

# ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TERHADAP KEBIJAKAN KENAIKAN PAJAK HIBURAN MENGGUNAKAN METODE SVM (SUPPORT VECTOR MACHINE)

Waldy Romadhona\*<sup>1)</sup>, Auliya Rahman Isnain\*<sup>2)</sup>

1. Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
2. Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

## Article Info

**Kata Kunci:** Analisis Sentimen; Pajak Hiburan; *Recursive Feature Elimination* (RFE); *Support Vector Machine* (SVM)

**Keywords:** Entertainment Tax; *Recursive Feature Elimination* (RFE); *Sentiment Analysis*; *Support Vector Machine* (SVM)

## Article history:

Received 2 September 2024

Revised 4 Oktober 2024

Accepted 14 November 2024

Available online 4 December 2024

## DOI :

<https://doi.org/10.29100/jipi.v9i4.5603>

\* Corresponding author.

Waldy Romadhona

E-mail address :

[waldy\\_romadhona1@teknokrat.ac.id](mailto:waldy_romadhona1@teknokrat.ac.id)

## ABSTRAK

Pemerintah Indonesia telah memberlakukan kenaikan pajak hiburan sebesar 40-75% atas aktivitas karaoke, diskotek, bar, dan mandi uap atau spa melalui UU No 1 Tahun 2022 tentang Hubungan Keuangan Antara Pemerintah Pusat dan Pemerintahan Daerah (HKPD). Kebijakan ini menuai beragam sentimen dari masyarakat, baik pro maupun kontra. Berdasarkan permasalahan tersebut, penulis melakukan analisis ini untuk mengetahui sentimen masyarakat pada kebijakan kenaikan pajak hiburan dengan menggunakan data yang didapatkan dari media sosial *twitter*. Metode yang dipakai adalah *Support Vector Machine* (SVM). Kemudian untuk mengukur kinerja klasifikasi SVM menggunakan metode RFE (*Recursive Feature Elimination*). Hasil penelitian menunjukkan bahwa pada metode SVM RFE (*Recursive Feature Elimination*) dengan nilai akurasi mencapai 95%, precision 99%, recall 94%, dan F1-Score 97%. Sedangkan hasil klasifikasi SVM tanpa menggunakan metode RFE dengan akurasi mencapai 93%, precision 85%, recall 94%, F1-Score 88%.

## ABSTRACT

The Indonesian government has imposed an entertainment tax increase of 40-75% on discotheques, karaoke, nightclubs, bars, and steam baths or spas through Law No. 1 of 2022 on HKPD. This policy has drawn various sentiments from the public, both pros and cons. Based on this problem, the author conducted this research to determine public sentiment towards the entertainment tax increase policy using data obtained from *twitter* social media. The method used is *Support Vector Machine* (SVM). Then to measure the performance of SVM classification using the RFE (*Recursive Feature Elimination*) method. The results showed that in the SVM RFE (*Recursive Feature Elimination*) method with an accuracy value of 95%, precision 95%, recall 99%, and F1-Score 97%. While the SVM classification results without using the RFE method with accuracy reaching 93%, precision 94%, recall 94%, F1-Score 95%.

## I. PENDAHULUAN

PADA awal Januari tahun 2024, Pemerintah telah memberlakukan tarif kenaikan pajak hiburan atas aktivitas karaoke, bar, klub malam, diskotek, dan mandi uap atau spa. Peningkatan tarif tersebut berkisar antara 40% sampai 75% dari yang semula paling tinggi sebesar 35%. Kebijakan tersebut dikelola dalam UU No 1 Tahun 2022 mengenai HKPD, dan baru mulai berlaku pada 2024. Pajak Hiburan merupakan bentuk pajak yang dikenakan pada penyelenggara kegiatan hiburan, yang harus dibayarkan oleh individu atau badan yang menikmati hiburan [1]. Badan atau individu yang menyelenggarakan kegiatan hiburan bertanggung jawab untuk membayar pajak hiburan tersebut kepada kas daerah.

Dasar hukum pengenaan pajak hiburan sebenarnya telah ada mulai UU No. 28 Tahun 2009 tentang PDRD. Kemudian kebijakan tersebut diperbarui dengan UU No. 1 Tahun 2022 tentang HKPD bahwa tarif Pajak Barang dan Jasa Tertentu (PBJT) atas jasa hiburan pada karaoke, diskotek, bar klub malam dan mandi uap/spa ditentukan paling minim 40 % dan paling maksimal 75 %. Tujuan pemerintah dengan kebijakan kenaikan pajak hiburan ini adalah untuk mengumpulkan lebih banyak uang untuk PAD, yang kemudian akan digunakan untuk mendanai proyek-proyek pembangunan daerah yang bermanfaat bagi masyarakat. Kenaikan tarif Pajak Hiburan ini berdampak kepada tekanan finansial bagi pelaku usaha. Dengan harga layanan yang mungkin dinaikkan untuk

menutupi beban pajak, juga berpotensi terhadap penurunan minat pelanggan. Konsumen yang merasa terbebani oleh harga tinggi mungkin beralih ke alternatif hiburan yang lebih terjangkau. Kebijakan kenaikan pajak hiburan ini tentunya mengundang beragam tanggapan dari masyarakat, pebisnis, dan pihak terkait lainnya.

