

**ANALISIS SENTIMEN PERSEPSI MASYARAKAT
TERHADAP PO ROSALIA INDAH MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

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



disusun oleh
ARDHIANSAPRATAMA
20.11.3488

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

**ANALISIS SENTIMEN PERSEPSI MASYARAKAT
TERHADAP PO ROSALIA INDAH MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN
SUPPORT VECTOR MACHINE (SVM)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ARDHIANSAPRATAMA

20.11.3488

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

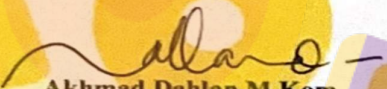
**ANALISIS SENTIMEN PERSEPSI MASYARAKAT
TERHADAP PO ROSALIA INDAH MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN
SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh
Ardhiansa Pratama

20.11.3488

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Januari 2025

Dosen Pembimbing,


Akhmad Dahlan M. Kom
NIK. 190302174

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PERSEPSI MASYARAKAT
TERHADAP PO ROSALIA INDAH MENGGUNAKAN
ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN
SUPPORT VECTOR MACHINE (SVM)**

yang disusun dan diajukan oleh

Ardhiansa Pratama

20.11.3488

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Januari 2025
Susunan Dewan Penguji

Nama Penguji

Ike Verawati, S.Kom., M.Kom
NIK. 190302237

Ali Mustopa, S.Kom., M.Kom
NIK. 190302192

Akhmad Dahlan, S.Kom., M.Kom
NIK. 190302174

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

Nama mahasiswa : Ardhiansa Pratama
NIM : 20.11.3488

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Persepsi Masyarakat Terhadap Po Rosalia Indah
Menggunakan Algoritma K-Nearest Neighbors (KNN) Dan Support Vector
Machine (SVM)**

Dosen Pembimbing : Akhmad Dahlan, 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, 23 Januari 2025

Yang Menyatakan,



Ardhiansa Pratama

HALAMAN PERSEMBAHAN

Pertama saya ucapkan puji syukur kepada Allah SWT atas segala nikmat dan kasih sayangnya. Skripsi ini penulis persembahkan untuk:

1. Bapak dan Ibu saya yang sangat saya sayangi, di mana telah memberikan kasih sayang yang berlimpah terhadap penulis sehingga bisa membuat penulis bertahan hidup sampai sekarang. Terimakasih telah menjadi orangtua yang sangat hebat.
2. Akhmad Dahlan, M.Kom. selaku dosen pembimbing saya yang sangat baik dan mau membimbing saya selama ini. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Bapak berikan. Semoga segala kebaikan Bapak mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.
3. Untuk Seli, Ghaluh, Raniya, Ardhiansa, Rangga, Dimas alias teman di bangku kuliah yang telah menemani *ups and downs* dari hidup penulis, terimakasih telah mendengarkan keluh-kesah serta mengerti dan memahami penulis setiap saat.

KATA PENGANTAR

Segala puji bagi Allah SWT atas limpahan rahmat, petunjuk dan karunia-Nya yang telah membantu peneliti menyelesaikan penyusunan skripsi ini. Skripsi ini disusun sebagai tahapan penting untuk memenuhi persyaratan kelulusan program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penyusunan skripsi ini tidak lepas dari doa, bantuan, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Keluarga tersayang, yang senantiasa memberikan doa terbaik, dukungan moral, materi, serta semangat dalam perjalanan penyusunan skripsi ini.
2. Akhmad Dahlan, M.Kom, selaku dosen pembimbing, atas dukungan, arahan, serta masukan berharga yang telah membimbing langkah-langkah penyusunan skripsi ini.
3. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.
4. Semua individu yang turut menyumbangkan gagasan, pandangan, dan sokongan dalam berbagai bentuk, yang telah membantu kelancaran perjalanan penyelesaian skripsi ini.

