



**ANALISIS SENTIMEN TERHADAP RESESI 2023 PADA  
MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE  
*NAÏVE BAYES* DENGAN SELEKSI FITUR *INFORMATION  
GAIN***

**SKRIPSI**

**HAFIDZ ASHABI MUHAMMAD**

**NIM. 1910511113**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN**

**JAKARTA**

**2023**



**ANALISIS SENTIMEN TERHADAP RESESI 2023 PADA  
MEDIA SOSIAL *TWITTER* MENGGUNAKAN METODE  
*NAÏVE BAYES* DENGAN SELEKSI FITUR *INFORMATION  
GAIN***

**SKRIPSI**

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

**Sarjana Komputer**

**HAFIDZ ASHABI MUHAMMAD**

**NIM. 1910511113**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN**

**JAKARTA**

**2023**

## PERNYATAAN ORISINALITAS

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

Nama : Hafidz Ashabi Muhammad

NIM : 1910511113

Tanggal : 24 Januari 2024

Judul Skripsi : **Analisis Sentimen Terhadap Resesi 2023 Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain***

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

Jakarta, 24 Januari 2024

Yang menyatakan,



Hafidz Ashabi Muhammad

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK  
KEPENTINGAN AKADEMIS**

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

Nama : Hafidz Ashabi Muhammad  
NIM : 1910511113  
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 berjudul:

**Analisis Sentimen Terhadap Resesi 2023 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 dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada Tanggal : 24 Januari 2024

Yang menyatakan,




Hafidz Ashabi Muhammad

## LEMBAR PENGESAHAN


Tugas Akhir ini diajukan oleh:


Nama : Hafidz Ashabi Muhammad  
NIM : 1910511113  
Program Studi : S1 Informatika  
Judul Tugas Akhir : Analisis Sentimen Terhadap Resesi 2023 Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain*


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

  
**Dr. Ermatita, M.Kom**  
Penguji I

  
**Kraugusteeliana, S.Kom., M.Kom., M.M.**  
Penguji II

  
**Nur Hafifah Matondang, S.Kom, MM**  
Pembimbing I

  
**Prof. Dr. Ir. Suprivanto, ST., M.Sc., IPM.**  
Dekan

  
**Dr. Widya Cholil, M.I.T.**  
Kepala Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 5 Januari 2024

**ANALISIS SENTIMEN TERHADAP RESESI 2023 PADA MEDIA SOSIAL  
TWITTER MENGGUNAKAN METODE NAÏVE BAYES DENGAN  
SELEKSI FITUR *INFORMATION GAIN***

**Hafidz Ashabi Muhammad**

**ABSTRAK**

Dalam menghadapi dinamika ekonomi global, resesi menjadi kritikal. Media sosial, khususnya Twitter, menjadi saluran utama untuk berekspresi dan berbagi pandangan terkait kondisi ekonomi. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap resesi tahun 2023 berdasarkan data yang diambil dari percakapan di media sosial Twitter. Metode analisis sentimen yang digunakan adalah Naïve Bayes, yang dikenal karena kesejukannya dalam penanganan data teks. Selain itu, seleksi fitur *Information Gain* diaplikasikan untuk memilih kata-kata atau fitur yang paling informatif dalam memprediksi sentimen. *Information Gain* memungkinkan kita untuk mengevaluasi kontribusi relatif dari setiap kata terhadap penentuan sentimen. Data yang digunakan dalam penelitian ini terdiri dari sejumlah besar tweet yang diambil pada periode yang relevan dengan konteks resesi 2023. Proses pengolahan data melibatkan tahap-tahap seperti tokenisasi, penghilangan stop words, dan normalisasi teks. 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*. 2. Dari hasil penelitian menunjukkan adanya peningkatan model Naïve Bayes apabila menggunakan seleksi fitur *Information Gain* dengan nilai pengambilan *top ranking* '>0.01' yaitu akurasi 0.96, *recall* 1, *precision* 0.93, *f1 score* 0.96 dan *specificity* 0.93 dibandingkan sebelumnya tanpa seleksi fitur *information gain* yaitu akurasi 0.92, *recall* 1, *precision* 0.85, *f1 score* 0.91 dan *specificity* 1.

