



**ANALISIS SENTIMEN VAKSIN COVID-19 MENGGUNAKAN EKSTRAKSI  
FITUR N-GRAM DAN ALGORITMA NAÏVE BAYES**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**TIA NUR HIDAYAH**

**NIM 1711016320030**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**Februari 2023**



**ANALISIS SENTIMEN VAKSIN COVID-19 MENGGUNAKAN EKSTRAKSI  
FITUR N-GRAM DAN ALGORITMA NAÏVE BAYES**

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh**

**TIA NUR HIDAYAH**

**NIM 1711016320030**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU**

**Februari 2023**

# SKRIPSI

## ANALISIS SENTIMEN VAKSIN COVID-19 MENGGUNAKAN EKSTRAKSI FITUR N-GRAM DAN ALGORITMA NAÏVE BAYES

Oleh:

**TIA NUR HIDAYAH**

**NIM 1711016320030**

Telah dipertahankan di depan Dosen Penguji pada Tanggal 24 Februari 2023

Susunan Dosen Penguji:

Pembimbing I

**Irwan Budiman S.T., M.Kom.**  
NIP. 197703252008121001

Dosen Penguji I

**Muliadi, S.Kom., M.Cs.**  
NIP. 197804222010121002

Pembimbing II

**Dwi Kartini, S.Kom., M.Kom.**  
NIP. 198704212012122003

Dosen Penguji II

**Muhammad Itqan Mazdadi, S.Kom., M.Kom.**  
NIP. 199006122019031013



Banjarnbaru, 30 Maret 2023

Koordinator PS Ilmu Komputer

**Irwan Budiman, S.T., M.Kom.**  
NIP. 197703252008121001

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 24 Februari 2023



**Tia Nur Hidayah**

**NIM. 1711016320030**

## ABSTRAK

**ANALISIS SENTIMEN VAKSIN COVID-19 MENGGUNAKAN EKSTRAKSI FITUR N-GRAM DAN ALGORITMA NAÏVE BAYES** (Oleh: Tia Nur Hidayah; Pembimbing: Irwan Budiman, S.T., M.Kom. dan Dwi Kartini, S.Kom., M.Kom.; 2023; 67 halaman)

Pandemi COVID-19 yang menyebabkan bencana di bidang kesehatan, ekonomi, dan masyarakat. Vaksin adalah salah satu harapan untuk menghentikan pandemi COVID-19. Sejak pemerintah pertama kali meluncurkan vaksin Sinovac untuk COVID-19, banyak komentar dari masyarakat terhadap vaksin COVID-19 salah satunya media sosial twitter. Analisis sentimen dilakukan bertujuan untuk memahami opini masyarakat tentang kasus vaksin COVID-19. Pada penelitian ini melakukan perbandingan menggunakan fitur kata *N-Gram* berupa *Unigram*, *Bigram*, *Trigram*, dan kombinasi sebagai fitur kata kemudian dilakukan klasifikasi menggunakan metode *Naïve Bayes*. Data yang digunakan adalah data tweet dari twitter yang berjumlah 1000 data dan terdiri dari 2 kelas yaitu positif dan negatif dengan masing-masing kelas sebanyak 500 data. Hasil dari perbandingan fitur kata *N-Gram* tersebut, diperoleh bahwa performa akurasi yang tinggi dihasilkan ketika dilakukan kombinasi fitur *N-Gram* dengan hasil akurasi sebesar 84,1% dari 18096 fitur, *Unigram* menghasilkan akurasi 82,1% dari 1805 fitur, *Bigram* menghasilkan akurasi 79,9% dari 7496 fitur, sementara untuk *Trigram* menghasilkan akurasi 67,9% dari 8795 fitur. Sehingga penelitian ini memberikan hasil kinerja terbaik pada pengujian dengan menggunakan kombinasi fitur kata *N-Gram*.

**Kata kunci:** *N-Gram*, *Naïve Bayes*, *Klasifikasi*, *Text Mining*, Twitter, Vaksin COVID-19

## ***ABSTRACT***

**ANALYSIS SENTIMENT VACCINE OF COVID-19 USING N-GRAM FEATURE EXTRACTION AND NAÏVE BAYES ALGORITHM** (By: Tia Nur Hidayah; Supervisor: Irwan Budiman, S.T., M.Kom. dan Dwi Kartini, S.Kom., M.Kom.; 2023; 67 pages)

The COVID-19 pandemic has caused a disaster in the health, economy, and social sectors. Vaccine is one of the hopes to stop the pandemic. Since the government launched the Sinovac vaccine, there have been many comments from the public about the vaccine in social media Twitter. Sentiment analysis was carried out with the aim of understanding public opinion about the COVID-19 vaccine case. In this study, a comparison was made using N-Gram word features in the form of Unigram, Bigram, Trigram, and combinations as word features and then classified using the Naïve Bayes method. The data used is tweet data from Twitter which totals 1000 data and consists of 2 classes, namely positive and negative, with 500 data for each class. The results of the comparison of the N-Gram word features, it is found that high accuracy performance is produced when combining N-Gram features with an accuracy of 84.1% of 18096 features, Unigram produces 82.1% accuracy of 1805 features, Bigram produces an accuracy of 79.9% of 7496 features, while for Trigram it produces 67.9% accuracy of 8795 features. So that this study gives the best performance results in testing using a combination of N-Gram word features.

