

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
FILM “AGAK LAEN” MENGGUNAKAN ALGORITMA K-
NEAREST NEIGHBORS (KNN)**

SKRIPSI

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



disusun oleh

RUSDI ALAM SYAH

20.11.3510

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
FILM “AGAK LAEN” MENGGUNAKAN ALGORITMA K-
NEAREST NEIGHBORS (KNN)**

SKRIPSI

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



disusun oleh

RUSDI ALAM SYAH

20.11.3510

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP FILM
"AGAK LAEN" MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBORS (KNN)**

yang disusun dan diajukan oleh

Rusdi Alam Syah

20.11.3510

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 16 Desember 2024

Agung Pambudi, S.T.,M.A.



Agung Pambudi, S.T.,M.A.
NTK. 190302012

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP FILM
"AGAK LAEN" MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBORS (KNN)**

yang disusun dan diajukan oleh
Rusdi Alam Syah

20.11.3510

Telah dipertaharkan di depan Dewan Penguji
pada tanggal 16 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Yuli Astuti, S.Kom., M.Kom.
NIK. 190302146



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



Agung Pambudi, S.T., M.A.
NIK. 190302012



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 16 Desember 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 : Rusdi Alam Syah
NIM : 20.11.3510

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Pengguna Twitter Terhadap Film “Agak Laen”
Menggunakan Algoritma K-Nearest Neighbors (KNN)**

Dosen Pembimbing : Agung Pambudi, S.T.,M.A

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, 16 Desember 2024

Yang Menyatakan,



The image shows an official stamp of Universitas AMIKOM Yogyakarta. The stamp is circular and contains the university's logo, the text 'UNIVERSITAS AMIKOM YOGYAKARTA', and the words 'METERAL TEMPEL'. Below the stamp is a handwritten signature in black ink.

Rusdi Alam Syah

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala nikmat dan kasih sayangnya. Skripsi ini penulis persembahkan untuk:

1. Kedua orang tua yang saya cintai , yang telah memberikan kasih sayang yang berlimpah terhadap penulis sehingga bisa membuat penulis bertahan hidup sampai sekarang. Terimakasih telah menjadi orangtua yang sangat hebat.
2. Keluarga besar yang selalu memberikan dukungan moral, kasih sayang, dan doa terbaik sampai saat ini.
3. Bapak Agung Pambudi, S.T.,M.A. selaku dosen pembimbing saya. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Bapak berikan. Semoga segala kebaikan Bapak mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.
4. Untuk Galih, Mas'ud, Sevilla alias teman di bangku kuliah yang telah menemani, terimakasih telah mendengarkan keluh-kesah penulis setiap saat.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk menyelesaikan program studi Informatika Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan ini tidak lepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Orang tua, keluarga, serta teman-teman yang selalu memberikan dukungan moral, doa, dan semangat yang tiada henti.
2. Bapak Agung Pambudi, S.T., M.A selaku pembimbing, atas arahan, dukungan, dan ilmu yang telah diberikan selama proses penyusunan skripsi ini.
3. Seluruh dosen dan staf Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, yang telah memberikan ilmu dan dukungan administratif selama masa perkuliahan.
4. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.

Yogyakarta, 2 Desember 2024

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.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	13

2.2.1	<i>Twitter</i>	13
2.2.2	<i>Text Mining</i>	13
2.2.3	Analisis Sentimen	13
2.2.4	<i>Python</i>	14
2.2.5	<i>K-Nearest Neighbors</i>	14
2.2.6	<i>Preprocessing</i>	15
BAB III METODE PENELITIAN		19
3.1	Objek Penelitian	19
3.2	Alur Penelitian	19
3.2.1	Observasi	20
3.2.2	Studi Pustaka	20
3.2.3	Pengumpulan Data	20
3.2.4	<i>Data Preprocessing</i>	21
3.2.5	<i>Labeling</i>	26
3.2.6	Klasifikasi Algoritma KNN	27
3.2.7	<i>Confussion Matrix</i>	27
3.3	Alat dan Bahan	27
3.3.1	Data Penelitian	28
3.3.1	Alat	28
BAB IV HASIL DAN PEMBAHASAN		29
4.1	Pengumpulan Data	29
4.2	<i>Data Pre-Processing</i>	30
4.2.1	<i>Case Folding</i>	30
4.2.2	<i>Cleaning</i>	31
4.2.3	Normalisasi	32

