

**KOMPARASI ALGORITMA PADA ANALISIS SENTIMEN
KOMENTAR YOUTUBE TERHADAP KERETA
CEPAT WHOOSH**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 - Informatika



disusun oleh

MOCHAMAD RIFALDI MUKIM

21.11.3931

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**KOMPARASI ALGORITMA PADA ANALISIS SENTIMEN
KOMENTAR YOUTUBE TERHADAP KERETA
CEPAT WHOOSH**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 - Informatika



disusun oleh

MOCHAMAD RIFALDI MUKIM

21.11.3931

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI ALGORITMA PADA ANALISIS SENTIMEN KOMENTAR
YOUTUBE TERHADAP KERETA CEPAT WHOOSH**

yang disusun dan diajukan oleh

Mochamad Rifaldi Mukim

21.11.3931

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 31 Juli 2025

Dosen Pembimbing,



Ferian Fauzi Abdullah S. Kom., M. Kom.
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
KOMPARASI ALGORITMA PADA ANALISIS SENTIMEN KOMENTAR
YOUTUBE TERHADAP KERETA CEPAT WHOOSH

yang disusun dan diajukan oleh

Mochamad Rifaldi Mukim
21.11.3931

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

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

Rumini, S.Kom., M.Kom.
NIK. 190302246

Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Mochamad Rifaldi Mukim
NIM : 21.11.3931

Menyatakan bahwa Skripsi dengan judul berikut:

Komparasi Algoritma Pada Analisis Sentimen Komentar Youtube Terhadap Kereta Cepat Whoosh

Dosen Pembimbing : Ferian Fauzi Abdulloh S. Kom., 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, 31 Juli 2025

Yang Menyatakan,



Mochamad Rifaldi Mukim

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “KOMPARASI ALGORITMA PADA ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP KERETA CEPAT WHOOSH”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi S1-Informatika di Universitas Amikom Yogyakarta.

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang telah memberikan Rahmat, Hidayah, serta PetunjukNya sehingga penulis bisa melewati segala rintangan dan halangan.
2. Bapak prof. Dr. M. Suyanto, M.M. selaku rector Universitas Amikom Yogyakarta.
3. Bapak Prof. Dr. Kusri, M.Kom. selaku Dekan fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Ferian Fauzi Abdulloh, M.Kom. selaku dosen pembimbing saya yang telah memberikan waktu, saran, kritik, motivasi dan bimbingan dalam skripsi ini sehingga dapat menyelesaikan skripsi selesai dengan baik.
5. Bapak dan Ibu selaku kedua orang tua penulis yang telah mendidik dan memfasilitasi segala keperluan yang diperlukan oleh penulis.
6. Teman-teman dan rekan-rekan kelas 21 IF 02 di Universitas Amikom Yogyakarta, yang selalu memberikan semangat, dukungan, dan kebersamaan selama proses penelitian dan penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, penulis dengan senang hati menerima saran dan kritik yang membangun dari pembaca demi perbaikan di masa mendatang. Semoga skripsi ini bermanfaat bagi semua pihak yang membutuhkan, khususnya dalam pengembangan ilmu pengetahuan di bidang teknologi informasi.

