



**SENTIMEN ANALISIS PUBLIK TERHADAP VIRUS COVID-19 PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

**SERAFIM CLARA
1710511024**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2021**



**SENTIMEN ANALISIS PUBLIK TERHADAP VIRUS COVID-19 PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES**

SKRIPSI

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

**SERAFIM CLARA
1710511024**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2021**

PERNYATAAN ORISINALITAS

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

Nama : Serafim Clara

NIM : 1710511024

Tanggal : 30 Juni 2021

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

Jakarta, 30 Juni 2021

Yang Menyatakan,



(Serafim Clara)

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Serafim Clara
NIM : 1710511024
Program Studi : Informatika
Judul : Sentimen Analisis Public Terhadap Virus Covid-19 Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes

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



Dr. Ermatita, M.Kom.

Ketua Penguji



Anita Muliawati, S.Kom., M.TI

Anggota Penguji



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

Pembimbing I



Mayanda Mega Santoni, S.Kom., M.Kom.

Pembimbing II



Dr. Ermatita, M.Kom.

Dean



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

Ketua Program Studi

Ditetapkan di : Jakarta
Tanggal Pengesahan : 21 Juli 2021



**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini:

Nama : Serafim Clara
NIM : 1710511024
Fakultas : Ilmu Komputer
Program Studi : Informatika

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

**SENTIMEN ANALISIS PUBLIK TERHADAP VIRUS COVID-19 PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE 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 data (database), 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 : 30 Juni 2021



Yang Menyatakan,
(Serafim Clara)

SENTIMEN ANALISIS PUBLIK TERHADAP VIRUS COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAÏVE BAYES

Serafim Clara

ABSTRAK

Twitter merupakan media sosial sebagai tempat untuk saling betukar infomasi, menyampaikan pendapat maupun kritik secara langsung. Covid-19 merupakan salah satu wabah yang menyebar sangat cepat hingga hampir seluruh dunia merasakan dampaknya. Hal ini tentu saja langsung menggegerkan dunia maya hingga menjadi trending nomor satu di *twitter*. Opini tersebut dapat di analisis menjadi analisis sentimen yang dikelompokkan menjadi sentimen positif dan sentimen negatif, lalu menggunakan algoritma klasifikasi *Naïve Bayes*. Data diambil tanggal 11 Desember 2020 pada 2 daerah yaitu DKI Jakarta dan Surabaya menggunakan API pada *twitter* dan juga fitur *geolocation* dengan #Covid-19. Hasil yang didapatkan yaitu data pada wilayah DKI Jakarta memiliki nilai akurasi paling tinggi pada sample 2 yaitu sebesar 78% sedangkan data pada wilaya Surabaya mendapatkan nilai akurasi tertinggi pada sample 1 yaitu sebesar 83%.

Kata kunci: *twitter, naïve bayes, geolocation*

ANALYSIS SENTIMENT PUBLIC OF COVID-19 ON SOCIAL MEDIA TWITTER USING NAIVE BAYES METHODE

Serafim Clara

ABSTRACT

Twitter is a social media platform where people may directly share information, express opinions and criticism. Covid-19 is an epidemic that spreads so quickly that almost the whole world feels its impact. This immediately stirred the virtual world and become the number one trending on twitter. This opinion can be analyzed with sentiment analysis which grouped them into positive sentiment and negative sentiment, then using the Naïve Bayes classification algorithm. The data was taken on December 11, 2020 in 2 regions, namely DKI Jakarta and Surabaya using the API on Twitter and also the geolocation feature with #Covid-19. The results obtained that the data from the DKI Jakarta region has the highest accuracy value in sample 2, with 78%, while the data from the Surabaya region has the highest accuracy value in sample 1 with 83%.

Keywords: *twitter, naïve bayes, geolocation*

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke pada Tuhan Yang Maha Esa atas segala karunia-Nya, sehingga Skripsi ini berhasil diselesaikan. Penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua serta keluarga yang selalu memberikan dukungan dan doa sehingga dapat menyelesaikan skripsi ini.
2. Ibu Yuni Widiastiwi, S.Kom., M.Si. dan Ibu Mayanda Mega Santoni, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan saran yang bermanfaat.
3. Ibu Yuni Widiastiwi, S.Kom., M.Si. selaku Ketua Program Studi Informatika Universitas Pembangunan Nasional Veteran Jakarta.
4. Ibu Dr. Ermatita, M. Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
5. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta terima kasih atas ilmu-ilmu yang bermanfaat.
6. Teman-teman Informatika 2017, rekan-rekan KSM Robotika, dan seluruh rekan mahasiswa yang tidak dapat penulis sebutkan namanya satu persatu, terima kasih atas saran dan dukungan yang telah diberikan.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 25 Juni 2021