4.2.4	<i>Tokenizing</i>	33
4.2.5	<i>Stopwords Removal</i>	34
4.2.6	<i>Stemming</i>	35
4.3	<i>Labeling</i>	36
4.4	Hasil Persebaran Kata.....	38
4.5	Klasifikasi KNN.....	39
4.5.1	Proses KNN Pada Penelitian.....	39
4.5.2	Evaluasi <i>Confusion Matrix</i>	43
BAB V PENUTUP.....		46
5.1	Kesimpulan.....	46
5.2	Saran.....	47
REFERENSI.....		48


DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 <i>Confussion Matrix</i>	17
Tabel 4.1 Sampel Dataset	29
Tabel 4.2 Hasil <i>Case Folding</i>	30
Tabel 4.3 Hasil <i>Cleaning</i>	31
Tabel 4.4 Hasil Normalisasi	33
Tabel 4.5 Hasil <i>Tokenizing</i>	34
Tabel 4.6 Hasil <i>Stopwords Removal</i>	35
Tabel 4.8 Hasil <i>Labeling</i>	37
Tabel 4.9 Persebaran Data	37
Tabel 4.9 Nilai k terbaik	40
Tabel 4.10 <i>Classification Report Test Size 0.3</i>	42
Tabel 4.11 <i>Classification Report Test Size 0.2</i>	42
Tabel 4.12 <i>Classification Report Test Size 0.1</i>	43
Tabel 4.10 <i>Confussion Matrix test_size 0.1</i>	43

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	19
Gambar 3.2 Alur Pengumpulan Data	21
Gambar 3.3 Alur <i>Preprocessing</i>	22
Gambar 3.3 Alur <i>Case Folding</i>	23
Gambar 3.5 Alur <i>Cleaning</i>	23
Gambar 3.6 Alur Normalisasi	24
Gambar 3.7 Alur <i>Tokenizing</i>	25
Gambar 3.8 Alur <i>Stopwords</i>	25
Gambar 3.9 Alur <i>Stemming</i>	26
Gambar 4.1 <i>Source Code Case Folding</i>	32
Gambar 4.2 <i>Source Code Cleaning</i>	30
Gambar 4.3 <i>Source Code Normalisasi</i>	32
Gambar 4.4 <i>Source Code Tokenizing</i>	33
Gambar 4.5 <i>Source Code Stopwords</i>	35
Gambar 4.6 <i>Source Code Stemming</i>	36
Gambar 4.7 <i>Source Code Labeling</i>	37
Gambar 4.8 <i>Pie Chart</i> Persentase Persebaran Data	38
Gambar 4.9 Persebaran Kata Positif	38
Gambar 4.10 Persebaran Kata Negatif	39
Gambar 4.11 <i>K-Value</i>	40

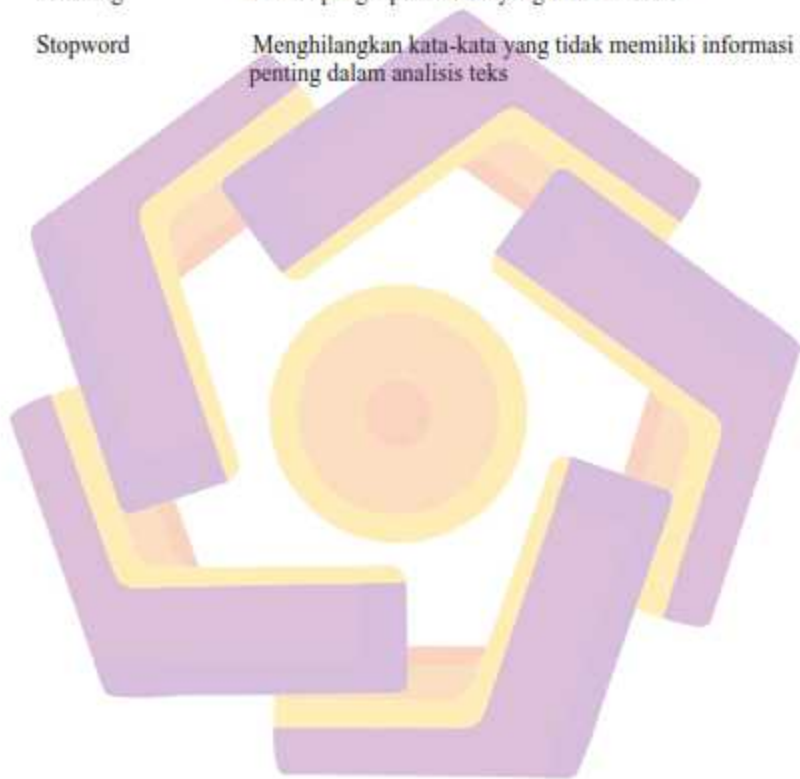
DAFTAR LAMBANG DAN SINGKATAN



KNN	K-Nearest Neighbors
API	Application Programming Interface
CSV	Comma-Separated Values
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
NLP	Natural Language Processing
AI	Artificial Intelligence
URL	Uniform Resource Locator
HTML	HyperText Markup Language
VADER	Valance Aware Dictionary and Sentiment Reasoner