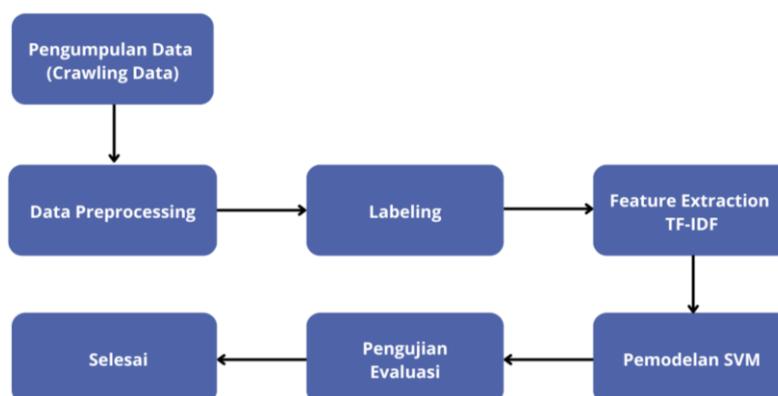
Berdasarkan latar belakang tersebut, diperlukan analisis sentimen untuk mengidentifikasi bagaimana tanggapan masyarakat terhadap kebijakan kenaikan Pajak Hiburan khususnya media sosial *twitter* [2]. Analisis sentimen akan membantu dalam menganalisis komentar *twitter* yang berisi komentar positif negatif dan netral. Metode yang dipakai pada analisa ini ialah SVM. Metode SVM ialah salah satu teknik yang seringkali dipakai dalam mengklasifikasikan data, khususnya data teks dalam mengukur tingkat keakuratan data pada teknik metode yang digunakan [3]. SVM sangat efektif dalam menangani berbagai jenis data teks yang sering kali tidak terstruktur, seperti data teks dari media sosial *twitter*. Data dari *twitter* umumnya memiliki banyak karakter dan frasa unik yang bisa menjadi tantangan bagi banyak algoritma dalam mengklasifikasikan data. SVM dapat memetakan data teks ini ke dalam ruang fitur berdimensi tinggi memungkinkan pemisahan yang lebih baik antara kelas sentimen yang berbeda. SVM memiliki kemampuan generalisasi yang baik sehingga memungkinkan performanya tetap optimal meski data latih terbatas.

Dalam penelitian sebelumnya, metode SVM terbukti mempunyai akurasi klasifikasi yang lebih besar dari metode lain. Pada penelitian yang dilakukan untuk menganalisis sentimen terhadap review aplikasi kredivo dengan menggunakan *confusion matrix* [4]. Hasil dari pengujian pada 10.000 ulasan pemakai aplikasi Kredivo di *Google Play Store*, dengan pemisahan 80:20 antara data pelatihan dan pengujian, mengungkapkan skor *f1* 85%, akurasi 87%, presisi 84%, dan penarikan 85%. Adapun penelitian lainnya [5] yang menggabungkan metode SVM dengan kernel *linear RFE (Recursive Feature Elimination)* memiliki tingkat akurasi klasifikasi yang menjanjikan. Berdasarkan hasil dari penelitian tersebut yang membahas mengenai sentiment terhadap Analisis Sentimen E-Wallet di *Twitter* menyatakan bahwa model klasifikasi SVM RFE menyatakan hasil terbaik dengan nilai akurasi 81%, *precision* 86%, *recall* 76% dan *f1-score* 78%.

Tujuan dari penelitian ini adalah untuk menganalisis sentimen masyarakat terhadap kebijakan kenaikan pajak hiburan menggunakan metode *Support Vector Machine (SVM)*. Penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang bagaimana masyarakat merespons kebijakan tersebut, baik secara positif atau negatif, dengan menganalisis data teks dari media sosial *twitter*. Melalui analisis sentimen ini, penelitian ini diharapkan dapat mengidentifikasi pola-pola sentimen yang dominan dan faktor-faktor yang mempengaruhi opini publik. Selain itu, penelitian ini berupaya untuk memberikan pemahaman yang lebih mendalam bagi pembuat kebijakan tentang persepsi masyarakat terhadap kebijakan kenaikan pajak hiburan. dalam merancang kebijakan dan mengambil keputusan yang tepat dalam menentukan kebijakan dalam pemerintah agar diterima oleh masyarakat.

