

**ANALISIS PERBANDINGAN EKSTRAKSI FITUR TEKS PADA
SENTIMEN ANALISIS KENAIKAN HARGA BBM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

BRIGA DARMAWAN

19.11.2831

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**ANALISIS PERBANDINGAN EKSTRAKSI FITUR TEKS PADA
SENTIMEN ANALISIS KENAIKAN HARGA BBM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

BRIGA DARMAWAN

19.11.2831

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN EKSTRAKSI FITUR TEKS PADA
SENTIMEN ANALISIS KENAIKAN HARGA BBM**

yang disusun dan diajukan oleh

Briga Darmawan

19.11.2831

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Juli 2023

Dosen Pembimbing,



Arif Dwi Laksito, M.Kom

NIK. 190302150

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS PERBANDINGAN EKSTRAKSI FITUR TEKS PADA
SENTIMEN ANALISIS KENAIKAN HARGA BBM

yang disusun dan diajukan oleh

Briga Darmawan

19.11.2831

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Juli 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Donni Prabowo, M.Kom
NIK. 190302253

Ike Verawati, M.Kom
NIK. 190302237

Arif Dwi Laksito, M.Kom
NIK. 190302150



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Juli 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Briga Darmawan
NIM : 19.11.2831

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Perbandingan Ekstraksi Fitur Teks Pada Sentimen Analisis
Kenaikan Harga BBM**

Dosen Pembimbing : Arif Dwi Laksito, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 10 Juli 2023

Yang Menyatakan,



Briga Darmawan

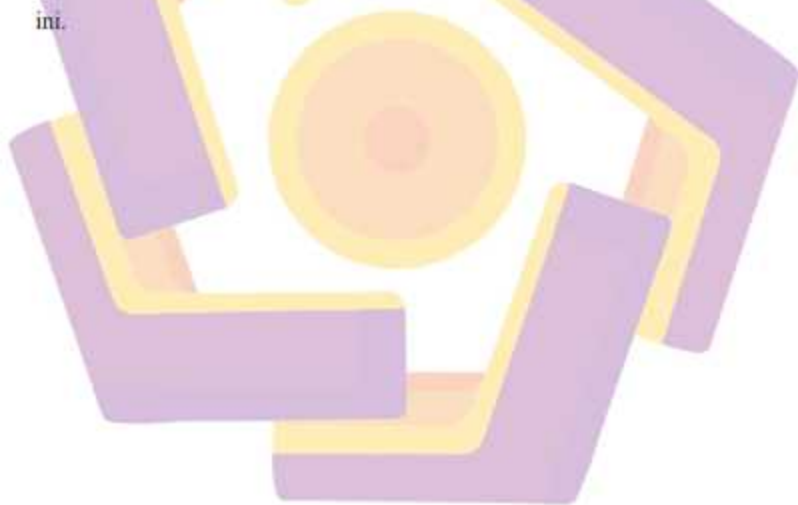
HALAMAN PERSEMBAHAN

Alhamdulillahirobbil'alamin atas segala rahmat yang telah diberikan kepada kita semua.

Terima kasih kepada kedua orang tua saya yang sudah memberikan dukungan dan doa sehingga saya dapat menyelesaikan skripsi ini dengan baik dan lancar. Tentunya dukungan dari mereka sangat berarti bagi saya.

Terima kasih kepada bapak Dosen Pembimbing saya, Pak Arif Dwi Laksito yang selalu membimbing dan mengarahkan saya dengan sabar, selalu mendukung proses pembuatan skripsi saya baik itu secara langsung atau tidak langsung.

Terima kasih kepada teman-teman saya dan semua pihak yang terlibat dalam proses pembuatan skripsi saya sehingga saya dapat menyelesaikan skripsi ini.



KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah memberikan Rahmat, Hidayah, dan Karunia-Nya kepada kita semua sehingga penulis dapat menyelesaikan naskah skripsi yang berjudul “ANALISIS PERBANDINGAN EKSTRAKSI FITUR TEKS PADA SENTIMEN ANALISIS KENAIKAN HARGA BBM”. Naskah skripsi ini disusun sebagai salah satu syarat untuk memenuhi mata kuliah skripsi dan syarat menyelesaikan program Strata-1 di jurusan Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya bantuan, dukungan, dan bimbingan dari berbagai pihak selama penyusunan naskah skripsi ini. Pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta
2. Windha Mega Pradnya D, M.Kom selaku Kaprodi Informatika
3. Arif Dwi Laksito, M.Kom selaku Dosen Pembimbing
4. Orang Tua Penulis, Bapak Handoko dan Ibu Juwaridah
5. Kakak Perempuan penulis, Hanidah Cahya
6. Sahabat-sahabat dan teman seperjuangan

Penulis menyadari bahwa penulisan naskah skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu penulis dengan kerendahan hati mengharapkan saran dan kritik untuk membangun dan mengembangkan naskah skripsi ini.