Yogyakarta, 23 Januari 2025

Penulis

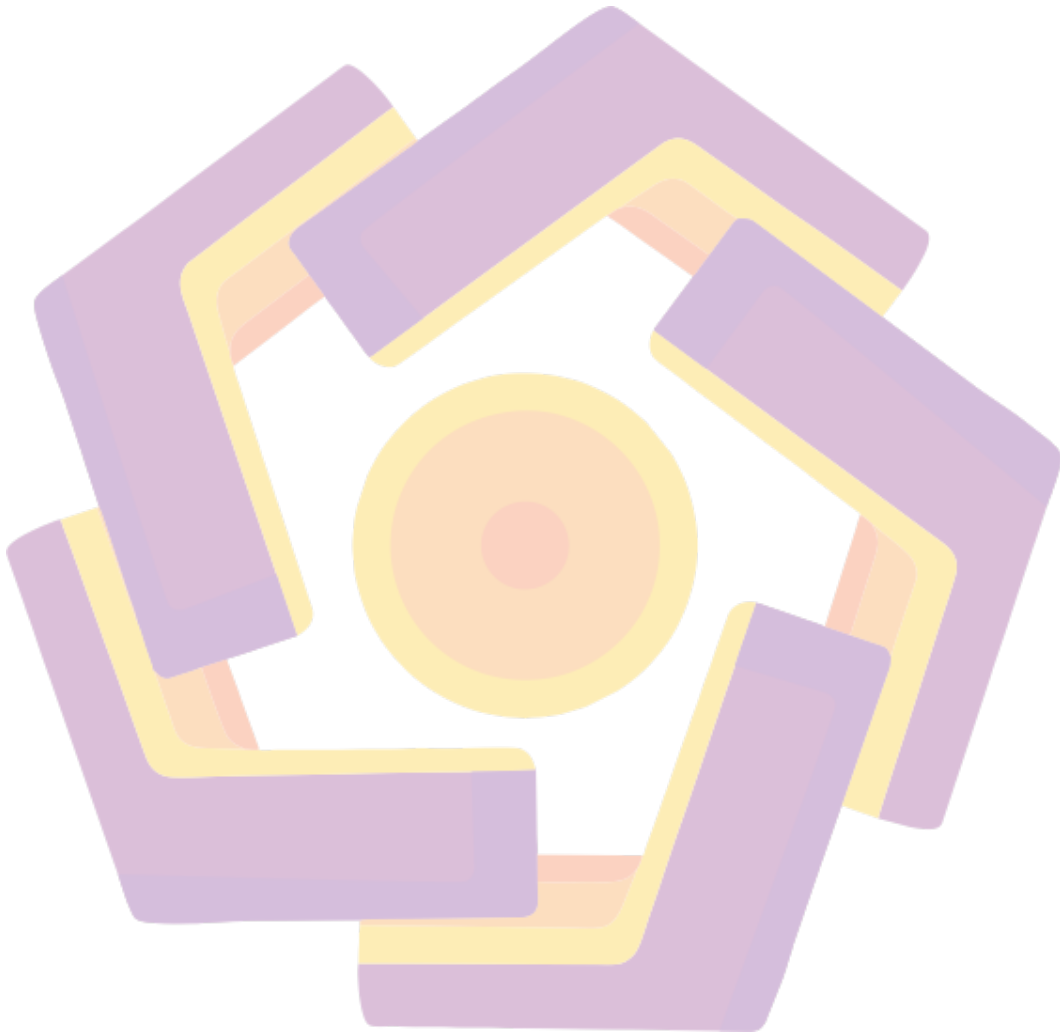
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	14
2.2.1. Rosalia Indah	14
2.2.2. Crawling Data	14

