

**SENTIMEN ANALISIS TENTANG TWEET BERISI OPINI PENYEBARAN
VIRUS COVID-19 INDONESIA DI TWITTER MENGGUNAKAN
SUPPORT VECTOR MACHINE**

SKRIPSI



disusun oleh

ADYTYO WAHYU NUGROHO

17.11.1660

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

2021

**SENTIMEN ANALISIS TENTANG TWEET BERISI OPINI PENYEBARAN
VIRUS COVID-19 INDONESIA DI TWITTER MENGGUNAKAN
SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

ADYTYO WAHYU NUGROHO

17.11.1660

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

2021

PERSETUJUAN

SKRIPSI

**SENTIMEN ANALISIS TENTANG TWEET BERISI OPINI
PENYEBARAN VIRUS COVID-19 INDONESIA DI TWITTER
MENGUNAKAN SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

ADYTYO WAHYU NUGROHO

17.11.1660

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 9 Desember 2021

Dosen Pembimbing,

Norhkmah, M.Kom
NIK. 190302245

PENGESAHAN

SKRIPSI

**SENTIMEN ANALISIS TENTANG TWEET BERISI OPINI
PENYEBARAN VIRUS COVID-19 INDONESIA DI TWITTER
MENGUNAKAN SUPPORT VECTOR MACHINE**

yang dipersiapkan dan disusun oleh

ADYTYO WAHYU NUGROHO

17.11.1660

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Desember 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yuli Astuti, M.Kom
NIK. 190302146

Rini Indrayan, ST, M.Eng
NIK. 190302417

Norhikmah, M.Kom
NIK. 190302245

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Desember 2021

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), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis & suatu insentif pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau/cara diterbitkan oleh orang lain, kecuali yang secara tertulis dicantumkan 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, 9 Desember 2021



KADITYO WAHYU NUROHO

NIM.11.11.1000

MOTTO

“Selalu percaya pada impian Anda, karena jika tidak, Anda masih akan memiliki harapan.”

- Mahatma Gandhi

“Usaha dan keberanian tidak cukup tanpa adanya tujuan dan arah perencanaan”

- John F. Kennedy

“Kesuksesan tergantung pada persiapan sebelumnya. Tanpa persiapan pasti akan gagal”

- Conficius

“Keberhasilan bukan milik orang pintar. Keberhasilan milik mereka yang terus berusaha”

- B.J.Habibie

PERSEMBAHAN

Alhamdulillah dengan kerja keras serta doa, skripsi ini dapat diselesaikan dengan baik. Segala puji dan syukur bagi Allah SWT yang tiada henti memberikan keberkahan. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung, yaitu untuk:

1. Kedua orang tua yang sudah memberikan *support* dan selalu mendoakan tiada hentinya untuk mengerjakan skripsi ini.
2. Dosen pembimbing saya Ibu Norhikmah, M.Kom. yang telah membimbing saya dari awal hingga akhir pembuatan skripsi ini.
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan ilmu selama masa kuliah.
4. Teman-teman satu bimbingan saya yang memberikan motivasi, semangat dan bantuan untuk menyelesaikan skripsi ini.
5. Teman-teman kelas 17-IF-11 yang telah menemani dan memberikan *Support* untuk menyelesaikan skripsi ini

KATA PENGANTAR

Puji syukur penulis panjatkan kepada kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shalawat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun ummat nya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul **"Sentimen Analisis tentang Tweet Berisi Opini Penyebaran Virus COVID-19 Indonesia Di Twitter Menggunakan Support Vector Machine"** ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas Amikom Yogyakarta. Penyelesaian skripsi ini juga tidak lepas dari bantuan beberapa pihak, karena itu penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. M.Suyanto, MM. Selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta,S.Kom., M.Kom. Selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya Dhuhita, M.Kom. Selaku ketua Program Studi Informatika Univeristas AMIKOM Yogyakarta.
4. Ibu Norhikmah, M.Kom. Selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat dan serta waktunya selama penulisan skripsi ini.

