

**ANALISIS SENTIMEN VAKSINASI COVID-19 MENGGUNAKAN
SUPPORT VECTOR MACHINE**

SKRIPSI



disusun oleh

Rizky Adhi Nugroho

18.61.0145

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**ANALISIS SENTIMEN VAKSINASI COVID-19 MENGGUNAKAN
SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Rizky Adhi Nugroho

18.61.0145

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN VAKSINASI COVID-19 MENGGUNAKAN
SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

Rizky Adhi Nugroho

18.61.0145

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Desember 2021

Dosen Pembimbing,

Majid Rahardi, S.Kom., M.Eng

NIK. 190302393

PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN VAKSINASI COVID-19 MENGGUNAKAN
SUPPORT VECTOR MACHINE

yang disusun oleh

Rizky Adhi Nugroho

18.61.0145

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Desember 2021

Susunan Dewan Penguji

Nama Penguji

Majid Rahardi, S.Kom, M.Eng
NIK. 190302393

Arif Dwi Laksito, M.Kom
NIK. 190302150

Jeki Kuswanto, M.Kom
NIK. 190302456

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar Sarjana Komputer
Tanggal

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa skripsi ini merupakan karya saya sendiri (ASLI). Isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 5 Januari 2022



Rizky Adhi Nugroho

NIM. 18.61.0145

MOTTO

“All we can do is our best, and sometimes the best that we can do is start over”

(Peggy Carter)

PERSEMBAHAN

Dengan menyebut nama Allah yang Maha Pengasih dan Maha Penyayang. Puji syukur ke hadirat Allah swt. atas nikmat yang telah diberikan sehingga penulis mampu menyelesaikan skripsi ini.

Skripsi ini penulis persembahkan kepada Bapak Djoko Susanto dan Angreni W.S selaku kedua orang tua yang selalu memberikan doa dan semangat hingga saat ini.

Skripsi ini juga penulis persembahkan kepada kerabat, sahabat, dan teman-teman yang selalu membantu dan menemani hari-hari dari awal sampai akhir masa perkuliahan ini.

KATA PENGANTAR

Alhamdulillah, puji dan syukur peneliti panjatkan kepada Allah swt. yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian dengan judul “Analisis Sentimen Vaksinasi Covid-19 Menggunakan Support Vector Machine”. Shalawat serta salam peneliti curahkan kepada junjungan Nabi Muhammad SAW. beserta keluarganya, para sahabat, dan pengikutnya.

Peneliti menyelesaikan penelitian ini untuk memperoleh gelar sarjana Komputer. Dalam penyusunan, peneliti mendapatkan banyak bantuan, dukungan, dan doa sehingga penelitian ini dapat terselesaikan. Oleh karena itu, peneliti mengucapkan terima kasih kepada :

1. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
2. Ibu Windha Mega PD, M.Kom selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
3. Bapak Majid Rahardi, S.Kom., M. Eng selaku Dosen Pembimbing yang telah memberikan ilmu dan arahan selama penyusunan skripsi ini hingga akhirnya dapat terselesaikan.
4. Bapak Arif Dwi Laksito, M.Kom selaku Dosen Penguji I dan Bapak Jeki Kuswanto, M.Kom selaku Dosen Penguji II, terima kasih atas kritik dan sarannya kepada peneliti, serta bersedia meluangkan waktu untuk menguji sidang skripsi.
5. Bapak Djoko Susanto dan Ibu Angreni W.S selaku orang tua yang selalu memberikan doa dan dukungan hingga peneliti mampu menyelesaikan skripsi ini.
6. Teman-teman dan sahabat Informatika Universitas Amikom Yogyakarta khususnya angkatan 2018.
7. Seluruh pihak yang secara langsung maupun tidak langsung telah membantu, mendukung, serta mendoakan penulis dalam menyelesaikan

skripsi ini. Meski tidak tertulis namun tidak mengurangi rasa cinta dan terima kasih dari peneliti.

Penulis menyadari bahwa masih ada kesalahan dalam penyusunan skripsi ini. Maka dari itu penulis mengharapkan kritik dan saran yang membangun supaya menjadi bahan perbaikan bagi peneliti selanjutnya. Penulis juga berharap penelitian ini bermanfaat bagi siapapun yang membacanya.