2.2.3.	Twitter	15
2.2.4.	Text Mining	15
2.2.5.	Analisis Sentimen	16
2.2.6.	Preprocessing	16
2.2.6.1	Pembersihan Data	16
2.2.6.2	Case Folding	16
2.2.6.3	Tokenizing	17
2.2.6.4	Stopword Removal.....	17
2.2.6.5	Stemming	17
2.2.7.	Translate.....	17
2.2.8.	Automatic Labeling	17
2.2.9.	Term Frequency – inverse document frequency.....	18
2.2.10.	K-Nearest Neighbor (KNN).....	18
2.2.11.	Support Vector Machine (SVM).....	19
2.2.12.	Confusion Matrix	21
2.2.13.	Python	21
BAB III METODE PENELITIAN		22
3.1	Objek Penelitian.....	22
3.2	Alur Penelitian	22
3.2.1.	Persiapan Data	22
3.2.2.	Preprocessing	23
3.2.3.	Translate.....	25
3.2.4.	Labeling	25
3.2.5.	Pembobotan.....	26
3.2.6.	Pembagian Data	26

3.2.7.	Klasifikasi Algoritma KNN	26
3.2.8.	Klasifikasi Algoritma SVM	27
3.2.9.	Evaluasi.....	28
3.3	Alat dan Bahan.....	29
3.3.1.	Data penelitian	29
3.3.2.	Alat.....	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Persiapan Data	30
4.2	Preprocessing Text.....	30
4.2.1	Cleaning	30
4.2.2	Case Folding	31
4.2.3	Normalisasi	32
4.2.4	Tokenize.....	32
4.2.5	Stopword.....	33
4.2.6	Stemming.....	34
4.3	Translate.....	34
4.4	Labeling	35
4.5	Pembobotan TF-IDF	37
4.6	Klasifikasi	37
4.7	Implementasi Algoritma KNN.....	38
4.8.1	Evaluasi Algoritma KNN.....	40
4.8	Implementasi algoritma SVM.....	41
4.7.1	Evaluasi algoritma SVM.....	43
4.9	Analisis perbandingan hasil	44
BAB V PENUTUP		47