Yogyakarta, 10 Juli 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori	9
2.2.1 Analisis Sentimen	9
2.2.2 Machine Learning	9
2.2.3 Support Vector Machine	10
2.2.4 Ekstraksi Fitur	12
2.2.5 Confusion Matrix	12
BAB III METODE PENELITIAN	15
3.1 Objek Penelitian	15

3.2 Alur Penelitian	15
3.2.1 Crawling Data	16
3.2.2 Preprocessing	18
3.2.3 Ekstraksi Fitur	22
3.2.4 Split Data	22
3.2.5 Support Vector Machine	22
3.2.6 Metode Evaluasi dan Analisis	22
3.3 Alat dan Bahan	23
3.3.1 Data Penelitian	23
3.3.2 Alat/instrumen	23
3.3.3 Perangkat Keras	23
3.3.4 Perangkat Lunak	23
BAB IV HASIL DAN PEMBAHASAN	24
4.1 Crawling Data	24
4.2 Preprocessing	26
4.3 Ekstraksi Fitur	29
4.4 Split Data	31
4.5 Support Vector Machine	31
4.6 Metode Evaluasi dan Analisis	32
BAB V PENUTUP	37
5.1 Kesimpulan	37
5.2 Saran	37
REFERENSI	38

DAFTAR TABEL

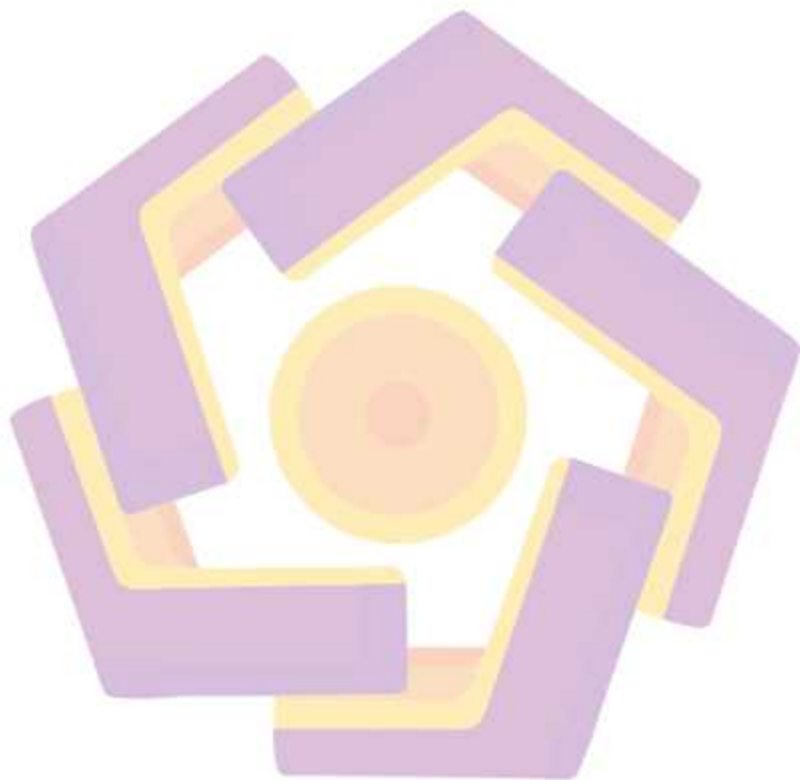
Tabel 2.1 Keaslian Penelitian	6
Tabel 3.1 Dataset Twitter.....	17
Tabel 3.2 <i>Cleaning Text</i>	18
Tabel 3.3 <i>Tokenization</i>	19
Tabel 3.4 <i>Stop Removal</i> atau <i>Stopwords</i>	20
Tabel 3.5 <i>Case Folding</i>	21
Tabel 3.6 <i>Labeling Data</i>	21
Tabel 4.1 Distribusi Data Setiap Sentimen	29
Tabel 4.2 Perhitungan TF-IDF.....	30
Tabel 4.3 Pehitungan <i>Bag of Word</i>	30
Tabel 4.4 Hasil Perbandingan Ekstraksi Fitur	36

