

**IDENTIFIKASI BUZZER BERBASIS WEBSITE PADA
KOMENTAR TIKTOK MENGGUNAKAN ALGORITMA
MACHINE LEARNING**

SKRIPSI

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



disusun oleh

AKMAL FAUZAN RESTU AGUNG

21.11.4154

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**IDENTIFIKASI BUZZER BERBASIS WEBSITE PADA
KOMENTAR TIKTOK MENGGUNAKAN ALGORITMA
MACHINE LEARNING**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AKMAL FAUZAN RESTU AGUNG

21.11.4154

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**IDENTIFIKASI BUZZER BERBASIS WEBSITE PADA KOMENTAR
TIKTOK MENGGUNAKAN ALGORITMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Akmal Fauzan Restu Agung

21.11.4154

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Desember 2024

Dosen Pembimbing,



Nuri Cahyono, M.Kom
NIK. 190302278

HALAMAN PENGESAHAN

SKRIPSI

**IDENTIFIKASI BUZZER BERBASIS WEBSITE PADA KOMENTAR
TIKTOK MENGGUNAKAN ALGORTIMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Akmal Fauzan Restu Agung

21.11.4154

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Desember 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Theopilus Bayu Sasongko, S.Kom., M.Eng
NIK. 190302250

Dina Maulina, S.Kom., M.Kom
NIK. 190302375

Nuri Cahyono, M.Kom
NIK. 190302278

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 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 : Akmal Fauzan Restu Agung
NIM : 21.11.4154

Menyatakan bahwa Skripsi dengan judul berikut:

Identifikasi Buzzer Berbasis Website pada Komentar tiktok menggunakan algoritma machine learning

Dosen Pembimbing : Nuri Cahyono, 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, 19 Desember 2024

Yang Menyatakan,

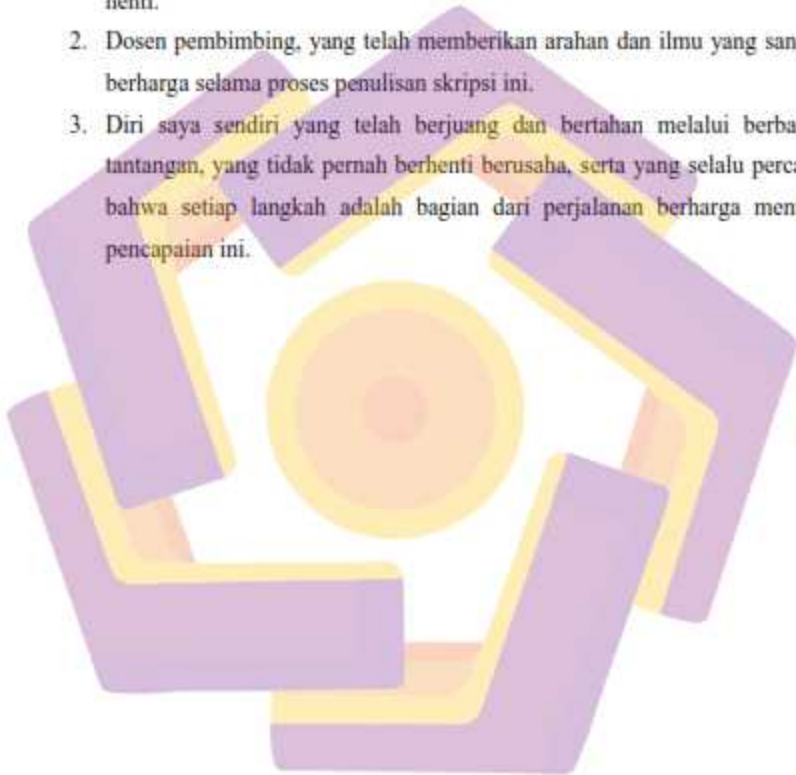


Akmal Fauzan Restu Agung

HALAMAN PERSEMBAHAN

Dengan rasa syukur dan penuh hormat, skripsi ini saya persembahkan untuk :

1. Ibu tercinta yang selalu memberikan doa, dukungan, dan kasih sayang tanpa henti.
2. Dosen pembimbing, yang telah memberikan arahan dan ilmu yang sangat berharga selama proses penulisan skripsi ini.
3. Diri saya sendiri yang telah berjuang dan bertahan melalui berbagai tantangan, yang tidak pernah berhenti berusaha, serta yang selalu percaya bahwa setiap langkah adalah bagian dari perjalanan berharga menuju pencapaian ini.



KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan Strata 1 di Program Studi Informatika di Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, saya mendapatkan banyak dukungan dan bimbingan dari berbagai pihak. Dengan penuh rasa hormat dan terima kasih, saya sampaikan apresiasi yang mendalam kepada ibu tercinta yang selalu memberikan doa, dukungan, dan kasih sayang tanpa henti. Rasa terima kasih yang tulus juga saya sampaikan kepada Bapak Nuri Cahyono, M.Kom, selaku dosen pembimbing, yang telah memberikan arahan dan ilmu yang sangat berharga selama proses penulisan skripsi ini.

Akhir kata, saya menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, saya terbuka terhadap kritik dan saran yang membangun untuk perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat bagi pembaca serta memberikan kontribusi bagi pengembangan ilmu pengetahuan dan memberikan dampak positif untuk masyarakat luas.

Yogyakarta, 19 Desember 2024

DAFTAR ISI

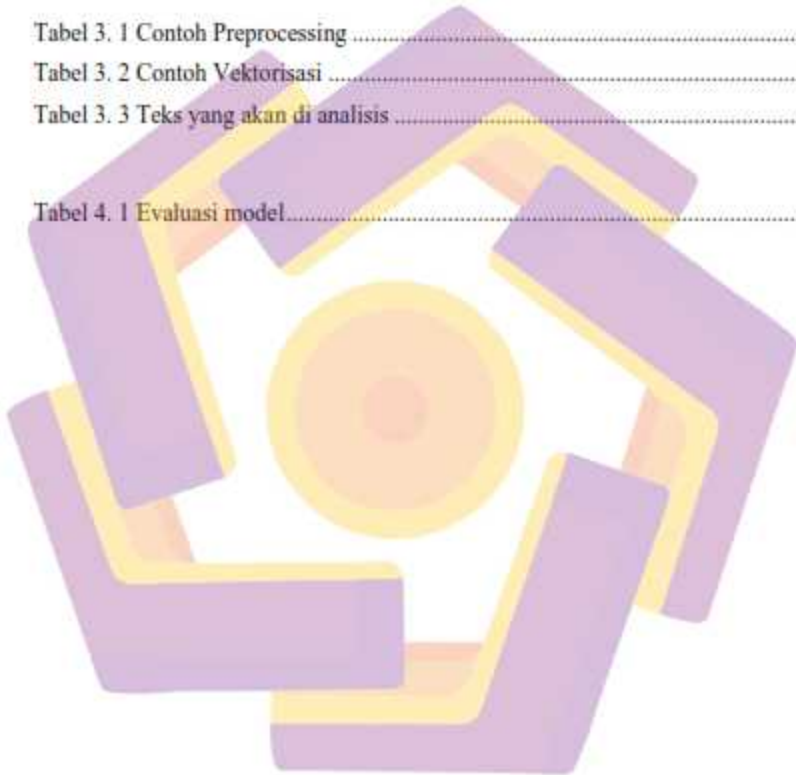
HALAMAN JUDUL	1
DAFTAR ISI.....	2
DAFTAR TABEL.....	10
DAFTAR GAMBAR.....	11
DAFTAR LAMPIRAN.....	13
DAFTAR ISTILAH.....	14
INTISARI.....	15
ABSTRACT.....	16
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Dasar Teori.....	13
2.2.1 Tiktok.....	13
2.2.2 Buzzer.....	13
2.2.3 Text Preprocessing.....	14
2.2.4 TF – IDF (Term Frequency - Inverse Document Frequency).....	15