5.1	Kesimpulan	47
5.2	Saran	48
REFERENSI		49

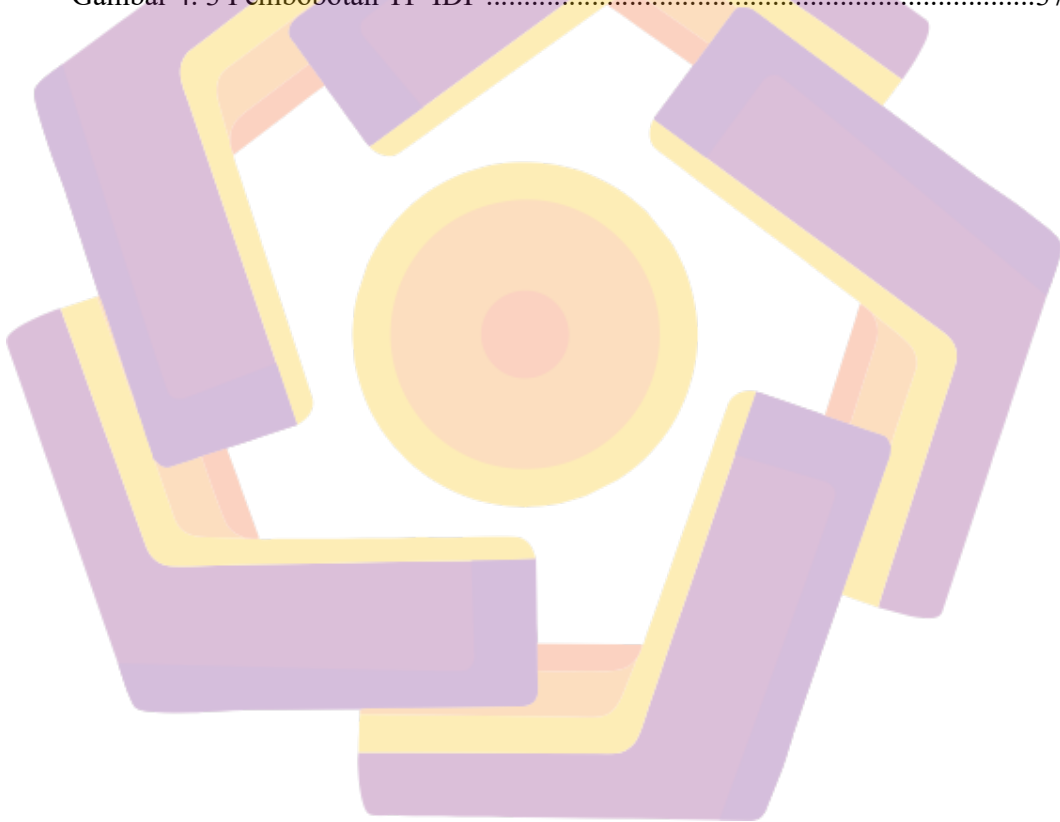


DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 Confusion Matrix	21
Tabel 4. 1 Cleaning	31
Tabel 4. 2 Case Folding	31
Tabel 4. 3 Normalisasi	32
Tabel 4. 4 Tokenize	33
Tabel 4. 5 Stopword	33
Tabel 4. 6 Stemming	34
Tabel 4. 7 Translate	35
Tabel 4. 8 Hasil Automatic Labelling	36
Tabel 4. 9 Hasil Nilai K Terbaik	38
Tabel 4. 10 Parameter Terbaik	40
Tabel 4. 11 Confusion Matrix KNN	40
Tabel 4. 12 Classification Report KNN	41
Tabel 4. 13 Parameter Pemilihan Kernel rbf	41
Tabel 4. 14 Hasil Kernel Rbf SVM	42
Tabel 4. 15 Confusion Matrix SVM	43
Tabel 4. 16 Classification Report SVM	44
Tabel 4. 17 Hasil KNN dengan KNeighborsClassifier	45
Tabel 4. 18 Hasil SVM dengan kernel Rbf	45

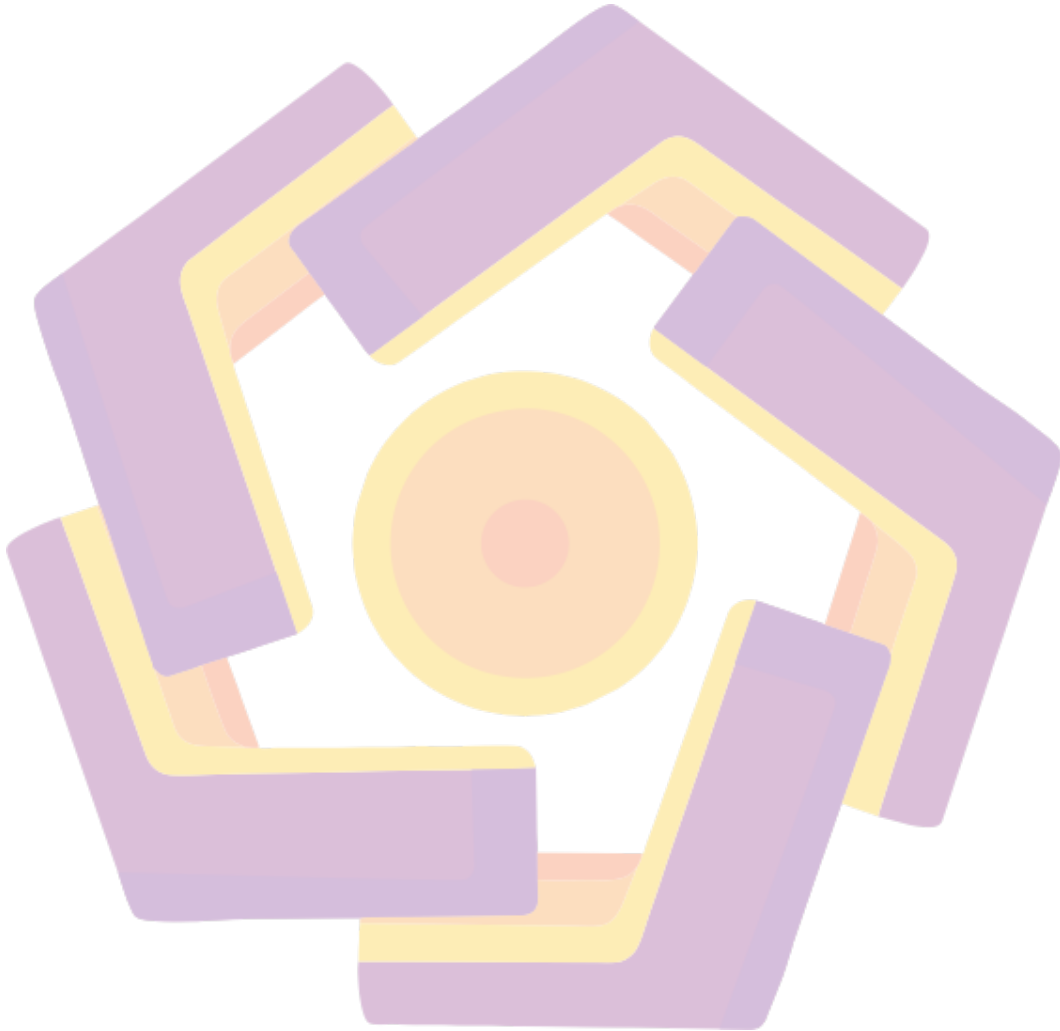
DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	22
Gambar 3. 2 Alur Algoritma KNN	27
Gambar 3. 3 Alur Algoritma SVM	28
Gambar 4. 1 Data Mentah Hasil Crawling.....	30
Gambar 4. 2 World Cloud.....	36
Gambar 4. 3 Pembobotan TF-IDF	37