5. Ibu Yuli Astuti, M.Kom dan Ibu Rini Indrayani, S.T, M.Eng. Selaku dosen penguji. Terima kasih atas saran yang sudah diberikan selama pengujian untuk memperbaiki penelitian menjadi lebih baik

Penulis menyadari skripsi ini masih terdapat kekurangan. Maka, Penulis menerima kritik dan saran yang membangun serta teguran dari semua pihak. Penulis menerima dengan lapang dada untuk kesempurnaan karya selanjutnya. Semoga skripsi yang sederhana ini dapat memberikan manfaat, khususnya bagi penulis dan pembaca pada umumnya. Apabila terdapat kesalahan semoga Allah SWT melimpahkan maghfirah-Nya. *Aamiin.*

Yogyakarta, 9 Desember 2021



Adytyo Wahyu Nugroho

DAFTAR ISI

| | |
|--|----------|
| JUDUL | I |
| PERSETUJUAN..... | II |
| PERNYATAAN..... | III |
| PENGESAHAN..... | III |
| MOTTO | V |
| PERSEMBAHAN..... | VI |
| KATA PENGANTAR..... | VII |
| DAFTAR ISI..... | IX |
| DAFTAR TABEL | XII |
| DAFTAR GAMBAR..... | XIII |
| INTISARI | XVI |
| ABSTRACT..... | XVII |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 LATAR BELAKANG..... | 1 |
| 1.2 RUMUSAN MASALAH | 4 |
| 1.3 BATASAN MASALAH..... | 4 |
| 1.4 MAKSUD DAN TUJUAN PENELITIAN | 5 |
| 1.5 MANFAAT PENELITIAN..... | 5 |
| 1.6 METODE PENELITIAN | 5 |
| 1.6.1 Metode Pengumpulan Data | 5 |
| 1.6.2 Labeling Dataset..... | 6 |
| 1.6.3 Preprocessing Data..... | 6 |
| 1.6.4 Pengujian Metode Support Vector Machine..... | 6 |
| 1.6.5 Evaluasi..... | 7 |
| 1.7 SISTEMATIKA PENULISAN | 7 |
| BAB II LANDASAN TEORI | 9 |

| | | |
|--------|--|-----------|
| 2.1 | TINJAUAN PUSTAKA..... | 9 |
| 2.2 | LANDASAN TEORI..... | 15 |
| 2.2.1 | <i>Natural Language Processing</i> | 15 |
| 2.2.2 | <i>Twitter</i> | 15 |
| 2.2.3 | <i>Analisis Sentimen</i> | 16 |
| 2.2.4 | <i>Text Mining</i> | 18 |
| 2.2.5 | <i>Pre-Processing</i> | 18 |
| 2.2.6 | <i>Support Vector Machine</i> | 22 |
| 2.2.7 | <i>Seleksi Fitur (Weight By SVM)</i> | 26 |
| 2.2.8 | <i>Term Frequency-Inverse Document Frequency</i> | 27 |
| 2.2.9 | <i>Klasifikasi</i> | 28 |
| 2.2.10 | <i>TextBlob</i> | 29 |
| 2.2.11 | <i>Pelabelan Data</i> | 30 |
| 2.2.12 | <i>Validasi dan Evaluasi</i> | 30 |
| 2.2.13 | <i>Confusion Matrix</i> | 31 |
| 2.2.14 | <i>Python</i> | 33 |
| 2.2.15 | <i>Cross-Validation</i> | 33 |
| | BAB III METODE PENELITIAN..... | 34 |
| 3.1 | ALAT DAN BAHAN PENELITIAN..... | 34 |
| 3.1.1 | <i>Perangkat Keras</i> | 34 |
| 3.1.2 | <i>Perangkat Lunak</i> | 34 |
| 3.2 | ALUR PENELITIAN..... | 35 |
| 3.3 | PENGUMPULAN DATA..... | 36 |
| 3.4 | LABELING DATASET..... | 36 |
| 3.5 | PRE-PROCESSING..... | 36 |
| 3.6 | DATA TRAINING DAN DATA TEST..... | 37 |
| 3.7 | IMPLEMENTASI ALGORITME KLASIFIKASI..... | 38 |
| 3.8 | EVALUASI..... | 38 |
| | BAB IV HASIL DAN PEMBAHASAN..... | 39 |
| 4.1 | PENCARIAN TRENDING..... | 39 |

