

**ANALISIS SENTIMEN DAN *SUPPORT VECTOR MACHINE* (SVM)
PADA OPINI PENGGUNA TWITTER TERHADAP VAKSIN
BOOSTER DI INDONESIA**

TUGAS AKHIR

Diajukan sebagai salah satu persyaratan guna memperoleh gelar Ahli Madya



Oleh:

NADIA SALSABILA HARLI

19037048

**PROGRAM STUDI DIPLOMA III STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2022**

PERSETUJUAN UJIAN TUGAS AKHIR

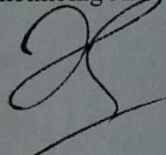
**ANALISIS SENTIMEN DAN *SUPPORT VECTOR MACHINE* (SVM)
PADA OPINI PENGGUNA TWITTER TERHADAP
VAKSIN BOOSTER DI INDONESIA**

Nama : Nadia Salsabila Harli
NIM/Tahun Masuk : 19037048/2019
Program Studi : DIII Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 24 Agustus 2022

Disetujui oleh:

Pembimbing Akademik



Dra. Nonong Amalita, M.Si
NIP. 196906151993032001

HALAMAN PENGESAHAN LULUS UJIAN TUGAS AKHIR

Nama : Nadia Salsabila Harli
TM/NIM : 2019/19037048
Program Studi : DIII Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

**ANALISIS SENTIMEN DAN *SUPPORT VECTOR MACHINE* (SVM)
PADA OPINI PENGGUNA TWITTER TERHADAP
VAKSIN BOOSTER DI INDONESIA**

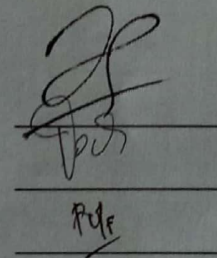
Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Tugas Akhir
Program Studi DIII Statistika Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 24 Agustus 2022

Tim Penguji

Nama

1. Ketua : Dra. Nonong Amalita, M.Si
2. Anggota : Dodi Vionanda, M.Si, Ph.D
3. Anggota : Fadhilah Fitri, S.Si, M.Stat



Handwritten signatures of the examiners, with horizontal lines below each signature.


SURAT PERNYATAAN TIDAK PLAGIAT

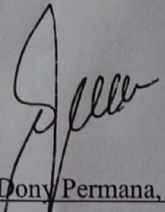
Saya yang bertanda tangan di bawah ini:

Nama : Nadia Salsabila Harli
NIM/TM : 19037048/2019
Program Studi : DIII Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan, bahwa Tugas Akhir saya dengan judul "**Analisis Sentimen dan Support Vector Machine (SVM) pada Opini Pengguna Twitter Terhadap Vaksin Booster di Indonesia**" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan Negara.

Demikianlah pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

 Diketahui oleh,
Ketua Departemen Statistika


Dr. Dony Permana, M.Si
NIP. 197501272006041001

Saya yang menyatakan,


Nadia Salsabila Harli
19037048

Analisis Sentimen dan *Support Vector Machine* (SVM) pada Opini Pengguna Twitter Terhadap Vaksin Booster di Indonesia

Nadia Salsabila Harli

ABSTRAK

Covid-19 (*Coronavirus disease 2019*) adalah wabah penyakit menular baru yang disebabkan oleh *coronavirus*. Penyebaran virus ini sangat cepat keseluruh negara di dunia. Pada bulan Maret 2020 Covid-19 masuk dan menyerang Indonesia. Penyebaran virus Covid-19 di Indonesia semakin meningkat setiap harinya. Salah satu upaya yang dilakukan pemerintah dalam mengurangi penyebaran virus ini dengan penggunaan vaksin. Dalam pelaksanaan vaksinasi, vaksin booster masih jauh dari target yang diharapkan pemerintah. Keberadaan vaksin booster ini mendapat berbagai pendapat dari masyarakat Indonesia yang disampaikan melalui media sosial Twitter. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia tentang vaksinasi booster menggunakan klasifikasi *Support Vector Machine*.

Jenis penelitian ini adalah penelitian terapan. Data yang digunakan adalah *tweet* berbahasa Indonesia yang membahas mengenai vaksin booster. Data *tweet* diperoleh dari dataset twitter dengan menggunakan teknik *crawling* data pada Python. Kemudian dilakukan tahap *preprocessing* data yang terdiri dari: *cleaning*, *case folding*, *tokenizing*, *stopwords*, dan *stemming*. Klasifikasi sentimen dibagi menjadi sentimen positif dan sentimen negatif.

