

**ANALISIS METODE PEMILIHAN FITUR PADA ANALISIS
SENTIMEN TERKAIT ISU VAKSINASI COVID-19**



TESIS

MUHAMMAD FAJAR

ENTERPRISE IT INFRASTRUCTURE

192420037

PROGRAM STUDI TEKNIK INFORMATIKA – S2

PROGRAM PASCASARJANA

UNIVERSITAS BINA DARMA

PALEMBANG

2023

**ANALISIS METODE PEMILIHAN FITUR PADA ANALISIS
SENTIMEN TERKAIT ISU VAKSINASI COVID-19**

**Tesis ini diajukan sebagai salah satu syarat
untuk memperoleh gelar**

MAGISTER KOMPUTER



MUHAMMAD FAJAR

ENTERPRISE IT INFRASTRUCTURE

192420037

PROGRAM STUDI TEKNIK INFORMATIKA – S2

PROGRAM PASCASARJANA

UNIVERSITAS BINA DARMA

PALEMBANG

2023

Halaman Pengesahan Pembimbing Tesis

Judul : ANALISA METODE PEMILIHAN FITUR PADA ANALISIS SENTIMEN TERKAIT ISU VAKSINASI COVID-19

Oleh MUHAMMAD FAJAR NIM 192420037 Tesis ini telah disetujui dan
disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi
ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina
Darma pada 31 Maret 2023 dan telah dinyatakan LULUS

Mengetahui,
Program Studi Teknik I
Universitas Bina Darna
Ketua,

Palembang, 2023

Pembimbing,



Zaid Amin, M.Kom., Ph.D

Tri Basuki Kurniawan, S.Kom., M.Eng., Ph.D

Halaman Pengesahan Penguji Tesis

Judul : ANALISA METODE PEMILIHAN FITUR PADA ANALISIS
SENTIMEN TERKAIT ISU VAKSINASI COVID-19

Oleh MUHAMMAD FAJAR NIM 192420037Tesis ini telah disetujui dan
disahkan oleh Tim Penguji Program Studi Teknik Informatika – S2 konsentrasi
ENTERPRISE IT INFRASTRUCTURE, Program Pascasarjana Universitas Bina
Darma pada 31 Maret 2023 dan telah dinyatakan LULUS

Palembang, 31 Maret 2023

Mengetahui,

Program Pascasarjana

Universitas Bina Darma

Direktur,

Tim Penguji :

Penguji I,



Prof. Hj. Isnawijayani, M.Si., Ph.D.

Tri Basuki Kurniawan, S.Kom., M.Eng., Ph.D.

Penguji II,

M. Izman Herdiansyah, M.M., Ph.D.

Penguji III,

Dr. Edi Surya Negara, M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini

Nama : MUHAMMAD FAJAR
NIM : 192420037

Dengan ini menyatakan bahwa:

1. Karya tulis Saya Tesis ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik Magister di Universitas Bina Darma;
2. Karya tulis ini murni gagasan, rumusan dan penelitian Saya sendiri dengan arahan tim pembimbing;
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar pustaka;
4. Karena yakin dengan keaslian karya tulis ini, Saya menyatakan bersedia Tesis yang Saya hasilkan di unggah ke internet;
5. Surat Pernyataan ini Saya tulis dengan sungguh-sungguh dan apabila terdapat penyimpangan atau ketidakbenaran dalam pernyataan ini, maka Saya bersedia menerima sanksi dengan aturan yang berlaku di perguruan tinggi ini.

