

**PERBANDINGAN ALGORITMA SUPPORT VECTOR  
MACHINE (SVM) DAN NAIVE BAYES CLASSIFIER DALAM  
ANALISIS SENTIMEN TERHADAP PENERAPAN APLIKASI  
MYPERTAMINA PADA MEDIA SOSIAL TWITTER**

**SKRIPSI**

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



disusun oleh

**Fadillah Nur Istiqomah**

**19.11.3058**

Kepada

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

**PERBANDINGAN ALGORITMA SUPPORT VECTOR  
MACHINE (SVM) DAN NAIVE BAYES CLASSIFIER DALAM  
ANALISIS SENTIMEN TERHADAP PENERAPAN APLIKASI  
MYPERTAMINA PADA MEDIA SOSIAL TWITTER**

**SKRIPSI**

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



disusun oleh

**Fadillah Nur Istiqomah**

**19.11.3058**

Kepada

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

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)  
DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN  
TERHADAP PENERAPAN APLIKASI MYPERTAMINA PADA MEDIA  
SOSIAL TWITTER**

yang disusun dan diajukan oleh

**Fadillah Nur Istiqomah  
19.11.3058**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 27 Juli 2023

**Dosen Pembimbing,**



**Theopilus Bayu Sasongko, S.Kom, M.Eng  
NIK. 190302375**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)  
DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN  
TERHADAP PENERAPAN APLIKASI MYPERTAMINA PADA MEDIA  
SOSIAL TWITTER**

yang disusun dan diajukan oleh

**Fadillah Nur Istiqomah**  
19.11.3058

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 27 Juli 2023

**Susunan Dewan Penguji**

**Nama Penguji**

Arif Dwi Laksito, M.Kom  
NIK. 190302150

Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.  
NIK. 190302235

Theopilus Bayu Sasongko, S.Kom, M.Eng  
NIK. 190302375

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 27 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 : Fadillah Nur Istiqomah  
NIM : 19.11.3058

Menyatakan bahwa Skripsi dengan judul berikut:

**PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN TERHADAP PENERAPAN APLIKASI MYPERTAMINA PADA MEDIA SOSIAL TWITTER**

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom, M.Eng

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, 27 Juli 2023

Yang Menyatakan,



Fadillah Nur Istiqomah

## HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT dengan kemurahan dan ridho-Nya, sehingga skripsi ini dapat ditulis dengan baik dan lancar hingga selesai. Penulisan skripsi ini sungguh tidak akan mungkin dapat terselesaikan tanpa adanya bantuan dan dukungan dari segala pihak. Peneliti ingin mengucapkan banyak rasa terima kasih dan mempersembahkan skripsi kepada :

1. Diri saya sendiri yang telah berjuang dan bertahan untuk menyelesaikan penulisan skripsi ini dengan sangat baik.
2. Kedua orang tua dan seluruh anggota keluarga yang selalu memberi bantuan, dukungan, dan nasehat sehingga penulis dapat menyelesaikan skripsi ini dengan sangat baik.
3. Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng selaku dosen pembimbing yang selalu memberi saran dan masukan dalam penulisan skripsi ini.
4. Yunita dan Siti selaku teman dekat penulis dan rekan seperjuangan penulis di AMIKOM yang selalu memberi saran dan menemani penulis selama proses penulisan skripsi.
5. Teman – teman anggota grup Taman Surga selaku teman dekat penulis yang selalu memberi semangat dan menemani penulis selama proses penulisan skripsi ini sehingga dapat selesai dengan sangat baik.
6. Kim Seokjin dan Lee Haechan sebagai salah satu penyemangat dalam hidup penulis sehingga skripsi ini dapat selesai dengan sangat baik.
7. BTS, NCT, dan kpop idol lainnya yang selalu memberi motivasi, semangat, dan hiburan kepada penulis melalui karya – karya dan konten video yang diberikan kepada penulis.
8. Teman – teman kelas 19 Informatika 08 selaku rekan seperjuangan penulis selama menempuh pendidikan.
9. Semua pihak yang berperan dalam proses penulisan skripsi ini yang tidak bisa disebutkan satu persatu.

