

**KLASIFIKASI SENTIMEN TERHADAP DATA *TEXT*  
JEJARING SOSIAL DENGAN TOPIK VAKSIN  
COVID-19 MENGGUNAKAN *SUPPORT VECTOR  
MACHINE (SVM)***



Oleh

**Ilham Solehan  
09011281621034**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA  
2021**

**KLASIFIKASI SENTIMEN TERHADAP DATA *TEXT*  
JEJARING SOSIAL DENGAN TOPIK VAKSIN  
COVID-19 MENGGUNAKAN *SUPPORT VECTOR  
MACHINE (SVM)***

**TUGAS AKHIR**

**Diajukan Untuk Melengkapi Salah Satu Syarat**

**Memperoleh Gelar Sarjana Komputer**



**Oleh:**

**ILHAM SOLEHAN**

**09011281621034**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA  
2021**

**LEMBAR PENGESAHAN**

**KLASIFIKASI SENTIMEN TERHADAP DATA *TEXT*  
JEJARING SOSIAL DENGAN TOPIK VAKSIN COVID-19  
MENGUNAKAN *SUPPORT VECTOR MACHINE (SVM)***

**TUGAS AKHIR**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**

Oleh:

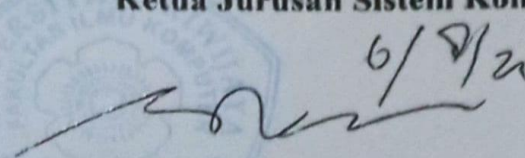
**ILHAM SOLEHAN**

**09011281621034**

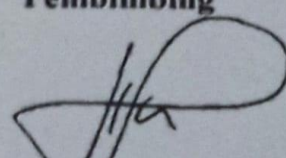
**Indralaya, Juli 2021**

Mengetahui,

**Ketua Jurusan Sistem Komputer**

  
**Dr. Ir. H. Sukemi, M.T.**  
**NIP. 196612032006041001**

**Pembimbing**

  
**Huda/Ubaya, S.T., M.T.**  
**NIP. 198106162012121003**



## HALAMAN PERSETUJUAN

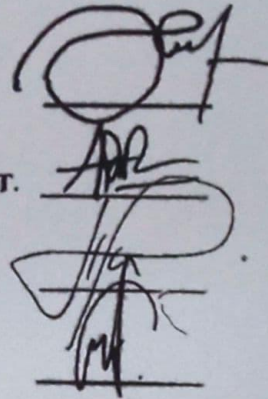
Telah diuji dan lulus pada :

Hari : Jum'at

Tanggal : 2 Juli 2021

Tim Penguji :

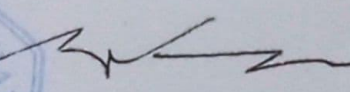
1. Ketua : Ahmad Fali Oklilas, S.T., M.T.
2. Sekretaris : Aditya Putra P Prasetyo, S.Kom., M.T.
3. Pembimbing : Huda Ubaya, S. T., M.T.
4. Penguji : Ahmad Zarkasi, S.T., M.T.



Mengetahui, 6/8/21

Ketua Jurusan Sistem Komputer



  
Dr. Ir. H. Sukemi, M.T.  
NIP. 196612032006041001



## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Ilham Solehan  
NIM : 09011281621034  
Judul : Klasifikasi Sentimen Terhadap Data *Text* Jejaring Sosial dengan Topik Vaksin Covid-19 Menggunakan *Support Vector Machine* (SVM)

**Hasil Penyecekan *Software iThenticate/Turnitin* : 5 %**

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Indralaya, Juli 2021



**Ilham Solehan**  
**NIM. 09011281621034**

## KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh,

Puji dan syukur penulis panjatkan kepada Allah Subhanahu Wa Ta'ala, atas segala karunia, berkat, dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Skripsi ini dengan judul “Klasifikasi Sentimen Terhadap Data *Text* Jejaring Sosial dengan Topik Vaksin Covid-19 Menggunakan *Support Vector Machine* (SVM)”.

Dalam laporan ini penulis menjelaskan mengenai bagaimana cara mengklasifikasi *Sentimen* dengan menggunakan *Support Vector Machine* disertai data-data yang diperoleh penulis saat melakukan pengujian. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan dan referensi untuk para akademisi dan peneliti lain yang juga berada atau sedang menekuni bidang *data text*, khususnya pengusaha dan pemerintah.