## II. METODE PENELITIAN

Ada beberapa metodologi yang dipakai pada analisis ini yang secara umum meliputi langkah-langkah seperti pengumpulan data, *preprocessing*, *labeling*, ekstraksi fitur, pemodelan SVM dan evaluasi model. Tahap penelitian ditampilkan dalam bentuk diagram alir pada gambar dibawah ini.



Gambar. 1. Tahapan Penelitian

### A. Crawling Data

*Crawling* adalah proses pengumpulan data dari berbagai sumber di internet. Pada tahapan ini, proses pengambilan data menggunakan teknik *crawling* melalui *platform* media sosial *twitter*. Dataset yang dihimpun memakai kata kunci seperti “pajakhiburan”, “pajakhiburan40%”, “hiburan\_pajak” dan dataset kemudian disimpan dengan format *.csv* (*command separated values*) [6].

### B. Data Preprocessing

*Preprocessing* merupakan langkah penting dalam melakukan pembersihan, merapikan dan mengubah data mentah supaya menjadi data yang lebih terstruktur dan lebih mudah dipahami. *Preprocessing* dilakukan untuk memastikan bahwa dataset siap digunakan untuk memudahkan dalam proses analisis sentiment [4]. Terdapat beberapa tahap dalam *preprocessing* data antara lain:

- 1) *Cleaning Data*, merupakan tahapan dalam melakukan pembersihan data seperti menghapus data duplikat dan menghilangkan karakter yang tidak penting seperti koma, titik, tanda tanya, tanda seru, dan emoji. [7]
- 2) *Case Folding*, ialah proses untuk merubah semua huruf dalam teks jadi huruf kecil dengan maksud membuat karakter data jadi seragam [8]. *Case Folding* memastikan bahwa tidak ada perbedaan dalam representasi teks antara huruf besar dan huruf kecil [9].
- 3) *Tokenizing*, ialah langkah untuk memisahkan kalimat atau data jadi bagian-bagian yang lebih kecil yang dikatakan token atau unit pemisah [10].
- 4) *Removal Stopword*, merupakan proses menghapus kata-kata umum yang sering muncul tapi tidak mempunyai nilai yang penting, seperti kata "yang", "dan", "atau", "dari", "di", "ke", dan lainnya. Tahap ini disebut juga dengan *Filtering*. [11]
- 5) *Stemming*, proses ini dipakai untuk menghilangkan kata yang memiliki imbuhan jadi kata dasarnya.. Tujuannya adalah meningkatkan konsistensi dan efisiensi dalam melakukan analisis. [12]

### C. Labelling Data

Labeling adalah tahap penting dalam analisis sentimen untuk menentukan kategori sentimen [13]. Pada langkah pelabelan ini, data akan dibagi jadi 2 kategori sentimen, yaitu positif dan negatif. Pada tahap labeling untuk mengetahui hasil sentimen, data akan diproses secara otomatis memakai *Library* pada Bahasa pemrograman *Python*.

### D. Feature Extraction

Pada penelitian ini, teknik pembobotan yang dipakai ialah TF-IDF. Teknik ini merupakan proses yang dilakukan untuk menentukan bobot pada setiap kata agar meningkatkan kemampuan proses dalam analisis sentimen [14]. TF-IDF ialah tehnik yang efektif untuk mencari kata-kata dalam setiap dokumen dengan memperhitungkan frekuensi kata tersebut dalam dokumen tersebut dan seberapa umum kata tersebut di seluruh koleksi dokumen [15]. Metode TF-IDF memberi bobot pada sebuah kata sesuai dengan frekuensi istilahnya, yang mengukur frekuensi kemunculannya dalam dokumen tertentu, dan frekuensi dokumen terbalik, yang menunjukkan kelangkaannya atau lebih sering terjadi dalam dokumen [16]. Rumus umum TF-IDF [17] dapat dilihat sebagai berikut :

$$W_{dt} = t_{fdt} \times I_{dft} \quad (1)$$

$W_{dt}$  : bobot dokumen ke-d pada kata ke-t

$t_{fdt}$  : banyak kata yang dicari pada sebuah dokumen

$i_{dft}$  : *inversed document frequency* ( $\log = \frac{n}{df}$ )

$N$  : total dokumen

$df$  : banyak dokumen yang mempunyai kandungan kata yang dicari

### E. Split Data (Pembagian Data)

Tahap berikutnya sesudah *Feature Extraction* ialah pembagian data. Data akan dibagi jadi dua bagian dengan skala 70:30. Data latihan (*training*) dengan perbandingan 70% dan data (*testing*) dengan perbandingan 30%. Dengan membagi data menjadi dua bagian ini, peneliti dapat mengevaluasi seberapa baik model dapat mengukur kinerja model secara objektif. [18]