Demikian Surat Pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 31 Maret 2023
Yang Membuat Pernyataan,



MUHAMMAD FAJAR
NIM : 192420037

ABSTRAK

Analisis sentimen atau penambangan opini adalah studi komputasi tentang pendapat, sentimen, evaluasi, sikap, suasana hati, dan emosi seseorang. Analisis sentimen adalah salah satu bidang penelitian paling aktif dalam pemrosesan bahasa alami, penambangan data, pencarian informasi, dan penambangan web. Salah satu masalah yang teridentifikasi dalam proses analisis sentimen adalah banyaknya data atau properti teks. Dalam analisis sentimen, setiap kata atau istilah dikumpulkan menjadi properti atau dimensi, membentuk tabel data. Karena banyaknya istilah, hal ini menyebabkan proses menjadi terlalu lama dan membutuhkan komputer dengan kekuatan atau kemampuan yang luar biasa. Selain itu, hal ini dapat menyebabkan penurunan kualitas model karena data yang terlalu besar juga akan memberikan nilai bias yang signifikan. Tidak semua term memiliki kontribusi atau hubungan terhadap keputusan atau label berupa nilai positif, negatif, dan netral. Untuk itu metode seleksi fitur akan digunakan dalam penelitian ini untuk menyeleksi fitur atau istilah yang lebih berkontribusi terhadap keputusan atau label. Hal ini juga diharapkan dapat meningkatkan kualitas model prediksi yang akan dibentuk. Pada penelitian ini, penulis akan melanjutkan penelitian dari peneliti lain dengan menambahkan proses seleksi fitur, seperti dua algoritma dari metode terfilter, *chi-square*, dan *information gain*, serta satu algoritma dari metode terbungkus (*wrapped*) yaitu Algoritma Genetika (GA). Hasil percobaan menunjukkan bahwa hasil GA yang diperoleh memiliki nilai akurasi paling tinggi dibandingkan dengan metode lainnya.

Kata Kunci : Analisis Sentimen, Pemilihan Fitur, Metode Filtered, Metode Wrapped

ABSTRACT

Sentiment analysis or opinion mining is a computational study of a person's opinions, sentiments, evaluations, attitudes, moods, and emotions. Sentiment analysis is one of the most active research areas in natural language processing, data mining, information retrieval, and web mining. One of the problems identified in the sentiment analysis process is the massive amount of data or text properties. In sentiment analysis, each word or term is collected into properties or dimensions, forming a data table. Due to the vast number of terms, this causes the process to take too long and requires a computer with tremendous power or ability. In addition, this can lead to a decrease in the quality of the model because data that is too large will also provide a significant bias value. Not all terms have contributions or relationships to decisions or labels in the form of positive, negative, and neutral values. For this reason, the feature selection method will be used in this study to select features or terms that contribute more to decisions or labels. It is also hoped that this can increase the quality of the prediction model that will be formed. In this study, the author will continue the research from another researcher by adding a feature selection process, such as two algorithms from the filtered method, chi-square, and information gain, and one algorithm from the wrapped method, which is Genetic Algorithms (GA). The experiment result shows that the GA obtained result has the highest accurate value compared to the other methods.

Keywords : *Sentiment Analysis, Feature Selection, Filtered Method, Wrapped Method*

MOTTO DAN HALAMAN PERSEMPAHAN

فَلَمَّا يَأْتِ إِلَيْهِ رِبُّكُمْ أُتَكَذِّبُونَ ﴿١٣﴾

Maka nikmat Tuhanmu yang manakah yang kamu dustakan ?

(QS: Ar-Rahman 13)

*“Jangan pernah menyerah sebelum mencoba, tetap berusaha semampunya,
lakukan yang terbaik jika ingin ada perubahan dihidupmu”*

(Penulis)

Karya besar ini kupersembahkan kepada :

- Allah SWT atas anugrah, hidayah, inayah dan maghfirahmu
- Ayahku Syafruddin Nasih (Alm) dan Ibuku Fatimah yang tercinta
- Saudara saudariku serta keluarga besar-ku
- Civitas akademika Magister Teknik Informatika Universitas Bina Darma
- Almamaterku

DAFTAR RIWAYAT HIDUP



Nama : Muhammad Fajar
Tempat Tanggal Lahir : Palembang, 31 Juli 1994
Jenis Kelamin : Laki-Laki
Agama : Islam
Status : Belum Menikah
Alamat : Perum. Bukit Sejahtera Blok Ab No.10, Palembang
Email : fajarcrzcrz@gmail.com

Riwayat Pendidikan Formal :