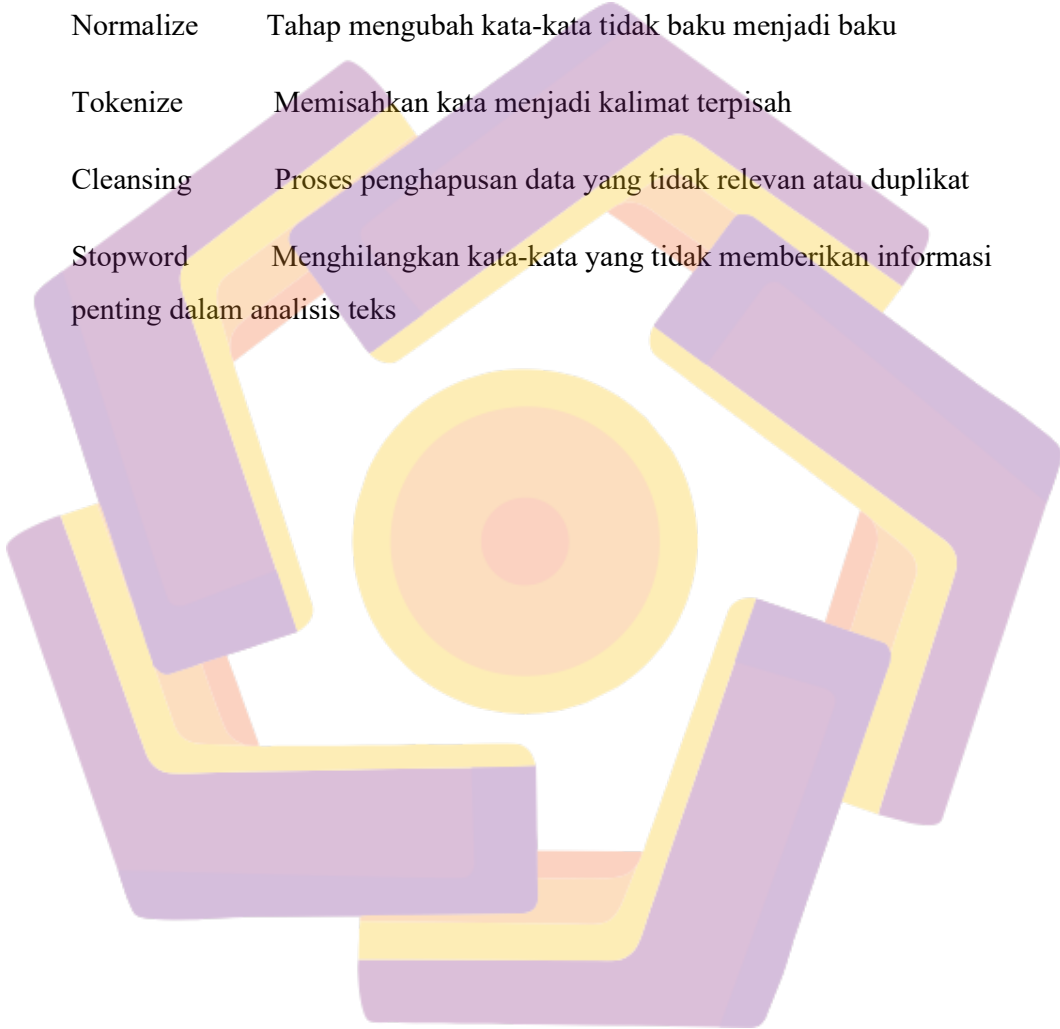
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
KNN	K-Nearest Neighbors
Po	Perusahaan Otobus



