



**ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA
PADA MEDIA SOSIAL *YOUTUBE* MENGGUNAKAN
METODE NAÏVE BAYES DAN SELEKSI FITUR PARTICLE
SWARM OPTIMIZATION**

SKRIPSI

TAUFIK ADI PRASETYO

1810511076

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022



**ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA
PADA MEDIA SOSIAL *YOUTUBE* MENGGUNAKAN
METODE NAÏVE BAYES DAN SELEKSI FITUR PARTICLE
SWARM OPTIMIZATION**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

TAUFIK ADI PRASETYO

1810511076

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2022

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar,

Nama : Taufik Adi Prasetyo
NIM : 1810511076
Tanggal : 20 Juni 2022

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan say ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 24 Juni 2022

Yang Menyatakan,



(Taufik Adi Prasetyo)

PERNYATAAN PERSETUJUAN PUBLIKASI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta saya yang bertandatangan di bawah ini:

Nama : Taufik Adi Prasetyo
NIM : 1810511076
Fakultas : Ilmu Komputer
Program Studi : S1 Informatika

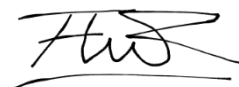
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non-Exclusive Royalty Free Right) atas karya ilmiah saya yang berjudul:

ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA PADA MEDIA SOSIAL *YOUTUBE* MENGGUNAKAN METODE NAÏVE BAYES DAN SELEKSI FITUR PARTICLE SWARM OPTIMIZATION

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan kata (Basis data), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada Tanggal : 24 Juni 2022

Yang Menyatakan,



(Taufik Adi Prasetyo)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut :

Nama : Taufik Adi Prasetyo
NIM : 1810511076
Program Studi : S1 Informatika
Judul Skripsi : ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA
PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN METODE
NAÏVE BAYES DAN SELEKSI FITUR PARTICLE SWARM
OPTIMIZATION

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



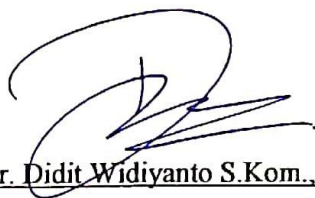
Yuni Widiastiwi, S.Kom., M.Si.

Penguji I



Nurul Chamidah, S.Kom., M.Kom.

Penguji II




Dr. Didit Widiyanto S.Kom., M.Si.

Pembimbing I



Desta Sandya Prasvita, S.Kom., M.Kom.

Pembimbing II



Dr. Ermatita, M. Kom.

Dekan



Desta Sandya Prasvita, S.Kom., M.Kom.

Kaprodi

Ditetapkan di : Jakarta

Tanggal Ujian : Selasa, 12 Juli 2022



ANALISIS SENTIMEN TERHADAP VAKSIN NUSANTARA PADA MEDIA SOSIAL *YOUTUBE* MENGGUNAKAN METODE NAÏVE BAYES DAN SELEKSI FITUR PARTICLE SWARM OPTIMIZATION

Taufik Adi Prasetyo

Abstrak

Youtube adalah salah satu dari sekian banyak media sosial yang dapat digunakan untuk memberikan komentar yang dapat diakses masyarakat Indonesia, komentar yang ditulis dapat berupa keluhan, saran, maupun kritik yang membangun mengenai suatu topik, salah satunya mengenai Vaksin Nusantara. Penelitian kali ini, menggunakan salah satu video di *youtube* yang berjudul “Peneliti Utama Jawab Kontroversi Vaksin Nusantara - ROSI (1)” dan di unggah oleh akun KOMPASTV. Komentar yang diambil menggunakan komentar pertama yang diketik bukan merupakan balasan dari suatu komentar, Data komentar tersebut diambil menggunakan *Integrated Development Environment (IDE) Google Apps Script*. Setelah itu dilakukan pelabelan untuk data komentar tersebut, dan dilakukan praproses teks menggunakan beberapa metode seperti, *Cleaning, case folding, tokenization, normalization, stopword removal*, dan *stemming*, kemudian kata-kata tersebut diberikan bobot dengan menggunakan pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*). Setelah itu dilakukan seleksi fitur menggunakan PSO (*Particle Swarm Optimization*) hingga 1000 iterasi dan didapatkan jumlah fitur yang digunakan sejumlah 933 fitur, lalu untuk melanjutkan pembuatan model klasifikasi, dilakukan sampling menggunakan SMOTE, yang awalnya 645 data negatif, dan 354 positif, menjadi berimbang 645 negatif, dan 645 positif, didapatkan hasil klasifikasi menggunakan *Naïve Bayes* dengan akurasi 82,5%, nilai presisi 78,7%, dan nilai *recall* 89,1%.