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR SIMBOL	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Luaran yang Diharapkan	3
1.7 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	5
2.1 Covid-19	5
2.2 Twitter	5
2.3 Text Mining	5
2.4 Analisis Sentimen	5
2.5 Naïve Bayes	6
2.6 Pembobotan Kata (TF-IDF)	7
2.7 Preprocessing	7
2.8 Penelitian terkait	7
BAB III METODE PENELITIAN	10
3.1 Tahapan penelitian	10

3.1.1	Identifikasi Masalah	11
3.1.2	Studi Literatur.....	11
3.1.3	Pengumpulan Data.....	11
3.1.4	Pelabelan Data.....	11
3.1.5	<i>Preprocessing</i> Data	12
3.1.6	Pembobotan Kata.....	13
3.1.7	Proses Klasifikasi <i>Naïve Bayes</i>	14
3.1.8	Evaluasi.....	14
3.2	Perangkat Penelitian.....	15
3.3	Jadwal Penelitian	16
BAB IV HASIL DAN PEMBAHASAN		17
4.1	Data	17
4.2	<i>Labeling</i>	18
4.3	<i>Text Preprocessing</i>	23
4.4	Pembobotan Kata (TF-IDF).....	29
4.5	Klasifikasi <i>Naïve Bayes</i>	31
4.6	Evaluasi.....	36
BAB V PENUTUP.....		40
5.1	Kesimpulan.....	40
5.2	Saran	40
DAFTAR PUSTAKA.....		41
RIWAYAT HIDUP		43
LAMPIRAN.....		44




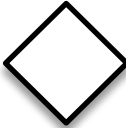
DAFTAR GAMBAR

Gambar 1. Tahapan Penelitian	10
Gambar 2. Hasil pengambilan data dengan format csv.....	17
Gambar 3. Skala Nilai Kappa	23

DAFTAR TABEL

Tabel 1. Rangkuman Penelitian	9
Tabel 2. Confusion Matrix	14
Table 3. Jadwal Penelitian	16
Table 4. Contoh hasil pelabelan data tweet <i>negatif</i>	18
Table 5. Contoh hasil pelabelan data tweet positif	18
Table 6. Tweet berlabel positif dan negatif.....	19
Table 7. Contoh hasil Case Folding pada data tweet negatif	24
Table 8. Contoh hasil Case Folding pada data tweet Positif.....	24
Table 9. Contoh hasil Cleansing pada data tweet Negatif	25
Table 10. Contoh hasil Cleansing pada data tweet Positif	25
Table 11. Contoh hasil Stemming pada data tweet Negatif	26
Table 12. Contoh hasil Stemming pada data tweet positif.....	26
Table 13. Contoh hasil Stopword Removal pada data tweet Negatif.....	27
Table 14. Contoh hasil Stopword Removal pada data tweet positif	27
Table 15. Contoh hasil Tokenisasi pada data tweet Negatif	28
Table 16. Contoh hasil Tokenisasi pada data tweet positif.....	28
Table 17. Document (Tweet)	29
Table 18. Pembobotan Kata TF-IDF.....	30
Table 19. Contoh tweet data latih	31
Table 20. Sample Data Uji.....	34
Table 21. <i>Confusion Matrix</i> data latih 75% dan data uji 25%	36
Table 22. Hasil dari 3 percobaan dengan dataset Jakarta.....	37
Table 23. Hasil dari 3 percobaan dengan dataset surabaya.....	38
Table 24. Perbandingan hasil antara Jakarta dan Surabaya	39

DAFTAR SIMBOL

Simbol <i>Flowchart</i>			
NO	Simbol	Nama Simbol	Keterangan
1		<i>Terminator</i> (<i>start, end</i>)	Simbol yang menggambarkan dimulainya atau berakhirnya suatu kegiatan berakhir.
2		<i>Flow Line</i>	Simbol yang menggambarkan hubungan dari simbol satu ke simbol lainnya dalam suatu proses kegiatan.
3		<i>Process</i>	Simbol yang menggambarkan suatu pengolahan yang dijalankan.
4		<i>Decision</i>	Simbol yang menggambarkan pilihan dari suatu proses.