

# Public Sentiment Analysis Toward JKT48 On Twitter Using The Support Vector Machine

Chaidir Chalaf Islamy<sup>1\*</sup>, I Gusti Bagus Putra Pradwi<sup>2\*\*</sup>

\* Teknik Informatika, Universitas 17 Agustus 1945 Surabaya

\*\* Teknik Informatika, Universitas 17 Agustus 1945 Surabaya  
igustiputra99@gmail.com<sup>1</sup>, chaidirc@untag-sby.ac.id<sup>2</sup>

## Article Info

### Article history:

Received ...

Revised ...

Accepted ...

### Keyword:

JKT48, Analisis Sentimen,  
Support Vector Machine,  
Twitter, Idol

## ABSTRACT

This study analyzes public sentiment towards JKT48 on Twitter using the Support Vector Machine (SVM) algorithm. Data was collected by crawling tweets related to JKT48 management (JOT) and member scandals throughout January–December 2024. The analysis stages include data preprocessing (case folding, cleaning, normalization, tokenization, stopword removal, stemming), sentiment labeling using the Inset Lexicon dictionary, TF-IDF weighting, SVM model training, and evaluation using a confusion matrix. The classification results show high accuracy, namely 91.98%–92.03% for the test data and 95.27%–95.37% for the training data. Negative sentiment dominates public discussions, especially on scandal issues, while positive sentiment indicates fanaticism and loyalty from fans. Wordcloud is used to visualize the dominant words in each sentiment category. This study contributes to the study of sentiment analysis in the realm of local idol entertainment, as well as providing useful insights in decision-making and public communication by idol management.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Perkembangan media sosial mendorong perubahan yang signifikan pada pola komunikasi dan distribusi informasi. Platform seperti Twitter menjadi salah satu media utama dalam berbagi konten secara instan, baik teks, gambar maupun video. Kemampuan dalam mempercepat penyebaran informasi menjadi Twitter sebagai ruang interaktif yang dimanfaatkan untuk membahasa berbagai isu hiburan seperti film, dan budaya idol [1].

Fenomena idol sendiri merupakan bagian budaya populer yang berkembang pesat yang Dimana pada sekelompok publik figur mempromosikan melalui media massa untuk tampil sebagai penari, penyanyi pada panggung hiburan. Budaya ini tidak hanya menghibur tetapi membentuk hubungan sosial antara idol dan penggemarnya [2]. JKT48 merupakan idol asal Indonesia yang mengadopsi konsep “idol you can meet” yang artinya penggemar dapat berinteraksi langsung melalui pertunjukkan theater dan berbagai aktivitas [3]. Saat ini, JKT48 memiliki puluhan anggota aktif dari berbagai generasi, serta menerapkan aturan dengan Bernama

*Golden Rules* untuk menjaga citra dan profesionalisme, termasuk larangan berkencan [4]. Beberapa hal mengenai kebijakan hingga tindakan tepat yang dikelola oleh Manajemen JKT48 menjadi fondasi penting untuk menjaga struktur, etika maupun keberlangsungan grup. Namun, seiring berkembangnya media sosial, khususnya Twitter berbagai opini publik bermunculan terkait dugaan pelanggaran aturan oleh beberapa member. Hal ini dapat menimbulkan perdebatan pada kalangan penggemar. Dalam konteks ini, analisis sentimen diperlukan untuk memahami persepsi publik terhadap isu yang berkembang.

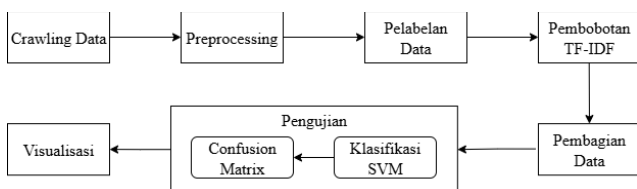
Penelitian yang dilakukan [1] berfokus pada analisis sentimen terhadap idol BTS di media sosial twitter menggunakan dua algoritma klasifikasi yaitu *naïve baiyes* dan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa algoritma *naïve baiyes* menghasilkan akurasi sebesar 79%, sedangkan *Support Vector Machine* memiliki akurasi lebih tinggi yaitu sebesar 81%, temuan ini mengindikasikan bahwa metode SVM memiliki performa yang baik dalam klasifikasi data teks berbagi opini pengguna. Penelitian lain oleh [5] menggunakan SVM untuk analisis

sentimen pengguna *Twitter* terkait konser K-Pop dengan dataset sebanyak 841 *tweet* berhasil mencapai akurasi sebesar 76,64% dengan 587 data positif dan 368 data negative, menunjukkan bahwa mayoritas komentar bernada positif dan menegaskan keandalan SVM untuk analisis sentimen teks di domain hiburan. Selanjutnya, studi oleh [6] mengevaluasi performa beberapa jenis kernel SVM pada 1.000 *tweet* *Twitter*. Menggunakan proses *cross-validation* 10-fold dengan pengulangan tiga kali, model SVM linear grid mencatat akurasi tertinggi sebesar 89,36%, dibandingkan kernel radial grid (84,97%) dan SVM standar (84,19%). Hasil ini mencerminkan kemampuan SVM dalam menangkap nuansa emosional dan pola sentimen pada media sosial.

