



**KLASIFIKASI DATA KOMENTAR COVID 19 PADA  
TWITTER MENGGUNAKAN METODE KLASIFIKASI  
NAIVE BAYES**

**SKRIPSI**

**NAUFALDY PURWANTO**

**1910511128**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

**2023**



**KLASIFIKASI DATA KOMENTAR COVID 19 PADA  
TWITTER MENGGUNAKAN METODE KLASIFIKASI  
NAIVE BAYES**

**SKRIPSI**

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

**Sarjana Komputer**

**NAUFALDY PURWANTO**

**1910511128**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA  
2023**

## PERNYATAAN ORISINALITAS

Tugas skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan benar.

Nama : Naufaldy Purwanto

NIM : 1910511128

Tanggal : 11 Januari 2024

Program Studi : SI Informatika

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

Jakarta, 11 Januari 2024

Yang menyatakan



(Naufaldy Purwanto)

## **PERNYATAAN PERSETUJUAN PUBLIKASI**

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

Nama : Naufaldy Purwanto

NIM : 1910511128

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:

### **KLASIFIKASI DATA KOMENTAR COVID 19 PADA TWITTER MENGUNAKAN METODE KLASIFIKASI NAIVE BAYES**

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 : 11 Januari 2024

Yang Menyatakan,



(Naufaldy Purwanto)

## LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Naufaldy Purwanto  
NIM : 1910511128  
Program Studi : S1 Informatika  
Judul : Klasifikasi Data Komentar *Covid19* Pada *Twitter*  
Menggunakan Metode Klasifikasi *Naïve Bayes*

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



**(Dr. Widva Cholil, S.Kom., M.I.T.)**  
Penguji 1



**(Musthofa Galih Pradana, S.Kom., M.Kom.)**  
Penguji 2



**(Javanta, S.Kom., M.Si.)**  
Dosen Pembimbing



**(Prof. Dr. Ir. Suprivanto, ST., M.Sc., IPM)**  
Dekan Fakultas Ilmu Komputer



**(Dr. Widva Cholil, M.I.T)**  
Ketua Program Studi SI Informatika

Ditetapkan di : Jakarta

Tanggal persetujuan : 11 Januari 2024

# KLASIFIKASI DATA KOMENTAR COVID 19 PADA TWITTER MENGUNAKAN METODE KLASIFIKASI NAÏVE BAYES

NAUFALDY PURWANTO

## ABSTRAK

Banyak sekali pengguna *Twitter* yang mencurahkan pendapat melalui *tweet* yang mereka kirimkan pada media social *Twitter*. Salah satu *Tweet* yang menjadi perbincangan pada tahun 2019 silam yaitu mengenai Pandemi *Covid19*. Banyak sekali *tweet* yang membahas tentang Pandemi *Covid19* tersebut. Mereka selalu membahas tentang bagaimana dampak positif dan negatif dari Pandemi *Covid19*. Dari permasalahan tersebut, dilakukanlah penelitian mengenai analisis sentiment pada data *Twitter* yang berkaitan dengan Pandemi *Covid19* dan menggunakan metode *Naïve Bayes*. Data yang diambil yaitu data dari *tweet* menggunakan library *tweet-harvest* yang sudah terintegrasi dengan *Twitter API*. Sebelum melakukan proses klasifikasi data, data tersebut akan dilakukan pemberian label yaitu label positif dan label negative secara manual. Setelah itu, dilakukan praproses dengan tahapan *data cleaning*, *tokenization*, *case folding*, *stopword removal*, *stemming*. Kemudian melakukan pembobotan kata menggunakan *CountVectorizer*. Tahap selanjutnya dilakukan perbandingan hasil pembagian data 80% data latih dan 20% data uji dengan 70% data latih dan 30% data uji. Dengan mengklasifikasikan menggunakan metode *Naïve Bayes*. Hasilnya pembagian data dengan 70% data latih dan 30% data uji menghasilkan tingkat akurasi tertinggi yaitu 91%.

**Kata Kunci :** Analisis Sentimen, Klasifikasi, Pandemi *Covid19*, *Naïve Bayes Classification*, *Twitter*

# **CLASSIFICATION OF COVID 19 COMMENT DATA ON TWITTER USING THE NAÏVE BAYES CLASSIFICATION METHOD**

**NAUFALDY PURWANTO**

## **ABSTRACT**

*A lot of Twitter users express their opinions through the tweets they send on Twitter social media. One of the tweets that became a conversation in 2019 was about the Covid19 Pandemic. Lots of tweets discussing the Covid19 Pandemic. They always discuss how the positive and negative impacts of the Covid19 Pandemic. From these problems, research was conducted on sentiment analysis on Twitter data related to the Covid19 Pandemic and using the Naïve Bayes method. The data taken is data from tweets using the tweet-harvest library that has been integrated with the Twitter API. Before carrying out the data classification process, the data will be labeled, namely positive labels and negative labels manually. After that, preprocessing is carried out with the stages of data cleaning, tokenization, case folding, stopword removal, stemming. Then do word weighting using CountVectorizer. The next stage is to compare the results of data division of 80% training data and 20% test data with 70% training data and 30% test data. By classifying using the Naïve Bayes method. The results of data sharing with 70% training data and 30% test data resulted in the highest accuracy rate 91%.*

**Keywords** : *Sentiment Analysis, Classification, Covid19 Pandemic, Naïve Bayes Classification, Twitter*

## KATA PENGANTAR

Puji dan syukur dipanjatkan atas kehadiran Allah SWT. Atas karunia-Nya penulis dapat menyelesaikan tugas akhir dengan baik. Dalam penyelesaian tugas akhir ini tidak lepas dari bantuan banyak pihak yang telah memberikan bantuan dan dukungan. Untuk itu peneliti mengucapkan banyak terima kasih kepada :