Preprocessing yang telah dilakukan menghasilkan data bersih sebanyak 707 *tweet* dari 1500 *tweet* data awal. Berdasarkan proses pelabelan yang dilakukan diperoleh data *tweet* sentimen positif sebanyak 572 data dan data *tweet* sentimen negatif sebanyak 135 data. Dari klasifikasi analisis sentimen opini pengguna twitter terhadap vaksin booster menggunakan *Support Vector Machine* didapatkan tingkat akurasi sebesar 85,9%.

Kata Kunci: Vaksin Booster, Twitter, Analisis Sentimen, *Support Vector Machine*

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillah, rasa syukur penulis ucapkan kehadiran Allah SWT atas berkat rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “**Analisis Sentimen dan *Support Vector Machine* (SVM) pada Opini Pengguna Twitter Terhadap Vaksin Booster di Indonesia**”. Shalawat beserta salam tidak lupa penulis sampaikan kepada Nabi Muhammad SAW sebagai suri tauladan bagi seluruh umat.

Penulisan Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan pada Program Studi Diploma III Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang. Dalam menyelesaikan Tugas Akhir ini, tidak lepas dari bimbingan dan bantuan berbagai pihak. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan terimakasih kepada:

1. Ibu Dra. Nonong Amalita, M.Si., pembimbing Tugas Akhir sekaligus Sekretaris Departemen Statistika FMIPA UNP.
2. Bapak Dodi Vionanda, M.Si, Ph.D., sebagai Dosen Penguji sekaligus Koordinator Program Studi DIII Statistika FMIPA UNP.
3. Ibu Fadhilah Fitri, M.Stat., sebagai Dosen Penguji.
4. Bapak Dr. Dony Permana, M.Si., Kepala Departemen Statistika FMIPA UNP.
5. Bapak dan Ibu Dosen dan Tenaga Kependidikan Departemen Statistika FMIPA UNP.

6. Teristimewa untuk kedua orang tua yang telah memberikan semangat, nasehat, dukungan dan bantuan moril maupun materi kepada penulis.
7. Semua sahabat, teman serta rekan-rekan yang selalu memberikan semangat serta dukungan kepada penulis.

Semoga dorongan, bimbingan dan bantuan yang diberikan kepada penulis menjadi amal ibadah yang diridhai Allah SWT. Penulis menyadari bahwa Tugas Akhir ini masih terdapat kekurangan. Oleh karena itu, masukan dan saran yang bersifat membangun dari berbagai pihak sangat diharapkan sehingga menjadikan Tugas Akhir ini menjadi lebih baik. Akhir kata, penulis berharap semoga Tugas Akhir ini dapat memberikan sumbangsan pemikiran bagi setiap pembaca dan dapat bermanfaat bagi semua pihak.

Padang, 16 Agustus 2022

Nadia Salsabila Harli

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vi
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN.....	viii
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Batasan Masalah.....	5
C. Rumusan Masalah	5
D. Tujuan Penelitian	6
E. Manfaat Penelitian	6
BAB II KERANGKA TEORITIS.....	7
A. Covid-19.....	7
1. Pengertian Covid-19.....	7
2. Penyebab Covid-19	8
3. Penularan Covid-19	9
4. Tanda dan Gejala.....	9
5. Vaksin Covid-19.....	10
B. <i>Text Mining</i>	11
C. Analisis Sentimen	13
D. Twitter.....	16

E. <i>Machine Learning</i>	17
F. <i>Support Vector Machine (SVM)</i>	18
G. <i>Teknik Validasi Model</i>	21
BAB III METODOLOGI PENELITIAN.....	24
A. Jenis Penelitian.....	24
B. Data dan Sumber Data	24
C. Variabel Penelitian	24
D. Teknik Pengumpulan Data.....	25
E. Teknik Analisis Data.....	28
BAB IV HASIL DAN PEMBAHASAN	30
A. Analisis Sentimen	30
1. <i>Pre-processing</i> data.....	30
2. Pelabelan Data.....	39
B. Deskripsi Data	40
C. Klasifikasi <i>Support Vector Machine</i>	44
D. Evaluasi Model.....	47
E. Pembahasan.....	50
BAB V PENUTUP.....	51
A. Kesimpulan	51
B. Saran.....	51
DAFTAR PUSTAKA	53
LAMPIRAN	55

