



**IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*  
UNTUK ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BBM  
PERTAMINA PADA MEDIA SOSIAL *TWITTER***

**SKRIPSI**

**IKHLASUL AMAL  
1910511125**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA  
2023**



**IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*  
UNTUK ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BBM  
PERTAMINA PADA MEDIA SOSIAL *TWITTER***

**SKRIPSI**

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

**Sarjana Komputer**

**IKHLASUL AMAL**

**1910511125**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA  
2023**

## PERNYATAAN ORISINALITAS

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

Nama : Ikhlasul Amal

NIM : 1910511125

Tanggal : 13 Juli 2023

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

Jakarta, 13 Juli 2023

Yang Menyatakan,



METERAL TEMPEL  
1000  
CBAKX098979875  
(Ikhlasul Amal)

## PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertandatangan di bawah ini:

Nama : Ikhlasul Amal

NIM : 1910511125

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 yang berjudul:

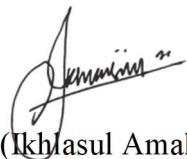
**IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)*  
UNTUK ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BBM  
PERTAMINA PADA MEDIA SOSIAL *TWITTER***

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 kata (basis data), 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 : 13 Juli 2023

Yang Menyatakan,

  
(Ikhlasul Amal)

## LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Ikhlasul Amal  
NIM : 1910511125  
Program Studi : S1 - Informatika  
Judul Tugas Akhir : Implementasi Algoritma Support Vector Machine (SVM) Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Media Sosial Twitter

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi S1 - Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.



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

Penguji I



(Rio Wirawan, S.Kom., MMSI.)

Penguji II



(Dr. Ermatita, M.Kom.)

Dekan



(Jayanta, S.Kom., M.Si.)

Pembimbing



(Dr. Widya Choli, M.I.T)

Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 3 Juli 2023



# IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE (SVM)* UNTUK ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BBM PERTAMINA PADA MEDIA SOSIAL *TWITTER*

IKHLASUL AMAL

## ABSTRAK

Pada 3 September 2022, pemerintah resmi mengumumkan kenaikan harga BBM berjenis Solar, Pertalite, dan Pertamax. Berbagai tanggapan dan keluhan masyarakat ditumpahkan, salah satunya lewat media sosial *Twitter*. Pengguna selaku masyarakat banyak sekali melakukan *tweet* di *Twitter* dengan kata kunci atau tagar (#) *bbm naik*, tentu banyak sekali data tersebut yang harus ditampung, maka dari itu diperlukan analisis sentimen untuk mengetahui sentimen pengguna *twitter* terhadap kenaikan harga BBM Pertamina di *Twitter* serta mengetahui perbandingan performa algoritma *Support Vector Machine (SVM)* menggunakan data dari pelabelan otomatis dan pelabelan manual. Data yang diambil yaitu data *tweet* dengan menggunakan *library python twint* yang dibantu dengan aplikasi *Docker Desktop* dan *VSCode*. Kemudian dilakukan praproses dengan tahapan pembersihan data, *case folding*, *normalization*, *tokenization*, *stopword removal*, dan *stemming*. Setelah itu, data tersebut diberi label positif dan negatif secara otomatis menggunakan *Lexicon Based* diperoleh 2963 positif dan 1114 negatif, dan secara manual oleh 2 anotator dengan perhitungan *kappa statistic* diperoleh 2838 positif dan 1239 negatif, kemudian pembobotan kata menggunakan *Term Frequency – Inverse Document Frequency (TF-IDF)*. Lalu data dibagi menjadi 80% data latih dan 20% data uji secara acak. Hasil klasifikasi dari data pelabelan otomatis dan manual menggunakan algoritma *Support Vector Machine (SVM)* masing-masing memiliki nilai akurasi sebesar 83% dan 81%, presisi sebesar 86% dan 84%, *recall* sebesar 92% dan 91%, *specificity* sebesar 58% dan 56%, dan *F1-score* sebesar 89% dan 88%.

**Kata kunci:** Kenaikan Harga BBM, *Twitter*, Analisis Sentimen, Klasifikasi, *Support Vector Machine (SVM)*

**IMPLEMENTATION OF SUPPORT VECTOR MACHINE (SVM)  
ALGORITHM FOR SENTIMENT ANALYSIS OF INCREASING  
PERTAMINA FUEL PRICE ON TWITTER SOCIAL MEDIA**