Yogyakarta, 31 Juli 2025



Mochamad Rifaldi Mukim



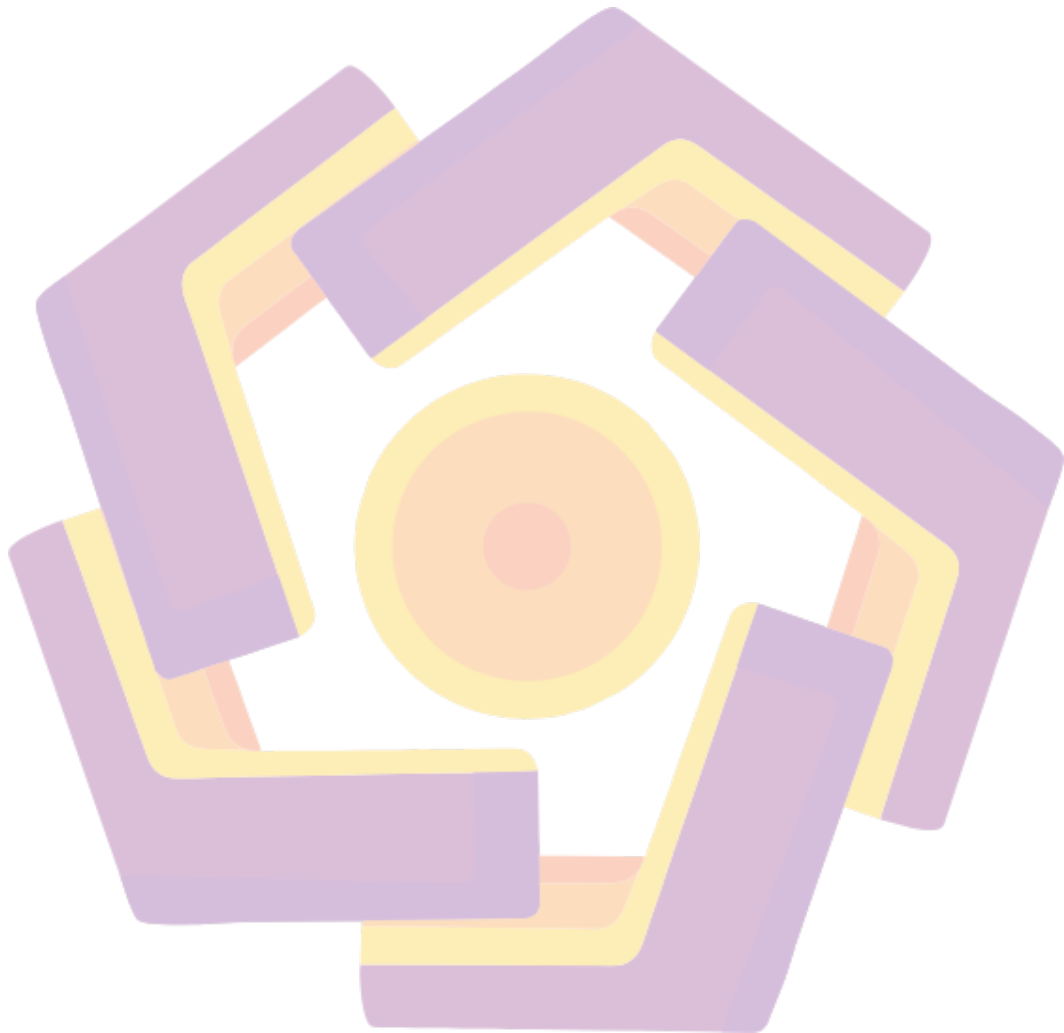
DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH.....	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	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.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	13
2.2.1 Analisis Sentimen	13

2.2.2 Support Vector Machine (SVM).....	13
2.2.3 Bidirectional Encoder Representations from Transformers (BERT)....	15
2.2.4 Perbandingan SVM dan BERT dalam Analisis Sentimen.....	16
2.2.5 Evaluasi Model Akurasi.....	16
BAB III METODE PENELITIAN	18
3.1 Alur Penelitian.....	18
3.1.1 Scraping Data.....	19
3.1.2 Labeling Data.....	19
3.1.3 Exploratory Data Analysis (EDA).....	19
3.1.4 Preprocessing Data.....	20
3.1.5 Pemodelan menggunakan SVM.....	22
3.1.6 Contoh Klasifikasi Sentimen Menggunakan SVM.....	24
3.1.6 Pemodelan menggunakan BERT	27
3.1.7 Evaluasi Model	29
3.1.8 Contoh Klasifikasi Sentimen Menggunakan BERT	29
3.2 Alat dan Bahan	31
3.2.1 Alat/instrumen	31
3.2.2 Bahan Penelitian	32
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Pengumpulan Data	33
4.2 Pelabelan Sentimen	34
4.3 Exploratory Data Analysis	35

4.3.1	Pemeriksaan Struktur Dataset	35
4.3.2	Distribusi Label Sentimen.....	36
4.3.3	Visualisasi Distribusi Sentimen	36
4.3.4	Kesimpulan dari Analisis Awal	36
4.4	Preprocessing Data	37
4.4.1	Normalisasi Data.....	37
4.4.2	Case Folding	39
4.4.3	Tokenizing	40
4.4.4	Stopword Removal	41
4.4.5	Stemming.....	43
4.4	Modeling menggunakan SVM	44
4.4.1	Proses Training dan Testing.....	45
4.4.2	Hasil Evaluasi Model.....	45
4.5	Modeling Menggunakan BERT	47
4.5.1	Proses Training dan Testing.....	48
4.5.2	Hasil Evaluasi Model.....	48
4.6	Perbandingan Hasil	50
4.7	Pembahasan Hasil	51
BAB V PENUTUP		54
5.1	Kesimpulan.....	54
5.2	Saran.....	55
REFERENSI		56