## KATA PENGANTAR

Puji Syukur bagi Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi dengan judul “PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN NAIVE BAYES CLASSIFIER DALAM ANALISIS SENTIMEN TERHADAP PENERAPAN APLIKASI MYPERTAMINA PADA MEDIA SOSIAL TWITTER” ini dengan baik. Selama proses penulisan skripsi ini, penulis banyak mendapatkan bantuan, pengarahan, dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini sebagai wujud rasa hormat, penulis mengucapkan terima kasih yang sebesar besarnya :

- 1 Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
- 2 Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
- 3 Bapak Theopilus Bayu Sasongko, S.Kom, M.Eng selaku dosen pembimbing yang selalu memberi saran dan masukan dalam penulisan skripsi ini.
- 4 Bapak Arif Dwi Laksito, M.Kom dan Dr. Ferry Wahyu Wibowo, S.Si., M.Cs selaku dosen penguji yang telah memberi saran dan masukan dalam penulisan skripsi ini.
- 5 Kedua orangtua, kakak, dan keluarga penulis yang senantiasa mendoakan dan memberikan dukungan kepada penulis selama penulisan skripsi ini.
- 6 Semua pihak yang senantiasa menjadi teman dan menghibur penulis sehingga skripsi ini dapat selesai dengan baik.

Penulis menyadari bahwa skripsi ini masih terdapat banyak kekurangan baik dalam isi maupun cara penyajian. Oleh karena itu, penulis mengharapkan kritik dan saran guna perbaikan di masa mendatang.

Yogyakarta, 27 Juli 2023

Penulis

## DAFTAR ISI

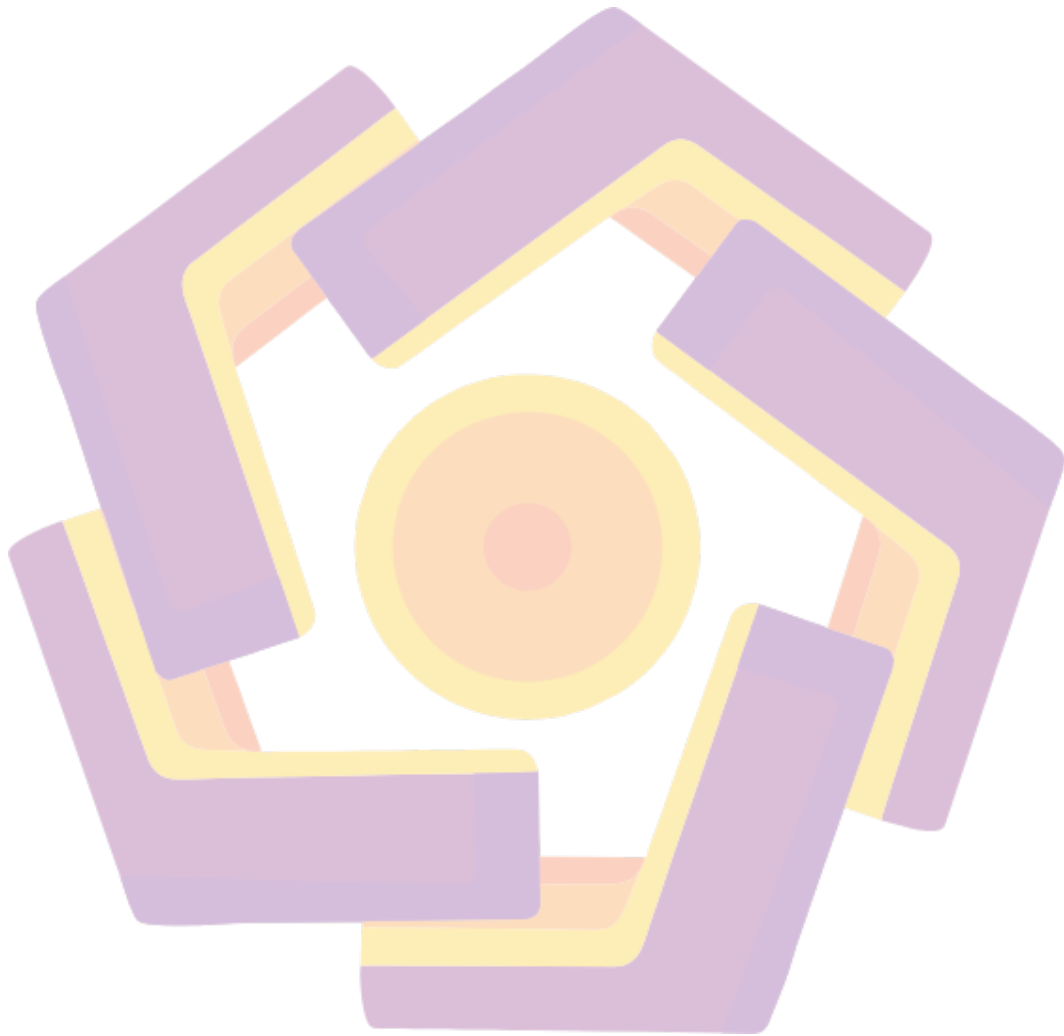
HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xii
INTISARI.....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	5
1.3 Batasan Penelitian .....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penelitian .....	6
BAB II LANDASAN TEORI.....	8
2.1 <i>Studi Literatur</i> .....	8
2.2 Dasar Teori .....	19
2.2.1 Data Mining .....	19
2.2.1.1 Pengertian Data Mining .....	19
2.2.1.2 Kegunaan Data Mining .....	20