Pada penyusunan skripsi ini, penulis banyak mendapatkan ide dan saran serta bantuan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah Subhanahu Wa Ta'ala dan terimakasih kepada yang terhormat:

1. Orang tua beserta keluarga yang selalu mendoakan, memberikan dukungan baik moril maupun materil, serta motivasi dan semangat selama hidup penulis.
2. Bapak Jaidan Jauhari, S.Pd. M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. H.Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Iman Saladin B. Azhar, S.Kom., M.MSI. selaku Pembimbing Akademik
5. Bapak Huda Ubaya, M.T selaku Pembimbing Tugas Akhir.
6. Jerry Pratama Saputra dan Rizki Ronaldo sebagai partner dan teman dari awal pengerjaan tugas akhir, baik dalam pencarian literatur, pemrograman, serta diskusi mengenai riset dan lainnya.
7. Teman - teman dari Jurusan Sistem Komputer yang tidak bisa disebutkan satu-persatu. Khususnya seluruh teman-teman dari kelas SK16B Indralaya.

8. Para sahabat saya yang ada di kosan, Hisyam Sanusi, Nauval Faris Fashlullah, Ichwanul Hakim, Rap Nur Muhammad, Madyus Randika & Muhammad Sultan Alif yang selalu memotifasi saya untuk cepat lulus.
9. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa Skripsi ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi dikemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pemikiran dalam peningkatan mutu pembelajaran dan penelitian.

Indralaya, Juli 2021

Penulis

Ilham Solehan  
NIM.09011281621034

# KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

**Ilham Solehan (09011281621034)**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : [ilhamsunsri@gmail.com](mailto:ilhamsunsri@gmail.com)

## ABSTRAK

Pada Agustus 2020, pemerintah mengumumkan telah menemukan vaksin virus *corona* dan mendistribusikannya kepada publik untuk menghentikan pandemi. Kali ini kebijakan pemerintah tersebut memicu banyak reaksi publik berupa sentimen negatif maupun positif tentang vaksin tersebut. Semakin banyaknya jumlah sentimen yang beredar di masyarakat mengenai pandemi yang sedang berlangsung dan tingginya antusias masyarakat terhadap produk kebijakan berupa Vaksin Covid-19 pemerintah kali ini, maka klasifikasi sentimen masyarakat terhadap vaksin ini dirasa perlu. Sentimen masyarakat direkam dan dikumpulkan dari suatu jejaring sosial, yaitu jejaring sosial twitter dimana terdapat banyak sentimen masyarakat tentang topik Vaksin Covid-19, Klasifikasi sentimen tersebut akan menggunakan *Natural Language Processing (NLP) & Support Vector Machine (SVM)* sebagai metode klasifikasinya. Dari percobaan yang dilakukan dengan *split* data 80% data pelatihan dengan 20% data pengujian menghasilkan nilai *Precision, Recall* dan *F1-Score* yang paling tinggi, sedangkan untuk percobaan dengan *split* data 20% data pelatihan dengan 80% data pengujian menghasilkan nilai *Precision, Recall* dan *F1-Score* yang paling rendah. Klasifikasi data dapat terlaksana dengan baik menggunakan metode *Support Vector Machine*, sehingga dapat membuat sistem yang mampu untuk mengklasifikasi sentimen dengan topik Vaksin Covid-19. Hasil prediksi data *text* sentimen menggunakan metode *Support Vector Machine* dapat dijadikan sebagai bahan pertimbangan atau acuan terhadap pengambilan keputusan pemerintah dalam mengambil kebijakan mengenai topik Vaksin Covid-19.

**Kata Kunci :** *Natural Language Processing, Support Vector Machine, Vaksin Covid-19, sentimen, data text.*



**CLASSIFICATION OF SENTIMENTS ON SOCIAL NETWORK TEXT DATA  
WITH COVID-19 VACCINE TOPIC USING SUPPORT VECTOR MACHINE  
(SVM)**

**Ilham Solehan (09011281621034)**

*Dept. of Computer Engineering, Faculty of Computer Science, Sriwijaya  
University*