Yogyakarta, 6 Januari 2022



Rizky Adhi Nugroho

INTISARI

Dalam beberapa tahun terakhir, perkembangan teknologi membuat pengumpulan data sosial dapat dilakukan melalui sosial media. Data sosial berupa tanggapan atau opini masyarakat dapat diteliti untuk berbagai keperluan. Seiring dengan berkembangnya wabah pandemi Covid-19, banyak tanggapan dan juga berita yang dibagikan melalui sosial media. Mulai dari individu, organisasi media, hingga pemerintah memberikan tanggapan mengenai kondisi terkini. Vaksinasi Covid-19 yang baru ini digalakkan pemerintah menimbulkan pro dan kontra dari berbagai kalangan.

Pada penelitian ini dilakukan analisa sentimen tentang vaksin Covid-19 menggunakan metode *Support Vector Machine*. Pengumpulan data diambil melalui twitter dengan tahun 2021 menggunakan library *Tweepy*. Setelah itu, dilakukan preprocessing data melalui *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses SMOTE juga diterapkan dalam penyeimbangan kelas dataset. Pengujian model dilakukan dengan perhitungan *precision*, *recall*, *accuracy*, dan *F-1 Score*. Selain itu, dilakukan perbandingan terhadap empat macam kernel *Support Vector Machine* dengan parameter *tuning*.

Hasil penelitian ini menunjukkan sentimen yang cenderung positif terhadap vaksinasi Covid-19 di Indonesia. Hal tersebut ditunjukkan dengan sentimen positif sebesar 56.80%, sentimen netral sebanyak 33.75%, dan sentimen negatif sebanyak 9.45%. Dalam hasil uji model, hasil evaluasi akurasi mendapatkan rata-rata sekitar 90% dengan akurasi tertinggi 92% diperoleh pada kernel RBF. Hasil akurasi dari kernel *linear* dan *polynomial* sama, yaitu 90%. Hasil akurasi lainnya diperoleh sebesar 89% pada kernel *sigmoid*.

Kata Kunci : covid-19, vaksinasi, *support vector machine*, *twitter*

ABSTRACT

For past few years, technological developments have made social data collection possible through social media. Social data in the form of public responses or opinions can be researched for various purposes. Along with the development of the Covid-19 pandemic outbreak, many responses and also news were shared through social media. Starting from individuals, media organizations, to the government gave their responses regarding the current conditions. The new Covid-19 vaccination promoted by the government has raised pros and cons from various circles.

In this study, sentiment analysis about the Covid-19 vaccine was carried out using the Support Vector Machine method. Data collection was taken via twitter with the year 2021 using the Tweepy library. After that, data is being preprocessed through cleaning, case folding, tokenizing, filtering, and stemming. Model testing is done by calculating precision, recall, accuracy, and F1-Score. Moreover, comparisons are made of four types of Support Vector Machine kernels with tuning parameters.

The results of this study indicate that sentiment tends to be positive towards Covid-19 vaccination in Indonesia. It is indicated by positive sentiment of 56.80%, neutral sentiment of 33.75%, and negative sentiment of 9.45%. For the model test results, the accuracy evaluation results get an average about 90% with the highest accuracy of 92% obtained in the RBF kernel. The results of the accuracy of the linear and polynomial kernels are the same, which is 90%. Another accuracy result is 89% on the sigmoid kernel.

Keywords : covid-19, vaccination, support vector machine, twitter

DAFTAR ISI

PERSETUJUAN	ii
PENGESAHAN	iii
KEASLIAN.....	iv
MOTTO	v
PERSEMBAHAN.....	vi
KATA PENGANTAR	vii
INTISARI.....	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Penelitian	4
1.6 Sistematika Penulisan.....	4
BAB II.....	6
2.1 Tinjauan Pustaka	6
2.2 Sentimen Analisis.....	11
2.3 Preprocessing.....	13
2.4 TF-IDF.....	15
2.5 Support Vector Machine	16
2.6 Performance Evaluation Measure.....	17
2.7 Synthetic Minority Oversampling Technique	19
BAB III.....	21
3.1 Data	21
3.2 Analisis Sistem	21