| | | |
|-----------------------------|---|-----------|
| 4.2 | PENGUMPULAN DATA | 40 |
| 4.3 | <i>DATASET</i> | 40 |
| 4.4 | PELABELAN DATASET | 42 |
| 4.5 | PREPROCESSING DATA | 44 |
| 4.6 | PEMBAGIAN DAN PEMBOBOTAN KATA DALAM DATASET | 58 |
| 4.7 | PENGUJIAN METODE SUPPORT VECTOR MACHINE KERNEL LINEAR (SVM LINEAR) | 61 |
| 4.8 | PENGUJIAN METODE SUPPORT VECTOR MACHINE KERNEL RBF (<i>RADIAL BASIS FUNCTION</i>) | 69 |
| 4.9 | PENGUJIAN METODE SUPPORT VECTOR MACHINE KERNEL POLYNOMIAL .. | 77 |
| 4.10 | EVALUASI HASIL MENGGUNAKAN <i>CONFUSION MATRIX</i> | 84 |
| BAB V PENUTUP | | 92 |
| 5.1 | KESIMPULAN | 92 |
| 5.2 | SARAN | 92 |
| DAFTAR PUSTAKA | | 93 |

DAFTAR TABEL

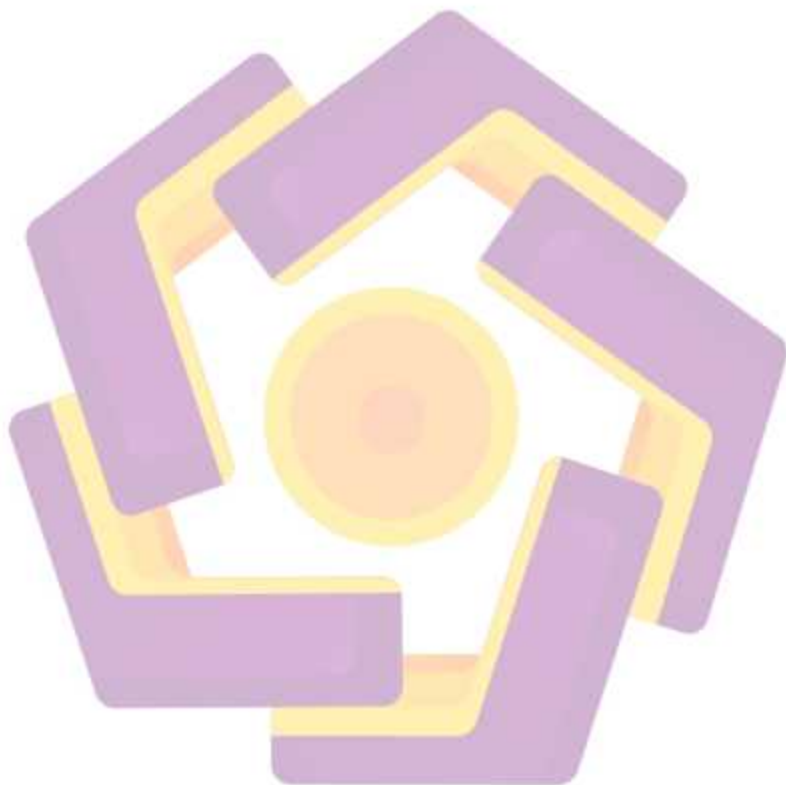
| | |
|--|----|
| Tabel 2 1 Perbandingan penelitian..... | 13 |
| Tabel 2 2 Tabel confusion matrix | 31 |
| Tabel 2 3 Tabel evaluasi matrix | 32 |
| Tabel 4 1 Ulasan Positif dan Negatif | 41 |
| Tabel 4 2 Output hasil case folding | 45 |
| Tabel 4 3 Output tokenizing..... | 49 |
| Tabel 4 4 Output stopword removal | 52 |
| Tabel 4 5 Output hasil normalisasi..... | 54 |
| Tabel 4 6 Output hasil stemming..... | 56 |
| Tabel 4 7 Akurasi menggunakan parameter | 72 |
| Tabel 4 8 Confusion matrix Linear | 86 |
| Tabel 4 9 Confusion matrix SVM RBF..... | 88 |
| Tabel 4 10 Confusion Matrix SVM Poly..... | 90 |
| Tabel 4 11 Tabel Perbandingan akurasi SVM..... | 91 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1 Formula SVM | 24 |
| Gambar 3. 1 Alur penelitian | 35 |
| Gambar 4. 1 Google Trends tentang Covid-19 Indonesia..... | 39 |
| Gambar 4. 2 Pengumpulan data..... | 40 |
| Gambar 4. 3 Dataset | 41 |
| Gambar 4. 4 Import Library..... | 42 |
| Gambar 4. 5 Labeling dataset | 43 |
| Gambar 4. 6Pie chart labeling | 43 |
| Gambar 4. 7 Hasil labeling..... | 44 |
| Gambar 4. 8Kode untuk case folding..... | 45 |
| Gambar 4. 9 Kode untuk Tokenizing..... | 47 |
| Gambar 4. 10 Kode untuk Tokenizing..... | 48 |
| Gambar 4. 11 Kode untuk menampilkan frekuensi..... | 50 |
| Gambar 4. 12 Kode untuk Stopword removal | 51 |
| Gambar 4. 13 Kode untuk Normalisasi..... | 53 |
| Gambar 4. 14 Kode untuk Stemming..... | 55 |
| Gambar 4. 15 Pembagian data..... | 58 |
| Gambar 4. 16 Kode untuk wordcloud..... | 58 |
| Gambar 4. 17 Wordcloud kata Negatif dan Positif..... | 59 |
| Gambar 4. 18 Pembagian data test dan training | 59 |
| Gambar 4. 19 Pembagian data untuk test dan training | 60 |
| Gambar 4. 20 Pembobotan kata menggunakan TF-IDF..... | 60 |
| Gambar 4. 21 Kode untuk Support vector Machine Linear | 62 |
| Gambar 4. 22 Kode untuk menampilkan hasil SVM Linear | 62 |
| Gambar 4. 23 Hasil SVM Linear | 63 |
| Gambar 4. 24 Hasil SVM Linear dan Eror Rate | 63 |
| Gambar 4. 25 Menghitung akurasi SVM Linear menggunakan parameter C | 64 |
| Gambar 4. 26 Hasil akurasi menggunakan parameter C | 64 |