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi SVM	11
Gambar 2.2 <i>Confusion Matrix</i> 3x3	13
Gambar 3.1 Alur Penelitian	15
Gambar 4.1 API Twitter	24
Gambar 4.2 Cara <i>Crawling Data</i>	24
Gambar 4.3 Pengambilan Kolom	25
Gambar 4.4 Hasil Pengambilan Data CSV	26
Gambar 4.5 Kode Program <i>Cleaning Text</i>	27
Gambar 4.6 Kode Program <i>Tokenization</i>	27
Gambar 4.7 Kode Program <i>Stop Removal</i>	27
Gambar 4.8 Kode Program <i>Case Folding</i>	28
Gambar 4.9 <i>Labeling Data</i> Manual	28
Gambar 4.10 Mendekati Kata Bors	31
Gambar 4.11 Kode Program Split Data	31
Gambar 4.12 Kode Program TF-IDF	32
Gambar 4.13 Kode Program <i>Bag of Word</i>	32
Gambar 4.14 Kode Program <i>FastText</i>	32
Gambar 4.15 Hasil pengukuran TF-IDF	33
Gambar 4.16 Hasil Pengukuran <i>Bag of Word</i>	33
Gambar 4.17 Hasil Pengukuran <i>FastText</i>	34
Gambar 4.18 Hasil <i>Score</i> TF-IDF	34
Gambar 4.19 Hasil <i>Score</i> BoW	35
Gambar 4.20 Hasil <i>Score</i> <i>FastText</i>	35

DAFTAR LAMBANG DAN SINGKATAN

TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
BoW	<i>Bag of Word</i>
SVM	<i>Support Vector Machine</i>



INTISARI

BBM merupakan bahan bakar yang digunakan kendaraan bermotor. Penggunaan BBM terus meningkat sejalan dengan pertumbuhan ekonomi di Indonesia. Ketersediaan BBM terus berkurang membuat harga BBM mengalami peningkatan yang menimbulkan banyak pendapat di kalangan masyarakat khususnya sosial media twitter. Fokus penelitian ini melakukan analisis sentimen terhadap Kenaikan BBM yang datanya didapat melalui twitter dengan jumlah data 1667 data. Tujuan dari penelitian ini melakukan perbandingan metode ekstraksi fitur yang memiliki kinerja paling baik seperti TF-IDF, *Bag of word*, FastText diuji dengan algoritma *machine learning Support Vector Machine* untuk klasifikasi teks. Untuk tahap penelitian yang pertama melakukan *crawling data twitter*, *pre-processing data*, analisis sentimen dengan ekstraksi fitur, pembuatan model dengan algoritma *machine learning Support Vector Machine*, dan kemudian dilakukan pengujian dan perbandingan model *confusion matrix* pada setiap ekstraksi fitur. Hasil dari penelitian ini menunjukkan bahwa penggunaan ekstraksi fitur *Bag of Word* memiliki kinerja lebih baik dengan nilai *accuracy* 65.3% dibandingkan model ekstraksi fitur TF-IDF dan *FastText* dengan nilai *accuracy* masing-masing 62.6% dan 59.3%.

Kata kunci: Analisis Sentimen, SVM, TF-IDF, BoW, FastText



ABSTRACT

BBM is the fuel used by motorized vehicles. The use of fuel is increasing in line with economic growth in Indonesia. The availability of fuel which continues to decrease, makes the prices rise, creating a lot of opinions among the public, especially on Twitter social media. This study focuses on carrying out a sentiment analysis of the increase in fuel prices, the data obtained via Twitter with a total of 1667 data. This study aims to compare feature extraction methods with the best performance, such as TF-IDF, Bag of Word, and FastText, evaluated with the Support Vector Machine machine learning algorithm. The study starts with crawling Twitter data, pre-processing, feature extraction, developing machine learning with Support Vector Machine algorithm, and then evaluating and comparing the confusion matrix model for each feature extraction. The results of this study indicate that that using the Bag of Word feature extraction has better performance with an accuracy value of 65.3% compared to the TF-IDF and FastText feature extraction models with an accuracy value of 62.6% and 59.3% respectively.

Keyword: Sentiment Analysis, SVM, TF-IDF, BoW, FastText.