*Email : [ilhamsunsri@gmail.com](mailto:ilhamsunsri@gmail.com)*

**ABSTRACT**

*In August 2020, the government announced it had found a coronavirus vaccine and distributed it to the public to stop the pandemic. This time the government's policy triggered many public reactions in the form of negative and positive sentiments about the vaccine. The greater number of sentiments circulating in the community regarding the ongoing pandemic and the high public enthusiasm for the government's Covid-19 Vaccine policy product this time, the classification of public sentiment towards this vaccine is deemed necessary. Community sentiments are recorded and collected from a social network, namely the Twitter social network where there are many public sentiments on the topic of the Covid-19 Vaccine. The classification of these sentiments will use Natural Language Processing (NLP) & Support Vector Machine (SVM) as the classification method. From the experiments conducted with 80% data split of training data with 20% of test data resulted in the highest Precision, Recall and F1-Score values, while for experiments with 20% data split of training data with 80% of test data yielded Precision, Recall and The lowest F1-Score. Data classification can be carried out properly using the Support Vector Machine method, so that it can create a system that is able to classify sentiments on the topic of the Covid-19 Vaccine. The results of the prediction of sentiment text data using the Support Vector Machine method can be used as consideration or reference for government decision making in making policies regarding the Covid-19 vaccine topic.*

**Keywords:** *Natural Language Processing, Support Vector Machine, Covid-19 Vaccine, sentiment, text data.*

## DAFTAR ISI

	<b>Halaman</b>
Halaman Judul .....	i
Halaman Pengesahan .....	ii
Halaman Persetujuan .....	iii
Halaman Pernyataan .....	iv
Kata Pengantar .....	v
Abstrak .....	vii
<i>Abstract</i> .....	viii
Daftar Isi .....	ix
Daftar Gambar .....	xiv
Daftar Tabel .....	xvii
Daftar Lampiran .....	xviii
Bab I. Pendahuluan .....	1
1.1 Latar Belakang .....	1
1.2. Tujuan .....	3
1.3. Perumusan Masalah .....	3
1.4 Batasan Masalah .....	3
1.5 Metodologi Penelitian .....	4

1.6 Sistematika Penulisan .....	5
Bab II. Tinjauan Pustaka .....	6
2.1. Analisis Sentimen .....	6
2.2. Text Mining .....	7
2.3. <i>Natural Language Processing</i> (NLP) .....	8
2.4. <i>Text Preprocessing</i> .....	10
2.4.1. Pembersihan Dokumen ( <i>Cleansing</i> ).....	10
2.4.2. <i>Case Folding</i> .....	11
2.4.3. <i>Tokenizing</i> .....	11
2.4.4. <i>Stemming</i> .....	11
2.4.5. Normalisasi .....	11
2.5. Ekstraksi Fitur .....	11
2.5.1. Model Bag of words .....	12
2.5.2. Pembobotan Kalimat .....	12
2.6. <i>Machine Learning</i> .....	13
2.7. <i>Support Vector Machine</i> .....	14
2.7.1. Proses Pembelajaran .....	15
2.8. <i>K-Fold Cross Validation</i> .....	16
2.9. <i>Confussion Matrix</i> .....	17
2.10. <i>Precision, Recall, F-measure</i> .....	18
Bab III. Metodologi Penelitian .....	20



3.1. Metodologi Penelitian .....	20
3.2. Kerangka Kerja .....	21
3.3. Pengambilan Data .....	22
3.4. <i>Preprocessing</i> .....	26
3.4.1. <i>Cleaning</i> .....	27
3.4.2. <i>Case Folding</i> .....	29
3.4.3. <i>Tokenizing</i> .....	30
3.4.4. <i>Stemming</i> .....	32
3.4.5. Pemisahaan feature dan label .....	34
3.5. Ekstraksi fitur .....	35
3.6. <i>Cross Validation</i> .....	35
3.7. <i>Confussion Matrix</i> .....	35
 Bab IV. Pengujian dan Analisa.....	 37
4.1 Pendahuluan.....	37
4.2 <i>Dataset Text</i> .....	37
4.3 Proses <i>Preprocessing</i> .....	37
4.4 Ekstraksi Fitur .....	41
4.5 Percobaan dan Pengujian <i>Support Vector Machine</i> .....	42
4.5.1. Percobaan menggunakan split dataset	
sebanyak 80%:20 .....	45
4.5.1.1 Analisis percobaan dengan data training	

