



**ANALISIS SENTIMEN PENGARUH COVID-19 TERHADAP
KEHIDUPAN BERMASYARAKAT PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN *K-NEAREST NEIGHBOR***

SKRIPSI

PASCAL ALDWIN HERNANDO

1710511044

PROGRAM STUDI INFORMATIKA PROGRAM SARJANA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

2021



**ANALISIS SENTIMEN PENGARUH COVID-19 TERHADAP
KEHIDUPAN BERMASYARAKAT PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN *K-NEAREST NEIGHBOR***

SKRIPSI

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

PASCAL ALDWIN HERNANDO

1710511044

PROGRAM STUDI INFORMATIKA PROGRAM SARJANA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN

JAKARTA

2021

PERNYATAAN ORISINALITAS

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

Nama : Pascal Aldwin Hernando

NIM : 1710511044

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,

A 1000 Rupiah postage stamp is shown with a handwritten signature in black ink over it. The stamp features the number '1000' and the words 'METERA TEMPEL' and 'POSTAL INDONESIA'. The signature is written in a stylized, cursive font.

(Pascal Aldwin Hernando)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Tugas Akhir berikut:

Nama : Pascal Aldwin Hernando
NIM : 1710511044
Program Studi : Informatika
Judul : Analisis Sentimen Pengaruh Covid-19 Terhadap
Kehidupan Bermasyarakat Pada Media Sosial Twitter
Menggunakan *K-Nearest Neighbor*

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,



Iin Ernawati, S.Kom., M.Si.

Ketua Penguji



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

Anggota Penguji



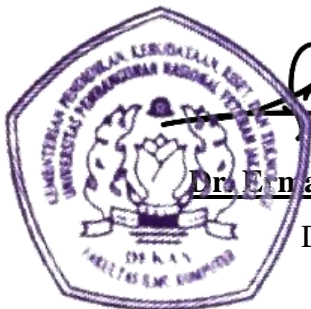
Yuni Widiastiwi, S.Kom., M.Si.

Pembimbing I



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

Pembimbing II



Dr. Ernawati, M.Kom.

Dekan



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 : Pascal Aldwin Hernando
NIM : 1710511044
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:

**ANALISIS SENTIMEN PENGARUH COVID-19 TERHADAP
KEHIDUPAN BERMASYARAKAT PADA MEDIA SOSIAL TWITTER
MENGUNAKAN *K-NEAREST NEIGHBOR***

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,



(Pascal Aldwin Hernando)

**ANALISIS SENTIMEN PENGARUH COVID-19 TERHADAP
KEHIDUPAN BERMASYARAKAT PADA MEDIA SOSIAL TWITTER
MENGUNAKAN *K-NEAREST NEIGHBOR***

Pascal Aldwin Hernando

ABSTRAK

Coronavirus Disease (COVID-19) merupakan sebuah pandemi yang terjadi karena terjadinya penyebaran coronavirus yang baru saja ditemukan. COVID-19 menyebar melalui air liur/*saliva* ataupun obyek yang keluar dari hidung dari seseorang yang terinfeksi batuk atau bersin. Dengan munculnya wabah ini, membuat seluruh dunia termasuk Indonesia mengalami kemunduran dalam berbagai bidang kehidupan. Dengan adanya COVID-19 ini, membuat pemerintah menetapkan *lockdown* pada berbagai daerah yang menyebabkan aktifitas setiap orang menjadi terbatas dari rumah saja. Untuk mengungkapkan keadaan selama *lockdown*, maka orang-orang menggunakan media sosial untuk sebagai perantara, dimana salah satunya adalah *Twitter*. Dengan banyaknya penggunaan *twitter*, membuat data yang berhubungan tentang COVID-19 untuk dilakukan *sentiment analysis* menggunakan metode klasifikasi *K-Nearest Neighbor*. Dengan merujuk pada daerah tertentu, yaitu Jakarta dan Surabaya dilakukan pengambilan data untuk dilakukan pengujian dengan KNN. Hasil dari pengujian menampilkan bahwa pada data Jakarta mendapatkan akurasi terbaik pada 78.95% dengan nilai $K=9$, sementara pada data Surabaya mendapatkan akurasi terbaik pada 76.47% dengan nilai $K=11,17,19,20$.

