



**ANALISIS SENTIMEN TERHADAP PPKM DARURAT PADA
MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE
NAÏVE BAYES DENGAN SELEKSI FITUR *INFORMATION
GAIN***

SKRIPSI

ALBET DWI PANGESTU

1810511014

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

2022



**ANALISIS SENTIMEN TERHADAP PPKM DARURAT PADA
MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE
NAÏVE BAYES DENGAN SELEKSI FITUR *INFORMATION
GAIN***

SKRIPSI

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

ALBET DWI PANGESTU

1810511014

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2022**

PERNYATAAN ORISINALITAS

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

Nama : Albet Dwi Pangestu

NIM : 1810511014

Tanggal : 02 Juni 2022

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

Jakarta, 02 Juni 2022

Yang Menyatakan



(Albet Dwi Pangestu)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademis Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Albet Dwi Pangestu

NIM : 1810511014

Fakultas : Ilmu Komputer

Program Studi : 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 PPKM Darurat Pada Media Sosial *Twitter*
Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain***

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 02 Juni 2022
Yang menyatakan,



(Albet Dwi Pangestu)

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa proposal berikut:

Nama : Albet Dwi Pangestu
NIM : 1810511014
Program Studi : SI Informatika
Judul : Analisis Sentimen Terhadap PPKM Darurat Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain*

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


Javanta, S.Kom., M.Si
Penguji 1


Ati Zaidiah, S.Kom, MTL
Penguji 2


Iin Ernawati, S.Kom., M.Si
Pembimbing 1


Nurul Chamidah, S.Kom., M.Kom.
Pembimbing 2



Dr. Ermatita, M.Kom.
Dekan


Desta Sandya Pravista, S.Kom., M.Kom.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal pengesahan : 03 Juli 2022



ANALISIS SENTIMEN TERHADAP PPKM DARURAT PADA MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE *NAÏVE BAYES* DENGAN SELEKSI FITUR *INFORMATION GAIN*

ALBET DWI PANGESTU

ABSTRAK

Twitter merupakan media sosial yang digunakan oleh masyarakat sebagai media berkomunikasi dan pengutaraan pendapat. Semenjak pandemik COVID-19 melanda Indonesia, pemerintah telah banyak mengeluarkan kebijakan-kebijakan untuk menekan penyebaran COVID-19, salah satunya adalah PPKM Darurat. Banyak opini masyarakat yang mengkritik ataupun mendukung kebijakan tersebut di media sosial, khususnya *twitter*. Penelitian ini bertujuan untuk membangun model analisis sentimen terhadap PPKM Darurat di media sosial *twitter* dengan tagar #ppkmdarurat. Pada penelitian ini akan menggunakan metode *Naïve Bayes* serta metode *Information Gain* sebagai seleksi fitur. Penelitian ini juga akan melakukan perbandingan penggunaan *Information Gain* dan tidak menggunakan *Information Gain* sebagai seleksi fitur. Pengumpulan data akan dilakukan *crawling* menggunakan bahasa pemrograman R dan terintegrasi pada *API* yang telah disediakan oleh *twitter*. Setelah dilakukan *filtering* data menjadi 770 dengan berlabel 335 positif dan 335 negatif. Hasil dari pengujian model klasifikasi *Naïve Bayes* terjadi peningkatan performa apabila menggunakan seleksi fitur *Information Gain* dengan nilai pengambilan *top ranking* '>0.0001' yaitu akurasi 0.81, *recall* 0.82, *precision* 0.84, *f1 score* 0.83 dan *specificity* 0.79 dibandingkan sebelumnya yaitu akurasi 0.79, *recall* 0.81, *precision* 0.81, *f1 score* 0.81 dan *specificity* 0.76.

