

**ANALISIS SENTIMEN VAKSINASI COVID-19 PADA *TWITTER*
MENGUNAKAN METODE
NAIVE BAYES CLASSIFER DAN ALGORITMA *GENETIKA***



SKRIPSI

HIDYATUN KHASANAH

1803040061

**TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
NOVEMBER 2022**

**ANALISIS SENTIMEN VAKSINASI COVID-19 PADA TWITTER
MENGUNAKAN METODE
NAIVE BAYES CLASSIFER DAN ALGORITMA GENETIKA**



SKRIPSI

**Diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer**

HIDYATUN KHASANAH

1803040061

**TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN SAINS
UNIVERSITAS MUHAMMADIYAH PURWOKERTO
NOVEMBER 2022**

HALAMAN PERSETUJUAN

Skripsi yang diajukan oleh:

Nama : Hidayatun Khasanah

NIM : 1803040061

Program Studi : Teknik Informatika

Fakultas : Teknik dan Sains

Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

Judul : Analisis Sentimen Vaksinasi Covid-19 Pada Twitter

Menggunakan Metode *Naive Bayes Classifier(NBC)* dan
Algoritma *Genetika*

Telah disetujui untuk diajukan dalam seminar hasil

Purwokerto, November 2022

PEMBIMBING


Muhammad Hamka, S.T., M.Kom

NIK.2160517

HALAMAN PENGESAHAN

Skripsi yang diajukan oleh:

Nama : Hidyatun Khasanah
 NIM : 1803040061
 Program Studi : Teknik Informatika
 Fakultas : Teknik dan Sains
 Perguruan Tinggi : Universitas Muhammadiyah Purwokerto
 Judul : Analisis Sentimen Vaksinasi Covid-19 Pada Twitter
 Menggunakan Metode *Naïve Bayes Classifier* (NBC)
 dan Algoritma *Genetika*

Telah berhasil dipertahankan di hadapan dewan penguji dan diterima sebagai bagian
 persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer (S.Kom)
 pada Program Studi Teknik Informatika, Fakultas Teknik dan Sains, Universitas
 Muhammadiyah Purwokerto.

DEWAN PENGUJI

Penguji 1 (Pembimbing) : Muhammad Hamka, S.T., M.Kom.
 Penguji 2 : Feri Wibowo, S.Kom., M.Cs.
 Penguji 3 : Hindayati Mustafidah, S.Si M.Kom.,

Ditetapkan di : Purwokerto
 Tanggal :



Mengetahui
 Dekan Fakultas Teknik dan Sains
 Ir. Teguh Marnendi, S.T., M.T., ASEAN.Eng., ACPE, IPM
 NIK. 2160172

HALAMAN PERNYATAAN ORISINILITAS

Proposal Skripsi yang diajukan oleh:

Nama : Hidyatun Khasanah
NIM : 1803040061
Program Studi : Teknik Informatika
Fakultas : Teknik dan Sains
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

Menyatakan dengan benar-benarnya bahwa skripsi ini adalah karya saya dan semua sumber bail yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta bukan hasil penjiplakan dari karya orang lain.

Demikian pernyataan ini saya buat dan apabila kelak dikemudian hari terbukti ada unsur penjiplakan, saya bersedia mempertanggungjawabkan sesuai dengan ketentuan yang berlaku.

Purwokerto, November 2022

membuat pernyataan



Hidyatun Khasanah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Muhammadiyah Purwokerto dan demi pengembangan ilmu pengetahuan, saya yang bertandatangan dibawah ini:

Nama : Hidyatun Khasanah
NIM : 1803040061
Program Studi : Teknik Informatika
Fakultas : Teknik dan Sains
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto
Hasil Karya : Skripsi

Menyetujui untuk memberikan Hak Bebas Royalti Noneklusif (Non-exclusive Royalty-Free Right) kepada Universitas Muhammadiyah Purwokerto atas karya ilmiah saya yang berjudul:

Analisis Sentimen Vaksinasi Covid-19 Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier (NBC)* dan Algoritma *Genetika*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Muhammadiyah Purwokerto berhak menyimpan, mengalih media/ mengalih formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya dengan tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sebenarnya.