#### F. Pemodelan Support Vector Machine (SVM)

Pada penelitian ini, model klasifikasi yang digunakan adalah *Support Vector Machine* (SVM). Dalam penelitian ini, metode SVM (*Support Vector Machine*) dipilih karena keunggulannya dalam menangani data berukuran besar dan kompleks serta kemampuannya untuk menghasilkan margin pemisahan yang optimal antara kelas-kelas. Dibandingkan dengan Naive Bayes, SVM lebih efektif dalam menangani data yang tidak terstruktur serta memberikan akurasi yang lebih tinggi. Sementara itu, metode lain seperti Random Forest juga merupakan metode yang kuat dan mampu menangani ketidakseimbangan kelas dengan baik, namun SVM sering kali lebih efisien dalam penggunaan memori dan lebih unggul ketika jumlah fitur yang sangat besar. Oleh karena itu, SVM dipilih karena menghasilkan akurasi tinggi dan kemampuan generalisasi yang kuat lebih diutamakan, meskipun dengan biaya komputasi yang lebih besar. SVM merupakan algoritma yang terbukti efektif dalam menangani berbagai tugas klasifikasi, termasuk analisis sentimen. [19]. SVM melakukan klasifikasi sentimen dengan memisahkan data menggunakan *hyperplane* di ruang fitur berdimensi tinggi. *Hyperplane* adalah garis pemisah yang memaksimalkan margin antara dua kelas sentimen yang berbeda. Data yang paling dekat dengan *hyperplane* disebut *support vectors* untuk menentukan posisi *hyperplane* yang optimal. Tujuan dari SVM adalah untuk menemukan *hyperplane* yang memaksimalkan margin, yaitu jarak terdekat antara data dari kedua kelas. Dengan memaksimalkan margin, SVM dapat menghasilkan klasifikasi yang lebih akurat. Selain itu, pendekatan *Recursive Feature Elimination* (RFE) juga digunakan dalam penelitian ini untuk membantu meningkatkan kinerja model SVM dengan memilih fitur-fitur yang paling relevan untuk klasifikasi. RFE bekerja dengan cara menghapus fitur yang paling tidak penting dan membangun model pada fitur-fitur yang tersisa. Proses ini dilakukan berulang kali hingga mencapai sejumlah fitur yang optimal [20]. Dengan demikian, RFE membantu mengurangi dimensionalitas data, menghilangkan fitur-fitur yang tidak relevan atau redundan, dan membantu meningkatkan akurasi serta efisiensi pada model SVM.

#### G. Evaluasi Model

Tujuan mengevaluasi model adalah untuk mengetahui seberapa baik model dapat memprediksi nada sepotong teks. peneliti di sini membandingkan beberapa metode memakai SVM dan kemudian memanfaatkan RFE untuk membuat SVM lebih baik. Kemudian *Confusion Matrix* dipakai untuk menjumlahkan nilai *accuracy*, *recall*, *precision* dan *f1-score*.

TABEL I  
 CONFUSION MATRIX

Actual	Prediction	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Keterangan :

*True Positive* (TP) : Jumlah data dimana model memprediksi kelas positif dengan benar.

*False Positive* (FP) : Jumlah data dimana model salah memprediksi kelas positif.

*True Negative* (TN) : Jumlah data dimana model yang secara benar diprediksi sebagai negatif.

*False Negative* (FN): Jumlah data dimana model salah memprediksi kelas Negatif.

Rumus untuk menjumlahkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* dapat dilihat yaitu :

- 1) *Accuracy* mengukur seberapa sering model klasifikasi benar-benar memprediksi kelas yang benar. Rumus perhitungan akurasi dapat dilihat pada Persamaan 1.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2)$$

- 2) *Precision* mengukur seberapa banyak prediksi positif yang dihasilkan oleh model yang benar-benar benar. Rumus perhitungan akurasi dapat dilihat pada Persamaan 2.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

3) *Recall* mengukur seberapa banyak kelas positif yang benar diprediksi oleh model. Rumus perhitungan akurasi dapat dilihat pada Persamaan 3.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