| | |
|---|----|
| Gambar 4. 27 Output ROC-Curve, Time proses dan Akurasi SVM Parameter C | 65 |
| Gambar 4. 28 Nilai SVM Linear sebelum Cross-validation | 66 |
| Gambar 4. 29 Kode untuk Cross-validation..... | 66 |
| Gambar 4. 30 Hasil dari cross-validation menggunakan parameter C | 67 |
| Gambar 4. 31 Hasil GridSearchCV untuk SVM Linear | 68 |
| Gambar 4. 32 Kode SVM kernel Radial basis function | 69 |
| Gambar 4. 33 Hasil SVM kernel Radial basis function..... | 70 |
| Gambar 4. 34 Nilai akurasi dan Error rate SVM RBF | 70 |
| Gambar 4. 35 Nilai C untuk SVM RBF | 71 |
| Gambar 4. 36 Nilai Akurasi SVM RBF menggunakan parameter C | 71 |
| Gambar 4. 37 Nilai ROC-Curve untuk SVM kernel RBF | 72 |
| Gambar 4. 38 Hasil lama proses dan akurasi SVM RBF parameter C | 73 |
| Gambar 4. 39 Hasil SVM RBF sebelum cross-validation | 74 |
| Gambar 4. 40 Kode untuk melakukan cross-validation untuk kernel RBF.. | 74 |
| Gambar 4. 41 Hasil akurasi dari cross-validation SVM RBF | 75 |
| Gambar 4. 42 Hasil dari GridSearchCV untuk SVM RBF | 76 |
| Gambar 4. 43 Hasil akurasi dari SVM Polynomial | 77 |
| Gambar 4. 44 Mengukur nilai Akurasi menggunakan Degree untuk SVM Polynomial | 78 |
| Gambar 4. 45 Hasil akurasi SVM Poly menggunakan degree | 78 |
| Gambar 4. 46 Hasil ROC-Curve untuk kernel Polynomial | 79 |
| Gambar 4. 47 Hasil akurasi dan lama proses menggunakan SVM Poly | 80 |
| Gambar 4. 48 Nilai akurasi SVM poly sebelum cross-validation | 81 |
| Gambar 4. 49 Kode untuk melakukan cross-validation SVM Poly | 81 |
| Gambar 4. 50 Hasil cross-validation SVM Polynomial menggunakan degree | 82 |
| Gambar 4. 51 Hasil dari GridSearchCV Poly | 83 |
| Gambar 4. 52 Kode untuk menampilkan akurasi menggunakan Confusion matrix | 84 |
| Gambar 4. 53 Confusion matrix untuk SVM Linear | 85 |