Dibuat : Purwokerto

Pada Tanggal :



ig Menyatakan

Hidyatun Khasanah

HALAMAN PERSEMBAHAN

Puji syukur saya panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penyusunan skripsi yang berjudul “Analisis Sentimen Vaksinasi Covid-19 Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier(NBC)* dan Algoritma *Genetika*” dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana Komputer. Penulis menyadari dalam penyelesaian skripsi ini banyak pihak yang terlibat langsung maupun tidak langsung atas selesainya skripsi ini, oleh karena itu penulis mengucapkan terima kasih kepada:

1. Bapak Feri Wibowo, S.Kom., M.Cs., selaku Ketua Program Studi Teknik Informatika Universitas Muhammadiyah Purwokerto.
2. Bapak Muhammad Hamka, S.T., M.Kom., selaku dosen pembimbing yang telah memberikan kontribusi berupa bimbingan, motivasi, kritik dan saran atas terselesaikannya skripsi ini.
3. Terimakasih kepada dosen penguji yang telah memberikan kritik dan saran berharganya demi terselesaikannya skripsi ini.
4. Semua dosen program studi Teknik Informatika, yang tidak bisa disebutkan. Terimakasih atas ilmunya, semoga bisa saya amalkan dikemudian hari.
5. Untuk Keluarga Tercinta, bapak Suwanto, mama Tasmiami, mba Hikmah, mba Hidayah, adiku Assyfa, dan adeku Suhail yang ganteng sendiri yang sudah menjadi motivasi terbesar untuk menyelesaikan skripsi dan selalu menyirami kasih sayang, semangat dan selalu mendoakanku.
6. Untuk semua teman-teman seperjuangan Teknik Informatika.
7. Serta semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu dalam menyelesaikan laporan tugas akhir ini.

HALAMAN MOTTO

“Tidak ada yang tidak mungkin selagi berusaha dan berdo’a”



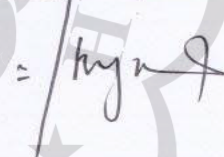
KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya kepada penulis, sehingga penulis bisa menyelesaikan skripsi ini yang berjudul “Analisis Sentimen Vaksinasi Covid-19 Pada *Twitter* Menggunakan Metode *Naïve Bayes Classifier(NBC)* dan Algoritma *Genetika*”. Shalawat dan juga salam senantiasa tercurah kepada junjungan kita semua Nabi Muhammad SAW yang mengantarkan manusia dari kegelapan ke zaman yang terang benderang.

Laporan skripsi ini disusun untuk memperoleh gelar Sarjana Komputer pada program studi Teknik Informatika Fakultas Teknik dan Sains Universitas Muhammadiyah Purwokerto.

Penulis menyadari atas ketidaksempurnaan penyusunan laporan skripsi ini. Namun penulis tetap berharap laporan ini akan memberikan manfaat bagi para pembaca. Demi kemajuan penulis, penulis juga mengharapkan adanya masukan berupa kritik atau saran yang berguna.

Yang Menyatakan



Hidyatun Khasanah

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN ORISINILITAS	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIK.....	vi
HALAMAN PERSEMBAHAN.....	vii
HALAMAN MOTTO	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	x
DAFTAR LAMPIRAN.....	xv
INTISARI.....	xvi
ABSTRACT.....	xvii
BAB 1 PENDAHULUAN	1
A. Latar Belakang.....	1
B. Perumusan Masalah.....	3
C. Batasan Masalah.....	4
D. Tujuan Penelitian.....	4
E. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
A. Penelitian Terdahulu.....	5
B. Landasan Teori	9
1. Analisis Sentimen	9
2. Vaksinasi Covid-19	9
3. Twitter	10
4. Teks Mining.....	10
5. TwitterAPI (Application Programming Interface).....	11
6. Crawling	11
7. Machine Learning.....	12
8. NLTK	13