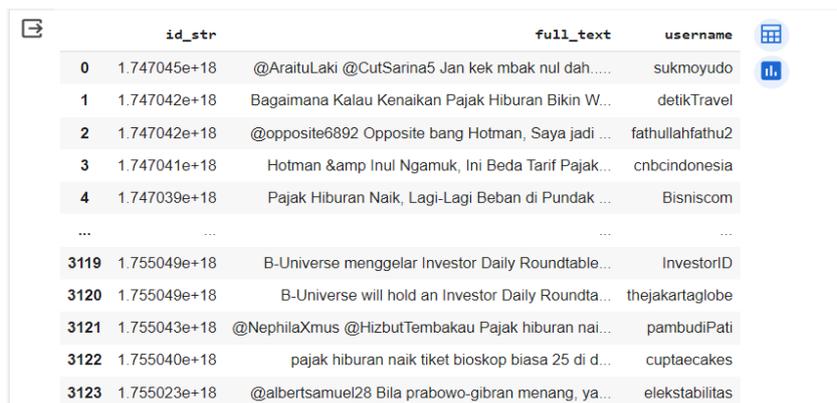
4) *F1 Score* adalah perbandingan nilai dari *precision* dan *recall* yang menunjukkan keseimbangan antara kedua metrik tersebut. *F1 Score* mencapai nilai maksimum ketika *precision* dan *recall* sama. Rumus perhitungan akurasi dapat dilihat pada Persamaan 4.

$$F1\ Score = 2x \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

### III. HASIL DAN PEMBAHASAN

#### A. Pengumpulan Data

Pada analisis ini, proses pengambilan data menggunakan teknik *Crawling* melalui platform media sosial *twitter* dengan memakai bahasa pemrograman *Python* memakai *library* *Harvest*. Dari hasil *crawling twitter*, dataset yang berhasil diperoleh sebanyak 3123 *tweet* berbahasa indonesia yang diambil dalam rentang waktu pengambilan data dimulai dari tanggal 22 Januari sampai dengan 20 Februari 2024 dengan menggunakan kata kunci “pajakhiburan”, “pajakhiburan40%”. Hasil *crawling* data bisa ditinjau pada Gambar 2.



	id_str	full_text	username
0	1.747045e+18	@AraituLaki @CutSarina5 Jan kek mbak nul dah....	sukmoyudo
1	1.747042e+18	Bagaimana Kalau Kenaikan Pajak Hiburan Bikin W...	detikTravel
2	1.747042e+18	@opposite6892 Opposite bang Hotman, Saya jadi ...	fathullahfathu2
3	1.747041e+18	Hotman & Inul Ngamuk, Ini Beda Tarif Pajak...	cnbcindonesia
4	1.747039e+18	Pajak Hiburan Naik, Lagi-Lagi Beban di Pundak ...	Bisniscom
...	...	...	...
3119	1.755049e+18	B-Universe menggelar Investor Daily Roundtable...	InvestorID
3120	1.755049e+18	B-Universe will hold an Investor Daily Roundta...	thejakartaglobe
3121	1.755043e+18	@NephilaXmus @HizbutTembakau Pajak hiburan nai...	pambudiPati
3122	1.755040e+18	pajak hiburan naik tiket bioskop biasa 25 di d...	cuptaecakes
3123	1.755023e+18	@albertsamuel28 Bila prabowo-gibran menang, ya...	elekstabilitas

Gambar. 2. Hasil Crawling Data Twitter

#### B. Pre-processing Data

Pada langkah ini, sebelum data diklasifikasikan, data harus melalui tahap persiapan atau pre-processing. Tahap ini menampilkan hasil dari tahap *pre-processing*, seperti *cleaning*, *case folding*, *tokenize*, *removal stopword* dan *stemming*.

##### 1) Cleaning Data

Pada tahap ini, *Cleaning Data* berfungsi untuk menghilangkan tanda baca, karakter, *hashtag*, *emoticon*, *url* dan menghapus data duplikat.

TABEL II  
CLEANING DATA

Sebelum Cleaning Data	Hasil Cleaning Data
Pajak hiburan naik -&gt; margin tipis -&gt; harga dinaikin. Pemerintah kita tercinta ini memang gak suka liat masyarakatnya bahagia.	Pajak hiburan naik gt margin tipis gt harga dinaikin Pemerintah kita tercinta ini memang gak suka liat masyarakatnya bahagia
@akhmadxfauzi @abigailimuriaa Kalau begitu langkah pemerintah sudah tepat memberikan pajak tinggi utk hiburan malam? Ya syukurlah kalau begitu. Biasa yg teriak ga setuju langkah ini pengusaha kelab malam dll (yg akan dibebankan ke pengunjungnyaâ€) mereka akan teriak pastiâ€	Kalau begitu langkah pemerintah sudah tepat memberikan pajak tinggi utk hiburan malam Ya syukurlah kalau begitu Biasa yg teriak ga setuju langkah ini pengusaha kelab malam dll yg akan dibebankan ke pengunjungnya mereka akan teriak pasti

## 2) Case Folding

Tahap selanjutnya adalah *Case Folding* untuk merubah huruf besar (*uppercase*) menjadi huruf kecil (*lowercase*). Hasil tahap *Case Folding* bisa dilihat pada Tabel III.