DAFTAR TABEL

TABEL	Halaman
1. Kriteria Kualitas Akurasi	23
2. Variabel Penelitian	25
3. Contoh dari Hasil Crawling Data	27
4. Contoh Hasil Pre-Processing Remove Mention Username dan Retweet	31
5. Contoh Hasil Pre-Processing Remove Url	32
6. Contoh Hasil Pre-Processing Remove Angka dan Tanda Baca	33
7. Contoh Hasil Pre-Processing Case Folding	34
8. Contoh Hasil Pre-Processing Tokenizing	35
9. Contoh Hasil Pre-Processing Stopwords	37
10. Contoh Hasil Pre-Processing Stemming	38
11. Hasil Pelabelan <i>Tweet</i>	40
12. Hasil Confusion Matrix	47
13. Nilai presisi, recall dan F1-Score Evaluasi Model	49

DAFTAR GAMBAR

GAMBAR	Halaman
1. Persentase Pelaksanaan Vaksin Covid 19 di Indonesia Hingga 6 Juli 2022.....	2
2. Contoh Support Vector Machine	19
3. Confusion Matrix	22
4. Tampilan Akun yang Telah Terdaftar di Developer Twitter	26
5. Kode Program Proses Crawling Data <i>Tweet</i>	27
6. Kode Program Remove Username.....	31
7. Kode Program Remove Url.....	32
8. Kode Program Remove Angka dan Tanda Baca.....	33
9. Kode Program Remove Data Duplikat	34
10. Kode Program Tahapan Case Folding	34
11. Kode Program Tahapan Tokenizing	35
12. Kode Program Tahapan Stopwords	36
13. Kode Program Tahapan Stemming	38
14. Kode Program Perhitungan Skor Sentimen	39
15. Kode Program Pelabelan Sentimen.....	40
16. Hasil Persentase Labeling Data <i>Tweet</i>	41
17. WordCloud Hasil <i>Pre-Processing</i>	42
18. WordCloud Hasil Pre-Processing Sentimen Positif.....	43
19. WordCloud Hasil Pre-Processing Sentimen Negatif	44
20. Kode Program Pembagian Data Training dan Data Testing.....	45
21. Kode Program Pembobotan Kata dengan TF IDF	46
22. Pembobotan Kata pada <i>Tweet</i>	46
23. Kode Program Klasifikasi <i>Support Vector Machine</i>	46
24. Kode Program Confusion Matrix.....	47
25. Perhitungan Akurasi.....	48

DAFTAR LAMPIRAN

LAMPIRAN	Halaman
1. Kode Program Python untuk <i>Preprocessing</i> Data	55
2. Kode Program Pelabelan Data	59
3. Kode Program Visualisasi Data	60
4. Kode Program Klasifikasi <i>Support Vector Machine</i>	61
5. Data Tweet yang Sudah Dilakukan Pelabelan	63
6. Data Testing dan Prediksi Menggunakan <i>Support Vector Machine</i>	79

BAB I

PENDAHULUAN

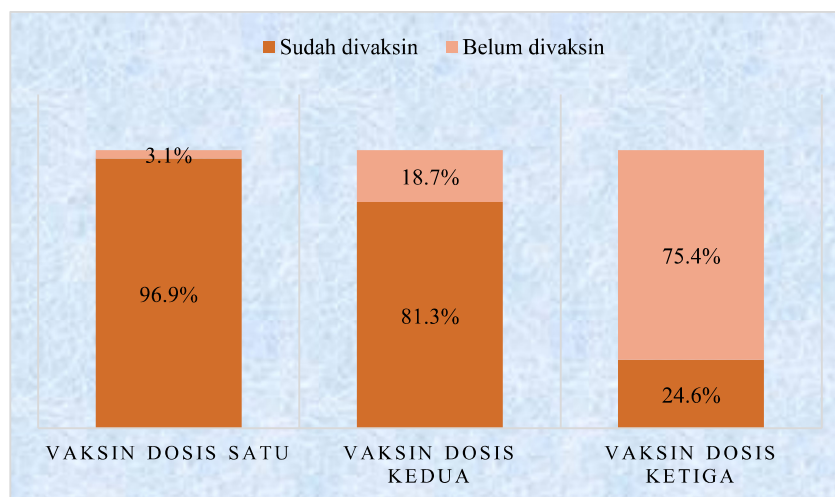
A. Latar Belakang

Covid 19 (*Coronavirus disease 2019*) adalah wabah penyakit menular baru yang disebabkan oleh *coronavirus* dan secara resmi telah ditetapkan menjadi pandemi global pada tanggal 11 Maret 2020 oleh *World Health Organization* (WHO). Harahap (2021) menyatakan bahwa penyebaran virus ini terjadi pertama kali di kota Wuhan, China, pada akhir tahun 2019. Virus ini pertama kali muncul di pasar hewan dan makanan laut di Wuhan. Masyarakat yang terinfeksi oleh virus ini ternyata terkait dengan pasar hewan tersebut. Virus ini semakin menyebar luas keseluruh dunia dengan jumlah kasus positif sebanyak 537 juta kasus dan jumlah kematian 6,3 juta jiwa pertanggal 16 Juni 2022 (Center for Systems Science and Engineering, 2022)