**IKHLASUL AMAL**

**ABSTRACT**

*On September 3, 2022, the government officially announced an increase in the price of diesel, Pertalite, and Pertamax fuels. Various responses and complaints from the community were shed, one of them via social media Twitter. Users as a community, there are lots of tweets on Twitter with the keyword or hashtag (#) bbm goes up, of course, there is a lot of data that must be accommodated, then from that required sentiment analysis to find out the sentiments of Twitter users towards Pertamina's fuel price increase on Twitter and find out the performance comparison of the Support Vector Machine (SVM) algorithm using data from automatic labeling and manual labeling. The data taken is tweet data using the twint python library assisted by the Docker application Desktop and VSCode. Then do preprocessing with stages of data cleaning, case folding, normalization, tokenization, stopword removal, and stemming. After that, the data is labeled positive and negative automatically using the Lexicon Based obtained by 2963 positive and 1114 negative, and manually by 2 the annotator by calculating the kappa statistic obtained 2838 positive and 1239 negative, then weighting words using Term Frequency – Inverse Document Frequency (TF-IDF). Then the data is divided into 80% training data and 20% random test data. The classification results of automatic and manual labeling data using the Support Vector Machine (SVM) algorithm each have accuracy values of 83% and 81%, precision of 86% and 84%, recall of 92% and 91%, specificity of 58% and 56%, and F1-score of 89% and 88%.*

**Keywords:** *Fuel Price Increase, Twitter, Sentiment Analysis, Classification, Support Vector Machine (SVM)*

## KATA PENGANTAR

Dengan mengucap puji dan syukur kehadiran Allah Subhanahu Wa Ta'ala, karena Ridho dan Rahmat-Nya skripsi dengan judul “Implementasi Algoritma *Support Vector Machine (SVM)* Untuk Analisis Sentimen Terhadap Kenaikan Harga BBM Pertamina Pada Media Sosial *Twitter*” dapat diselesaikan dengan baik.

Disadari sepenuhnya bahwa penyusunan dan materi skripsi ini masih banyak kekurangannya, oleh karena itu dengan segala kerendahan hati penulis sangat terbuka untuk menerima kritikan dan saran pembaca dalam penyempurnaan isi skripsi ini.

Akhir kata, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih dan penghargaan yang setinggi-tinggi-nya kepada :

1. Yth. Bapak Dr. Ir. Reda Rizal, B.Sc., M.Si., IPU. ASEAN Eng. dan Ibu Lily Kuswiati selaku orang tua penulis yang selalu memberikan doa dan dukungan sehingga penulis dapat menyelesaikan skripsi ini.
2. Yth. Bapak Jayanta, S.Kom., M.Si. selaku dosen pembimbing penulisan skripsi ini.
3. Yth. Ibu Nurul Chamidah, S.Kom., M.Kom. dan Ibu Neny Rosmawarni, M.Kom. selaku dosen pembimbing akademik.
4. Yth. Ibu Dr. Widya Cholil, M.I.T. selaku Kaprodi S1 Informatika.
5. Yth. Ibu Dr. Ermatita, M.Kom. selaku dekan Fakultas Ilmu Komputer.
6. Teman-teman Informatika 2019 yang telah berjuang bersama dalam setiap proses perkuliahan serta saling memberikan semangat untuk dapat menyelesaikan skripsi.

Jakarta, Juni 2023

Penyusun

Ikhlasul Amal



## DAFTAR ISI

COVER.....	i
JUDUL.....	ii
PERNYATAAN ORISINALITAS.....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iv
LEMBAR PENGESAHAN.....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR SIMBOL.....	xiv
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Perumusan Masalah.....	3
1.3. Ruang Lingkup Penelitian.....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Luaran Penelitian.....	4
1.7. Sistematika Penulisan.....	4
BAB 2 TINJAUAN PUSTAKA.....	6
2.1. <i>Twitter</i> .....	6
2.2. <i>Text Mining</i> .....	6
2.3. Analisis Sentimen.....	6
2.4. Praproses Teks.....	7
2.4.1. Pembersihan Data.....	7
2.4.2. <i>Case Folding</i> .....	7
2.4.3. <i>Normalization</i> .....	7
2.4.4. <i>Tokenization</i> .....	8
2.4.5. <i>Stopword Removal</i> .....	8
2.4.6. <i>Stemming</i> .....	8
2.5. Pelabelan <i>Lexicon Based</i> .....	8
2.6. <i>Kappa Statistic</i> .....	9
2.7. <i>Term Frequency – Inverse Document Frequency (TF – IDF)</i> .....	10