Berdasarkan berbagai studi sebelumnya, metode *Support Vector Machine (SVM)* telah terbukti unggul dalam melakukan klasifikasi sentimen dengan tingkat akurasi tinggi yang melampaui metode lainnya. Oleh karena itu, peneliti memilih SVM sebagai algoritma untuk mengklasifikasikan opini public di platform *Twitter*. Penelitian ini bertujuan untuk memberikan pemahaman mendalam mengenai persepsi publik terhadap skandal anggota dan kebijakan manajemen grup idol JKT48. Dengan menganalisis berbagai cuitan dari pengguna *Twitter*, penelitian mengklasifikasikan opini menjadi sentimen positif, negatif dan netral menggunakan algoritma *Support Vector Machine (SVM)*. Metode ini dipilih karena telah terbukti memberikan akurasi tertinggi dalam klasifikasi data berbasis teks, khususnya dalam konteks media sosial. Dengan pendekatan analisis sentimen, peneliti tidak hanya evaluasi kinerja algoritma SVM dalam klasifikasi teks, tetapi juga menyediakan gambaran umum mengenai dinamika hubungan idol, penggemar dan manajemen. Hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap Keputusan yang bermanfaat dalam kebijakan komunikasi antara penggemar dan manajemen di Tengah sorotan public yang semakin digital.

## II. METODE

Bab ini menjelaskan sistematis dalam penelitian yang dilakukan dengan menggambarkan tahapan pada penelitian dimulai dari pengumpulan data, praprosesing data, pelabelan, pembobotan hingga pada proses klasifikasi dan evaluasi model hingga visualisasi.



Gambar 1. Alur Penelitian

### A. Crawling Data

Crawling Data merupakan Teknik pengumpulan data otomatis berdasarkan kata kunci maupun topik yang diberikan pengguna dengan disebut crawler dirancang algoritma

tertentu untuk pemindaian web dan membaca teks untuk mendapatkan informasi [7]. Crawling data ini menggunakan *tweet harvest* dengan tools library yang diciptakan oleh Helmi Satria digunakan untuk mengambil data pada Platform *Twitter* dengan tools API dibantu Bahasa Pemrograman Python [8].

### B. Preprocessing Data

Setelah tahap crawling data, tahap preprocessing data dilakukan untuk membersihkan hingga mengelola data teks untuk diolah lebih lanjut pada algoritma, preprocessing tersebut dilakukan untuk meningkatkan kualitas pada data [9].

- *Case Folding*

Metode untuk mengubah seluruh karakter dalam teks menjadi huruf kecil. Pada proses ini bertujuan untuk mengurangi variasi yang tidak diperlukan pada analisis. Dalam hal ini digunakan `lower()` melalui Bahasa pemrograman python[10].

- *Cleaning*

Metode yang bertujuan untuk memastikan data teks bersih, terstruktur. Pada proses ini menghapus unsur yang tidak diperlukan seperti symbol, tanda baca, tautan URL, angka dan karakter khusus yang terdapat pada hasil crawling. Digunakan library python berupa *regular expression (regex)* untuk penghapusan karakter khusus [9].

- *Normalisasi*

Proses mengubah kata tidak baku menjadi bentuk baku untuk mengatasi penggunaan Bahasa baku, singkatan dan slang ejaan umum pada media sosial. Normalisasi dilakukan untuk mencocokkan kata pada teks dengan kamus normalisasi untuk mudah dianalisis [9].

TABEL 1  
NORMALISASI DATA

Slang	Formal
Terlalu	terlalu
dkasih	dikasih
distu	disitu
ngbrol	bicara
ngrasa	merasa

- *Tokenization*

Proses memisahkan kalimat menjadi unit-unit terkecil kata atau karakter menjadi komponen per kata disebut token. Hal ini menyederhanakan teks sebagai kumpulan kata-kata terpisah.[9]

- *Stopwords Removal*

Proses menghapus kata-kata umum seperti “dengan”, “di”, “ke” yang sering muncul tetapi tidak memberikan kontribusi pada sentiment. Kata – kata ini tidak memiliki informasi penting untuk pemahaman konteks kalimat. Proses ini dilakukan menggunakan library pada python yang bernama *sastrawi* yang secara otomatis identifikasi dan menghapus

kata-kata penghubungnya. Hal ini menggunakan library NLTK dan Sastrawi Indonesian yang difungsikan pada teks yang sudah melalui tokenisasi[9].