LAMPIRAN.....	60
---------------	----



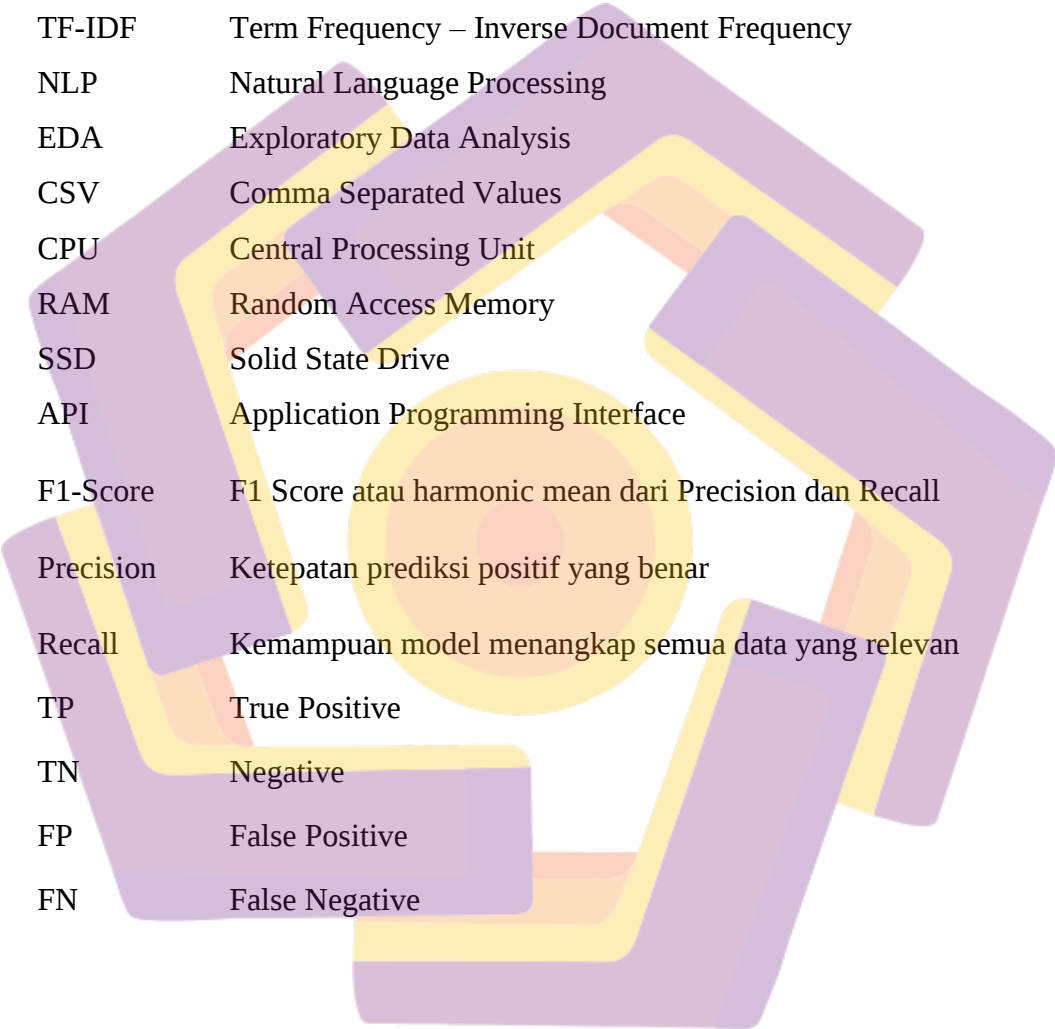
DAFTAR TABEL

Tabel 2. 1 - Keaslian Penelitian	10
Tabel 3. 1 - Normalisasi	20
Tabel 3. 2 - Contoh Data Diubah	24
Tabel 3. 3 - Tabel Tfidfvectorizer mengubah komentar menjadi vector	25
Tabel 3. 4 - Tabel Prediksi Sentimen Menggunakan Model SVM.....	26
Tabel 3. 5 - Evaluasi Kinerja Model	31
Tabel 4. 1 - Hasil Scraping	33
Tabel 4. 2 - Jumlah Data yang Dipakai.....	34
Tabel 4. 3 - Sentimen Positif.....	34
Tabel 4. 4 - Sentimen Negatif	35
Tabel 4. 5 - Hasil Normalisasi	37
Tabel 4. 6 - Hasil Case Folding	40
Tabel 4. 7 - Hasil Tokenizing.....	41
Tabel 4. 8 - Hasil Stopword Removal.....	42
Tabel 4. 9 - Hasil Stemming	44
Tabel 4. 10 - Hasil Classification Report SVM	45
Tabel 4. 11 – Hasil Classification Report BERT	48
Tabel 4. 12 - Perbandingan Hasil Evaluasi Model SVM dan BERT	51