Kata Kunci : *Youtube, Analisis Sentimen, Naïve Bayes, Particle Swarm Optimization*

SENTIMEN ANALYSIS OF THE NUSANTARA VACCINE ON YOUTUBE SOCIAL MEDIA USING NAÏVE BAYES METHOD AND PARTICLE SWARM OPTIMIZATION FEATURES

Taufik Adi Praasetyo

Abstract

Youtube is one of the many social media that can be used to provide comments that can be accessed by various groups of people, including the people of Indonesia, comments written can be in the form of complaints, suggestions, or constructive criticism on a topic, one of which is about the Nusantara Vaccine. This research uses a video on YouTube entitled “Peneliti Utama Jawab Kontroversi Vaksin Nusantara - ROSI (1)” and uploaded by the KOMPASTV account. Comments taken using the first comment typed are not a reply to a comment, the comment data is retrieved using the Integrated Development Environment (IDE) Google Apps Script. After that, labeling for the comment data is carried out, and text preprocessing is carried out using several methods such as, Cleaning, case folding, tokenization, normalization, stopword removal, and stemming, then the words are given weights using TF-IDF (Term Frequency – Inverse Document Frequency). After that, feature selection was carried out using PSO (Particle Swarm Optimization) for up to 100 iterations and the number of features used was 933 features, then to continue making the classification model, sampling was carried out using SMOTE, which initially 645 negatif data, and 354 positive, became balanced 645 negatif, and 645 positive, the classification results obtained with an accuracy of 82.5%, a precision value of 78.7%, and a recall value of 89.1%.

Kata Kunci : Youtube, Analisis Sentimen, Naïve Bayes, Particle Swarm Optimization

KATA PENGANTAR

Puji syukur sebesar sebesarnya penulis panjatkan kehadirat Allah SWT atas segala karunia dan ridha-Nya, sehingga Skripsi ini berhasil diselesaikan. Tak lupa, penulis ingin mengucapkan banyak terima kasih kepada:

1. Almarhum ayah, ibu saya, dan keluarga yang selalu memberikan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Dr. Didit Widiyanto, S.Kom, M.Si. dan Bapak Desta Sandya Prasvita, S.Komp, M.Kom, selaku dosen pembimbing yang telah memberikan saran yang bermanfaat selama proses pembuatan Proposal hingga menyelesaikan skripsi.
3. Bapak Desta Sandya Prasvita, S.Komp, M.Kom, selaku dosen pembimbing akademik.
4. Bapak Desta Sandya Prasvita, S.Komp, M.Kom, selaku Kaprodi Informatika yang telah memberikan informasi mengenai tugas akhir.
5. Ibu Dr. Ermatita, M.Kom. selaku dekan Fakultas Ilmu Komputer.
6. Teman-teman Informatika 2018 yang telah berjuang bersama dan membantu saya untuk dapat menyelesaikan Skripsi. Terutama kepada Sheila, Yohanes, Ero, Adjie, Arief, Luthfi, Arvi, Bagas, dan Hasan.
7. Teman-teman KSM Android UPNVJ Periode 2020/2021 dan Periode 2021/2022.

Dan semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.

Jakarta, 24 Juni 2022

Taufik Adi Prasetyo

DAFTAR ISI

COVER	i
HALAMAN JUDUL.....	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT.....	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup Masalah	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Luaran yang Diharapkan	4
1.7 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 <i>Text Mining</i>	6

2.2	Data Mining.....	6
2.3	Analisis Sentimen.....	7
2.4	<i>Youtube</i>	7
2.5	Vaksin Nusantara	7
2.6	PSO (<i>Particle Swarm Optimization</i>) dan PySwarms	7
2.7	SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	8
2.8	<i>Naïve Bayes</i>	9
2.9	<i>Text Preprocessing</i>	11
2.9.1	<i>Cleaning</i>	11
2.9.2	<i>Case Folding</i>	11
2.9.3	<i>Tokenizing</i>	11
2.9.4	<i>Stopword Removal</i>	11
2.9.5	<i>Stemming</i>	12
2.10	<i>Term Frequency – Inverse Document Frequency (TF – IDF)</i>	12
2.11	Evaluasi.....	13
2.12	Studi Literatur	14
BAB III METODOLOGI PENELITIAN.....		17
3.1	Kerangka Pikir.....	17
3.1.1	Identifikasi Masalah	18
3.1.2	Studi Literatur	18
3.1.3	Pengumpulan Data	18
3.1.4	Pelabelan Data.....	18
3.1.5	<i>Text Preprocessing</i>	19
3.1.6	Pembobotan Kata dengan TF-IDF	20
3.1.7	PSO (<i>Particle Swarm Optimization</i>).....	21