2.2.5	Cosine Similarity	16
2.2.6	SMOTE	17
2.2.7	Standard Scaler	18
2.2.8	Machine Learning	18
2.2.9	Evaluasi Model	28
2.2.10	Confusion Matrix	29
2.2.11	Cross Validation	31
2.2.12	Flask	32
2.2.13	Docker	33
2.2.14	Google Cloud Platform	34
BAB III METODE PENELITIAN		37
3.1	Alur Penelitian	37
3.1.1	Data Gathering	38
3.1.2	Data Cleaning	38
3.1.3	Text Preprocessing	38
3.1.4	TF - IDF (Term Frequency - Inverse Document Frequency)	40
3.1.5	Cosine Similarity	41
3.1.6	Clustering menggunakan K - Means	42
3.1.7	Split Data	43
3.1.8	SMOTE	44
3.1.9	Modeling	44
3.1.10	Evaluasi	44
3.1.11	Cross Validation	45
3.2	Rancangan Web	46
3.3	Build Docker Image	47

3.4	Deployment Google Cloud Platform	48
3.5	Alat dan Bahan	48
BAB IV HASIL DAN PEMBAHASAN		49
4.1	Data Gathering	49
4.2	Data Cleaning	49
4.3	Text Preprocessing	50
4.4	Vektorisasi menggunakan TF-IDF	51
4.5	Cosine Similarity	52
4.6	Clustering menggunakan K-Means	54
4.7	Split Data	58
4.8	SMOTE	58
4.9	Modeling	59
4.10	Evaluasi Model	59
4.11	Cross Validation	63
4.12	Implementasi Web	65
4.13	Build Docker Image	70
4.13	Deployment Google Cloud Platform	71
BAB V PENUTUP		77
5.1	Kesimpulan	77
5.2	Saran	77
REFERENSI		79
LAMPIRAN		83

DAFTAR TABEL

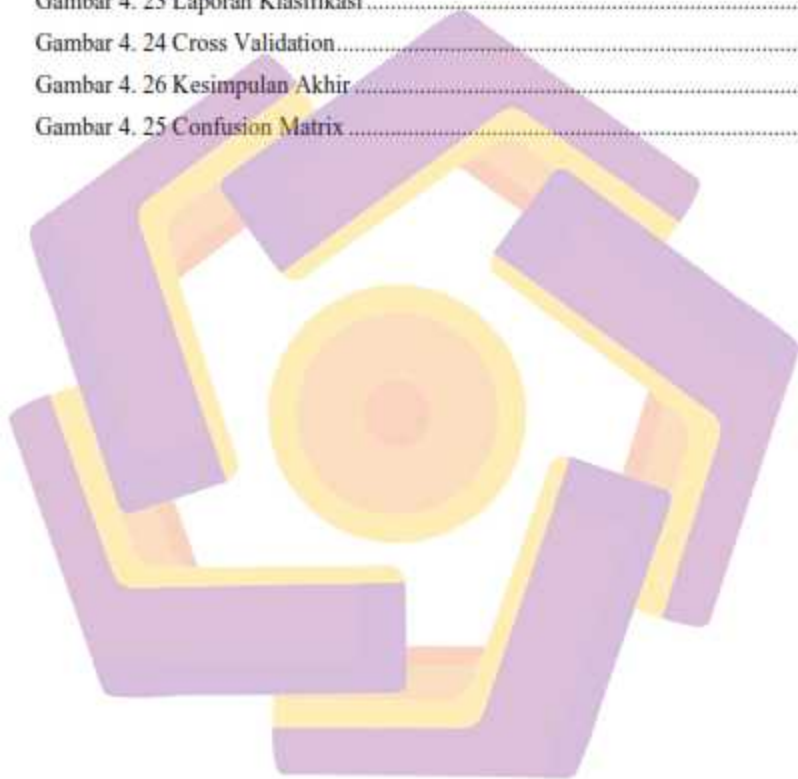
Tabel 2. 1 Literatur Review	8
Tabel 3. 1 Contoh Preprocessing	39
Tabel 3. 2 Contoh Vektorisasi	40
Tabel 3. 3 Teks yang akan di analisis	41
Tabel 4. 1 Evaluasi model.....	59