Kata kunci: Analisis sentimen, PPKM Darurat, *Twitter*, *Naïve Bayes*, *Information Gain*

ANALYSIS OF EMERGENCY PPKM ON TWITTER SOCIAL MEDIA USING NAÏVE BAYES METHOD WITH INFORMATION GAIN FEATURE SELECTION

ALBET DWI PANGESTU

ABSTRACT

Twitter is a social media used by the public as a medium for communicating and expressing opinions. Since the COVID-19 pandemic hit Indonesia, the government has issued many policies to suppress the spread of COVID-19, one of which is *PPKM Darurat*. Many public opinions criticize or support this policy on social media, especially Twitter. This study aims to build a sentiment analysis model for *PPKM Darurat* on Twitter social media with the hashtag #ppkmdarurat. In this study, the Naïve Bayes method and the Information Gain method will be used as feature selection. This study will compare the use of Information Gain and not use Information Gain as a feature selection. Data collection will be crawled using the R programming language and integrated into the API provided by Twitter. After filtering the data, it became 770 labeled 335 positive and 335 negative. The results of testing the Naïve Bayes classification model showed an increase in performance when using the Information Gain feature selection with a top ranking value of '> 0.0001', namely accuracy 0.81, recall 0.82, precision 0.84, f1 score 0.83 and specificity 0.79 compared to the previous one, namely accuracy 0.79, recall 0.81, precision 0.81, f1 score 0.81 and specificity 0.76.

Keywords: Sentiment analysis, *PPKM Darurat*, Twitter, Naïve Bayes, Information Gain

KATA PENGANTAR

Segala syukur dan puji hanya bagi Tuhan Yesus Kristus, oleh karena anugerah-Nya yang melimpah, kemurahan dan kasih setia yang besar akhirnya penulis dapat menyelesaikan Tugas Akhir/skripsi yang berjudul “**Analisis Sentimen Terhadap PPKM Darurat Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain***” dengan baik.

Dalam penyelesaian tugas akhir ini, penulis tidak lupa mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan masukan kepada penulis, pihak-pihak tersebut adalah:

1. Kedua orang tua penulis, Bambang Sugiarto (bapak), Iti Tumi Asni (ibu), Frengky Eko Saputro (kakak) dan Ruben Jaya Kusuma (adik), serta keluarga yang selalu memberi dorongan dan doa sehingga dapat menyelesaikan Skripsi ini.
2. Ibu Iin Ernawati, S.kom., M.Si., dan Ibu Nurul Chamidah, S.Kom., M.Kom., selaku dosen pembimbing yang telah memberikan saran dan masukan yang sangat bermanfaat.
3. Ibu Dr. Ermatita, M. Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta terima kasih atas ilmu-ilmu yang selama ini sudah diajarkan.
5. Kekasih saya Nengsi Pakpahan yang selalu menemani dan memotivasi saya dalam menempuh skripsi.
6. Teman-teman perkumpulan Pelabelan Geng yang selalu menemani dan memberikan dukungan.
7. Seluruh pihak yang terlibat dalam kelancaran pembuatan tugas akhir ini dan yang belum disebutkan di atas, penulis ucapkan terima kasih.

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

Jakarta, 24 Mei 2022

Penulis

DAFTAR ISI

ANALISIS SENTIMEN TERHADAP PPKM DARURAT PADA MEDIA SOSIAL <i>TWITTER</i> MENGGUNAKAN METODE <i>NAÏVE BAYES</i> DENGAN SELEKSI FITUR <i>INFORMATION GAIN</i>.....	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iii
LEMBAR PENGESAHAN.....	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Ruang Lingkup.....	3
1.5 Luaran yang Diharapkan	3
1.6 Manfaat Penelitian	4
1.7 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 <i>Text Mining</i>	6

2.2	Analisis Sentimen	6
2.3	<i>Twitter</i>	7
2.4	PPKM Darurat	7
2.5	<i>Preprocessing</i>	7
2.5.1	<i>Cleaning</i>	8
2.5.2	<i>Case Folding</i>	8
2.5.3	Normalisasi Bahasa	8
2.5.4	<i>Stopword Removal</i>	8
2.5.5	<i>Stemming</i>	8
2.5.6	Tokenisasi	8
2.6	Pembobotan Term (TF-IDF)	8
2.7	Seleksi Fitur	9
2.8	Information Gain	10
2.9	Klasifikasi Naïve Bayes	11
2.10	Evaluasi	13
2.11	Review Riset Relevan	14
BAB III METODE PENELITIAN		16
3.1	Tahapan Penelitian	16
3.1.1	Identifikasi Masalah	16
3.1.2	Studi Literatur	17
3.1.3	Pengumpulan Data	17
3.1.4	<i>Filtering</i>	17
3.1.5	Pelabelan Data	17
3.1.6	<i>Preprocessing</i>	17
3.1.7	Pembobotan Term (TF-IDF)	19
3.1.8	Seleksi Fitur Information Gain	19