2.2.2 Analisis Sentimen .....	21
2.2.3 Text Mining .....	23
2.2.4 Pre-Processing Data.....	23
2.2.4.1 <i>Special Character Removal</i> .....	24
2.2.4.2 <i>Case Folding</i> .....	25
2.2.4.3 <i>Indonesian Stemming</i> .....	25
2.2.4.4 <i>Indonesian Stopword Removal</i> .....	26
2.2.4.5 <i>Tokenization</i> .....	26
2.2.4.6 <i>Normalization</i> .....	26
2.2.5 <i>Term Frequency - inverse Document Frequency (TF-IDF)</i> .....	26
2.2.6 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i> .....	27
2.2.7 Algoritma <i>Support Vector Machine (SVM)</i> .....	28
2.2.8 Naive-Bayes Classifier .....	31
2.2.9 <i>Confusion Matrix</i> .....	35
2.2.10 Twitter.....	37
2.2.11 Crawling Data.....	39
2.2.12 MyPertamina.....	39
2.2.13 Kurva ROC .....	39
<b>BAB III METODE PENELITIAN.....</b>	<b>41</b>
3.1 <i>Objek Penelitian</i> .....	41
3.2 Alat dan Bahan Penelitian .....	41
3.2.1 Perangkat Keras .....	41
3.2.2 Perangkat Lunak .....	41
3.3 Metode Pengumpulan Data .....	42
3.4 Alur Penelitian.....	42

3.4.1 Analisis Sistem .....	42
3.4.1.1 Analisis Kebutuhan Sistem .....	43
3.4.1.1.1 Kebutuhan Fungsional .....	43
3.4.1.1.2 Kebutuhan Non-fungsional .....	43
3.4.1.2 Analisis Kebutuhan Data.....	44
3.4.2 Perancangan Sistem .....	44
3.4.2.1 Dataset .....	47
3.4.2.2 Pre-processing Data .....	47
3.4.2.3 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i> .....	52
3.4.2.4 Pembuatan Model ( <i>Modelling</i> ) .....	52
3.4.2.5 Evaluasi .....	54
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>56</b>
4.1 Implementasi Pengambilan Data & Text Pre-processing.....	56
4.2 Implementasi Pembobotan Text dengan TF-IDF .....	63
4.3 Implementasi Perhitungan Algoritma Naïve Bayes Classifier .....	65
4.4 Implementasi Pembagian Data dan SMOTE.....	69
4.5 Implementasi Algoritma Klasifikasi.....	71
4.5.1 Algoritma Support Vector Machine (SVM) .....	71
4.5.2 Algoritma Naïve Bayes Classifier .....	72
4.6 Hasil dan Evaluasi .....	74
4.6.1 Hasil .....	74
4.6.2 Evaluasi .....	75
4.5.1 Classification Report.....	75
4.5.1 Kurva ROC .....	77
4.5.1 Confusion Matrix .....	78