Gambar 4. 54 Confusion matrix untuk SVM kernel RBF 87
Gambar 4. 55 Confusion matrix untuk SVM Polynomial..... 89



INTISARI

Perkembangan teknologi informasi yang masif ini mempermudah dalam kehidupan masyarakat dalam berbagai bidang, salah satunya adalah sosial media, Sosial media yang masyarakat gunakan banyak mendapatkan informasi mengenai berita atau peristiwa yang sedang terjadi di Indonesia, salah satunya adalah sosial media *Twitter* yang banyak memberikan informasi untuk masyarakat Indonesia salah satunya adalah informasi mengenai *Covid-19* yang sedang marak terjadi di wilayah Indonesia. Analisis sentimen merupakan salah satu cabang dari *Natural Language Processing (NLP)* yang dapat membantu mengetahui sentimen yang terjadi pada masyarakat. Penelitian ini menggunakan data berupa *tweet* untuk melakukan sentimen analisis yang didapat pada sosial media *Twitter*.

Penelitian ini memanfaatkan salah satu algoritme dari *Supervised Learning* yaitu *Support Vector Machine* dalam penelitian ini menggunakan tiga (3) kernel untuk *Support Vector Machine* masing masing adalah *Linear*, *Radial basis function* dan *Polynomial* untuk mencari pada kernel berapa menghasilkan nilai akurasi tertinggi.

Dari percobaan yang dilakukan dengan menggunakan pembagian data untuk *training* sebanyak 70% dan untuk data *testing* sebanyak 30% dari total data 6000 data, dihasilkan nilai akurasi untuk metode *Support Vector Machine* pada kernel *Linear* menghasilkan nilai akurasi sebesar 89% dan untuk kernel *Radial basis function* akurasi sebesar 90% dan untuk kernel *Polynomial* menghasilkan akurasi sebesar 88%. Sehingga disimpulkan untuk ke-tiga (3) kernel untuk pengujian metode *Support Vector Machine* pada kernel *Radial basis function* menghasilkan nilai akurasi terbaik.

Kata Kunci: Sentimen Analisis, *Support Vector Machine (SVM)*, *Twitter*, *Covid-19*.

ABSTRACT

This massive development of information technology makes it easier for people's lives in various fields, one of them is social media, social media that people use a lot to get information about news or events that are happening in Indonesia, one of which is social media Twitter which provides a lot of information for the people of Indonesia, one of which is information about Covid-19 which is currently rife in the territory of Indonesia Sentiment analysis is a branch of Natural Language Processing (NLP) which can help determine the sentiments that occur in society. This study uses data in the form of tweets to carry out sentiment analysis obtained on Twitter social media.

This research utilizes one of the Supervised Learning algorithms, namely Support Vector Machine. In this study, three (3) kernels are used for the Support Vector Machine, each of which is Linear, Radial basis function and Polynomial, to find which kernel produces the highest accuracy value.

From the experiments carried out using data sharing for training as much as 70% and for testing data as much as 30% of the total data of 6000 data, the resulting accuracy value for the Support Vector Machine method on the Linear kernel produces an accuracy value of 89% and for the Radial kernel base function accuracy by 90% and for the Polynomial kernel it produces an accuracy of 88%. So it is concluded for the three (3) kernels for testing the Support Vector Machine method on the Radial basis function kernel to produce the best accuracy value.

Keyword: *Sentiment Analysis, Support Vector Machine (SVM), Twitter, Covid-19.*