DAFTAR GAMBAR

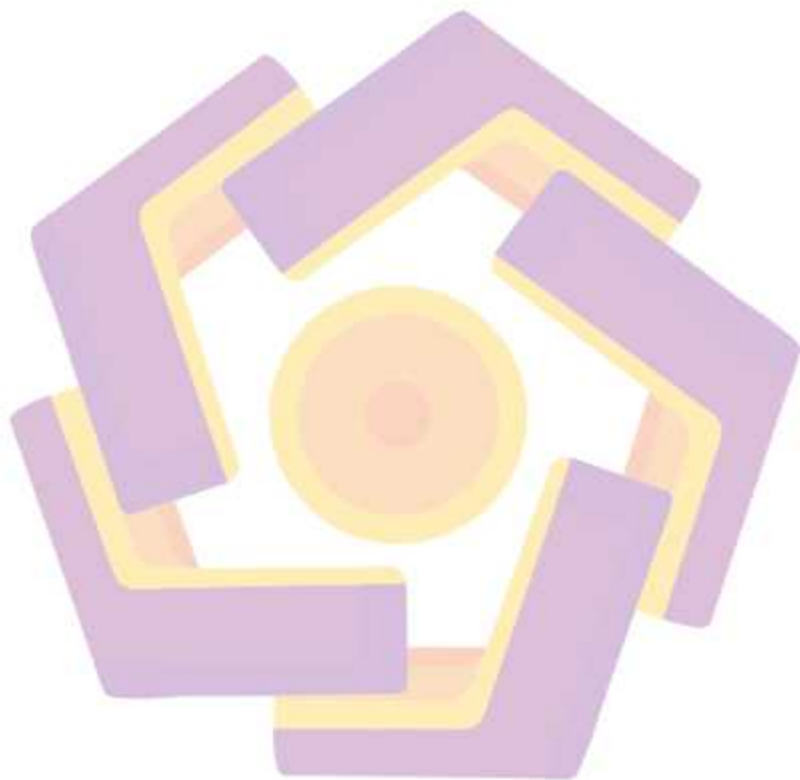
Gambar 2. 1 Logo TikTok	13
Gambar 2. 2 Logistic Regression	20
Gambar 2. 3 Decision Tree	22
Gambar 2. 4 Random Forest	23
Gambar 2. 5 Support Vector Machine	24
Gambar 2. 6 Confusion Matrix	29
Gambar 2. 7 Flask	32
Gambar 2. 8 Docker	33
Gambar 2. 9 Google Cloud Platform	34
Gambar 3. 1 Alur Penelitian	37
Gambar 4. 1 Hasil scraping data	49
Gambar 4. 2 Sebelum data cleaning	50
Gambar 4. 3 Setelah data cleaning	50
Gambar 4. 4 Sebelum text preprocessing	50
Gambar 4. 5 Setelah text preprocessing	51
Gambar 4. 6 Hasil vektorisasi	51
Gambar 4. 7 Hasil cosine similarity	52
Gambar 4. 8 Hasil analisis	54
Gambar 4. 9 Cluster natural comment	55
Gambar 4. 10 Cluster Buzzer	56
Gambar 4. 11 Statistik setiap Cluster	57
Gambar 4. 12 Confusion matrix Logistic Regresion	61
Gambar 4. 13 Confusion matrix Random Forest	61
Gambar 4. 14 Confusion matrix Decision Tree	62
Gambar 4. 15 Confusion matrix SVM	63
Gambar 4. 16 Cross Validation	64
Gambar 4. 17 Upload file	65

Gambar 4. 18 Visualisasi hasil clustering	66
Gambar 4. 19 Pola komentar buzzer	67
Gambar 4. 20 Contoh Komentar	67
Gambar 4. 21 Opsi akhir Clustering	68
Gambar 4. 22 Upload file untuk modeling	68
Gambar 4. 23 Laporan Klasifikasi	69
Gambar 4. 24 Cross Validation	69
Gambar 4. 26 Kesimpulan Akhir	70
Gambar 4. 25 Confusion Matrix	70



