkan perangkat lunak berbasis *Graphical User Interface* (GUI). Perangkat lunak ini dibuat untuk melakukan prediksi menggunakan model yang telah dibuat dengan teknik SVM. Pengguna dapat memasukkan suatu data yang berbentuk kalimat untuk pengujian sentimen ke dalam perangkat lunak dan perangkat lunak mengeluarkan hasil prediksi sentimen yang terdapat pada kalimat tersebut.

1.2 Rumusan Masalah

Berikut rumusan masalah pada penelitian ini:

1. Bagaimana cara mengambil data *Twitter* tentang vaksinasi *Covid-19* di Indonesia?
2. Bagaimana cara untuk melakukan analisis sentimen tentang vaksinasi *Covid-19* di Indonesia?
3. Bagaimana cara membangun perangkat lunak yang dapat menampilkan hasil dari analisis sentimen tentang vaksinasi *Covid-19* di Indonesia?

1.3 Tujuan

Berikut tujuan dari penelitian ini:

1. Mempelajari teknik pengumpulan data *Twitter* tentang vaksinasi *Covid-19* di Indonesia dengan menggunakan API *Twitter* dan mengimplementasikan teknik pengambilan data tentang vaksinasi *Covid-19* di Indonesia dengan menggunakan *library tweepy* dalam *Python*.
2. Mempelajari dan mengimplementasikan teknik *Natural Language Processing* untuk pra-pengolahan data.
3. Membangun model dengan mengimplementasikan metode klasifikasi *Support Vector Machine* (SVM) yang akan digunakan untuk menguji sentimen tentang vaksinasi *Covid-19* di Indonesia. Membangun perangkat lunak dengan menggunakan *Python* yang mengimplementasikan model yang dibuat dengan teknik *Support Vector Machine*.

1.4 Batasan Masalah

Berikut batasan masalah dari penelitian ini:

1. Data *tweet* yang diambil menggunakan bahasa Indonesia.
2. Pengambilan data tentang vaksinasi *covid-19* di Indonesia berada pada rentang bulan Januari 2022 hingga Maret 2022.

1.5 Metodologi

Berikut metodologi penelitian yang dilakukan:

1. Melakukan studi literatur mengenai materi yang akan digunakan dalam penelitian ini. Berikut materi yang dipelajari sebagai acuan dalam penelitian:
 - (a) Studi literatur tentang *Twitter*
 - (b) Studi literatur tentang analisis sentimen
 - (c) Studi literatur tentang konsep *data mining*
 - (d) Studi literatur tentang konsep *machine learning*
 - (e) Studi literatur tentang konsep pada teknik *natural language processing*
 - (f) Studi literatur tentang teknik ekstraksi fitur teks
 - (g) Studi literatur tentang teknik klasifikasi *Support Vector Machine*
 - (h) Studi literatur tentang evaluasi teknik klasifikasi
2. Melakukan pengumpulan data dari platform *Twitter* tentang vaksinasi *Covid-19* di Indonesia.
3. Melakukan eksplorasi teknologi yang akan digunakan dengan mempelajari cara pemakaian *library* pada *Python*.
4. Melakukan persiapan data seperti pembersihan data, normalisasi data, pelabelan data, ekstraksi fitur data.
5. Melakukan perancangan model dan evaluasi model.
6. Melakukan pengujian model yang telah dibuat.
7. Melakukan analisis pada hasil pengujian.
8. Melakukan implementasi perangkat lunak.
9. Melakukan peluncuran model pada perangkat lunak.
10. Melakukan eksperimen pengujian perangkat lunak yang dibuat.
11. Menulis dokumen skripsi.

1.6 Sistematika Pembahasan

Berikut sistematika dari penelitian ini:

1. **Bab 1 Pendahuluan**

Pada bagian ini akan dibahas mengenai latar belakang dari penelitian, rumusan masalah, tujuan penelitian, batasan masalah penelitian, metodologi penelitian dan sistematika pembahasan dari penelitian ini.

2. **Bab 2 Landasan Teori**

Pada bagian ini akan dibahas mengenai teori-teori yang dipakai dalam penelitian. Teori-teori yang akan dipakai yaitu *Twitter*, Analisis Sentimen, *Data Mining*, *Machine Learning*, Teknik *Natural Language Processing*, Teknik Ekstraksi Fitur, Teknik Klasifikasi *Support Vector Machine*, dan Evaluasi Teknik Klasifikasi.

3. **Bab 3 Studi dan Eksplorasi Tools**

Pada bagian ini akan dibahas mengenai tahap dari penelitian dan eksplorasi *tools* yang akan digunakan dalam penelitian ini. Eksplorasi *tools* digunakan sebagai gambaran awal dari tahap-tahap untuk melakukan pembangunan model.

4. **Bab 4 Pembangunan Model dan Eksperimen**

Pada bagian ini akan dibahas mengenai analisis masalah, deskripsi dataset, juga proses pembangunan model, visualisasi data, eksperimen model dan melakukan analisis terhadap model yang dibuat.

5. **Bab 5 Implementasi Perangkat Lunak Dan Pengujian**

Pada bagian ini akan dibahas mengenai pembuatan *mockup* perangkat lunak, implementasi perangkat lunak, dan pengujian perangkat lunak dengan implementasi model *Support Vector Machine* yang telah dibuat.

6. **Bab 6 Kesimpulan dan Saran**

Pada bagian ini akan dibahas mengenai kesimpulan yang didapat dalam penelitian dan saran untuk penelitian selanjutnya.