DAFTAR GAMBAR

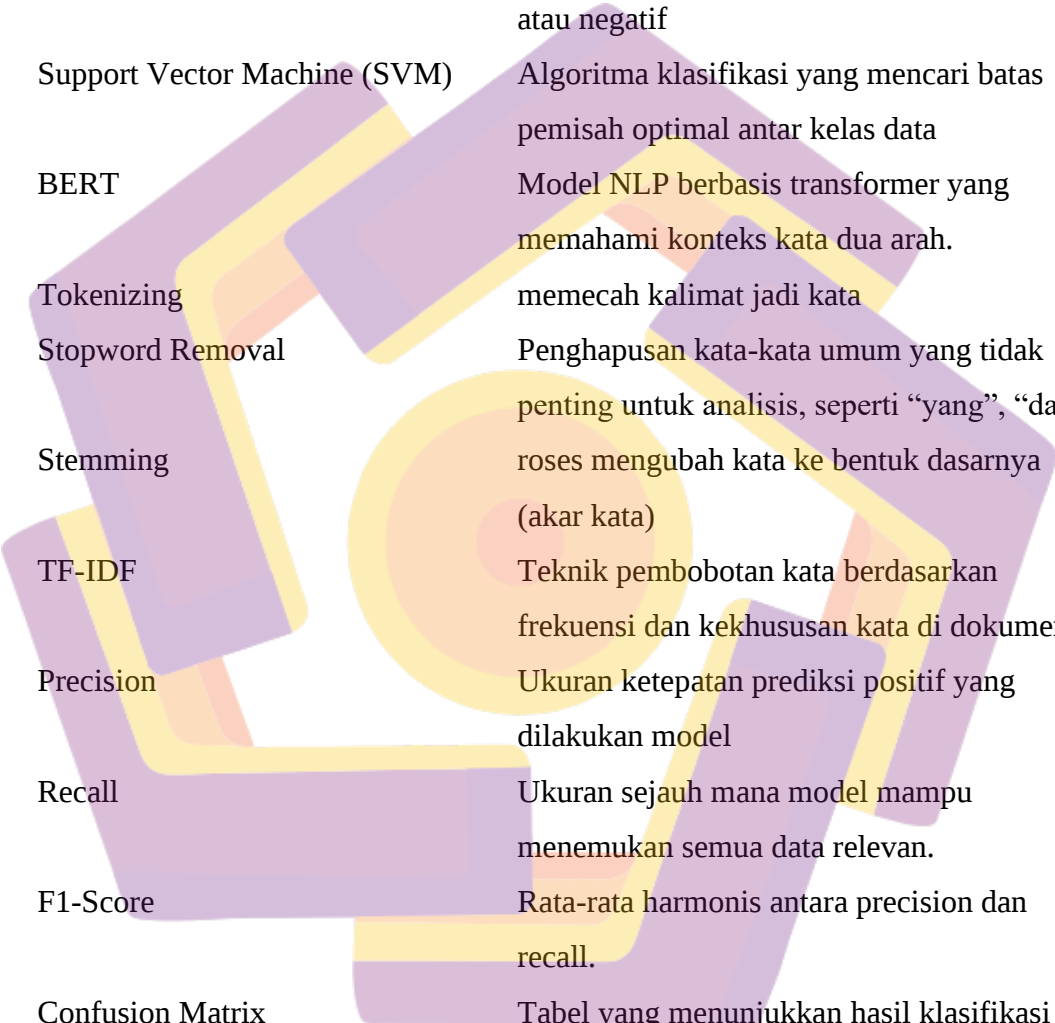
Gambar 3. 1 - Alur Penelitian	18
Gambar 3. 2 - Alur Case Folding	21
Gambar 3. 3 - Alur Tokenizing	21
Gambar 3. 4 - Alur Stopword Removal	22
Gambar 3. 5 - Alur Stemming	22
Gambar 4. 1 - Kode Program Tahapan Normalisasi	38
Gambar 4. 2 - Kode Program Case Folding	39
Gambar 4. 3 - Kode Program Tokenizing	40
Gambar 4. 4 - Kode Program Stopword Removal	42
Gambar 4. 5 - Kode Program Stemming	43
Gambar 4. 6 - Confusion Matrix SVM	46
Gambar 4. 7 - Confusion Matrix BERT	49
Gambar 4. 8 - Wordcloud Positif	52
Gambar 4. 9 - Wordcloud Negatif	52