DAFTAR LAMPIRAN

Lampiran 1 Kode dan analisis Manual	83
Lampiran 2 kode Implementasi ke Flask	91



DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Buzzer	Individu atau akun yang memperluas jangkauan pesan di media sosial untuk tujuan tertentu.
Text Preprocessing	Tahap pengolahan teks untuk membersihkan dan menyiapkan data agar siap dianalisis.
Cosine Similarity	Metode untuk menghitung kemiripan antara dua vektor teks berdasarkan sudut kosinus.
Scaling	Proses mengubah skala atau rentang data, sering kali dengan menstandarkan atau menormalkan nilai, sehingga fitur data memiliki proporsi yang seragam.
Standard Scaler	Metode standarisasi data sehingga memiliki rata-rata 0 dan deviasi standar 1.
Machine Learning	Cabang ilmu komputer yang menggunakan algoritma untuk mendeteksi pola dalam data.
Supervised Learning	Pendekatan dalam machine learning yang menggunakan data berlabel untuk melatih model.
Unsupervised Learning	Pendekatan dalam pembelajaran mesin yang tidak menggunakan data berlabel, bertujuan untuk menemukan struktur tersembunyi atau pola dari data yang diberikan.
Modeling	Proses dalam machine learning di mana model dibuat, dilatih, dan diuji menggunakan data agar dapat membuat prediksi atau klasifikasi pada data baru.
Cross Validation	Metode evaluasi model dengan membagi data ke beberapa bagian untuk pelatihan dan pengujian.

INTISARI

Media sosial telah menjadi platform informasi yang bebas untuk berekspresi, namun dengan meningkatnya penggunaannya, muncul fenomena buzzer yang memanfaatkan media sosial untuk tujuan tertentu, seperti penyebaran informasi yang terorganisir. TikTok, sebagai salah satu platform yang paling populer, memiliki potensi besar dalam mempengaruhi opini publik melalui komentar-komentar yang diberikan oleh penggunanya. Penelitian ini bertujuan untuk mengidentifikasi buzzer pada komentar TikTok dengan menggunakan algoritma machine learning dan mengimplementasikan ke dalam sebuah website. Metode yang digunakan mencakup pengumpulan data komentar TikTok, pra-proses teks, analisis teks menggunakan Cosine Similarity, clustering, dan penerapan algoritma klasifikasi seperti Random Forest, Support Vector Machine (SVM), Decision Tree, dan Logistic Regression untuk mengidentifikasi pola perilaku buzzer. Untuk evaluasi model, digunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, serta dilakukan cross-validation untuk mengidentifikasi potensi overfitting pada model. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemahaman dinamika interaksi di media sosial serta pengembangan sistem deteksi buzzer secara otomatis dan efektif. Sistem ini kemudian diimplementasikan dalam bentuk aplikasi web Flask yang di-deploy ke Google Cloud untuk mempermudah akses dan penggunaan oleh pengguna secara global.

Kata kunci: Buzzer, Machine learning, Cosine Similarity, Clustering, Random Forest, Support Vector Machine, F1-score, cross-validation, Flask, Google Cloud.

ABSTRACT

Social media has become a free platform for information expression, but with its increasing use comes the phenomenon of buzzers who utilize social media for specific purposes, such as organized information dissemination. TikTok, as one of the most popular platforms, has great potential in influencing public opinion through comments provided by its users. This research aims to identify buzzers in TikTok comments using machine learning algorithms and implement it into a website. The methods used include collecting TikTok comment data, text preprocessing, text analysis using Cosine Similarity, clustering, and applying classification algorithms such as Random Forest, Support Vector Machine (SVM), Decision Tree, and Logistic Regression to identify buzzer behavior patterns. For model evaluation, evaluation metrics such as accuracy, precision, recall, and F1-score are used, and cross-validation is performed to identify potential overfitting in the model. The results of this research are expected to contribute to the understanding of interaction dynamics in social media as well as the development of an automatic and effective buzzer detection system. This system is then implemented in the form of a Flask web application that is deployed to Google Cloud to facilitate access and use by users globally.

Keyword : *Buzzer, Machine learning, Cosine Similarity, Clustering, Random Forest, Support Vector Machine, F1-score, cross-validation, Flask, Google Cloud.*