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

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap PO Rosalia Indah dengan menggunakan data dari Twitter. Data dikumpulkan melalui proses *crawling* dengan kata kunci "rosalia indah" dan "po rosalia indah", menghasilkan 2.665 tweet dalam rentang waktu Desember 2023 hingga Maret 2024. Setelah data diperoleh, dilakukan tahap preprocessing yang mencakup *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, *stemming*, dan translasi ke Bahasa Inggris. Data kemudian diberi bobot menggunakan metode TF-IDF sebelum dilakukan analisis sentimen.

Proses klasifikasi sentimen dilakukan dengan dua algoritma, yaitu *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN). Dataset dibagi menjadi 710 data latih dan 178 data uji menggunakan rasio 80:20, yang memberikan hasil terbaik. Berdasarkan hasil penelitian, algoritma SVM menunjukkan akurasi lebih tinggi (79%) dibandingkan KNN (75%). SVM lebih stabil dalam menangani semua kategori sentimen, terutama pada kelas Netral yang memiliki distribusi data lebih kecil. Sementara itu, KNN lebih efektif dalam menangani kelas dengan distribusi data dominan, seperti Positif dan Negatif.

Penelitian ini memberikan wawasan penting bagi PO Rosalia Indah untuk memahami persepsi masyarakat dan meningkatkan kualitas layanan mereka. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi pengembangan metode analisis sentimen di masa mendatang, khususnya dalam penerapan algoritma pembelajaran mesin.

Kata Kunci: Analisis Sentimen, Twitter, PO Rosalia Indah, SVM, KNN

ABSTRACT

This research aims to analyze public sentiment towards PO Rosalia Indah using data from Twitter. Data was collected through a crawling process with the keywords "rosalia indah" and "po rosalia indah", yielding 2,665 tweets during the period from December 2023 to March 2024. After the data was gathered, a preprocessing stage was conducted, including cleaning, case folding, normalization, tokenization, stopword removal, stemming, and translation to English. The data was then weighted using the TF-IDF method before sentiment analysis.

The sentiment classification process was conducted using two algorithms: Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). The dataset was split into 710 training data and 178 testing data using an 80:20 ratio, which produced the best results. Based on the findings, the SVM algorithm demonstrated higher accuracy (79%) compared to KNN (75%). SVM exhibited better stability across all sentiment categories, particularly in handling the Neutral class with smaller data distribution. Meanwhile, KNN was more effective in handling classes with dominant data distribution, such as Positive and Negative.

This research provides valuable insights for PO Rosalia Indah to understand public perceptions and improve the quality of their services. Additionally, this study is expected to serve as a reference for the development of sentiment analysis methods in the future, particularly in the application of machine learning algorithms.

Keywords: Sentiment Analysis, Twitter, PO Rosalia Indah, SVM, KNN