TABEL III  
CASE FOLDING

Cleaning Data	Hasil Case Folding
Pajak hiburan naik gt margin menipis gt harga dinaikin Pemerintah kita tercinta ini memang gak suka liat masyarakatnya bahagia	pajak hiburan naik gt margin menipis gt harga dinaikin pemerintah kita tercinta ini memang gak suka liat masyarakatnya bahagia
Kalau begitu langkah pemerintah sudah tepat memberikan pajak tinggi utk hiburan malam Ya syukurlah kalau begitu Biasa yg teriak ga setuju langkah ini pengusaha kelab malam dll yg akan dibebankan ke pengunjungnya mereka akan teriak pasti	kalau begitu langkah pemerintah sudah tepat memberikan pajak tinggi utk hiburan malam ya syukurlah kalau begitu biasa yg teriak ga setuju langkah ini pengusaha kelab malam dll yg akan dibebankan ke pengunjungnya mereka akan teriak pasti

## 3) Tokenizing

Tahap tokenize data berfungsi untuk memecahkan kata-kata pada kalimat, paragraf atau halaman menjadi token, kata individual atau potongan kata tunggal. Hasil *Tokenizing* bisa dilihat pada Tabel IV.

TABEL IV  
TOKENIZING

Case Folding	Hasil Tokenizing
pajak hiburan naik gt margin menipis gt harga dinaikin pemerintah kita tercinta ini memang gak suka liat masyarakatnya bahagia	['pajak', 'hiburan', 'naik', 'gt', 'margin', 'menipis', 'gt', 'harga', 'dinaikin', 'pemerintah', 'kita', 'tercinta', 'ini', 'memang', 'gak', 'suka', 'liat', 'masyarakatnya', 'bahagia']
kalau begitu langkah pemerintah sudah tepat memberikan pajak tinggi utk hiburan malam ya syukurlah kalau begitu biasa yg teriak ga setuju langkah ini pengusaha kelab malam dll yg akan dibebankan ke pengunjungnya mereka akan teriak pasti	['kalau', 'begitu', 'langkah', 'pemerintah', 'sudah', 'tepat', 'memberikan', 'pajak', 'tinggi', 'utk', 'hiburan', 'malam', 'ya', 'syukurlah', 'kalau', 'begitu', 'biasa', 'yg', 'teriak', 'ga', 'setuju', 'langkah', 'ini', 'pengusaha', 'kelab', 'malam', 'dll', 'yg', 'akan', 'dibebankan', 'ke', 'pengunjungnya', 'mereka', 'akan', 'teriak', 'pasti']

## 4) Removal Stopword

Pada langkah ini, kata yang dianggap tidak penting akan dihilangkan, tujuannya adalah untuk memudahkan proses klasifikasi data. Hasil *Removal Stopword* bisa dilihat pada Tabel V.

TABEL V  
REMOVAL STOPWORD

Tokenizing	Hasil Removal Stopword
['pajak', 'hiburan', 'naik', 'gt', 'margin', 'menipis', 'gt', 'harga', 'dinaikin', 'pemerintah', 'kita', 'tercinta', 'ini', 'memang', 'gak', 'suka', 'liat', 'masyarakatnya', 'bahagia']	['pajak', 'hiburan', 'naik', 'gt', 'margin', 'menipis', 'gt', 'harga', 'dinaikin', 'pemerintah', 'tercinta', 'memang', 'gak', 'suka', 'liat', 'masyarakatnya', 'bahagia']
['kalau', 'begitu', 'langkah', 'pemerintah', 'sudah', 'tepat', 'memberikan', 'pajak', 'tinggi', 'utk', 'hiburan', 'malam', 'ya', 'syukurlah', 'kalau', 'begitu', 'biasa', 'yg', 'teriak', 'ga', 'setuju', 'langkah', 'ini', 'pengusaha', 'kelab', 'malam', 'dll', 'yg', 'akan', 'dibebankan', 'ke', 'pengunjungnya', 'mereka', 'akan', 'teriak', 'pasti']	['langkah', 'pemerintah', 'tepat', 'memberikan', 'pajak', 'tinggi', 'hiburan', 'malam', 'syukurlah', 'biasa', 'teriak', 'ga', 'setuju', 'langkah', 'pengusaha', 'kelab', 'malam', 'dibebankan', 'pengunjungnya', 'teriak']

## 5) Stemming

Tahap terakhir ialah *stemming*, yaitu menghilangkan kata imbuhan menjadi kata dasarnya. Hasil dari *stemming* bisa dilihat pada Tabel VI.