- *Stemming*

Proses menghilangkan awal dan akhiran pada kata mendapatkan bentuk dasar tanpa mengubah makna dengan bertujuan menyederhanakan teks sehingga variasi kata dapat dianalisis untuk satu entitas, seperti mengubah kata “cintaan” jadi “cinta dan “bacaan” jadi “baca”. Pada hal ini digunakan library sastrawi yang dirancang untuk Bahasa Indonesia dengan pendekatan berbasis aturan [9].

### C. Pelabelan Data

Setelah tahap preprocessing data, maka dilakukan pelabelan data menggunakan Inset Lexicon merupakan kamus lexicon Berbahasa Indonesia yang dikembangkan tahun 2018[11]. INSET terdiri dari 3.609 data kata positif dan 6.609 data negative dengan masing-masing bobot skor sentimen disebut polarity score berkisar negatif (-5) – (-1), 0 Netral, 1-5 Positif. Setiap kata yang sudah melalui stemming. Jika total bobot lebih dari 0 maka diberi label positif, jika kurang dari 0 maka diberi label negatif kemudian jika nilai 0 maka dianggap netral [12].

### D. Pembobotan TF-IDF

Setelah tahap pelabelan data, data teks dikonversi menjadi representasi numerik dengan metode *Term frequency-inverse Document Frequency (TF-IDF)* merupakan Teknik pembobotan kata yang digunakan untuk mengukur seberapa penting sebuah kata terhadap pada tweet tersebut. [9].

$$idf_t = \log \left( \frac{N}{d_{ft}} \right) \quad (1)$$

$$W = \frac{d}{dt} = \text{Bobot TF-IDF} \quad (2)$$

Keterangan:

$TF_{dt}$  = Jumlah frekuensi kata yang muncul sebuah dokumen.

$IDF_{ft}$  = Jumlah frekuensi dokumen frekuensi dokumen tiap kata.

$idf_t$  = Jumlah frekuensi dokumen tiap kata.

$N$  = Jumlah total dokumen.

### E. Pembagian Data

Pada tahap selanjutnya, data dibagi menjadi dua bagian dengan rasio 80:20, merujuk ke 80% sebagai data latih dan 20% sebagai data uji. Rasio ini merujuk kepada data latih digunakan untuk membangun model klasifikasi sentimen dengan algoritma *Support Vector Machine (SVM)*, sementara

data uji digunakan untuk mengukur kinerja model pada label positif, negatif dan netral[10].

### F. Support Vector Machine

Pada penelitian ini menggunakan kerner linear karena bersifat linier. Model dimaksimalkan margin antar kelas positif, negatif dan netral, agar pemisahan data lebih optimal. Model hasil pelatihan ini digunakan prediksi sentimen pada data uji dan evaluasi akurasi. Margin maksimum antara kelas tersebut diupayakan agar pemisahan tidak tumpang tindih. Dari hal itu Hyperplane bekerja pada ruang berdimensi tinggi untuk memastikan pemisahan yang efektif sehingga model mampu memberikan prediksi yang baik.[13]

### G. Confusion Matrix

Evaluasi dilakukan menggunakan Confusion Matrix untuk mengukur akurasi, presisi, recall dan F1-Score. Metode ini membandingkan hasil prediksi dan data aktual sebagai penilaian performa dari model. Proses evaluasi menggunakan library *scikit-learn dan seaborn*[14].

Pada hal ini confusion matrix berbentuk tabel untuk menunjukkan data yang diklasifikasikan. Berikut tabel confusion matrix:

TABEL II  
CONFUSION MATRIX

Predicted Class	True Class		
	Positif	Negatif	Netral
Positif	True Positif	True Negatif	True Netral
Negatif	True Positif	True Negatif	True Netral
Netral	True Positif	True Negatif	True Netral

Keterangan:

$$Accuracy = ((TP_{Positive} + TP_{Negative} + TP_{Neutral})) / Total \text{ Prediksi}$$

$$Precision_i = (TP_i) / (TP_i + FP_i)$$

$$Recall_i = (TP_i) / (TP_i + FN_i)$$

$$F1_{macro} = (1/3) \times \Sigma (2 \times ((Precision_i \times Recall_i)) / (Precision_i + Recall_i))$$

