

**PERBANDINGAN IMPLEMENTASI STEMMING NAZIEF &  
ADRIANI SETIONO DAN ARIFIN PADA SENTIMEN  
ANALISIS PENGGUNA TWITTER TERHADAP TIKTOK  
SHOP MENGGUNAKAN NAIVE BAYES DAN SUPPORT  
VECTOR MACHINE**

**SKRIPSI**

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



disusun oleh  
**AGI RIZKI ANANDA**  
20.11.3495

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2024**

**PERBANDINGAN IMPLEMENTASI STEMMING NAZIEF &  
ADRIANI DAN ARIFIN SETIONO PADA SENTIMEN  
ANALISIS PENGGUNA TWITTER TERHADAP TIKTOK  
SHOP MENGGUNAKAN NAIVE BAYES DAN SUPPORT  
VECTOR MACHINE**

**SKRIPSI**

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



disusun oleh  
**AGI RIZKI ANANDA**  
20.11.3495

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN IMPLEMENTASI STEMMING NAZIEF & ADRIANI  
DAN ARIFIN SETIONO PADA SENTIMEN ANALISIS PENGGUNA  
TWITTER TERHADAP TIKTOK SHOP MENGGUNAKAN NAIVE  
BAYES DAN SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

**Agi Rizki Ananda**  
20.11.3495

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 30 Juli 2024

Dosen Pembimbing,



Windha Mega Pradnya D. M.Kom  
NIK. 190302185

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PERBANDINGAN IMPLEMENTASI STEMMING NAZIEF & ADRIANI DAN  
ARIFIN SETIONO PADA SENTIMEN ANALISIS PENGGUNA TWITTER  
TERHADAP TIKTOK SHOP MENGGUNAKAN NAIVE BAYES DAN SUPPORT  
VECTOR MACHINE**

yang disusun dan diajukan oleh

**Agi Rizki Ananda**

**20.11.3495**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 30 Juli 2024

**Susunan Dewan Penguji**

**Nama Penguji**

**Robert Marco, M.T., Ph.D**

**NIK. 190302228**

**Donni Prabowo, M.Kom**

**NIK. 190302253**

**Windha Mega Pradnya D, M.kom**

**NIK. 190302185**

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 30 Juli 2024

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D.**

**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Agi Rizki Ananda

NIM : 20.11.3495

Menyatakan bahwa Skripsi dengan judul berikut:

**Perbandingan Implementasi Stemming Nazief & Adriani dan Arifin Setiono pada Sentimen Analisis Pengguna Twitter Terhadap Tiktok Shop Menggunakan Naive Bayes dan Support Vector Machine**

Dosen Pembimbing : Windha Mega Pradnya D, 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, 30 Juli 2024

Yang Menyatakan,



Agi Rizki Ananda

## HALAMAN PERSEMBAHAN

Dengan segala puji syukur kepada Tuhan yang Maha Esa dan atas dukungan dan doa dari orang-orang tercinta, akhirnya skripsi ini dapat dirampungkan dengan baik dan tepat pada waktunya. Oleh karena itu, dengan rasa bangga dan bahagia saya ucapkan rasa syukur dan terimakasih saya kepada:

1. Allah SWT, atas segala rahmat, hidayah, dan karunia-Nya yang tiada terhingga, yang telah memberikan kekuatan dan kemampuan sehingga skripsi ini dapat diselesaikan.
2. Kedua orang tua saya, Almarhum Bapak Budiyanto dan Ibu Waginem yang selalu memberikan doa, dukungan, serta kasih sayang yang tiada henti. Skripsi ini juga saya dedikasikan untuk Almarhum Ayah saya yang belum sempat saya berikan kebahagiaan dan rasa bangga. Skripsi ini sebagai tanda bahwa perjuangan orang tua saya untuk memberikan pendidikan tinggi untuk anaknya tidak sia-sia.
3. Kakak, Adik dan saudara dari Penulis yang selalu memberikan dukungan, doa, saran dan nasihat.
4. Para dosen dan pembimbing, yang telah memberikan ilmu, arahan, dan motivasi selama proses penyusunan skripsi ini. Terima kasih atas segala dedikasi dan kesabaran dalam membimbing saya.
5. Sahabat – sahabat saya tercinta, Dandi, Tiyo, Fafa, Bayu, Adi, Dimas dan Andre yang telah memberikan semangat, dukungan, serta kerjasama selama masa perkuliahan dan penyusunan skripsi ini. Terima kasih atas kebersamaan dan persahabatannya.
6. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan dan dukungan dalam berbagai bentuk selama penyusunan skripsi ini. Terima kasih atas segala kontribusinya. Semoga skripsi ini dapat memberikan manfaat dan menjadi sumbangsih bagi perkembangan ilmu pengetahuan.



## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul "Perbandingan Implementasi Algoritma Stemming Nazief & Adriani dan Arifin Setiono pada Analisis Sentimen Pengguna Twitter terhadap TikTok Shop" ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Amikom Yogyakarta.

Pada kesempatan ini, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada berbagai pihak yang telah memberikan bantuan, bimbingan, dan dukungan selama penyusunan skripsi ini. Ucapan terima kasih yang tulus penulis sampaikan kepada:

1. Ibu Windha Mega Pradnya D, M.Kom., selaku dosen pembimbing, yang telah memberikan bimbingan, arahan, dan motivasi kepada penulis sehingga skripsi ini dapat terselesaikan dengan baik.
2. Tim Dosen Penguji, yang telah memberikan masukan dan kritik yang membangun pada saat ujian skripsi, sehingga skripsi ini menjadi lebih baik.
3. Orang tua tercinta, yang selalu memberikan doa, dukungan moral, dan materiil kepada penulis sehingga dapat menyelesaikan pendidikan ini dengan baik.
4. Teman-teman seperjuangan di Program Studi Ilmu Komputer, yang selalu memberikan semangat dan membantu dalam berbagai hal selama proses penyusunan skripsi ini.
5. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan dan dukungan dalam penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari berbagai pihak demi kesempurnaan skripsi ini. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca dan dapat menjadi sumbangsih yang berguna bagi perkembangan ilmu pengetahuan.

Yogyakarta, 30 Juli 2024

Penulis

## DAFTAR ISI

HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	vii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
DAFTAR LAMPIRAN .....	xii
DAFTAR LAMBANG DAN SINGKATAN .....	xiii
DAFTAR ISTILAH .....	xiv
INTISARI .....	xv
<i>ABSTRACT</i> .....	xvi
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan .....	4
1.6.1 BAB I : Pendahuluan .....	4
1.6.2 BAB II : Tinjauan Pustaka .....	5
1.6.3 BAB III : Metode Penelitian .....	5
1.6.4 BAB IV : Implementasi dan Pembahasan .....	5
1.6.5 BAB V : Penutup .....	5
<b>BAB II TINJAUAN PUSTAKA</b> .....	6
2.1 Studi Literatur .....	6
2.2 Dasar Teori .....	12
2.2.1 Information Retrieval .....	12
2.2.2 Analisis Sentimen .....	12
2.2.3 Preprocessing .....	12
2.2.4 Stemming .....	12



2.2.5	Translate	15
2.2.6	Automatic Labelling	16
2.2.7	Pembobotan	16
2.2.8	Naive Bayes Classifier	17
2.2.9	Support Vector Machine (SVM)	18
2.2.10	Python	20
2.3	Metode Pengujian	20
2.3.1	Akurasi	20
2.3.2	Confusion Matrix	20
2.3.3	Waktu Proses	21
<b>BAB III METODE PENELITIAN</b>		22
3.1	Objek Penelitian	22
3.2	Alur Penelitian	22
3.1.1	Persiapan Data	23
3.2.2	Preprocessing	23
3.2.3	Stemming Algoritma Nazief & Adriani	24
3.2.4	Stemming Algoritma Arifin Setiono	25
3.2.5	Translate	26
3.2.6	Labeling	26
3.2.7	Pembobotan	26
3.2.8	Klasifikasi	27
3.2.9	Evaluasi	27
3.3	Kebutuhan Sistem	28
3.3.1	Perangkat Lunak	28
3.3.2	Perangkat Keras	28
<b>BAB IV HASIL DAN PEMBAHASAN</b>		29
4.1	Persiapan Data	29
4.2	Preprocessing Text	29
4.2.1	Cleansing	30
4.2.2	Case Folding	31
4.2.3	Normalize	31
4.2.4	Tokenize	32
4.2.5	Stopword	33
4.3	Penerapan Algoritma Stemming	34