80% dan testing 20% .....	47
4.5.2. Percobaan menggunakan split dataset	
sebanyak 70%:30 .....	50
4.5.2.1 Analisis percobaan dengan data training	
70% dan testing 30% .....	52
4.5.3. Percobaan menggunakan split dataset	
sebanyak 60%:40 .....	55
4.5.3.1 Analisis percobaan dengan data training	
60% dan testing 40% .....	57
4.5.4. Percobaan menggunakan split dataset	
sebanyak 50%:50 .....	60
4.5.4.1 Analisis percobaan dengan data training	
50% dan testing 50% .....	62
4.5.5. Percobaan menggunakan split dataset	
sebanyak 40%:60 .....	65
4.5.5.1 Analisis percobaan dengan data training	
40% dan testing 60% .....	67
4.5.6. Percobaan menggunakan split dataset	
sebanyak 30%:70 .....	70
4.5.6.1 Analisis percobaan dengan data training	
30% dan testing 70% .....	72
4.5.7. Percobaan menggunakan split dataset	





## DAFTAR GAMBAR

	<b>Halaman</b>
Gambar 2.1 Struktur Kalimat Bahasa Indonesia .....	9
Gambar 2.2 menunjukkan hyperline yang membagi jarak antara kelas.....	14
Gambar 2.3 Bentuk ruang vektor dengan dimensi yang lebih tinggi.....	15
Gambar 2.4 Matriks Konfolusi Ukuran 3x3.....	18
Gambar 3.1 Diagram alur penelitian .....	21
Gambar 3.2 Dataset yang diambil dari Drone Emprit Academy.....	23
Gambar 3.3 Total isi Dataset Sentimen Vaksin Covid-19.....	24
Gambar 3.4 Wordcloud sentimen negatif .....	25
Gambar 3.5 Wordcloud sentimen netral .....	25
Gambar 3.6 Wordcloud sentimen positif .....	26
Gambar 3.7 Diagram alur penelitian .....	27
Gambar 3.8 Kerangka kerja cleanning.....	28
Gambar 3.9 Proses Cleanning.....	28
Gambar 3.10 Sentimen yang diproses cleaning.....	29
Gambar 3.11 Kerangka kerja case folding.....	29
Gambar 3.12 Proses case folding.....	30
Gambar 3.13 Kerangka kerja tokenizing .....	30
Gambar 3.14 Proses tokenizing.....	31
Gambar 3.15 Sentimen yang diproses tokenisasi.....	31

Gambar 3.16 Kerangka kerja stemming.....	33
Gambar 3.17 Proses stemming.....	33
Gambar 3.18 Sentimen yang diproses stemming.....	34
Gambar 3.19 kata-kata yang menjadi scalar.....	35
Gambar 4.1 Alur Kerja dari Prerocessing.....	38
Gambar 4.2 Perbandingan sebelum dan setelah dilakukan Cleanning.....	39
Gambar 4.3 Perbandingan sebelum dan setelah dilakukan Stemming.....	40
Gambar 4.4 Perbandingan sebelum dan setelah dilakukan <i>Tokenization</i> .....	40
Gambar 4.5 Contoh data sentimen dari ekstraksi fitur.....	41
Gambar 4.6 Split data Pelatihan sebanyak 80% dataset.....	46
Gambar 4.7 Split data Pengujian sebanyak 20% dataset.....	46
Gambar 4.8 Hasil percobaan dengan data pelatihan 80% dan data pengujian 20%.....	48
Gambar 4.9 Split data Pelatihan sebanyak 70% dataset.....	51
Gambar 4.10 Split data Pengujian sebanyak 30% dataset.....	51
Gambar 4.11 Hasil percobaan dengan data pelatihan 70% dan data pengujian 30%.....	53
Gambar 4.12 Split data Pelatihan sebanyak 60% dataset.....	56
Gambar 4.13 Split data Pengujian sebanyak 40% dataset.....	56
Gambar 4.14 Hasil percobaan dengan data pelatihan 60% dan data pengujian 40%.....	58
Gambar 4.15 Split data Pelatihan sebanyak 50% dataset.....	61
Gambar 4.16 Split data Pengujian sebanyak 50% dataset.....	61
Gambar 4.17 Hasil percobaan dengan data pelatihan 50% dan data pengujian 50%.....	63
Gambar 4.18 Split data Pelatihan sebanyak 40% dataset.....	66
Gambar 4.19 Split data Pengujian sebanyak 60% dataset.....	66