3.1.9	Klasifikasi <i>Naïve Bayes</i>	21
3.1.10	Evaluasi	21
3.1.11	Skenario Pengujian	21
3.2	Alat yang Digunakan.....	22
3.3	Jadwal Penelitian.....	23
BAB IV HASIL DAN PEMBAHASAN.....		24
4.1	Data.....	24
4.2	<i>Filtering</i>	25
4.3	Pelabelan Data	26
4.4	<i>Preprocessing</i>	28
4.4.1	<i>Cleaning</i>	28
4.4.2	<i>Case Folding</i>	29
4.4.3	Normalisasi Bahasa.....	29
4.4.4	<i>Stopword Removal</i>	31
4.4.5	<i>Stemming</i>	32
4.4.6	Tokenisasi	32
4.5	Pembobotan <i>Term</i> (TF-IDF).....	34
4.6	Seleksi Fitur <i>Information Gain</i>	36
4.6.1	Pemisahan fitur sesuai label.....	36
4.6.2	Hitung nilai <i>entropy</i>	38
4.6.3	Hitung nilai <i>entropy</i> pada fitur	40
4.6.4	Hitung nilai <i>Information Gain</i>	42
4.6.5	Pengurutan serta pengambilan <i>top ranking</i>	43
4.7	Klasifikasi <i>Naïve Bayes</i>	44
4.8	Evaluasi	49
4.9	Skenario Pengujian	51

BAB V PENUTUP	53
5.1 Kesimpulan	53
5.2 Saran	54
DAFTAR PUSTAKA	55
RIWAYAT HIDUP	59
LAMPIRAN	60

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	16
Gambar 3.2 <i>Preprocessing</i>	18
Gambar 3.3 Proses <i>Information Gain</i>	19
Gambar 4.1 Hasil <i>Crawling</i>	24
Gambar 4.2 Visualisasi Perbandingan Data Label Positif dan Negatif.....	27
Gambar 4.3 <i>Corpus</i> Normalisasi Bahasa	30
Gambar 4.4 Visualisasi <i>Word Cloud</i> Data <i>Tweet</i> Label Positif.....	33
Gambar 4.5 Visualisasi <i>Word Cloud</i> Data <i>Tweet</i> Label Negatif.....	34

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	13
Tabel 3.1 Jadwal Penelitian.....	23
Tabel 4.1 Contoh <i>Tweet</i> pada Proses <i>Filtering</i>	25
Tabel 4.2 Contoh Pelabelan Data.....	26
Tabel 4.3 <i>Cleaning</i>	28
Tabel 4.4 <i>Case Folding</i>	29
Tabel 4.5 Normalisasi Bahasa.....	30
Tabel 4.6 <i>Stopword Removal</i>	31
Tabel 4.7 <i>Stemming</i>	32
Tabel 4.8 Tokenisasi	33
Tabel 4.9 Contoh Data Proses TF-IDF	34
Tabel 4.10 Perhitungan TF-IDF	35
Tabel 4.11 Contoh Hasil TF-IDF	37
Tabel 4.12 Pemisahan Fitur Sesuai Label	37
Tabel 4.13 Perhitungan Nilai <i>Entropy</i>	39
Tabel 4.14 Perhitungan Nilai <i>Entropy</i> pada Fitur	41
Tabel 4.15 Perhitungan Nilai <i>Information Gain</i>	43
Tabel 4.16 Pembagian Data <i>Training</i> dan <i>Testing</i>	44
Tabel 4.17 Hasil Perhitungan Peluang Setiap Kata Data <i>Training</i> Berdasarkan Kelas Positif dan Negatif	45
Tabel 4.18 Contoh Data <i>Testing</i>	46
Tabel 4.19 TF-IDF Data <i>Testing</i>	47
Tabel 4.20 Hasil Perhitungan Peluang Setiap Kata Data <i>Testing</i> Berdasarkan Kelas Positif dan Negatif	48
Tabel 4.21 <i>Confusion Matrix</i> Tanpa <i>Information Gain</i>	49
Tabel 4.22 <i>Confusion Matrix</i> Menggunakan <i>Information Gain</i>	50

Tabel 4.23 Hasil Skenario Pengujian	51
Tabel 4.24 Ringkasan Jumlah Fitur	53

DAFTAR LAMPIRAN

Lampiran 1 <i>Similarity Index</i> Skripsi	60
--	----