2.8.	Klasifikasi.....	11
2.8.1.	<i>Support Vector Machine (SVM)</i> .....	11
2.8.2.	Evaluasi.....	14
2.9.	Review Penelitian yang Relevan.....	16
BAB 3 METODOLOGI PENELITIAN.....		20
3.1.	Identifikasi Masalah.....	21
3.2.	Studi Literatur.....	21
3.3.	Pengumpulan Data.....	21
3.4.	Praproses.....	21
3.4.1.	Pembersihan data.....	22
3.4.2.	<i>Case Folding</i> .....	22
3.4.3.	<i>Normalization</i> .....	22
3.4.4.	<i>Tokenization</i> .....	22
3.4.5.	<i>Stopword Removal</i> .....	22
3.4.6.	<i>Stemming</i> .....	22
3.5.	Pelabelan Otomatis <i>Lexicon Based</i> .....	22
3.6.	Pelabelan Manual <i>Kappa Statistic</i> .....	23
3.7.	Pembobotan Kata <i>TF-IDF</i> .....	23
3.8.	Klasifikasi.....	23
3.9.	Evaluasi.....	24
3.10.	Alat Bantu Penelitian.....	24
3.10.1.	<i>Hardware</i> .....	24
3.10.2.	<i>Software</i> .....	24
3.11.	Jadwal Penelitian.....	25
BAB 4 HASIL DAN PEMBAHASAN.....		26
4.1.	Data.....	26
4.2.	Praproses.....	26
4.2.1.	Pembersihan data.....	26
4.2.2.	<i>Case Folding</i> .....	28
4.2.3.	<i>Normalization</i> .....	29
4.2.4.	<i>Tokenization</i> .....	31
4.2.5.	<i>Stopword Removal</i> .....	33
4.2.6.	<i>Stemming</i> .....	34
4.3.	Pelabelan Data.....	36
4.3.1.	Pelabelan Otomatis dengan <i>Lexicon Based</i> .....	36
4.3.2.	Pelabelan Manual dengan <i>Kappa Statistic</i> .....	39

4.3.3. Perbandingan Pelabelan Otomatis dengan Manual.....	41
4.4. Pembobotan Kata dengan TF-IDF .....	41
4.5. Klasifikasi.....	46
4.6. Evaluasi .....	48
BAB 5 PENUTUP .....	51
5.1. Kesimpulan.....	51
5.2. Saran.....	52
DAFTAR PUSTAKA .....	53


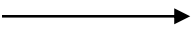

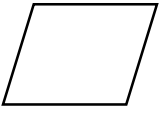
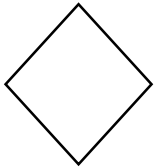
## DAFTAR TABEL

Tabel 2.1 Variasi Antar-Pengamat/Penilai.....	9
Tabel 2.2 Interpretasi Kappa.....	10
Tabel 2.3 Confusion Matrix.....	15
Tabel 2.4 Review Penelitian Terdahulu.....	16
Tabel 2.5 Perbedaan Penelitian Terdahulu dengan Penelitian Ini.....	18
Tabel 3.1 Jadwal Penelitian.....	25
Tabel 4.1 Hasil Sampel Pembersihan Data.....	27
Tabel 4.2 Hasil Sampel Case Folding.....	28
Tabel 4.3 Sampel Kata-Kata Normalization.....	30
Tabel 4.4 Hasil Sampel Normalization.....	30
Tabel 4.5 Hasil Sampel Tokenization.....	32
Tabel 4.6 Hasil Sampel Stopword Removal.....	33
Tabel 4.7 Hasil Sampel Stemming.....	35
Tabel 4.8 Hasil Sampel Pelebelan Otomatis.....	37
Tabel 4.9 Hasil Anotasi.....	39
Tabel 4.10 Perbandingan Hasil Pelabelan.....	41
Tabel 4.11 Hasil sampel nilai TF, IDF, dan TF-IDF.....	45
Tabel 4.12 Pembagian Data.....	47
Tabel 4.13 Proses Grid Search CV.....	47
Tabel 4.14 Confusion Matrix Hasil Pelabelan Otomatis.....	48
Tabel 4.15 Confusion Matrix Hasil Pelabelan Manual.....	49
Tabel 4.16 Perbandingan Nilai Evaluasi.....	50

## DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian.....	20
Gambar 4.1 Hasil Crawling Data Twitter .....	26
Gambar 4.2 Proses Pembersihan Data .....	27
Gambar 4.3 Proses Case Folding .....	28
Gambar 4.4 Proses Normalization .....	29
Gambar 4.5 Proses Tokenization .....	31
Gambar 4.6 Proses Stopword Removal .....	33
Gambar 4.7 Proses Stemming.....	35
Gambar 4.8 Proses Pelabelan Otomatis .....	36
Gambar 4.9 Hasil Pelabelan Otomatis .....	38
Gambar 4.10 Hasil Pelabelan Manual.....	41
Gambar 4.11 Kode Program Grid Search CV.....	48
Gambar 4.12 Hasil Parameter Terbaik Grid Search CV .....	48

## DAFTAR SIMBOL

<b>SIMBOL FLOWCHART</b>			
<b>NO</b>	<b>GAMBAR</b>	<b>NAMA</b>	<b>KETERANGAN</b>
1		<i>Terminator</i>	Menggambarkan dimulai dan berakhirnya suatu kegiatan.
2		<i>Flow Direction</i>	Menggambarkan hubungan antara simbol yang mengarahkan suatu jalannya proses dalam sistem.
3		<i>Process</i>	Menggambarkan deskripsi dari proses yang dijalankan.
4		<i>Data</i>	Menggambarkan suatu proses yang memiliki data masukan ( <i>input</i> ) atau keluaran ( <i>output</i> ).
5		<i>Decision</i>	Menggambarkan suatu proses keputusan yang perlu dibuat untuk mencapai langkah proses selanjutnya.