### H. Visualisasi

Visualisasi dilakukan menggunakan 2 pendekatan utama yaitu wordcloud dan bar chart. Wordcloud digunakan sebagai penggambaran kata-kata yang paling sering muncul dalam tweet berdasarkan sentimen positif, negatif dan netral. Selain itu visualisasi batang digunakan untuk menampilkan distribusi jumlah tweet berdasarkan kategori distribusi sentiment. Dari hal ini memberikan Gambaran umum tentang

persebaran opini dengan sentimen secara visual. Pada visualisasi dilakukan menggunakan library python wordcloud dan matplotlib.

### III. HASIL DAN PEMBAHASAN

Bab ini digunakan hasil dan pembahasan untuk penyajian data pada temuan temuan penelitian. Bagian ini bertujuan untuk menggambarkan proses analisis dan interpretasi data secara visualisasi guna untuk menunjukkan keberhasilan dalam proses metode dalam penelitian:

#### A. Crawling Data

Proses pengambilan data digunakan library tweet harvest yang dijalankan dengan Visual Studio Code dengan rentang Januari – Desember 2024 dengan kata kunci “JOT JKT48” didapat 6.217 *tweet* kemudian dibersihkan sehingga menjadi 1.992 *tweet*. Untuk 72 thread twitter dari @polisiidol digabungkan menjadi 1 didapat *tweet* 6.083 *tweet* kemudian dibersihkan sehingga menjadi 2.220 *tweet*.

#### B. Preprocessing Data

Preprocessing dilakukan untuk membersihkan dan pengelola teks agar siap digunakan untuk analisis. Pada tahap ini dibuat untuk meningkatkan kualitas pada data.

##### 1) Case Folding

Tahap ini bertujuan untuk mengubah seluruh huruf pada data crawling yang kemudian diambil kolom `full_text` yang terdiri `original_text` dan `case_folding` Seperti pada fungsi nya `case_folding` mengubah seluruh *tweet* menjadi huruf lowercase

TABEL III  
CASE FOLDING JOT JKT48

Original Text	Lowercased Text
harapan untuk jkt48 yaitu membuat para jot sadar bahwa TANPA FANS MEREKA BUKAN APA-APA DAN GAAKAN ADA UANG MENGALIR KALAU GAADA FANS YANG BULOL SAMA OSHINYA. jadi stop menyepelekan fans dan berikan imbalan terbaik	harapan untuk jkt48 yaitu membuat para jot sadar bahwa tanpa fans mereka bukan apa-apa dan gaakan ada uang mengalir kalau gaada fans yang bulol sama oshinya. jadi stop menyepelekan fans dan berikan imbalan terbaik

TABEL IV  
CASE FOLDING SKANDAL MEMBER JKT48

Original Text	Lowercased Text
@polisiidol2nd banyak bacot jing, itu udah keputusan manajemen yang lebih berhak	@polisiidol2nd banyak bacot jing, itu udah keputusan manajemen yang lebih berhak

##### 2) Cleaning Data

Pada tahap ini peneliti menggunakan cleaning data untuk menghapus simbol, tanda baca, tautan URL, #hastag, angka dan karakter khusus yang ada pada data crawling tersebut.

TABEL V  
CLEANING DATA SKANDAL MEMBER JKT48

Lowercased Text	Cleaning Text
@polisiidol2nd banyak bacot jing, itu udah keputusan manajemen yang lebih berhak	banyak bacot jing itu udah keputusan manajemen yang lebih berhak

TABEL VI  
CLEANING DATA JKT48

Lowercased_Text	Cleaning_Text
harapan untuk jkt48 yaitu membuat para jot sadar bahwa tanpa fans mereka bukan apa apa dan gaakan ada uang mengalir kalau gaada fans yang bulol sama oshinya jadi stop menyepelekan fans dan berikan imbalan terbaik.	harapan untuk jkt48 yaitu membuat para jot sadar bahwa tanpa fans mereka bukan apa apa dan gaakan ada uang mengalir kalau gaada fans yang bulol sama oshinya jadi stop menyepelekan fans dan berikan imbalan terbaik

### 3) Normalisasi Data

Tahapan normalisasi dilakukan untuk mengganti kata yang tidak baku, seperti singkatan dan slang diubah menjadi kata baku sesuai KBBI. Proses ini menggunakan pendekatan kamus yang disusun secara manual. Dan disesuaikan dengan konteks Bahasa umum pada media sosial twitter berkaitan dengan topik JKT48.