Indonesia pertama kali terinfeksi virus covid-19 pada tanggal 2 Maret 2020 yang menginfeksi 2 warga Depok, Jawa Barat. Berawal dari kasus tersebut penyebaran virus covid-19 semakin bertambah dengan sangat cepat setiap harinya (Laurensz & Eko Sedyono, 2021). Berdasarkan web resmi covid.id, kasus covid-19 di Indonesia pertanggal 1 Juli 2022 berjumlah 6,09 juta kasus dengan jumlah kematian sebanyak 156 ribu jiwa. Kondisi penyebaran virus yang semakin pesat menyebabkan dampak yang sangat besar bagi seluruh masyarakat. Pemerintah Indonesia menetapkan protokol kesehatan diseluruh kegiatan dimulai dari *sosial distancing* hingga *lockdown* total sehingga menghambat seluruh kegiatan masyarakat Indonesia (Yulita *et al.*, 2021).

Melihat permasalahan yang banyak terjadi, maka cara untuk mencegah virus ini adalah dengan penggunaan vaksin (Fitriana *et al.*, 2021). Vaksin adalah cara paling tepat untuk pencegahan penyakit menular. Vaksin bertujuan untuk menciptakan kekebalan kelompok (*herd immunity*) agar masyarakat menjadi lebih produktif dalam menjalankan aktivitas kesehariannya. Vaksin tidak hanya dilakukan untuk melindungi yang divaksin tetapi juga masyarakat luas yang berkemungkinan berinteraksi sehingga mengurangi penyebaran virus covid dimasyarakat (Harun & Ananda, 2021). Pelaksanaan vaksin di Indonesia dilakukan dari tanggal 13 Januari 2021 dengan Presiden Joko Widodo menjadi penerima vaksin pertama.

Berdasarkan data Kementerian Kesehatan Republik Indonesia pemerintah menargetkan yang akan divaksin sebanyak 208.265.720 jiwa. Berikut persentase masyarakat yang sudah divaksin dosis satu, vaksin dosis kedua, dan vaksin dosis ketiga hingga tanggal 6 Juli 2022.



Gambar 1. Persentase Pelaksanaan Vaksin Covid 19 di Indonesia Hingga 6 Juli 2022

Pada Gambar 1 dapat dilihat ternyata pelaksanaan vaksin dosis satu di Indonesia hampir 100% terpenuhi, untuk pelaksanaan vaksin dosis kedua sudah lebih dari 50% target terpenuhi, tetapi pada pelaksanaan vaksin dosis ketiga atau booster masih jauh dari yang diharapkan pemerintah yaitu masih sebanyak 24,6%.

Menurut Arami (2022) mengatakan bahwa vaksin dosis ketiga sangat dibutuhkan karena efektifitas pemberian vaksin dosis satu dan dosis kedua mengalami penurunan sebesar 80% terhadap infeksi yang disebabkan oleh virus varian *omicron*. Pemerintah telah melaksanakan vaksinasi Covid-19 dosis ketiga sejak 12 Januari 2022 melalui surat edaran kementerian kesehatan RI. Namun kebijakan ini menimbulkan pro dan kontra di masyarakat. Beberapa argumen menyatakan bahwa vaksin booster memberikan efek samping yang lebih besar dari vaksin sebelumnya, vaksin yang belum ampuh karena masih bisa terinfeksi covid setelah divaksin, dan vaksin yang dianggap sebagai akal-akalan pemerintah (Sihidi *et al.*, 2022). Sedangkan masyarakat yang pro akan vaksin booster mengatakan bahwa vaksin booster bagus untuk meningkatkan imun lebih kuat sehingga dapat terlindungi dari virus varian *omicron* (Arami *et al.*, 2022). Masyarakat cenderung mengungkapkan argumen dan kritiknya di media sosial, salah satunya twitter.

Twitter adalah salah satu media sosial yang banyak digunakan oleh masyarakat aktif dan tersebar luas didunia untuk mengungkapkan respon dan opini nya masing-masing terhadap sesuatu (Fitriana *et al.*, 2021). Twitter dapat menampilkan topik yang sedang sering dibicarakan. Di twitter masyarakat sering mengungkapkan apa yang sedang dilakukan dan dirasakan, saling berbagi

informasi yang diperoleh, dan mengungkapkan argumen tentang suatu topik. Isi *tweet* dijadikan sebagai sumber data opini masyarakat yang dapat dianalisis sehingga menjadi bahan evaluasi dan alat penentu kebijakan bagi pihak terkait.