TABEL VI  
 REMOVAL STEMMING

Removal Stopword	Hasil Stemming
['pajak', 'hiburan', 'naik', 'gt', 'margin', 'menipis', 'gt', 'harga', 'dinaikin', 'pemerintah', 'tercinta', 'memang', 'gak', 'suka', 'liat', 'masyarakatnya', 'bahagia']	pajak hibur naik gt margin tip gt harga dinaikin perintah cinta memang gak suka liat masyarakat bahagia
['langkah', 'pemerintah', 'tepat', 'memberikan', 'pajak', 'tinggi', 'hiburan', 'malam', 'syukurlah', 'biasa', 'teriak', 'ga', 'setuju', 'langkah', 'pengusaha', 'kelab', 'malam', 'dibebankan', 'pengunjungnya', 'teriak']	langkah perintah tepat beri pajak tinggi hibur malam syukur biasa teriak ga tuju langkah usaha kelab malam beban unjung teriak

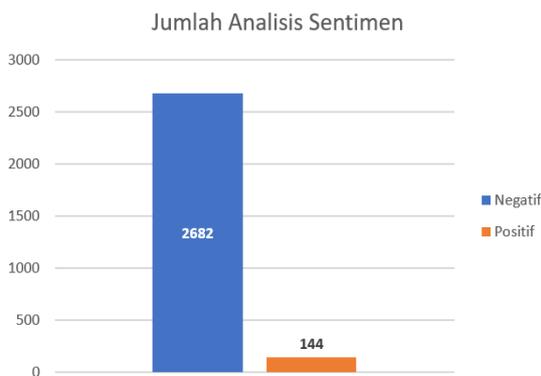
### C. Labelling Data

Setelah melalui tahap *pre-processing*, data yang telah dikumpulkan sebanyak 2827 dan akan melalui proses pelabelan dengan cara membagi data menjadi 2 kategori sentimen, yaitu positif dan negatif. Hasil pelabelan data bisa ditinjau pada Tabel VII.

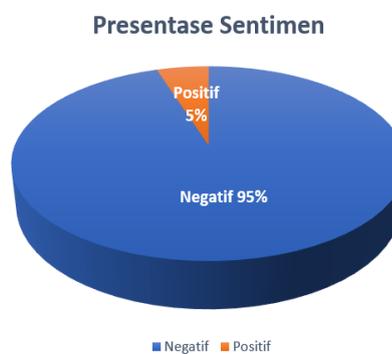
TABEL VII  
 HASIL LABELLING DATA

Teks	Kategori Label
pajak hibur naik gt margin tip gt harga dinaikin perintah cinta memang gak suka liat masyarakat bahagia	Negatif
langkah perintah tepat beri pajak tinggi hibur malam syukur biasa teriak ga tuju langkah usaha kelab malam beban unjung teriak	Positif

Selanjutnya, setelah melakukan proses labeling, data dibagi menjadi dua kategori label yaitu positif dan negative. Sentimen terdiri dari 2682 bernilai negatif dan 144 kumpulan label bernilai positif. Hasil jumlah sentimen pada komentar dan presentase nilai sentimen bisa ditinjau pada Gambar 3 dan Gambar 4.



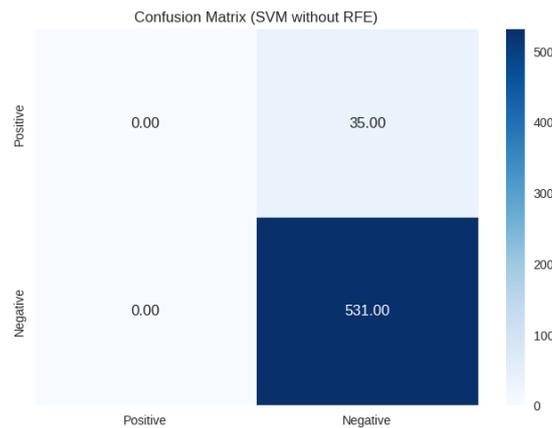
Gambar. 3. Jumlah Sentimen



Gambar. 4. Persentase Nilai Sentimen

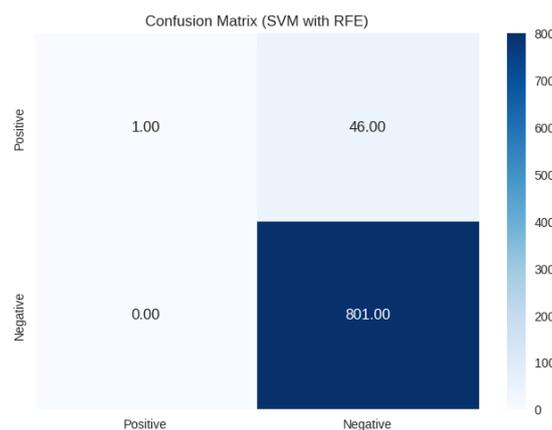
### D. Implementasi Support Vector Machine (SVM)