**Keywords:** N-Gram, Naïve Bayes, Klasifikasi, Text Mining, Twitter, Vaccine COVID-19

## PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan kepada Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen Vaksin COVID-19 Menggunakan Ekstraksi Fitur *N-Gram* Metode *Naïve Bayes*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiama.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, diantaranya:

1. Keluarga terutama kedua orang tua yang selalu memberikan dukungan, doa, dan bantuan dalam proses penyelesaian skripsi ini.
2. Bapak Irwan Budiman, S.T., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM sekaligus Dosen Pembimbing Akademik, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.
7. Teman-teman keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.

8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 24 Februari 2023



Tia Nur Hidayah



## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN .....	ii
PERNYATAAN.....	ii
ABSTRAK .....	iv
<i>ABSTRACT</i> .....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN.....	xiv
<b>BAB I PENDAHULUAN</b>	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan.....	3
1.5 Manfaat Penelitian.....	3
<b>BAB II TINJAUAN PUSTAKA</b>	
2.1 Kajian Terdahulu .....	4
2.2 Landasan Teori .....	6
2.2.1 Text Mining.....	6
2.2.2 Analisis Sentimen .....	9
2.2.3 Vaksin .....	10
2.2.4 Text Preprocessing .....	10
2.2.5 Ekstraksi Fitur .....	12

2.2.6	TF-IDF .....	13
2.2.7	N-Gram .....	15
2.2.8	<i>K-Fold Validation</i> .....	17
2.2.9	Naïve Bayes .....	18
2.2.10	<i>Confusion Matrix</i> .....	20
 <b>BAB III METODE PENELITIAN</b>		
3.1	Bahan Penelitian .....	21
3.2	Alat Penelitian .....	21
3.3	Variabel Penelitian.....	21
3.4	Prosedur Penelitian .....	21
 <b>BAB IV HASIL DAN PEMBAHASAN</b>		
4.1	Hasil.....	26
4.1.1	Pengumpulan Data .....	26
4.1.2	<i>Data Preprocessing</i> .....	27
4.1.3	Ekstraksi Fitur .....	36
4.1.4	<i>K-Fold Cross Validation</i> .....	45
4.1.5	Klasifikasi <i>Naïve Bayes</i> .....	46
4.1.6	Evaluasi.....	53
4.2	Pembahasan .....	57
 <b>BAB V PENUTUP</b>		
5.1	Kesimpulan.....	63
5.2	Saran .....	63
DAFTAR PUSTAKA .....		64
LAMPIRAN .....		68

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1 Keaslian Penelitian.....	5
Tabel 2 Rancangan Penelitian.....	6
Tabel 3 Contoh Penggunaan TF-IDF.....	15
Tabel 4 Ekstraksi Fitur N-Gram.....	17
Tabel 5 Tabel Confusion Matrix.....	20
Tabel 6 Tweet Positif.....	22
Tabel 7 Tweet Negatif.....	23
Tabel 8 Data Tweet Komentar.....	26
Tabel 9 Jumlah Data Perkelas.....	27
Tabel 10 Data Cleansing.....	27
Tabel 11 Data Hasil <i>Cleansing</i> .....	27
Tabel 12 Data Hasil <i>Case Folding</i> .....	29
Tabel 13 Kamus <i>Stopwords</i> .....	30
Tabel 14 Kamus <i>Stopwords</i> .....	30
Tabel 15 Data Penghapusan <i>Stopwords</i> .....	31
Tabel 16 Kamus <i>Slang Words</i> .....	32
Tabel 17 Data <i>Slang Words</i> .....	33
Tabel 18 Awalan (Prefiks).....	33
Tabel 19 Akhiran (Sufiks).....	34
Tabel 20 Sisipan (Infiks).....	34
Tabel 21 Gabungan Imbuhan (Konfiks).....	34
Tabel 22 Data <i>Stemming</i> .....	35
Tabel 23 Data <i>Stopword 2</i> .....	35
Tabel 24 Contoh hasil ekstraksi fitur TF-IDF dengan <i>Unigram</i> .....	36
Tabel 25 Contoh perhitungan ekstraksi fitur TF-IDF dengan <i>Unigram</i> .....	37
Tabel 26 Contoh hasil ekstraksi fitur TF-IDF dengan <i>Bigram</i> .....	39
Tabel 27 Contoh perhitungan ekstraksi fitur TF-IDF dengan <i>Bigram</i> .....	39
Tabel 28 Contoh hasil ekstraksi fitur TF-IDF dengan <i>Trigram</i> .....	41
Tabel 29 Contoh perhitungan ekstraksi fitur TF-IDF dengan <i>Trigram</i> .....	41