3.1.8	SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	21
3.1.9	Pembagian Data	22
3.1.10	Klasifikasi	22
3.1.11	Evaluasi.....	22
3.1.12	Visualisasi	23
3.2	Perangkat Penelitian	23
BAB IV HASIL DAN PEMBAHASAN		25
4.1	Data	25
4.1.1	Pelabelan Data.....	25
4.2	<i>Text Preprocessing</i>	26
4.2.1	<i>Cleaning</i>	26
4.2.2	Case Folding	27
4.2.3	<i>Tokenization</i>	29
4.2.4	Normalization.....	30
4.2.5	<i>Stopword Removal</i>	32
4.2.6	<i>Stemming</i>	34
4.3	Pembobotan TF-IDF.....	35
4.4	PSO (<i>Particle Swarm Optimization</i>).....	38
4.5	SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	40
4.5.1	Visualisasi Scatterplot.....	42
4.6	Klasifikasi <i>Naïve Bayes</i>	44
4.6.1	Proses Latih.....	44
4.6.2	Proses Uji	49
4.7	Evaluasi	52
4.8	Visualisasi	53

4.9 Analisis Perbandingan	55
BAB V PENUTUP	56
5.1 Kesimpulan.....	56
5.2 Saran	57
DAFTAR PUSTAKA	58
RIWAYAT HIDUP	62
LAMPIRAN	64

DAFTAR GAMBAR

Gambar 3.1 Kerangka Pikir.....	17
Gambar 3.2 <i>Tokenizing</i>	20
Gambar 4.1 Alur Pemrosesan <i>Cleaning</i>	26
Gambar 4.2 Alur pemrosesan <i>Case Folding</i>	27
Gambar 4.3 Alur Pemrosesan <i>Tokenization</i>	29
Gambar 4.4 Alur pemrosesan <i>Normalization</i>	31
Gambar 4.5 Alur pemrosesan <i>Stopword Removal</i>	33
Gambar 4.6 Alur pemrosesan <i>Stemming</i>	34
Gambar 4.7 Visualisasi Scatterplot <i>Imbalanced Data</i>	42
Gambar 4.8 Visualisasi Scatterplot <i>Undersampling</i>	43
Gambar 4.9 Visualisasi Scatterplot SMOTE	43
Gambar 4.10 <i>Wordcloud</i> kelas Positif	54
Gambar 4.11 <i>Wordcloud</i> kelas Negatif.....	54

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	13
Tabel 3.1 Kategori Tingkat Ketidakseimbangan	22
Tabel 4.1 Data Sampel <i>Cleaning Data</i>	27
Tabel 4.2 Data Sampel <i>Case Folding</i>	28
Tabel 4.3 Data Sampel Hasil <i>Tokenization</i>	29
Tabel 4.4 Data Sampel Kamus kata yang dibuat	31
Tabel 4.5 Data Sampel Hasil <i>Normalization</i>	32
Tabel 4.6 Data Sampel Hasil <i>Stopword Removal</i>	33
Tabel 4.7 Data Sampel Hasil <i>Stemming</i>	34
Tabel 4.8 Data Sampel	35
Tabel 4.9 Data Perhitungan TF-IDF	36
Tabel 4.10 Data iterasi optimizer PSO dengan jumlah Fitur	38
Tabel 4.11 Data Sampel fitur yang dibuang dari seleksi PSO	39
Tabel 4.12 Data Sampel fitur setelah seleksi PSO	39
Tabel 4.13 Data Nilai Sampel Dokumen	41
Tabel 4.14 Data Nilai Sampel Dokumen ditambah Dokumen Baru	42
Tabel 4.15 Pembagian Data yang Digunakan	44
Tabel 4.16 Data Latih Sampel.....	45
Tabel 4.17 Data Bobot dari Data Latih	45
Tabel 4.18 Probabilitas Data Latih Sampel.....	48
Tabel 4.19 Probabilitas Data Latih Sampel.....	49
Tabel 4.20 Data Bobot Data Uji.....	50
Tabel 4.21 Tabel Evaluasi <i>Confusion Matrix</i>	52
Tabel 4.22 Tabel Percobaan Penelitian	55

DAFTAR LAMPIRAN

Lampiran 1 Pelabelan Data	65
Lampiran 2 Daftar <i>Stopword</i>	117
Lampiran 3 Daftar Kamus <i>Normalization</i>	120