BAB V KESIMPULAN DAN SARAN.....	81
5.1 Kesimpulan.....	81
5.2 Saran .....	82
DAFTAR PUSTAKA .....	83



## DAFTAR TABEL

Tabel 2.1 <i>Studi Literatur</i> Penelitian Terdahulu.....	12
Tabel 2.2 Klasifikasi Nilai AUC.....	40
Tabel 3.1 Contoh <i>Special Character Removal</i> .....	49
Tabel 3.2 Contoh <i>Case Folding</i> .....	49
Tabel 3.3 Contoh <i>Normalization</i> .....	50
Tabel 3.4 Contoh <i>Tokenization</i> .....	50
Tabel 3.5 Contoh <i>Stopword Removal</i> .....	51
Tabel 3.6 Contoh <i>Stemming</i> .....	52
Tabel 4.1 Contoh Hasil <i>Cleaning</i> dan <i>Case Folding</i> .....	59
Tabel 4.2 Contoh Hasil <i>Normalization</i> .....	60
Tabel 4.3 Contoh Hasil <i>Tokenization</i> .....	61
Tabel 4.4 Contoh Hasil <i>Stopword Removal</i> .....	62
Tabel 4.5 Contoh Hasil <i>Stemming</i> .....	63
Tabel 4.6 Contoh Perhitungan TF-IDF .....	64
Tabel 4.7 Sampel Dataset .....	65
Tabel 4.8 Sampel Dataset Uji .....	66
Tabel 4.9 Hasil Proses Modelling .....	74
Tabel 4.10 CR Algoritma SVM.....	75
Tabel 4.11 CR Algoritma Naïve Bayes.....	76

## DAFTAR GAMBAR

Gambar 2.1 Alur kerja <i>Naïve Bayes</i> .....	35
Gambar 2.2 Tabel <i>Confusion Matrix</i> .....	36
Gambar 3.1 Alur Sistem Analisis Sentimen .....	45
Gambar 3.2 Alur <i>pre-processing</i> .....	48
Gambar 3.3 Alur Klasifikasi dengan Model .....	54
Gambar 4.1 Dataset Hasil <i>Crawling</i> .....	56
Gambar 4.2 Dataset Hasil <i>Labelling</i> .....	57
Gambar 4.3 Diagram Batang Dataset Hasil <i>Crawling</i> .....	58
Gambar 4.4 Dataset Hasil <i>Pre-Processing</i> .....	58
Gambar 4.5 Baris Kode <i>Split Data</i> .....	69
Gambar 4.6 Dataset <i>Train</i> Sebelum SMOTE .....	70
Gambar 4.7 Dataset <i>Train</i> Sesudah SMOTE .....	70
Gambar 4.8 Baris Kode <i>Modelling SVM</i> .....	72
Gambar 4.9 Baris Kode <i>Modelling Naïve Bayes</i> .....	73
Gambar 4.10 Kurva ROC Algoritma SVM .....	77
Gambar 4.11 Kurva ROC Algoritma <i>Naïve Bayes</i> .....	78
Gambar 4.12 Confusion Matrix Algoritma SVM .....	79
Gambar 4.13 Confusion Matrix Algoritma <i>Naïve Bayes</i> .....	80

## INTISARI

Aplikasi MyPertamina adalah aplikasi layanan keuangan digital dari Pertamina dan Badan Usaha Milik Negara (BUMN) yang digunakan untuk pembayaran bahan bakar minyak secara non-tunai di Stasiun Pengisian Bahan Bakar Umum (SPBU) Pertamina. Baru – baru ini, pihak Pertamina membuat kebijakan baru tentang penggunaan aplikasi MyPertamina menjadi aplikasi wajib bagi kendaraan bermotor yang memenuhi regulasi. Hal ini tentu menimbulkan banyak opini dari masyarakat terutama yang terdampak kebijakan ini. Banyak orang mulai berkomentar dan melakukan protes tentang kebijakan di beberapa jejaring sosial, salah satunya Twitter. Twitter adalah salah satu layanan jejaring sosial yang saat ini banyak digunakan oleh masyarakat seluruh dunia. Para pengguna menyampaikan komentar, pendapat dan protes mereka terhadap isu yang sedang hangat melalui cuitan tweet akun mereka.