Tabel 30 Contoh hasil ekstraksi fitur TF-IDF dengan Gabungan.....	43
Tabel 31 Contoh perhitungan ekstraksi fitur TF-IDF dengan Gabungan .....	43
Tabel 32 Pembagian data dengan <i>10-Fold Cross Validation</i> .....	46
Tabel 33 Frekuensi kemunculan term <i>Unigram</i> .....	47
Tabel 34 Probabilitas term kata <i>Unigram</i> .....	47
Tabel 35 Hasil kali probabilitas term kata <i>Unigram</i> .....	48
Tabel 36 Frekuensi kemunculan term <i>Bigram</i> .....	48
Tabel 37 Hasil kali probabilitas term kata <i>Bigram</i> .....	49
Tabel 38 Probabilitas term kata <i>Bigram</i> .....	49
Tabel 39 Frekuensi kemunculan term <i>Trigram</i> .....	50
Tabel 40 Hasil kali probabilitas term kata <i>Trigram</i> .....	50
Tabel 41 Probabilitas term kata <i>Trigram</i> .....	51
Tabel 42 Frekuensi kemunculan term gabungan .....	51
Tabel 43 Hasil kali probabilitas term kata gabungan.....	52
Tabel 44 Probabilitas term kata gabungan .....	53
Tabel 45 <i>Confusion matrix Naïve Bayes</i> menggunakan TF-IDF dengan <i>unigram</i> ....	54
Tabel 46 Perhitungan akurasi klasifikasi <i>Naïve Bayes</i> menggunakan TF-IDF dengan <i>unigram</i> .....	54
Tabel 47 <i>Confusion matrix Naïve Bayes</i> menggunakan TF-IDF dengan <i>bigram</i> .....	54
Tabel 48 Perhitungan akurasi klasifikasi <i>Naïve Bayes</i> menggunakan TF-IDF dengan <i>bigram</i> .....	55
Tabel 49 <i>Confusion matrix Naïve Bayes</i> menggunakan TF-IDF dengan <i>trigram</i> .....	55
Tabel 50 Perhitungan akurasi klasifikasi <i>Naïve Bayes</i> menggunakan TF-IDF dengan <i>trigram</i> .....	56
Tabel 51 <i>Confusion matrix Naïve Bayes</i> menggunakan TF-IDF dengan gabungan <i>N-Gram</i> .....	56
Tabel 52 Perhitungan akurasi klasifikasi <i>Naïve Bayes</i> menggunakan TF-IDF dengan gabungan <i>N-Gram</i> .....	57
Tabel 53 Akurasi setiap fold pada <i>Unigram</i> .....	59
Tabel 54 Akurasi setiap fold pada <i>Bigram</i> .....	60
Tabel 55 Akurasi setiap fold pada <i>Trigram</i> .....	60
Tabel 56 Akurasi setiap fold pada gabungan .....	60

Tabel 57 Perbandingan hasil akurasi .....	61
---	----

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Diagram interaksi <i>text mining</i> antar pendekatan .....	7
Gambar 2 Proses <i>text mining</i> .....	8
Gambar 3. Diagram Alur TF-IDF .....	14
Gambar 4 Representasi 10 <i>folds cross validation</i> .....	18
Gambar 5 Alur Penelitian.....	22
Gambar 6. Grafik hasil akurasi <i>Naïve Bayes</i> menggunakan fitur <i>N-Gram</i> .....	62

## DAFTAR LAMPIRAN

### **Lampiran**

Lampiran 1 Kamus Hidayah

Lampiran 2 Kamus *Stopwords* Pahrul

Lampiran 3 Slangwords Prakoso

Lampiran 4 *Source Code Cleansing*

Lampiran 5 *Source Code Case Folding*

Lampiran 6 *Source Code Removal of Stopwords*

Lampiran 7 *Source Code Slangwords Conversion*

Lampiran 8 *Source Code Stemming*

Lampiran 9 *Source Code Removal of Stopwords 2*

Lampiran 10 *Source Code Final*

Lampiran 11 *Source Code Unigram*

Lampiran 12 *Source Code Bigram*

Lampiran 13 *Source Code Trigram*

Lampiran 14 *Source Code Gabungan*