TABEL VII  
NORMALISASI DATA SKANDAL MEMBER JKT48

Cleaned Text	Normalisasi
banyak bacot jing itu udah keputusan manajemen yang lebih berhak	banyak banyak bicara anjing itu sudah keputusan manajemen yang lebih berhak

TABEL VIII  
NORMALISASI DATA JOT JKT48

Cleaned Text	Normalisasi
harapan untuk jkt48 yaitu membuat para jot sadar bahwa tanpa fans mereka bukan apa apa dan gaakan ada uang mengalir kalau gaada fans yang bulol sama oshinya jadi stop menyepelkan fans dan berikan imbalan terbaik	harapan untuk jkt48 yaitu membuat parah jot sadar bahwa tanpa penggemar mereka bukan apa apa dan gak akan ada uang mengalir kalau tidak ada penggemar yang bodoh sama oshinya jadi berhenti menyepelkan penggemar dan berikan imbalan terbaik

### 4) Tokenization

Tahapan tokenization untuk proses memisahkan kalimat menjadi unit-unit terkecil berupa kata atau karakter yang disebut token. Proses ini menyederhanakan teks menjadi kumpulan kata-kata terpisah sehingga memudahkan perhitungan dalam frekuensi kata dan identifikasi kata kunci yang muncul pada data.

TABEL IX  
TOKENIZATION SKANDAL MEMBER JKT48

Normalisasi	Tokenized Text
banyak banyak bicara anjing itu sudah keputusan manajemen yang lebih berhak	['banyak', 'banyak', 'bicara', 'anjing', 'itu', 'sudah', 'keputusan', 'manajemen', 'yang', 'lebih', 'berhak']

TABEL X  
TOKENIZATION JOT JKT48

Normalisasi	Tokenized Text
harapan untuk jkt48 yaitu membuat parah jot sadar bahwa tanpa penggemar mereka bukan apa apa dan gak akan ada uang mengalir kalau tidak ada penggemar yang bodoh sama oshinya jadi berhenti menyepelkan penggemar dan berikan imbalan terbaik	['harapan', 'untuk', 'jkt48', 'yaitu', 'membuat', 'parah', 'jot', 'sadar', 'bahwa', 'tanpa', 'penggemar', 'mereka', 'bukan', 'apa', 'apa', 'dan', 'gak', 'akan', 'ada', 'uang', 'mengalir', 'kalau', 'tidak', 'ada', 'penggemar', 'yang', 'bodoh', 'sama', 'oshinya', 'jadi', 'berhenti', 'menyepelkan', 'penggemar', 'dan', 'berikan', 'imbalan', 'terbaik']

### 5) Stopwords Removal

Stopword removal dilakukan untuk menghapus kata yang tidak berkontribusi. Namun, agar tidak menghilangkan makna pada kalimat, Peneliti tetap mempertahankan kata – kata negasi seperti “tidak”, “bukan”, “belum”, “jangan”, “tak” dan “tidaklah” karena pada kata tersebut tetap memiliki peran penting pada makna suatu pernyataan. daftar stopwords pada library sastrawi yang mencakup kata umum seperti “ada”, “adalah” dan lainnya.

TABEL XI  
STOPWORDS REMOVAL JOT JKT48

Normalisasi	Tokenized Text
['harapan', 'untuk', 'jkt48', 'yaitu', 'membuat', 'parah', 'parah', 'jot', 'sadar', 'bahwa', 'bahwa', 'tanpa', 'penggemar', 'penggemar', 'mereka', 'mereka', 'bukan', 'apa', 'apa', 'dan', 'gak', 'akan', 'gak', 'akan', 'ada', 'uang', 'ada', 'uang', 'mengalir', 'mengalir', 'kalau', 'tidak', 'ada', 'ada', 'penggemar', 'yang', 'penggemar', 'yang', 'bodoh', 'sama', 'oshinya', 'jadi', 'berhenti', 'jadi', 'berhenti', 'menyepelkan', 'menyepelkan', 'penggemar', 'harapan', 'jkt48', 'parah', 'jot', 'sadar', 'sadar', 'tanpa', 'tanpa', 'penggemar', 'penggemar', 'bukan', 'gak', 'uang', 'gak', 'uang', 'penggemar', 'berhenti', 'berhenti', 'mengalir', 'mengalir', 'bodoh', 'bodoh', 'menyepelkan', 'menyepelkan', 'penggemar', 'tidak', 'penggemar', 'tidak', 'oshinya', 'imbalan', 'oshinya', 'imbalan', 'terbaik'] 'dan', 'berikan', 'imbalan', 'terbaik']	['harapan', 'untuk', 'jkt48', 'yaitu', 'membuat', 'parah', 'jot', 'sadar', 'bahwa', 'tanpa', 'penggemar', 'mereka', 'bukan', 'apa', 'apa', 'dan', 'gak', 'akan', 'ada', 'uang', 'mengalir', 'kalau', 'tidak', 'ada', 'penggemar', 'yang', 'penggemar', 'yang', 'bodoh', 'sama', 'oshinya', 'jadi', 'berhenti', 'jadi', 'berhenti', 'menyepelkan', 'penggemar', 'harapan', 'jkt48', 'parah', 'jot', 'sadar', 'tanpa', 'penggemar', 'bukan', 'gak', 'uang', 'penggemar', 'berhenti', 'mengalir', 'bodoh', 'menyepelkan', 'penggemar', 'tidak', 'oshinya', 'imbalan', 'terbaik'] 'dan', 'berikan', 'imbalan', 'terbaik']