Kata Kunci: *COVID-19, Twitter, sentiment analysis, K-Nearest Neighbor.*

SENTIMENT ANALYSIS OF THE EFFECT OF COVID-19 ON COMMUNITY LIFE ON TWITTER SOCIAL MEDIA USING K- NEAREST NEIGHBOR

Pascal Aldwin Hernando

ABSTRACT

Coronavirus Disease (COVID-19) is a pandemic that occurs due to the spread of the newly discovered coronavirus. COVID-19 is spread through saliva or objects that come out of the nose of an infected person coughing or sneezing. With the emergence of this pandemic, the whole world, including Indonesia, has experienced a setback in various fields of life. With this COVID-19, the government has imposed a lockdown in various areas which has caused everyone's activities to be limited from home. To reveal the situation during the lockdown, people use social media as an intermediary, one of which is Twitter. With so many uses of twitter, making data related to COVID-19 for sentiment analysis using the classification K-Nearest Neighbor. By referring to certain areas, namely Jakarta and Surabaya, data were collected for testing with KNN. The results of the test show that the Jakarta data get the best accuracy at 79.85% with a value of $K = 9$, while the Surabaya data get the best accuracy at 76.47% with a value of $K = 11, 17, 19, 20$.

Keywords: *COVID-19, Twitter, sentiment analysis, K-Nearest Neighbor.*

KATA PENGANTAR

Puji dan syukur penulis ucapkan pada ke hadirat Tuhan Yang Maha Esa atas segala berkah dan karunianya sehingga skripsi ini dapat diselesaikan di tengah masa pandemi. Penulis ingin mengucapkan terimakasih kepada:

1. Keluarga penulis, teruntuk Ayah dan Ibu penulis, yang selalu memberikan dukungan dalam berbagai hal sehingga penulis dapat menyelesaikan skripsi ini.
2. Ibu Yuni Widiastiwi, S.Kom., M.Si., dan Ibu Mayanda Mega Santoni S.Kom., M.Kom. selaku dosen pembimbing penulis yang telah memberikan arahan, saran, serta dorongan dalam pengerjaan skripsi ini.
3. Ibu Iin Ernawati, S.Kom., M.Si., dan Ibu Nurul Chamidah, S.Kom., M.Kom. selaku dosen penguji yang telah memberikan arahan dan saran komprehensif.
4. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta yang telah memberi ilmu yang banyak dan bermanfaat.
5. Teman-teman dari Tim NaQoS yang telah memberi dukungan moral serta menyediakan waktu untuk melakukan diskusi tentang tugas akhir.
6. Teman-teman satu bimbingan yang selalu memberikan masukan, saran, dan semangat kepada penulis.
7. Teman-teman di Jurusan Informatika yang telah menjalankan masa-masa perkuliahan bersama penulis.

Penulis menyadari bahwa masih terdapat kekurangan dalam penulisan skripsi ini. Semoga skripsi ini bermanfaat.

Jakarta, Juli 2021



Pascal Aldwin Hernando

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	xi
DAFTAR TABEL	xii
BAB 1. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Ruang Lingkup Penelitian	2
1.4. Tujuan Penelitian	2
1.5. Manfaat Penelitian	3
1.6. Luaran Yang Diharapkan	3
1.7. Sistematika Penulisan	3
BAB 2. TINJAUAN PUSTAKA	5
2.1. <i>Coronavirus Disease 2019 (COVID-19)</i>	5
2.2. <i>Text Mining</i>	5
2.2.1. <i>Case Folding</i>	5
2.2.2. <i>Stemming</i>	6
2.2.3. <i>Filtering</i>	6