10. Algoritma Genetika	15
11. Python.....	18
12. SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	19
BAB III METODE PENELITIAN	20
A. Jenis Penelitian	20
B. Waktu dan Tempat	20
C. Metode Pengumpulan Data	20
D. Metode Analisis Data	21
1. Pengumpulan Data.....	22
2. Pre-processing	23
3. Pelabelan Data.....	24
4. <i>Ekstraksi Fitur Term Weighting (TF-IDF)</i>	24
5. SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	26
6. <i>Splitting Data</i>	26
7. <i>Naïve Bayes Multinomial</i>	26
8. Algoritma Genetika	27
9. Visualisasi	28
10. Hasil Akurasi Kedua Model.....	28
BAB IV HASIL DAN PEMBAHASAN	30
A. Data Penelitian.....	30
B. Preprocessing.....	31
C. Pelabelan Data	37
D. <i>Ekstraksi Fitur Term Weighting(TF-IDF)</i>	39
E. SMOTE For Imbalance Data	44
F. Splitting Data.....	45
G. Naïve Bayes Multinomial.....	46
E. Algoritma Genetika.....	48
F. Visualisasi	62
H. Hasil Akurasi Kedua Model.....	66
BAB V KESIMPULAN DAN SARAN.....	68
A. KESIMPULAN	68
B. SARAN.....	68

DAFTAR PUSTAKA 69
LAMPIRAN 72



DAFTAR TABEL

Tabel 4.1 Contoh Hasil <i>Crawling Data Opini Vaksinasi Covid19</i>	31
Tabel 4.2 Contoh Data <i>Tweet</i>	32
Tabel 4.3 Hasil <i>Cleaning Text</i>	33
Tabel 4.6 Hasil Proses <i>Stemming</i>	36
Tabel 4.7 Hasil Proses <i>Remove Stopword</i>	37
Tabel 4.8 Hasil Pelabelan Data <i>Twitter</i>	38
Tabel 4.10 Contoh Hasil Nilai TF dan Normalisasi pada 4 Buah Dokumen	41
Tabel 4.11 Contoh Hasil Perhitungan IDF dan TF-IDF pada 4 buah Dokumen	43
Tabel 4.13 Inisialisasi Populasi	55
Tabel 4.14 Inisialisasi Jumlah Gen	55
Tabel 4.15 Representasi Kromosom	56
Tabel 4.16 Hasil Perbandingan Nilai Fitnes Iterasi 1 dan Iterasi 2	62
Tabel 4.18 Contoh WordCloud Negatif	65
Tabel 4.19 Contoh WordCloud Netral	66

DAFTAR GAMBAR

Gambar 2.1 Pindah Silang Satu Titik.....	17
Gambar 3. 1 Alur Penelitian	22
Gambar 4.1 Presentase Data <i>Tweet</i>	38
Gambar 4.3 Hasil SMOTE <i>For Imbalance</i> Sentimen <i>Tweet</i>	45
Gambar 4.4 Hasil Klasifikasi Teks	46
Gambar 4.5 Hasil <i>Confusion Matrix</i> dari <i>Naïve Bayes Multinomial</i>	47
Gambar 4.6 Hasil <i>Accuracy, Precision, Recall</i> dan <i>F1-score</i>	48
Gambar 4.7 Hasil Nilai Fitnes 30 <i>Population</i> dan 20 <i>Generation</i>	49
Gambar 4.8 Hasil Nilai Fitnes 50 <i>Population</i> dan 20 <i>Generation</i>	50
Gambar 4.9 Hasil Nilai Fitnes 30 <i>Population</i> dan 30 <i>Generation</i>	51
Gambar 4.10 Plot Fitnes <i>Population</i> 50 dan 20 <i>Generation</i>	52
Gambar 4.11 Hasil Optimasi <i>NB_classifier-GA</i>	53
Gambar 4.12 Hasil <i>Accuracy, Precision, Recall</i> dan <i>F1-score</i>	54
Gambar 4.13 Visualisasi <i>Wordcloud Tweet</i> Positif	63
Gambar 4.14 Visualisasi <i>Wordcloud Tweet</i> Negatif.....	64
Gambar 4.15 Visualisasi <i>Wordcloud Tweet</i> Netral.....	65
Gambar 4.16 Perbandingan Akurasi Kedua Model	66