TABEL XII  
STOPWORDS REMOVAL SKANDAL MEMBER JKT48

Tokenized Text	Stopwords Removal
[[ 'banyak', 'banyak', 'bicara', 'anjing', 'itu', 'sudah', 'keputusan', 'manajemen', 'yang', 'lebih', 'berhak' ]]	[ 'bicara', 'anjing', 'manajemen', 'berhak', 'keputusan', 'manajemen', 'keputusan', 'keputusan' ]

6) *Stemming*

Stemming yaitu proses pengubahan kata berimbuhan menjadi bentuk dasar sesuai KBBI tanpa mengubah makna dengan bertujuan menyederhanakan teks sehingga variasi kata menjadi 1 entitas. Berikut tabelnya dari Stemming

TABEL XIII  
STEMMING SKANDAL MEMBER JKT48

Stopwords Removal	Stemming
[ 'bicara', 'anjing', 'keputusan', 'manajemen', 'berhak' ]	[ 'bicara', 'anjing', 'putus', 'manajemen', 'hak' ]

TABEL XIV  
STEMMING JOT JKT48

Stopwords Removal	Stemming
[ 'harapan', 'jkt48', 'parah', 'jot', 'sadar', 'tanpa', 'penggemar', 'bukan', 'gak', 'uang', 'mengalir', 'tidak', 'penggemar', 'bodoh', 'oshinya', 'berhenti', 'menyepelkan', 'penggemar', 'imbalan', 'terbaik' ]	[ 'harap', 'jkt48', 'parah', 'jot', 'sadar', 'tanpa', 'gemar', 'bukan', 'gak', 'uang', 'alir', 'tidak', 'gemar', 'bodoh', 'oshinya', 'henti', 'sepele', 'gemar', 'imbal', 'baik' ]

C. *Pelabelan Data*

Setelah tahap processing data, tahap selanjutnya yaitu melakukan pelabelan data menggunakan kamus inset lexicon yang sudah diberi polarity score.

TABEL XV  
PELABELAN DATA SKANDAL MEMBER JKT48

Stopwords Removal	Polarity Score	Indonesian Sentimen
[ 'bicara', 'anjing', 'putus', 'manajemen', 'hak' ]	-6	Negative
[ 'jot', 'kotor', 'tidak', 'milik', 'member', 'potensi' ]	-9	Negative
[ 'sumber', 'untung', 'kakak', 'sayang' ]	2	Positive

TABEL XVI  
PELABELAN DATA JOT JKT48

Stopwords Removal	Polarity Score	Indonesian Sentimen
[ 'harap', 'jkt48', 'parah', 'jot', 'sadar', 'tanpa', 'gemar', 'bukan', 'gak', 'uang', 'alir', 'tidak', 'gemar', 'bodoh', 'oshinya', 'henti', 'sepele', 'gemar', 'imbal', 'baik' ]	-5	Negative
[ 'urus', 'internal', 'jot', 'bukan', 'urus', 'gemar', 'tidak', 'jot', 'member', 'maaf', 'kakak', 'milik', 'pacar', 'milik', 'putusin', 'kakak', 'riwayat', 'hidup', 'bukan', 'bakat', 'terampil', 'member', 'pacar', 'apa', 'luar' ]	-9	Negative
[ 'tunjuk', 'proses', 'ketat', 'ribu', 'orang', 'daftar', 'terima', 'seleksi', 'berat', 'perempuan', 'suka', 'leceh', 'verbal', 'masuk', 'jkt', 'tidak', 'langsung' ]	-18	Negative