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

## **ABSTRACT**

*In facing global economic dynamics, recession becomes critical. Social media, especially Twitter, has become the main channel for expression and sharing views regarding economic conditions. This research aims to conduct sentiment analysis regarding the 2023 recession based on data taken from conversations on Twitter social media. The sentiment analysis method used is Naïve Bayes, which is known for its coolness in handling text data. In addition, Information Gain feature selection is applied to select the most informative words or features in predicting sentiment. Information Gain allows us to evaluate the relative contribution of each word to sentiment determination. The data used in this research consists of a large number of tweets taken in a period relevant to the context of the 2023 recession. The data processing process involves stages such as tokenization, removal of stop words, and text normalization. This research will also compare the use of Information Gain and not using Information Gain as feature selection. Data collection will be carried out by crawling using the R programming language and integrated with the API provided by Twitter. 2. The research results show that there is an improvement in the Naïve Bayes model when using the Information Gain feature selection with a top ranking value of '>0.01', namely accuracy 0.96, recall 1, precision 0.93, f1 score 0.96 and specificity 0.93 compared to before without information gain feature selection, namely accuracy 0.92, recall 1, precision 0.85, f1 score 0.91 and specificity 1.*

**Keywords:** *Sentiment Analysis, Twitter, Naïve Bayes, Information Gain*

## KATA PENGANTAR

Alhamdulillah, puji dan syukur peneliti ucapkan ke hadirat Allah SWT yang telah memberikan banyak rahmat, kesehatan, dan segala karunia-Nya sehingga proses penyusunan dan penyelesaian skripsi dengan judul “Analisis Sentimen Terhadap Resesi 2023 Pada Media Sosial *Twitter* Menggunakan Metode *Naïve Bayes* Dengan Seleksi Fitur *Information Gain*” dapat peneliti laksanakan dengan baik.

Skripsi ini disusun oleh peneliti guna memenuhi salah satu syarat untuk dapat memperoleh Gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jakarta. Dalam penyusunan skripsi ini, peneliti juga mendapat banyak dukungan, bimbingan, bantuan, dan kerja sama dari berbagai pihak baik dari segi moral, material, maupun spiritual sehingga peneliti dapat menyelesaikan skripsi ini dengan baik. Maka dari itu, peneliti mengucapkan terima kasih kepada:

1. Ibu, Ayah, Kakak, Adik, dan keluarga peneliti yang tidak hentinya memberikan dukungan, semangat, dan doa sehingga peneliti dapat menyelesaikan penelitian dan skripsi ini.
2. Nur Hafifah Matondang, S.Kom, MM sebagai dosen pembimbing yang telah membimbing dan senantiasa memberikan semangat dan doa untuk peneliti dalam menyusun skripsi.
3. Ibu Dr. Widya Cholil, M.I.T. sebagai ketua program studi Informatika.
4. Ibu Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM. sebagai Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jakarta.
5. Bapak/Ibu Dosen Fakultas Ilmu Komputer yang berdedikasi mengajar dan membagikan ilmu sehingga peneliti dapat memperluas wawasan serta ilmu pengetahuan selama menjadi mahasiswa di Fakultas Ilmu Komputer.
6. Sahabat-sahabat peneliti yang senantiasa mendukung, memberikan semangat, dan mendoakan peneliti dalam melakukan penelitian dan menyusun skripsi ini.



7. Teman-teman peneliti dan seluruh pihak lain yang tidak dapat disebutkan satu per satu atas kontribusi dalam penyusunan skripsi baik secara langsung maupun tidak langsung.

Peneliti menyadari bahwa pada penelitian dan skripsi ini masih terdapat banyak kekurangan dan kesalahan. Peneliti berharap skripsi ini dapat memberikan informasi dan pengetahuan yang berharga bagi pembaca.

Jakarta, 22 Januari 2023

Peneliti,

Hafidz Ashabi Muhammad

## DAFTAR ISI

	Halaman
PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI .....	iv
LEMBAR PENGESAHAN .....	v
ABSTRAK .....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	x
DAFTAR GAMBAR .....	xiv
DAFTAR TABEL.....	xv
BAB I .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Ruang Lingkup Penelitian.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian .....	4
1.6 Luaran yang Diharapkan .....	4
1.7 Sistematika Penulisan .....	4
BAB II.....	6
2.1 Analisis Sentimen.....	6
2.2 <i>Text Mining</i> .....	7
2.3 <i>Fleiss Kappa</i> .....	8
2.4 Resesi Ekonomi.....	9
2.5 Media Sosial <i>Twitter</i> .....	9
2.6 <i>Text Preprocessing</i> .....	10

2.5.1	<i>Tokenizing</i> .....	11
2.5.2	<i>Stopword Removal</i> .....	11
2.5.3	<i>Stemming</i> .....	11
2.5.4	<i>Data Cleansing</i> .....	11
2.5.5	<i>Normalization</i> .....	11
2.7	Pembobotan Term (TF-IDF).....	11
2.8	Seleksi Fitur.....	13
2.8.1	<i>Information Gain</i> .....	14
2.9	Klasifikasi <i>Naïve Bayes</i> .....	15
2.10	Pembagian Data Latih dan Data Uji.....	18
2.11	Evaluasi.....	19
2.12	Penelitian Terdahulu.....	21
BAB III.....		26
3.1	Tahapan Penelitian.....	26
3.1.1	Identifikasi Masalah.....	26
3.1.2	Studi Literatur.....	27
3.1.3	Pengumpulan Data.....	27
3.1.4	<i>Filtering</i> .....	27
3.1.5	Pelabelan Data.....	27
3.1.6	<i>Preprocessing</i> .....	28
3.1.7	Pembobotan Term (TF-IDF).....	29
3.1.8	Synthetic Minority Over-sampling Technique (SMOTE).....	30
3.1.9	Seleksi Fitur <i>Information Gain</i> .....	31
3.1.10	Klasifikasi <i>Naïve Bayes</i> .....	31
3.1.11	Evaluasi.....	32
3.2	Perangkat Penelitian.....	33