Proses analisis sentimen dari pendapat tentang aplikasi MyPertamina dapat dimulai dengan *crawling* data dari tweet (post dalam twitter) yang ditulis oleh pengguna. Setelah proses *crawling* berhasil mendapatkan data sesuai keinginan, akan dilakukan *pre-processing*. Selanjutnya, proses *modelling* akan dilakukan dengan penerapan algoritma Support Vector Machine (SVM) dan Naïve Bayes Classifier dan Confusion Matrix untuk mengetahui perbandingan performa dua algoritma yang digunakan. Output dari model berupa nilai akurasi, presisi, dan recall akan ditampilkan dalam bentuk grafik untuk mempermudah pembacaan hasil.

Penelitian ini menggunakan dataset hasil *crawling* cuitan (tweet) pada media sosial twitter dengan kata kunci ‘mypertamina’ sebanyak 2500 data yang merupakan tulisan pengguna twitter dalam kurun waktu mulai dari 15 juni hingga 25 juli 2022. Hasil dari proses *crawling* tersebut kemudian dilakukan proses *labelling* dengan 2 label yaitu kelas sentimen negatif dan kelas sentimen positif. Perbandingan jumlah data negatif dan positif tidak seimbang, biasa disebut *imbalanced* data. Sebanyak 1499 data dengan label negative dan 1001 data dengan label positive tersebut akan dilakukan proses SMOTE (*Synthetic Minority Over-sampling Technique*) agar data menjadi seimbang. Dari proses *modelling* yang dilakukan dengan pembagian 10% data *test* dan 90% data *training*, dihasilkan akurasi dengan algoritma Support Vector Machine (SVM) sebesar 84.8% sedangkan dengan algoritma Bernoulli Naïve Bayes sebesar 79.2%.

**Kata Kunci:** Analisis Sentimen, Support Vector Machine, Naïve Bayes, MyPertamina, Twitter

## ABSTRACT

*MyPertamina application is a digital financial service application from Pertamina and State-Owned Enterprises (BUMN) which is used for non-cash payments for fuel oil at Pertamina's Public Fuel Filling Stations (SPBU). Recently, Pertamina made a new policy regarding the use of MyPertamina to become a mandatory application for motorized vehicles that meet regulations. This certainly raises a lot of opinions from the public, especially those affected by this policy. Many people began to comment and protest about the policy on several social networks, one of which was Twitter. Twitter is a social networking service that is currently widely used by people all over the world. Users convey their comments, opinions and protests against the hot issue through their tweets.*

*The process of sentiment analysis of opinions about MyPertamina can be started by crawling data from tweets (posts in twitter) written by users. After the crawling process has succeeded in obtaining the desired data, pre-processing will be carried out. Furthermore, the modeling process will be carried out by applying the Support Vector Machine (SVM) algorithm and the Naïve Bayes Classifier and Confusion Matrix to compare the performance of the two algorithms used. The output of the model in the form of accuracy, precision, and recall values will be displayed in graphic form to facilitate reading the results. The results of the sentiment analysis process can be used as a reference for decision making in order to produce a fair decision.*

*This research uses a dataset of crawling tweets on Twitter with the keyword 'mypertamina' as many as 2500 data which are posts by Twitter users in the time range from 15 June to 25 July 2022. The results of the crawling process are then carried out by a labeling process with 2 labels namely negative sentiment class and positive sentiment class. Comparison of the number of negative and positive data is not balanced, usually called imbalanced data. A total of 1499 data with negative labels and 1001 data with positive labels will be processed by SMOTE (Synthetic Minority Over-sampling Technique) so that the data becomes balanced. From the modeling process which was carried out by dividing 10% test data and 90% training data, the accuracy with the Support Vector Machine (SVM) algorithm was 84.8% while with the Bernoulli Naïve Bayes algorithm it was 79.2%.*

**Keyword : Sentiment Analysis, Support Vector Machine, Naïve Bayes, MyPertamina, Twitter**