D. *Pembobotan TF-IDF*

Tahap selanjutnya yaitu pembobotan teks menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini memberikan bobot numerik pada tiap kata berdasarkan frekuensi dalam dokumen dan tingkat keunikannya. Pembobotan dilakukan menggunakan library *tfidfVectorizer* dan *Sklearn* pada *python*. Berikut tabel pembobotannya:

TABEL XVII  
TF-IDF SKANDAL MEMBER JKT48

Term	Weight
tidak	0.046583
anjing	0.056272
member	0.025301

TABEL XVIII  
TF-IDF JOT JKT48

Term	Weight
tidak	0.046583
Jkt48	0.056272
member	0.025301

### E. Pembagian Data

Setelah data diberi label menggunakan kamus inset lexicon, tahap selanjutnya yaitu pembagian dataset menjadi dua bagian yaitu data latih dan data uji. Dataset dibagi dengan rasio 80:20 dimana data 80% digunakan data latih untuk membangun model klasifikasi sentiment dan 20% sisa digunakan untuk data uji sebagai evaluasi performa model.

### F. Support Vector Machine

Support Vector Machine digunakan untuk metode klasifikasi pada penelitian ini karena kemampuan tinggi dalam memisahkan kelas yang optimal pada data teks dengan bekerja hyperplane terbaik untuk pemisahan kelas sentimen positif, negatif dan netral. Model diimplementasikan dengan linearSVC dari Pustaka scikit-learn dengan parameter  $C=0.05$  untuk dataset JOT JKT48, kemudian  $C=0.1$  untuk dataset skandal member JKT48 dimana berfungsi mengatur regularisasi agar model tidak overfitting terhadap data yang dilatih, kemudian dilakukan penyeimbang kelas melalui `class_weight` dimana kelas positive diberi bobot 1.3 dibanding negatif dan Netral yaitu 1.0.

### G. Confusion Matrix JOT JKT48

Evaluasi dilakukan dengan *Confusion Matrix* dengan matrix akurasi, presisi, recall dan F1-Score terhadap 816 data uji pada tweet yang bertopik JOT JKT48 dengan baik yakni sebesar 92,03%, presisi, recall dan F1-Score 92,20%, 92.03% dan 91,94%.

TABEL XIX  
CONFUSION MATRIX DATA UJI JOT JKT48

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.95	0.84	0.89	272
Netral	0.91	0.98	0.94	272
Positif	0.91	0.94	0.93	272
Rata-Rata	0.92	0.92	0.92	816

Hasil menunjukkan bahwa kelas netral recall tertinggi, menunjukkan Sebagian besar tweet diklasifikasikan secara akurat. Sementara itu kelas negative memiliki presisi tinggi 95% namun recall lebih rendah 84% yang mengindikasikan tweet negative diklasifikasikan sebagai netral. Untuk kelas positif, model mampu mengenali tweet secara efektif dengan presisi 91% dan recall 94%. Evaluasi juga dilakukan kepada data test dilakukan untuk sejauh mana model dapat mengenali pola sentimen terdapat pada data latih pada tweet yang bertopik JOT JKT48.

Hasil evaluasi pada data latih menunjukkan performa model yang baik, dengan presisi tinggi negative 97% dan recall 91%. Kelas netral recall tertinggi 99% dan F1-Score 97% mengindikasikan model efektif dalam identifikasi kelas

Netral. Untuk kelas positif, presisi dan recall masing-masing sebesar 94% dan 96% menunjukkan akurasi tinggi dalam klasifikasi tweet positif.

### H. Confusion Matrix Skandal Member JKT48

Evaluasi dilakukan terhadap 873 data bertopik skandal member JKT48 yang sudah dilabel sentimen. Model tersebut menunjukkan mampu bekerja dengan baik dengan akurasi 91,98%, presisi 92,64%, recall sebesar 91,98% dan F1-Score 91,87%. Hal tersebut menunjukkan model mampu bekerja dengan baik indentifikasi hal tersebut.

TABEL XX  
CONFUSION MATRIX DATA UJI SKANDAL MEMBER JKT48

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.97	0.80	0.88	291
Netral	0.85	0.98	0.91	291
Positif	0.95	0.97	0.96	291
Rata-Rata	0.93	0.92	0.92	873

Hasil menunjukkan bahwa sentimen positif diklasifikasikan dengan baik dengan presisi 95%, recall 97% dan F1-Score 96%. Kelas netral memiliki netral tertinggi 98%, namun presisi 85%, hal tersebut menandakan bahwa tweet dari kelas lain diklasifikasikan sebagai netral. Sementara itu, kelas negatif memiliki presisi 97%, tetapi recall rendah 80% yang menunjukkan bahwa model masih kurang optimal dalam mengenali seluruh tweet bernuansa negatif.