4.4	Translate .....	37
4.5	Labeling .....	38
4.6	Pembobotan TF-IDF .....	40
4.7	Klasifikasi .....	41
4.8	Evaluasi Algoritma SVM .....	42
4.8.1	Evaluasi algoritma SVM pada Nazief & Adriani .....	42
4.8.2	Evaluasi algoritma SVM pada Arifin Setiono .....	43
4.9	Evaluasi Algoritma Naive Bayes .....	44
4.9.1	Evaluasi algoritma Naive Bayes pada Nazief & Adriani .....	44
4.9.2	Evaluasi algoritma Naive Bayes pada Arifin Setiono .....	45
4.10	Analisis perbandingan Hasil Evaluasi .....	46
<b>BAB V PENUTUP</b> .....		49
5.1	Kesimpulan .....	49
5.2	Saran .....	50
<b>REFERENSI</b> .....		51

## DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian.....	23
Tabel 2.2. Confison Matrix 3 Kelas.....	35
Tabel 4.1. Cleansing.....	30
Tabel 4.2. Case Folding.....	31
Tabel 4.3. Normalize.....	32
Tabel 4.4. Tokenize.....	32
Tabel 4.5. Stopword.....	33
Tabel 4.6. Hasil Stemming Nazief & Adriani.....	34
Tabel 4.7. Perbandingan Waktu Proses Stemming.....	37
Tabel 4.8. Translate.....	38
Tabel 4.9. Hasil Automatic Labeling Vader.....	38
Tabel 4.10. Confusion Matrix Kernel Linier Nazief & Adriani.....	42
Tabel 4.11. Classification Report Kernel Linier Nazief & Adriani.....	42
Tabel 4.12. Confusion Matrix Kernel Linier Arifin Setiono.....	43
Tabel 4.13. Classification Report Kernel Linier Arifin Setiono.....	43
Tabel 4.14. Confusion Matrix Complement Naive Bayes Nazief & Adriani.....	44
Tabel 4.15. Classification Report Complement Naive Bayes Nazief & Adriani.....	44
Tabel 4.16. Confusion Matrix Complement Naive Bayes Arifin Setiono.....	45
Tabel 4.17. Classification Report Complement Naive Bayes Arifin Setiono.....	45

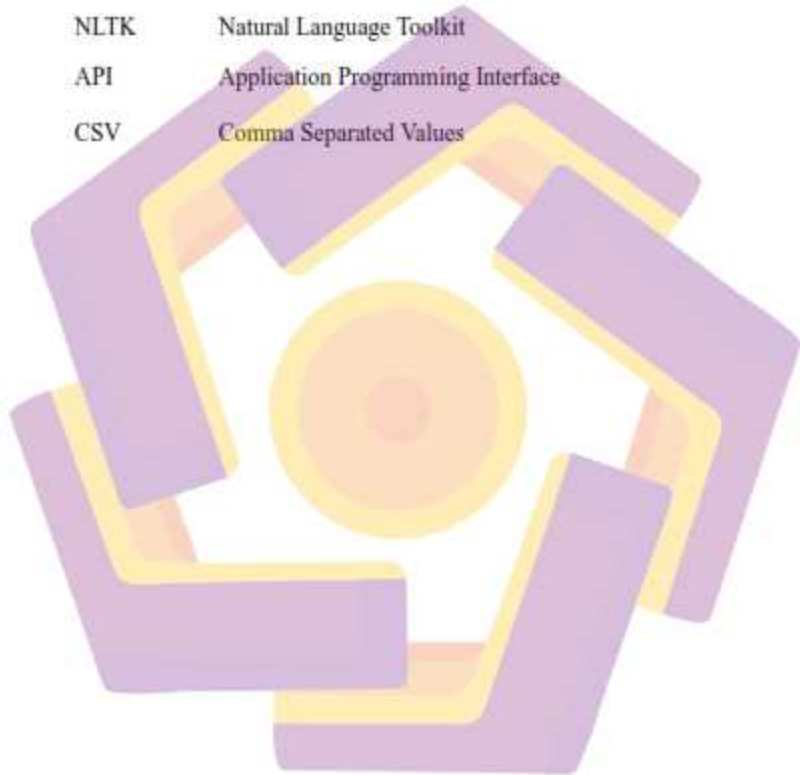
## DAFTAR GAMBAR

Gambar 3.1. Flowchart Alur Penelitian .....	38
Gambar 3.2. Flowchart Stemming Nazief & Adriani .....	40
Gambar 3.3. Flowchart Stemming Arifin Setiono .....	41
Gambar 4.1. Data Mentah Hasil Crawling.....	45
Gambar 4.2. Plot Persebaran Hasil Labelling Pada Nazief & Adriani .....	55
Gambar 4.3. Plot Persebaran Hasil Labelling Pada Arifin Setiono .....	56
Gambar 4.4. Hasil Pembobotan .....	57
Gambar 4.5. Perbandingan Akurasi Model.....	63
Gambar 4.6. Perbandingan F1-Score Model.....	64



## DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
IR	Information Retrieval
TF-IDF	Term Frequency-Inverse Document Frequency
VADER	Valence Aware Dictionary and Sentiment Reasoner
NLTK	Natural Language Toolkit
API	Application Programming Interface
CSV	Comma Separated Values





## DAFTAR ISTILAH

Slang	Kata gaul
Case Folding	Proses mengubah semua data teks menjadi huruf kecil (lowercase)
Normalize	Tahap mengubah kata-kata tidak baku menjadi baku
Tokenize	Memisahkan kata menjadi kalimat terpisah
Cleansing	Proses penghapusan data yang tidak relevan atau duplikat
Stopword	Menghilangkan kata-kata yang tidak memberikan informasi penting dalam analisis teks



## INTISARI

Perkembangan kebutuhan informasi yang pesat di era digital menjadikan internet sebagai sumber utama informasi, dengan stemming menjadi aspek penting dalam Information Retrieval (IR). Penelitian ini membandingkan dua algoritma stemming populer, yaitu Nazief & Adriani dan Arifin Setiono, dalam analisis sentimen pengguna Twitter terhadap TikTok Shop. Algoritma Nazief & Adriani menggunakan kamus kata dasar dan mendukung penyusunan kembali kata, sedangkan Arifin Setiono menggunakan 12 kombinasi untuk mengatasi overstemming. Dengan menggunakan metode klasifikasi Naive Bayes dan Support Vector Machine (SVM), penelitian ini mengevaluasi akurasi, presisi, recall, F1score, dan waktu proses stemming kedua algoritma tersebut. Data dikumpulkan dari Twitter dan terdiri dari 986 data tweet terkait TikTok Shop. Hasil penelitian menunjukkan bahwa untuk metode SVM, algoritma Nazief & Adriani memiliki akurasi 75.43%, sedangkan Arifin Setiono memiliki akurasi 74.29%. Sementara itu, untuk metode Naive Bayes, algoritma Nazief & Adriani memiliki akurasi 68% dan algoritma Arifin Setiono memiliki akurasi 64%. Memiliki waktu proses 275.1028 detik pada algoritma Nazief & Adriani dan 1.0558 detik pada algoritma Arifin Setiono. Penelitian ini memberikan wawasan berharga mengenai efektivitas kedua algoritma dalam analisis sentimen, yang bermanfaat bagi peneliti dalam memilih algoritma yang tepat dan sebagai referensi untuk penelitian selanjutnya.

**Kata kunci:** Stemming, Nazief & Adriani, Arifin Setiono, Analisis Sentimen, TikTok Shop.

## **ABSTRACT**

*The rapid development of information needs in the digital era has made the internet the main source of information, with stemming being an important aspect in Information Retrieval (IR). This research compares two popular stemming algorithms, namely Nazief & Adriani and Arifin Setiono, in analyzing Twitter user sentiment towards TikTok Shop. Nazief & Adriani's algorithm uses a basic word dictionary and supports word rearrangement, while Arifin Setiono uses 12 combinations to overcome overstemming. Using the Naive Bayes and Support Vector Machine (SVM) classification methods, this research evaluates the accuracy, precision, recall, F1-score and stemming process time of the two algorithms. Data was collected from Twitter and consists of 986 tweet data related to TikTok Shop. The research results show that for the SVM method, the Nazief & Adriani algorithm has an accuracy of 75.43%, while Arifin Setiono has an accuracy of 74.29%. Meanwhile, for the Naive Bayes method, Nazief & Adriani's algorithm has an accuracy of 68% and Arifin Setiono's algorithm has an accuracy of 64%. Has a processing time of 275.1028 seconds on the Nazief & Adriani algorithm and 1.0558 seconds on the Arifin Setiono algorithm. This research provides valuable insight into the effectiveness of both algorithms in sentiment analysis, which is useful for researchers in choosing the right algorithm and as a reference for further research.*

**Keywords:** *Stemming, Nazief & Adriani, Arifin Setiono, Sentiment Analysis, TikTok Shop.*