- Tahun 2000 – 2006 : SD Negeri 12 Palembang
- Tahun 2006 – 2009 : SMP Negeri 18 Palembang
- Tahun 2009 – 2012 : SMA Negeri 1 Palembang
- Tahun 2012 – 2018 : Sistem Komputer Universitas Sriwijaya
- Tahun 2019 – 2023 : Magister Teknik Informatika Universitas Bina Darma *

Riwayat Pendidikan Non Formal :

- Internship di PT. PLN (Persero) WS2JB (2015)
- Pelatihan Nilai-Nilai Olahraga 18th Asian Games Jakarta - Palembang 2018 (INASGOC)
- E-Course Mahir Desain Angkatan Ke-08 (2019)
- Pelatihan Jurnalistik (Sumeks) (2019)
- Program Fresh Graduate Academy Digital Talent Scholarship Tahun 2019 Tema Cyberscurity (Cisco), Softskill Development And Entrepreneurship
- Program Online Academy Digital Talent Scholarship Tahun 2020 Tema Programming: HTML, CSS, Javascript (Progate)
- Workshop Machine Learning Program Studi Teknik Informatika – S2 “Pemanfaatan Teknologi di Era Pandemi Covid-19” (2021)

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillah, atas segala nikmat yang diberikan oleh Allah SWT yang selalu memberikan berkah, rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul Penerapan Metode Pemilihan Fitur pada Analisis Sentimen terkait Isu Vaksinasi Covid-19.

Proposal tesis disusun sebagai salah satu syarat untuk memperoleh gelar Magister Teknik Informatika pada Universitas Bina Darma Palembang. Dalam penulisan tesis ini penulis telah melakukan semaksimal mungkin memberikan dan menyajikan yang terbaik. Tetapi Penulis juga menyadari bahwa proposal tesis ini masih jauh dari sempurna, hal ini dikarenakan masih terbatasnya pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun untuk kesempurnaan proposal tesis ini.

Pada kesempatan ini, tidak lupa penulis mengucapkan terimakasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasihat, dan pemikiran dalam menyelesaikan proposal tesis ini, terutama kepada :

1. **Dr. Sunda Ariana, M.Pd., M.M.**, Selaku Rektor Universitas Bina Darma Palembang.
2. **Prof. Hj. Isnawijayani, M.Si., Ph.D.** Selaku Direktur Program Pascasarjana Universitas Bina Darma Palembang.
3. **Zaid Amin, M.Kom., Ph.D.** selaku ketua Program Studi Magister Teknik Informatika Universita Bina Darma Palembang

4. **Tri Basuki Kurniawan, S.Kom., M.Eng., Ph.D.** selaku pembimbing yang telah memberikan bimbingan, ilmu, serta arahan dalam penulisan tesis ini.
5. **Pihak Sekretariat Pascasarjana Universitas Bina Darma Palembang** yang telah memberikan bimbingan pelayanan dengan baik
6. **Orang Tuaku**, terima kasih telah selalu memberikan dukungan dan doa setiap langkah saya sampai dengan hari ini.
7. **Rahma Ricadonna**, terima kasih selalu memberikan semangat, motivasi, dan doa dalam penyelesaian tesis ini.
8. **Rekan-rekan Bimbingan Pak Tri Basuki dan Angkatan MTI 21**, terima kasih bantuan serta infomasinya selama ini.

Palembang, 31 Maret 2023

Penulis

**Muhammad Fajar
192420037**

DAFTAR ISI

HALAMAN DEPAN	i
Halaman Pengesahan Pembimbing Tesis	ii
Halaman Pengesahan Penguji Tesis	iii
SURAT PERNYATAAN.....	iv
ABSTRAK	v
ABSTRACT	vi
MOTTO DAN HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah.....	4
1.3 Rumusan Masalah.....	5
1.4 Batasan Masalah.....	5
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian.....	6
1.7 Susunan dan Struktur Proposal Tesis	7
BAB II.....	9
2.1 Data Mining.....	9
2.2 <i>Text Mining</i>	10
2.2.1 Text Preprocessing	11
2.3 Klasifikasi	12
2.3.1 Bayes Classifier (NBC)	13
2.3.2 Logistic Regression	15
2.3.3 Random Forest	16
2.3.4 Support Vector Machine.....	16
2.4 Pemilihan Fitur	17