Gambar 4.20 Hasil percobaan dengan data pelatihan 40% dan data pengujian 60%.....	68
Gambar 4.21 Split data Pelatihan sebanyak 30% dataset.....	71
Gambar 4.22 Split data Pengujian sebanyak 70% dataset.....	71
Gambar 4.23 Hasil percobaan dengan data pelatihan 30% dan data pengujian 70%.....	73
Gambar 4.24 Split data Pelatihan sebanyak 20% dataset.....	76
Gambar 4.25 Split data Pengujian sebanyak 80% dataset.....	76
Gambar 4.26 Hasil percobaan dengan data pelatihan 20% dan data pengujian 80%.....	78
Gambar 4.27 Grafik transisi nilai dari pengujian dengan parameter Precision.....	81
Gambar 4.28 Grafik transisi nilai dari pengujian dengan parameter Recall.....	82
Gambar 4.29 Grafik transisi nilai dari pengujian dengan parameter F1-Score.....	82



## DAFTAR TABEL

	<b>Halaman</b>
Tabel 2.1 Tabel Confusion matrix berukuran $M = 2$ .....	17
Tabel 3.1 Klasifikasi multi-class yang memiliki 3 keluaran kelas confusion matrix.....	36
Tabel 4.1 Tabel Jumlah pembagian dataset pada setiap percobaan.....	42
Tabel 4.2 Matriks Konfolusi dengan uluran $M=3$ atau $3 \times 3$ .....	43
Tabel 4.3 Hasil perbandingan pengujian pada setiap split data yang di uji.....	81

## DAFTAR LAMPIRAN

**LAMPIRAN 1.** Program Klasifikasi Sentimen *Support Vector Machine* ..... 1-A

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Di awal tahun 2020, merebaknya virus jenis baru bernama Virus Corona (SARS-CoV-2) dan penyakit baru yang berasal dari virus tersebut bernama Coronavirus 2019 (COVID-19) telah menggemparkan dunia. Seperti kita ketahui bersama, sumber virusnya berasal dari Wuhan, Tiongkok. Virus itu ditemukan pada akhir Desember 2019. Sejauh ini, 65 negara telah dipastikan terinfeksi virus tersebut[1].

Bersamaan dengan pandemi virus Corona, krisis lain juga menampakkan diri dalam bentuk ketakutan dan kepanikan masyarakat, yang disebabkan oleh informasi yang tidak lengkap dan seringkali tidak akurat. Oleh karena itu, ada kebutuhan besar untuk menyelesaikan dan lebih memahami krisis informasi COVID-19 dan mengukur sentimen publik sehingga pesan yang benar dan keputusan kebijakan dapat dilaksanakan.

Pada Agustus 2020, pemerintah mengumumkan telah menemukan vaksin virus Corona dan mendistribusikannya kepada publik untuk menghentikan pandemi. Kali ini kebijakan pemerintah tersebut memicu banyak reaksi publik, dengan sentimen negatif maupun positif tentang vaksin tersebut[2].

Semakin banyaknya jumlah sentimen yang beredar di masyarakat mengenai pandemi yang sedang berlangsung dan tingginya antusias masyarakat terhadap produk kebijakan berupa Vaksin Covid-19 baru pemerintah kali ini, maka klasifikasi sentimen yang beredar di masyarakat terhadap vaksin ini dianggap perlu.

Pada masa pandemi kali ini sentimen masyarakat dapat direkam dan dikumpulkan dari suatu jejaring sosial yaitu jejaring sosial Twitter dimana terdapat banyak sentimen masyarakat tentang topik Vaksin Covid-19, klasifikasi sentimen tersebut akan menggunakan *Natural Language Processing* (NLP) & *Support Vector Machine* (SVM) sebagai metodenya.

*Support Vector Machine* (SVM) memiliki beberapa kelebihan dibandingkan metode klasifikasi lain yaitu: 1. Generalisasi, Generalisasi didefinisikan sebagai kemampuan suatu metode untuk mengklasifikasikan suatu pattern, yang tidak termasuk data yang dipakai dalam fase pembelajaran metode tersebut. 2. *Curse of Dimensionality*, *Curse of Dimensionality* didefinisikan sebagai masalah yang dihadapi suatu metode pattern recognition dalam mengestimasi parameter misal jumlah hidden neuron pada neural network, stopping criteria dalam proses pembelajaran, dsb dikarenakan jumlah sampel data yang relatif sedikit dibandingkan dimensional ruang vektor data tersebut. Semakin tinggi dimensi dari ruang vektor informasi yang diolah, membawa konsekuensi dibutuhkan jumlah data dalam proses pembelajaran. 3. *Feasibility* SVM dapat diimplementasikan relatif mudah, karena proses penentuan support vector dapat dirumuskan dalam QP problem *Quadratic Programming*.