1. Allah SWT atas segala rahmat, ridho, dan karunia-Nya.
2. Orang tua dan keluarga yang selalu memberikan dukungan dan doa kepada penulis agar dapat menyelesaikan tugas akhir ini.
3. Bapak Jayanta, S.Kom, M.Si, selaku Dosen Pembimbing yang telah memberikan bimbingan sehingga tugas akhir ini dapat diselesaikan.
4. Ibu Kraugusteeliana, S.Kom., M.Kom., M.M. selaku Wakil Dekan Bidang Akademik Fakultas Ilmu Komputer.
5. Ibu Dr. Widya Cholil, S.Kom, M.I.T. selaku Wakil Dekan Bidang Akademik Fakultas Ilmu Komputer.
6. Teman-teman Informatika Angkatan 2019 yang saling memberikan dukungan satu sama lain.

Disadari bahwa masih banyaknya kekurangan dari Proposal Tugas Akhir ini, baik dari materi maupun teknik penulisan, mengingat kurangnya pengetahuan dan pengalaman peneliti. Oleh karena itu, kritik dan saran yang membangun akan sangat berarti bagi peneliti.

Jakarta, 1 Januari 2024



Naufaldy Purwanto



## DAFTAR ISI

ABSTRAK .....	vi
ABSTRACT .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian .....	3
1.4 Manfaat Penelitian .....	3
1.5 Ruang Lingkup .....	3
1.6 Luaran yang diharapkan.....	4
1.7 Sistematika Penulisan .....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 <i>Twitter</i> .....	5
2.2 Analisis Sentimen .....	5
2.3 <i>REST API</i> .....	6
2.4 <i>Twitter REST API</i> .....	6
2.5 <i>Text Analysis</i> .....	6
2.6 <i>Text Mining</i> .....	7
2.7 Pra Proses Data .....	7
2.7.1 <i>Data Cleaning</i> .....	7
2.7.2 <i>Tokenization</i> .....	8
2.7.3 <i>Case Folding</i> .....	8
2.7.4 <i>Stopword Removal</i> .....	8
2.7.5 <i>Stemming</i> .....	8
2.8 <i>Lexicon Based</i> .....	8
2.9 <i>Count Vectorizer</i> .....	10
2.10 <i>Python</i> .....	10
2.11 Metode Klasifikasi .....	11
2.11.1 <i>Algoritma Naïve Bayes</i> .....	11
2.12 Review Penelitian yang Relevan .....	13

BAB III METODOLOGI PENELITIAN.....	16
3.1 Identifikasi Masalah.....	17
3.2 Studi Literatur.....	17
3.3 Akuisi Data.....	17
3.4 Praproses.....	17
3.4.1 Data Cleaning.....	18
3.4.2 Tokenization.....	18
3.4.3 Case Folding.....	18
3.4.4 Stopword Removal.....	18
3.4.5 Stemming.....	18
3.5 <i>Lexicon Based</i> .....	18
3.6 Pelabelan Data.....	18
3.7 Klasifikasi.....	19
3.8 Evaluasi.....	19
3.9 Alat Bantu Penelitian.....	20
3.10 Jadwal Penelitian.....	21
BAB IV HASIL DAN PEMBAHASAN.....	22
4.1 Data.....	22
4.2 Pra proses.....	22
4.2.1 <i>Data Cleaning</i> .....	23
4.2.2 <i>Tokenization</i> .....	25
4.2.3 <i>Stopword Removal</i> .....	26
4.2.4 <i>Stemming</i> .....	28
4.3 Pelabelan Data.....	30
4.3.1 Pelabelan Otomatis dengan <i>Lexicon Based</i> .....	30
4.4 Klasifikasi dengan <i>Naïve Bayes</i> .....	32
4.4.1 Hasil Klasifikasi.....	33
BAB V PENUTUP.....	36
5.1 Kesimpulan.....	36
5.2 Saran.....	37
DAFTAR PUSTAKA.....	38

## DAFTAR TABEL

Tabel 3.1 Confusion Matrix .....	19
Tabel 3.2 Jadwal Penelitian.....	21
Tabel 4.1 Hasil dari pembersihan data .....	24
Tabel 4.2 Tokenization.....	26
Tabel 4.3 Stopword .....	27
Tabel 4.4 Stemming .....	29

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	16
Gambar 4.1 <i>Crawling data</i> .....	22
Gambar 4.2 Pembersihan data .....	23
Gambar 4.3 Pembersihan simbol .....	23
Gambar 4.4 Pembersihan angka.....	23
Gambar 4.5 <i>Tokenization</i> .....	25
Gambar 4.6 <i>Tokenization dan Case Folding</i> .....	25
Gambar 4.7 <i>Stopword</i> .....	27
Gambar 4.8 Proses <i>stopword removal</i> .....	27
Gambar 4.9 <i>Stemming</i> .....	28
Gambar 4.10 <i>Stemming Code</i> .....	29
Gambar 4.11 <i>Lexicon Based</i> .....	30
Gambar 4.12 Nilai <i>Compound</i> .....	31
Gambar 4.13 Hasil <i>Lexicon Based</i> .....	31
Gambar 4.15 Pembagian data .....	32
Gambar 4.16 Pembobotan kata .....	32
Gambar 4.17 Sebelum penyeimbangan data.....	33
Gambar 4.18 Sesudah Penyeimbangan data .....	33
Gambar 4.19 Hasil Klasifikasi Uji 1 .....	34
Gambar 4.19 Hasil Klasifikasi Uji 2 .....	35