2.2.4.	<i>Tokenization</i>	6
2.2.5.	Normalisasi Bahasa.....	6
2.3.	<i>Sentiment Analysis</i>	7
2.4.	<i>Kappa Value</i>	7
2.5.	<i>K-Nearest Neighbor</i>	8
2.6.	<i>Term Frequency – Inverse Document Frequency (TF-IDF)</i>	8
2.6.1.	<i>Term Frequency (TF)</i>	9
2.6.2.	<i>Document Frequency (DF)</i>	9
2.6.3.	<i>Inverse Document Frequency (IDF)</i>	9
2.7.	Normalisasi <i>Min-Max</i>	9
2.8.	<i>Confusion Matrix</i>	10
2.9.	Penelitian Terkait	11
BAB 3. METODOLOGI PENELITIAN.....		14
3.1.	Kerangka Pikir.....	14
3.1.1.	Identifikasi Masalah	14
3.1.2.	Studi Literatur	15
3.1.3.	Pengambilan Data	15
3.1.4.	Pelabelan Data.....	16
3.1.5.	<i>Preprocessing</i>	16
3.1.6.	Pembobotan TF-IDF	17
3.1.7.	<i>K-Nearest Neighbor</i>	17
3.1.8.	<i>Output Data</i>	18
3.1.9.	Dokumentasi	18
3.2.	Alat Bantu Penelitian.....	18
3.3.	Jadwal Penelitian	19
BAB 4. HASIL DAN PEMBAHASAN.....		21

4.1.	Pengumpulan Data	21
4.2.	Pelabelan Data	22
4.3.	<i>Preprocessing</i>	24
4.3.1.	Case Folding	25
4.3.2.	<i>Stemming</i>	26
4.3.3.	Filtering	26
4.3.4.	<i>Tokenization</i>	27
4.3.5.	Normalisasi Bahasa	28
4.4.	<i>Undersampling</i>	29
4.5.	Pembobotan TF-IDF	30
4.6.	K-Nearest Neighbor	32
4.6.1.	Data Split	32
4.6.2.	Model Train	32
4.6.3.	Predict Model	33
4.7.	Summary Data Jakarta dan Data Surabaya	45
BAB 5. KESIMPULAN DAN SARAN		47
5.1.	Kesimpulan	47
5.2.	Saran	48
DAFTAR PUSTAKA		49
DAFTAR RIWAYAT HIDUP		51
LAMPIRAN		52

DAFTAR GAMBAR

Gambar 1. Stemming	6
Gambar 2. Kerangka Pikir.....	14
Gambar 3. Akurasi K Percobaan 1 Jakarta	34
Gambar 4. Akurasi K Percobaan 2 Jakarta	36
Gambar 5. Akurasi K Percobaan 3 Jakarta	37
Gambar 6. Akurasi K Percobaan 1 Surabaya.....	40
Gambar 7. Akurasi K Percobaan 2 Surabaya.....	41
Gambar 8. Akurasi K Percobaan 2 Surabaya.....	43

DAFTAR TABEL

Tabel 1. Kappa Interpretation	8
Tabel 2. Confusion Matrix	10
Tabel 3. Penelitian Terkait	12
Tabel 4. Jadwal Kegiatan Penelitian	19
Tabel 5. Pengumpulan Data	22
Tabel 6. Contoh Pelabelan Data.....	22
Tabel 7. Hasil Pelabelan Data	24
Tabel 8. Contoh Preprocess Case Folding	25
Tabel 9. Contoh Preprocess Stemming	26
Tabel 10. Contoh Preprocess Filtering.....	27
Tabel 11. Contoh Preprocess Tokenization	27
Tabel 12. Contoh Preprocess Normalisasi Bahasa.....	28
Tabel 13. Hasil Pelabelan Setelah Preprocess	29
Tabel 14. Proses Undersampling.....	29
Tabel 15. Contoh Dokumen TF-IDF.....	30
Tabel 16. Contoh TF-IDF	30
Tabel 17. Contoh Normalisasi TF-IDF	31
Tabel 18. Contoh Perhitungan KNN.....	33
Tabel 19. Akurasi Percobaan 1 Jakarta	35
Tabel 20. Akurasi Percobaan 2 Jakarta	36
Tabel 21. Akurasi Percobaan 3 Jakarta	37
Tabel 22. Akurasi Data Jakarta	38
Tabel 23. Akurasi Percobaan 1 Surabaya	40
Tabel 24. Akurasi Percobaan 2 Surabaya	42
Tabel 25. Akurasi Percobaan 3 Surabaya	43
Tabel 26. Akurasi Data Surabaya.....	44
Tabel 27. Akurasi Jakarta dan Surabaya	45