DAFTAR LAMBANG DAN SINGKATAN



BERT	Bidirectional Encoder Representations from Transformers
SVM	Support Vector Machines
TF-IDF	Term Frequency – Inverse Document Frequency
NLP	Natural Language Processing
EDA	Exploratory Data Analysis
CSV	Comma Separated Values
CPU	Central Processing Unit
RAM	Random Access Memory
SSD	Solid State Drive
API	Application Programming Interface
F1-Score	F1 Score atau harmonic mean dari Precision dan Recall
Precision	Ketepatan prediksi positif yang benar
Recall	Kemampuan model menangkap semua data yang relevan
TP	True Positive
TN	Negative
FP	False Positive
FN	False Negative

DAFTAR ISTILAH



Analisis Sentimen	Proses mengklasifikasi opini dalam teks menjadi sentimen positif atau negatif
Support Vector Machine (SVM)	Algoritma klasifikasi yang mencari batas pemisah optimal antar kelas data
BERT	Model NLP berbasis transformer yang memahami konteks kata dua arah.
Tokenizing	memecah kalimat jadi kata
Stopword Removal	Penghapusan kata-kata umum yang tidak penting untuk analisis, seperti “yang”, “dan”
Stemming	proses mengubah kata ke bentuk dasarnya (akar kata)
TF-IDF	Teknik pembobotan kata berdasarkan frekuensi dan kekhususan kata di dokumen
Precision	Ukuran ketepatan prediksi positif yang dilakukan model
Recall	Ukuran sejauh mana model mampu menemukan semua data relevan.
F1-Score	Rata-rata harmonis antara precision dan recall.
Confusion Matrix	Tabel yang menunjukkan hasil klasifikasi benar dan salah dari model.
Exploratory Data Analysis (EDA)	Tahapan analisis awal untuk memahami karakteristik data

INTISARI

Penelitian ini bertujuan untuk membandingkan efektivitas dua algoritma dalam analisis sentimen komentar YouTube tentang proyek Kereta Cepat Whoosh. Algoritma yang diuji adalah Support Vector Machine (SVM) dan BERT (Bidirectional Encoder Representations from Transformers). Dataset terdiri dari 6.108 komentar, yang mencakup sentimen positif dan negatif. Data tersebut melalui beberapa tahap praproses, termasuk normalisasi, pelipatan huruf besar-kecil, tokenisasi, penghapusan stopwords, dan stemming. Model SVM dan BERT kemudian digunakan untuk mengklasifikasikan komentar menjadi sentimen positif dan negatif. Hasil evaluasi menunjukkan bahwa model SVM berkinerja lebih baik, mencapai akurasi 88,2%, dibandingkan dengan BERT dengan akurasi 87,3%. SVM juga lebih efektif dalam mendeteksi komentar negatif, dengan tingkat recall 89%. Di sisi lain, BERT berkinerja lebih baik dalam mendeteksi komentar positif dengan recall 90%, meskipun menghadapi tantangan dalam mengklasifikasikan komentar negatif. Perbedaan ini menunjukkan bahwa sementara BERT unggul dalam memahami konteks semantik, SVM lebih efisien untuk menangani data teks yang lebih sederhana. Secara keseluruhan, SVM terbukti lebih efektif dalam menganalisis sentimen pada komentar mengenai Kereta Cepat Whoosh.

Kata kunci: Analisis Sentimen, Kereta Cepat Whoosh, Support Vector Machine, BERT, Pra-pemrosesan Teks.

ABSTRACT

This study aims to compare the effectiveness of two algorithms in sentiment analysis of YouTube comments about the Kereta Cepat Whoosh project. The algorithms tested are Support Vector Machine (SVM) and BERT (Bidirectional Encoder Representations from Transformers). The dataset consists of 6,108 comments, which include both positive and negative sentiments. The data went through several preprocessing stages, including normalization, case folding, tokenization, stopword removal, and stemming. SVM and BERT models were then used to classify the comments into positive and negative sentiments. The evaluation results showed that the SVM model performed better, achieving an accuracy of 88.2%, compared to BERT with an accuracy of 87.3%. SVM was also more effective in detecting negative comments, with a recall rate of 89%. On the other hand, BERT performed better in detecting positive comments with a recall of 90%, although it faced challenges in classifying negative comments. This difference indicates that while BERT excels in understanding semantic context, SVM is more efficient for handling simpler text data. Overall, SVM proved to be more effective in analyzing sentiment on comments regarding Kereta Cepat Whoosh.

Keyword: Sentiment Analysis, Kereta Cepat Whoosh, Support Vector Machine, BERT, Text Preprocessing.