U. Rofiqoh, et al. Dalam jurnalnya yang berjudul “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Feature”[5]. Melakukan penelitian yang bermaksud untuk mengukur tingkat akurasi metode dalam melakukan klasifikasi sentimen. Dari penelitian tersebut menunjukkan bahwa metode *Support Vector Machine* menghasilkan akurasi yang cukup baik. Berdasar penelitian tersebut, penulis memutuskan untuk menggunakan *Support Vector Machine* sebagai metode klasifikasi sentimen dan menggunakan bahasa pemrograman Python. Dengan begitu klasifikasi sentimen yang berupa data text dapat dilaksanakan.

## 1.2 Tujuan

Tujuan dari penelitian ini adalah :

1. Penelitian ini memiliki tujuan untuk dapat mengklasifikasikan sentimen yang tersebar di masyarakat mengenai topik Vaksin Covid-19 menggunakan metode *support vector machine* (SVM).
2. Membangun sistem klasifikasi sentimen yang hasilnya dapat menjadi masukan sebagai acuan bagi Pemerintah terhadap kebijakan yang diambil mengenai Vaksin Covid-19 yang akan berdampak pada masyarakat.

## 1.3 Perumusan Masalah

Dalam penelitian ini inti permasalahan adalah belum adanya klasifikasi sentimen dari masyarakat mengenai pandangan mereka terhadap Vaksin Covid-19 yang menjadi program pemerintah dalam menanggulangi kasus pandemi virus Covid-19.

## 1.4 Batasan Masalah

Berdasarkan rumusan masalah di atas, akan dilakukan pembatasan masalah agar penelitian ini tetap terarah.

1. Penelitian ini menggunakan *dataset* dari Drone Emprit Academy (DEA) yaitu *dataset* API Twitter dengan topik Vaksin Covid-19 selama 4 bulan yaitu dari tanggal 24 Agustus 2020 sampai dengan tanggal 24 November 2020.
2. Penelitian ini menggunakan *Support Vector Machine* (SVM) untuk melakukan klasifikasi sentimen.
3. Penelitian ini bersifat simulasi menggunakan program dengan bahasa *python*.

## 1.5 Metodologi Penelitian

Tahapan dari Metodologi penelitian kali ini akan dijelaskan dalam poin berikut :

1. Rumusan masalah

Tahapan ini akan merumuskan masalah yang dihadapi dalam melakukan *preprocessing dataset*, ekstraksi fitur, serta penyelesaiannya.

2. Studi Pustaka.

Tahapan ini bertujuan untuk mencari literatur yang sesuai dalam rangka menunjang penelitian yang dilakukan, dengan cara mencari literatur yang tersebar menggunakan kata kunci yang spesifik dalam penelitian kali ini.

3. Perancangan.

Tahapan ini merupakan proses perancangan dari sistem yang akan dibangun. Tahapan ini membahas tentang runtutan proses dari pembuatan program untuk normalisasi *dataset*, ekstraksi fitur, serta klasifikasi.

4. Pengujian.

Tahapan ini merupakan proses pengujian dari sistem yang dibangun sesuai dengan rancangan yang dibuat.

5. Validasi

Tahapan ini diperlukan untuk mengecek apakah sistem telah mampu melakukan klasifikasi dari *dataset* yang diberikan.

6. Evaluasi

Tahap ini bertujuan untuk mengetes sistem yang dibuat menggunakan data tes yang telah disediakan agar dapat mengetahui seberapa akurat sistem mengklasifikasi data sentimen.

7. Analisis

Tahapan ini adalah proses menganalisis hasil dari sistem yang dibangun, Analisis ditampilkan dalam rangka mengukur seberapa bagus sistem yang dibangun dalam menjalankan fungsinya.

## 8. Kesimpulan dan Saran

Tahapan ini adalah proses memberikan kesimpulan berdasarkan hasil dari analisis dari sistem yang dibangun, dan saran bagi penelitian yang akan datang.