3.3 Penjadwalan Penelitian .....	34
BAB IV .....	35
4.1 Pengumpulan Data .....	35
4.2 <i>Filtering</i> .....	36
4.3 Pelabelan Data.....	36
4.4 <i>Preprocessing</i> .....	37
4.4.1 <i>Cleaning</i> .....	37
4.4.2 <i>Case Folding</i> .....	38
4.4.3 <i>Normalization</i> .....	39
4.4.4 <i>Stopword Removal</i> .....	40
4.4.5 <i>Stemming</i> .....	41
4.4.6 <i>Tokenizing</i> .....	42
4.5 Pembobotan Term (TF-IDF) .....	43
4.6 SMOTE ( <i>Synthetic Minority Oversampling Technique</i> ).....	45
4.7 Seleksi Fitur <i>Information Gain</i> .....	46
4.6.1 Pemisahan Fitur Sesuai Label .....	46
4.6.2 Hitung Nilai <i>Entropy</i> .....	49
4.6.3 Hitung Nilai <i>Entropy</i> Pada Fitur .....	52
4.6.4 Hitung Nilai <i>Information Gain</i> .....	54
4.6.5 Pengurutan Serta Pengambilan <i>Top Ranking</i> .....	55
4.8 Klasifikasi Naïve Bayes .....	56
4.8.1 <i>Data Training</i> .....	56
4.8.2 <i>Data Testing</i> .....	58
4.9 Evaluasi .....	61
4.10 Analisis Hasil Kegiatan .....	63
4.10.1 <i>Wordcloud</i> Sentimen Positif Terhadap Data <i>Tweet</i> Resesi.....	63

4.10.2 <i>Wordcloud</i> Sentimen Negatif Terhadap Data <i>Tweet</i> Resesi .....	64
BAB V .....	65
PENUTUP .....	65
5.1 Kesimpulan .....	65
5.2 Saran.....	66
DAFTAR PUSTAKA .....	67
LAMPIRAN .....	70

## DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian .....	26
Gambar 4.1 Hasil Crawling Data .....	35
Gambar 4.2 <i>Wordcloud</i> Sentimen Positif.....	64
Gambar 4.3 <i>Wordcloud</i> Sentimen Negatif .....	64

## DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i> .....	19
Tabel 3.1 Tabel Penjadwalan Penelitian.....	34
Tabel 4.1 Tabel Hasil Pelabelan Data.....	36
Tabel 4.2 <i>Cleaning</i> .....	38
Tabel 4.3 <i>Case Folding</i> .....	38
Tabel 4.4 <i>Normalization</i> .....	39
Tabel 4.5 <i>Stopword Removal</i> .....	40
Tabel 4.6 <i>Stemming</i> .....	41
Tabel 4.7 Tokenizing .....	42
Tabel 4.8 Contoh Data Proses TF- IDF .....	43
Tabel 4.9 Perhitungan TF-IDF .....	43
Tabel 4.10 Jumlah data sebelum dan sesudah SMOTE .....	46
Tabel 4.11 Contoh Hasil TF-IDF .....	46
Tabel 4.12 Pemisahan Fitur Sesuai Label .....	47
Tabel 4.13 Perhitungan Nilai <i>Entropy</i> .....	49
Tabel 4.14 Perhitungan Nilai <i>Entropy</i> Pada Fitur .....	52
Tabel 4.15 Perhitungan Nilai <i>Information Gain</i> .....	54
Tabel 4.16 Hasil Perhitungan Peluang Setiap Kata <i>Data Training</i> Berdasarkan Kelas Positif dan Negatif .....	57
Tabel 4.17 Contoh <i>Data Testing</i> .....	58
Tabel 4.18 TF-IDF <i>Data Testing</i> .....	59
Tabel 4.19 Hasil Perhitungan Peluang Setiap Kata <i>Data Testing</i> Berdasarkan Kelas Positif dan Negatif.....	60
Tabel 4.20 <i>Confusion matrix</i> tanpa seleksi fitur <i>Information Gain</i> .....	62
Tabel 4.21 <i>Confusion matrix</i> menggunakan seleksi fitur <i>Information Gain</i> .....	62