DAFTAR ISTILAH

Case Folding	Proses mengubah semua data teks menjadi huruf kecil
Tokenizing	Memisahkan kata menjadi kalimat terpisah
Cleaning	Proses penghapusan data yang tidak relevan
Stopword	Menghilangkan kata-kata yang tidak memiliki informasi penting dalam analisis teks



INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen terhadap film “Agak Laen” pada platform *Twitter* dengan menggunakan algoritma *K-Nearest Neighbors*. Data yang digunakan dalam penelitian ini terdiri dari 4.378 *tweet* yang diperoleh dari *Twitter* melalui proses *crawling data*. Analisis sentimen dilakukan untuk mengotomatisasi proses identifikasi dan pengelompokan opini atau emosi dari teks, yang dalam konteks ini berasal dari komentar pengguna di platform *Twitter*. Tahapan penelitian meliputi pengumpulan data, proses *preprocessing*, serta penerapan metode klasifikasi menggunakan algoritma KNN. Pengujian dilakukan dengan variasi ukuran data uji, yaitu 30%, 20%, dan 10%. Kinerja ketiga data uji kemudian dibandingkan berdasarkan beberapa metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Pada ukuran data uji 30%, akurasi model mencapai 82%, dengan *precision* positif sebesar 100% dan negatif 74%. *Recall* untuk komentar positif tercatat 66%, sedangkan negatif 100%. *F1-score* untuk komentar positif adalah 79% dan negatif 85%. Data uji 20% menghasilkan akurasi 84,42%, dengan *recall* positif 70% dan negatif 100%, serta *precision* positif 99% dan negatif 76%. *F1-score* komentar positif adalah 82%, sedangkan negatif 86%. Dengan ukuran data uji 10%, akurasi meningkat menjadi 85,93%. *Recall* komentar positif meningkat menjadi 73% dan negatif 99%, sedangkan *precision* positif 99% dan negatif 78%. *F1-score* juga mengalami peningkatan untuk komentar positif sebesar 84%, sedangkan negatif mencapai 87%. Hasil penelitian ini menunjukkan pentingnya keseimbangan antara proporsi data pelatihan dan pemilihan ukuran data uji yang tepat dapat menghasilkan performa yang lebih efektif dan akurat dalam mendeteksi sentimen pada komentar pengguna di platform *Twitter*.

Kata Kunci : Analisis Sentimen, K-Nearest Neighbors, Cross-Validation, Film Agak Laen, Machine Learning

ABSTRACT

This study aims to analyze sentiment toward the movie "Agak Laen" on the Twitter platform using the K-Nearest Neighbors (KNN) algorithm. The dataset consists of 4,378 tweets collected via data crawling from Twitter. Sentiment analysis was performed to automate the identification and categorization of opinions or emotions from text, specifically user comments on Twitter. The research process included data collection, preprocessing, and applying the KNN classification method. Testing was done with varying test data proportions: 30%, 20%, and 10%. With 30% test data, the model achieved an accuracy of 82%, with a precision of 100% for positive comments and 74% for negative comments. Recall for positive comments was 66%, while for negative comments, it was 100%. The F1-score was 79% for positive comments and 85% for negative comments. Using 20% test data, accuracy was 84.42%, with recall of 70% for positive comments and 100% for negative comments, and precision of 99% for positive and 76% for negative comments. The F1-score was 82% for positive comments and 86% for negative comments. With 10% test data, accuracy increased to 85.93%, recall for positive comments rose to 73%, and for negative comments, it was 99%. Precision for positive and negative comments was 99% and 78%, respectively, with an F1-score of 84% for positive comments and 87% for negative comments. These results highlight the importance of balancing training data proportions, selecting the optimal test data size, and ensuring the model's stability for accurate sentiment detection on Twitter.

Keyword: *Sentiment Analysis, K-Nearest Neighbors, Cross-Validation, Agak Laen Movie, Machine Learning*