### 1.6 Sistematika Penulisan

Dalam penulisan skripsi ini akan dibagi menjadi beberapa bab, sebagai berikut :

#### **BAB I PENDAHULUAN**

Pada bab satu akan berisi uraian sistematis mengenai pembahasan yang diteliti, meliputi berbagai aspek mulai dari latar belakang, manfaat, tujuan, sistematika penulisan, rumusan masalah, hingga metodologi.

#### **BAB II TINJAUAN PUSTAKA**

Pada bab dua akan berisi penjelasan mengenai landasan teori yang dapat mendukung pembahasan dari penelitian kali ini, Dasar teori tersebut meliputi setiap aspek yang akan digunakan dalam penelitian kali ini.

#### **BAB III METODOLOGI PENELITIAN**

Pada bab tiga akan berisi penjelasan dari bagaimana penelitian ini akan berlangsung.

#### **BAB IV HASIL DAN ANALISIS**

Pada bab empat akan berisi hasil dan analisis dari penelitian yang telah berlangsung.

#### **BAB V KESIMPULAN**

Pada bab lima akan berisi kesimpulan dari proses sebelumnya, yaitu hasil dan analisis.



## DAFTAR PUSTAKA

- [1] F. Aslam, T. M. Awan, J. H. Syed, A. Kashif, and M. Parveen, "Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak," *Humanit. Soc. Sci. Commun.*, vol. 7, no. 1, pp. 1–9, 2020, doi: 10.1057/s41599-020-0523-3.
- [2] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 public sentiment insights and machine learning for tweets classification," *Inf.*, vol. 11, no. 6, pp. 1–22, 2020, doi: 10.3390/info11060314.
- [3] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (Svm)," *J. Gaussian*, vol. 9, no. 3, pp. 376–390, 2020, doi: 10.14710/j.gauss.v9i3.28932.
- [4] Y. Zhao, S. Cheng, X. Yu, and H. Xu, "Chinese public's attention to the COVID-19 epidemic on social media: Observational descriptive study," *J. Med. Internet Res.*, vol. 22, no. 5, pp. 1–13, 2020, doi: 10.2196/18825.
- [5] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, pp. 1725–1732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.
- [6] W. Gata, "Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS," vol. 6, pp. 1–13, 2017.
- [7] A. Hussain *et al.*, "Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study," *J. Med.*

*Internet Res.*, vol. 23, no. 4, p. e26627, 2021, doi: 10.2196/26627.

- [8] X. Chen, H. Xie, G. Cheng, L. K. M. Poon, M. Leng, and F. L. Wang, "Trends and features of the applications of natural language processing techniques for clinical trials text analysis," *Appl. Sci.*, vol. 10, no. 6, pp. 1–36, 2020, doi: 10.3390/app10062157.
- [9] M. Allahyari *et al.*, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," *arXiv*, 2017.
- [10] M. Rusli, "Ekstraksi Fitur Menggunakan Model Word2Vec Pada Sentiment Analysis Kolom Komentar Kuisisioner Evaluasi Dosen Oleh Mahasiswa," *Klik - Kumpul. J. Ilmu Komput.*, vol. 7, no. 1, p. 35, 2020, doi: 10.20527/klik.v7i1.296.
- [11] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.
- [12] V. A. Oliinyk, V. Vysotska, Y. Burov, K. Mykich, and V. Basto-Fernandes, "Propaganda Detection in Text Data Based on NLP and machine learning," *CEUR Workshop Proc.*, vol. 2631, pp. 132–144, 2020.
- [13] M. Ihsan, E. Roza, and E. Widodo, "Analisis Sentimen Twitter terhadap Bom Bunuh Diri di Surabaya 13 Mei 2018 menggunakan Pendekatan Support Vector Machine," *Prism. Pros. Semin. Nas. Mat.*, vol. 2, pp. 416–426, 2019.
- [14] M. CINDO, D. P. Rini, and E. Ermatita, "Analisis Sentimen Pada Twitter Menggunakan Metode Maximum Entropy Dan Support Vector Machine," no. 2015, pp. 509–513, 2019.
- [15] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.

- [16] M. Classification, E. Morvant, and L. Ralaivola, "PAC-Bayesian Generalization Bound on Confusion Matrix for co," pp. 1–13, 2018.
- [17] S. R. Widyamahendra, "Analisis Sentimen Twitter Hasil Pilkada DKI Jakarta 2017 menggunakan Metode Support Vector Machine," 2018.