2.5	Penelitian Terdahulu.....	21
BAB III		26
3.1	Tahapan Penelitian	26
3.2	Tahap Perencanaan	27
3.3	Tahap Pengumpulan Data.....	27
3.4	Tahap Pengelolaan Data	28
3.5	Jadwal Pelaksanaan Penelitian.....	28
BAB IV		30
4.1	Teknik Analisis Data	30
4.2	Pengumpulan Data.....	30
4.3	Pengolahan Data.....	32
4.3.1	Pembersihan Data.....	32
4.3.2	Analisis Sebaran Label dan Cloud Word.....	34
4.3.3	Pembobotan dan Pembentukan Matrik	36
BAB V.....		38
5.1	Proses dan Hasil Pembentukan Model dan Pembahasan.....	38
5.1.1	Pemilihan Fitur Metode Fiter	45
5.1.2	Pemilihan Fitur Metode Wrapped	49
BAB VI		53
6.1	Kesimpulan	53
6.2	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN		59

DAFTAR GAMBAR

Gambar 2.1 Proses <i>text mining</i> (Even-Zohar, 2002).....	11
Gambar 2.2 Klasifikasi <i>Naive Bayes Classifier</i> (Harrington, 2012).....	14
Gambar 2.3 Klasifikasi <i>Support Vector Machine</i> (Rani et al., 2022)	17
Gambar 3.1 Tahap Penelitian.....	26
Gambar 4.1 Data asli. Sumber: Anggraini, Negara Harahap, & Kurniawan (2021). ...	31
Gambar 4.2 Data yang akan digunakan dalam penelitian ini	32
Gambar 4.3 Kode <i>Pre-Processing</i>	33
Gambar 4.4 Cara menggunakan library Sastrawi.....	33
Gambar 4.5 Coding untuk menghapus baris kosong.....	33
Gambar 4.6 Hasil <i>pre-processing</i>	34
Gambar 4.7 Sebaran label data pada dataset.....	35
Gambar 4.8 Visualisasi menggunakan library wordcloud dari teks.	36
Gambar 4.9 Coding untuk proses pembobotan menggunakan SVD.....	37
Gambar 5.1 Pembagian data training dan data testing.	39
Gambar 5.2 Sebaran label untuk data training.....	39
Gambar 5.3 Sebaran label untuk data testing.	40
Gambar 5.4 Kode python pembentukan model, prediksi dan pengukuran akurasi.	41
Gambar 5.5 Hasil proses pengukuran akurasi model Naïve Bayes.	41
Gambar 5.6 <i>Confusion Matrik</i> dari hasil model Naïve Bayes.	41
Gambar 5.7 <i>Confusion matrix</i> dari hasil model Logistic Regression dan SVM..	43
Gambar 5.8 <i>Confusion Matrix</i> dari hasil model Random Forest.	44
Gambar 5.9 Kode Python untuk menghitung nilai <i>chi-square</i> dan <i>information gain</i> . .	45

Gambar 5.10 Graph garis nilai akurasi untuk setiap algoritma pada jumlah fitur. 46

Gambar 5.11 Graph garis nilai akurasi berdasarkan pendekatan *information gain*. 49

Gambar 5.12 *Convergence Curve* untuk hasil akurasi metode *wrapper*..... 51



DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	22
Tabel 3.1 Jadwal Pelaksanaan Penelitian.....	29
Tabel 5.1 Nilai akurasi dan waktu proses model dari 4 algoritma.....	44
Tabel 5.2 Nilai akurasi dan waktu proses untuk <i>chi-square</i> model.	46
Tabel 5.3 Nilai akurasi dan waktu proses untuk <i>information gain</i> model	48
Tabel 5.4 Hasil akurasi dalam metode <i>wrapper</i>	50

DAFTAR LAMPIRAN

- SK Pembimbing
- Lembar Konsultasi Tesis
- Jurnal Seminar & Sertifikat Seminar
- Lembar Perbaikan Seminar