TABEL XXI  
CONFUSION MATRIX DATA LATIH SKANDAL MEMBER JKT48

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.98	0.90	0.94	1164
Netral	0.93	0.97	0.95	1164
Positif	0.95	0.99	0.97	1164
Rata-Rata	0.9533	0.9533	0.9533	3492

Hasil evaluasi menunjukkan data uji menunjukkan akurasi sebesar 95,27%, presisi 95,41%, recall 95,37% dan F1-Score 95,35% yang diidentifikasi bahwa model SVM mampu mengenali dan mengklasifikasi sentimen dengan baik. Kemudian Sentimen negative memiliki presisi dan recall sebesar 90% menunjukkan performa yang baik meskipun beberapa tweet belum dikenali secara tepat. Sentimen netral







## DAFTAR PUSTAKA

- [1] R. Noviana and I. Rasal, "PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER," *Jurnal Teknik dan Science*, vol. 2, no. 2, pp. 51–60, Jun. 2023.
- [2] Damasta Gefanly Anno and Dewi Damajanti Kusuma, "HUBUNGAN ANTARA FANATISME DENGAN PERILAKU KONSUMTIF PADA FANS JKT48 DI SURABAYA," *Character Jurnal Penelitian Psikologi Faculty of Psychology*, vol. 04, no. 04, pp. 13–18, Sep. 2020.
- [3] S. Zahra Maharani and A. Putri Az-Zahra, "Triwikrama: Jurnal Multidisiplin Ilmu Sosial STRATEGI KOMUNIKASI PEMASARAN PRODAK ERIGO TERHADAP INFLUENCE JKT48," 2024, [Online]. Available: <https://erigostore.co.id/>
- [4] S. I. Natio and S. Paramita, "Peran Public Relations Industri Musik dalam Membangun Reputasi (Analisis Terhadap DX Entertainment JKT48)," *Prologia*, vol. 4, no. 2, pp. 324–331, Oct. 2020.
- [5] Dessy Angelina, U. Hayati, and G. Dwilestari, "Penerapan Metode Support Vector Machine Pada Sentimen Analisis Pengguna Twitter Terhadap Konser K-Pop," *Kopertip : Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 7, no. 1, pp. 14–23, Feb. 2023, doi: 10.32485/kopertip.v7i1.251.
- [6] L. K. Ramasamy, S. Kadry, Y. Nam, and M. N. Meqdad, "Performance analysis of sentiments in Twitter dataset using SVM models," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2275–2284, Jun. 2021, doi: 10.11591/ijece.v11i3.pp2275-2284.
- [7] R. Puja, I. Putra, M. Akbar, and R. Amalia, "Analisis Sentimen Masyarakat Terhadap Kinerja Persatuan Sepakbola Seluruh Indonesia Menggunakan Metode Backpropagation," 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-ita/index>
- [8] S. A. Putra and A. Wijaya, "Analisis Sentimen Artificial Intelligence (AI) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," *Jurnal Sistem & Teknologi Informasi Komunikasi*, vol. 7, no. 1, pp. 21–28, Dec. 2023.
- [9] Ade Dwi Dayani, Yuhandri, and G. Widi Nurcahyo, "Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *Jurnal KomtekInfo*, pp. 1–10, Mar. 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [10] Hendiana, A. Irma Purnamasari, and I. Ali, "ANALISIS SENTIMEN KOMENTAR BERITA DETIK.COM MENGGUNAKAN ALGORITMA SUPORT VEKTOR MACHINE (SVM)," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 3, Jun. 2024, doi: <https://doi.org/10.36040/jati.v8i3.8421>.
- [11] D. Musfiroh, U. Khaira, P. Eko, P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, Apr. 2021.
- [12] M. F. N. Fathoni, E. Y. Puspaningrum, and A. N. Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Jurnal Informatika dan Sains Teknologi*, vol. 1, no. 3, pp. 62–76, Aug. 2024, doi: 10.62951/modem.v1i3.112.
- [13] D. Purnamasari, A. B. Aji, D. A.P Wulandari, F. A. Arie, N. Yanda, and U. Hidayati, *Pengantar Metode Analisis Sentimen*, 1st ed., vol. 1. Jakarta: Gunadarma, 2023.
- [14] V. Ramanathan and T. Meyyappan, "Prediction of Individual's Character in Social Media Using Contextual Semantic Sentiment Analysis," *Mobile Networks and Applications*, vol. 24, no. 6, pp. 1763–1777, Dec. 2019, doi: 10.1007/s11036-019-01388-3.