DAFTAR LAMPIRAN

LAMPIRAN 1. Hasil <i>Crwaling</i> data menggunakan software postman.....	72
LAMPIRAN 2. Kode Program.....	73



INTISARI

Program melaksanakan vaksinasi COVID-19 sudah menjadi kewajiban untuk rakyat Indonesia. Hal tersebut berpengaruh terhadap rakyat Indonesia yang memberikan sentimen untuk program vaksin di Indonesia. Masyarakat Indonesia biasanya berkomentar melalui media sosial seperti *twitter dan facebook*. Pada penelitian ini lebih fokus pada media sosial *Twitter*. Banyaknya jumlah data sentimen yang dapatkan dalam penelitian sebanyak 16.291 *tweet* dengan menggunakan 2 kunci yaitu “*vaksinasicovid19*” dan “*vaksincovid19*”. Data kemudian di kategorikan berdasarkan nilai kalkulasi menggunakan *lexicon* menghasilkan sentimen positif sebesar 63,7%, sentimen negatif sebesar 26,2% dan sentimen netral sebesar 10,1%. *Klasifikasi Naïve Bayes* yang digunakan yaitu klasifikasi teks *Naïve Bayes Multinomial* dan Algoritma *Genetika* dengan fungsi *Feature Selection*. Sebelum diklasifikasi data dilakukan Pembobotan dengan *TF-IDF*, *Smote For Ibalance* data dan data *tweet* di bagi menjadi 2 bagian yaitu data testing sebesar 20% dan data latih sebesar 80%. Hasil klasifikasi menggunakan *Naïve Bayes Multinomial* akurasi sebesar 82% dan hasil optimasi *GA feature selection* menghasilkan akurasi 82%. Dalam penelian ini pembahasan dan temuan yang dilakukan pengujian *Naïve Bayes* dioptimasi dengan *GA featur selection* menghasilkan akurasi yang paling optimum sebesar 82%.

Kata kunci: Vaksinasi covid-19, Sentimen, *Naïve Bayes*, *GA*, *TF-IDF*, *Smote*, Akurasi.

ABSTRACT

The program to carry out the COVID-19 vaccination has become an obligation for the people of Indonesia. This affects the Indonesian people who provide sentiment for the vaccine program in Indonesia. Indonesians usually comment via social media. Commonly used social media are Twitter and Facebook. This study focuses more on Twitter social media. The number of sentiment data obtained in the study was 16,291 tweets using 2 keys, namely "vaccinationcovid19" and "vaksinCovid19". The data was then categorized based on the calculatory value using lexicon resulting in positive sentiment of 63.7%, negative sentiment of 26.2% and neutral sentiment of 10.1%. The Naïve Bayes classification used is the Naïve Bayes Multinomial text classification and Genetic Algorithm with the Feature Selection function. Before classifying the data carried out Weighting with TF-IDF, Smote For Ibalance data and tweet data are divided into 2 parts, namely testing data of 20% and training data of 80%. The classification results using Naïve Bayes Multinomial accuracy of 82% and GA feature selection optimization results in 82% accuracy. In this study, the discussion and findings carried out by the Naïve Bayes test were optimized with GA featur selection resulting in the most optimum accuracy of 82%.

Keywords: Covid-19 vaccination, Sentiment, Naïve Bayes, GA, TF-IDF, Smote, Accuracy.