Text mining adalah sebuah teknik untuk menggali informasi dari sebuah teks dengan menemukan pola dari sekumpulan data. Analisis sentimen merupakan bagian dari *text mining* yang melakukan proses pemahaman, penarikan dan pengolahan data dalam bentuk teks untuk mendapatkan sentimen yang terkandung pada tiap opini (Hidayat *et al.*, 2021). Analisis sentimen dilakukan untuk mengetahui sikap emosional masyarakat dalam argumen yang diklasifikasikan sehingga argumen masuk pada kelompok sentimen yang sama (Nomleni, 2015). Banyak cara klasifikasi yang digunakan untuk analisis sentimen diantaranya *Decision Tree*, *Naïve bayes classifier*, *Support Vector Machine (SVM)*, dan *Random forest*. Berdasarkan penelitian terdahulu, SVM merupakan algoritma yang sering digunakan dan memberikan hasil yang cukup bagus dalam riset analisis sentimen. SVM menyediakan pilihan yang banyak dalam pembentukan model klasifikasinya terkait dengan fungsi kernel yang dapat digunakan untuk membangun modelnya (Zuriel & Fahrurozi, 2021).

Berdasarkan penelitian (Hikmah & Fauzan, 2020) tentang perspektif masyarakat terhadap vaksin booster di Indonesia menggunakan klasifikasi Naïve Bayes dengan akurasi 89% mengatakan bahwa vaksin booster kurang bagus menurut netizen Indonesia karena vaksin booster merupakan syarat mudik tahun 2022 dan efek yang didapat menyebabkan badan pegal-pegal, tangan tidak bisa diangkat. Menurut penelitian Rakhmawati *et al.*(2020) dalam klasifikasi pengguna

twitter terhadap pengadaan vaksin covid-19 mengatakan bahwa vaksin covid-19 memberikan dampak positif berdasarkan percakapan dan hasil sentimen positif yang lebih tinggi pada media sosial twitter.

Melihat banyaknya terjadi pro dan kontra di masyarakat terhadap program vaksin booster, maka dilakukan penelitian untuk melihat bagaimana respon dan opini masyarakat terhadap pelaksanaan vaksin booster di Indonesia dari media sosial twitter. Untuk menjawab permasalahan tersebut, dilakukan penelitian dengan judul “**Analisis Sentimen dan *Support Vector Machine* (SVM) pada Opini Pengguna Twitter Terhadap Vaksin Booster di Indonesia**”.

B. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Objek dalam penelitian adalah *tweet* dengan kata kunci vaksin booster
2. *Tweet* yang digunakan adalah *tweet* yang berbahasa Indonesia
3. Data *tweet* yang digunakan adalah historis *tweet* dari tanggal 12 Januari 2022 sampai dengan 5 Agustus 2022.

C. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah pada penelitian adalah sebagai berikut:

1. Bagaimana klasifikasi opini pengguna twitter terhadap Vaksin booster di Indonesia menggunakan Analisis Sentimen?
2. Bagaimana respon dan opini pengguna twitter terhadap Vaksin booster di Indonesia menggunakan Analisis Sentimen?

3. Bagaimana tingkat akurasi algoritma *Support Vector Machine* dalam pengklasifikasian sentimen pengguna twitter terhadap Vaksin booster?

D. Tujuan Penelitian

Adapun tujuan pada penelitian ini adalah sebagai berikut:

1. Mengetahui klasifikasi opini masyarakat pengguna twitter terhadap vaksin booster menggunakan Analisis Sentimen
2. Mengetahui respon dan opini pengguna twitter terhadap Vaksin booster di Indonesia menggunakan Analisis Sentimen
3. Mengetahui tingkat akurasi algoritma *Support Vector Machine* dalam klasifikasi sentimen pengguna twitter terhadap Vaksin booster

E. Manfaat Penelitian

Manfaat yang akan didapatkan dari hasil penelitian adalah:

1. Bagi pemerintah, untuk mengetahui opini masyarakat mengenai vaksin booster sebagai bahan pertimbangan dalam membuat kebijakan
2. Bagi peneliti, menambah wawaasan dan pengetahuan khususnya bagaimana efektifitas algoritma *Support Vector Machine* dalam pengklasifikasian
3. Bagi mahasiswa, dapat dijadikan referensi oleh mahasiswa untuk penelitian selanjutnya.