BAB IV	34
4.1 Pengambilan Data.....	34
4.2 Preprocessing.....	36
4.3 TF-IDF.....	44
4.4 Implementasi SMOTE.....	46
4.5 Data Training dan Data Testing	46
4.6 Implementasi Support Vector Machine.....	47
4.7 Evaluasi Model.....	47
BAB V.....	54
5.1 Kesimpulan.....	54
5.2 Saran.....	55
DAFTAR PUSTAKA	57

DAFTAR TABEL

Tabel 2.1 Rincian dari Penelitian Sebelumnya	8
Tabel 2.2 Confusion Matrix	19
Tabel 3.1 Contoh <i>Tweets</i> yang diambil dari <i>Twitter</i>	23
Tabel 3.2 Contoh Hasil dari Proses <i>Cleaning</i>	24
Tabel 3.3 Contoh Hasil dari Proses <i>Labelling</i>	25
Tabel 3.4 Contoh Hasil Proses <i>Case Folding</i>	26
Tabel 3.5 Contoh Hasil Proses <i>Tokenizing</i>	26
Tabel 3.6 Contoh Hasil Proses <i>Normalization</i>	27
Tabel 3.7 Contoh Hasil Proses <i>Filtering</i> atau <i>Stopwords</i>	28
Tabel 3.8 Contoh Hasil Proses <i>Stemming</i>	28
Tabel 3.9 Contoh Perhitungan Proses TF-IDF	29
Tabel 3.10 Persamaan setiap Kernel <i>Support Vector Machine</i>	31
Tabel 3.11 Contoh Penerapan <i>Confusion Matrix</i>	32
Tabel 4.1 Perbandingan sebelum dan sesudah <i>Case Folding</i>	37
Tabel 4.2 Perbandingan sebelum dan sesudah <i>Tokenizing</i>	38
Tabel 4.3 Perbandingan Sebelum dan Sesudah Proses Normalisasi	40
Tabel 4.4 Perbandingan Sebelum dan Sesudah Proses <i>Filtering</i>	42
Tabel 4.5 Perbandingan Sebelum dan Sesudah Proses <i>Stemming</i>	43
Tabel 4.6 Perbandingan tingkat akurasi dari masing-masing kernel	47
Tabel 4.7 Confusion Matrix Kernel Linear	48
Tabel 4.8 Hasil Evaluasi Performa Model Kernel <i>Linear</i>	49
Tabel 4.9 <i>Confusion Matrix</i> Kernel RBF	50
Tabel 4.10 Hasil Evaluasi Performa Model Kernel RBF	50
Tabel 4.11 Confusion Matrix Kernel Polynomial	51
Tabel 4.12 Hasil Evaluasi Performa Model Kernel <i>Polynomial</i>	52
Tabel 4.13 Confusion Matrix Kernel Sigmoid	53
Tabel 4.14 Hasil Evaluasi Performa Model Kernel <i>Sigmoid</i>	53

DAFTAR GAMBAR

Gambar 2.1 <i>Hyperplane</i> dengan Dua Dimensi	16
Gambar 3.1 Jalannya Proses Analisis Sentimen	22
Gambar 4.1 Halaman <i>Dashboard Developer Portal Twitter</i>	34
Gambar 4.2 <i>Library</i> untuk Proses <i>Crawling Tweets</i>	35
Gambar 4.3 Hasil <i>Crawling Tweets</i>	35
Gambar 4.4 Hasil <i>Cleaning</i> dan <i>Labelling</i>	36
Gambar 4.5 Hasil Proses <i>Case Folding</i>	37
Gambar 4.6 Hasil Proses <i>Tokenizing</i>	38
Gambar 4.7 Hasil Proses Normalisasi.....	39
Gambar 4.8 Hasil Proses <i>Filtering</i> atau <i>Stopwords</i>	41
Gambar 4.9 Hasil Proses <i>Stemming</i>	43
Gambar 4.10 Implementasi TF-IDF.....	44
Gambar 4.11 Hasil Proses TF-IDF pada Dokumen Dua.....	44
Gambar 4.12 Hasil Proses TF-IDF pada Dokumen Tiga.....	45
Gambar 4.13 Hasil Proses TF-IDF pada Dokumen Empat.....	45
Gambar 4.14 Implementasi SMOTE.....	46