Pada tahap pemodelan ini, data akan dibagi menjadi dua bagian sesuai dengan rasio 70:30 data pelatihan (70 persen) dan data (pengujian) (30 persen). Setelah data terbagi menjadi dua kelompok, langkah selanjutnya adalah mengimplementasikan algoritma klasifikasi pada dataset. Pada tahap ini, peneliti menggunakan algoritma *Support Vector Machine* (SVM) dengan teknik kernel *linear*. Selanjutnya, model algoritma SVM yang digunakan dibagi menjadi 2 skenario, skenario pertama adalah menggunakan metode SVM biasa dan skenario kedua menggunakan model SVM dengan seleksi fitur RFE (*Recursive Feature Elimination*). Tahap selanjutnya adalah mengevaluasi kinerja algoritma klasifikasi dengan menggunakan *Confusion Matrix*. Untuk mengetahui hasil *Confusion Matrix* dari model SVM dan seleksi fitur RFE bisa ditinjau pada Gambar 5 dan Gambar 6.



Gambar. 5. Confusion Matrix SVM

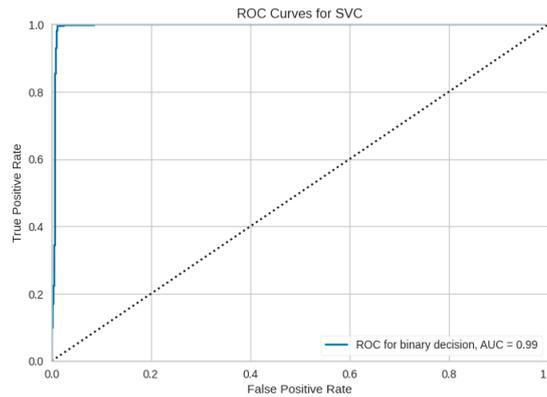
Pada Gambar 4 *confusion matrix* diatas memberi deskripsi yang jelas mengenai kinerja model klasifikasi dalam memprediksi sentimen. Jumlah *True Negative* (TN) yang tinggi menunjukkan bahwa model mampu mengidentifikasi dengan baik sentimen negatif yang sebenarnya negatif. Selain itu, absennya *False Positive* (FP) menunjukkan bahwa model tidak membuat banyak kesalahan dengan mengklasifikasikan sentimen negatif sebagai positif. Namun, adanya *False Negative* (FN) sebesar 35 menunjukkan bahwa model memiliki kesulitan dalam mengenali sentimen positif yang sebenarnya positif. Meskipun tidak ada *True Positive* (TP) yang tercatat dalam confusion matrix ini.



Gambar. 6. Confusion Matrix SVM dengan fitur RFE

Pada Gambar 6 *confusion matrix* menyatakan bahwa model klasifikasi mempunyai kinerja yang cukup baik dalam mengenali sentimen negatif. Dengan 801 *True Negative* (TN), yang merupakan data yang sebenarnya negatif dan juga diprediksi sebagai negatif, model menyatakan keterampilan yang baik untuk mengidentifikasi sentimen negatif yang sebenarnya negatif. Selain itu, nilai 0 *False Positive* (FP) menunjukkan bahwa tidak ada sampel sentimen negatif yang salah diprediksi sebagai positif. Namun, ada ruang untuk peningkatan dalam mengenali sentimen positif, karena terdapat 46 *False Negative* (FN). Meskipun hanya ada 1 *True Positive* (TP) yang tercatat.

Dari kedua *Confusion Matrix* tersebut dapat disimpulkan bahwa SVM dengan fitur RFE memiliki kinerja yang lebih baik daripada SVM tanpa fitur. SVM dengan fitur RFE membuat kesalahan yang lebih sedikit dalam menentukan prediksi sentimen. Hal ini menunjukkan bahwa seleksi fitur RFE dapat membantu meningkatkan kinerja model klasifikasi. Selain itu digunakan pula kurva ROC untuk mengevaluasi kinerja model klasifikasi pada berbagai *threshold*. Bisa ditinjau pada Gambar 6 dibawah ini yang menunjukkan kurva ROC dari model klasifikasi SVM.



Gambar 7. Kurva ROC

Kurva ROC merupakan alat yang dipakai untuk mengevaluasi kinerja model klasifikasi. Kurva ROC dengan model klasifikasi SVC dengan fitur RFE memiliki AUC yang lebih tinggi. AUC, yang merupakan singkatan dari area di bawah kurva, menggambarkan luas daerah di bawah kurva. Kurva ROC membandingkan grafik antara laju *true positive* (TPR), yang ditampilkan pada sumbu vertikal atau sumbu koordinat, dan laju *false positive* (FPR), yang direpresentasikan dalam bentuk kurva. [21]. Kurva ROC pada Gambar 6 menunjukkan bahwa model SVM RFE memiliki kinerja yang baik untuk klasifikasi biner. AUC (*Area Under the Curve*) adalah 0.99, yang sangat dekat dengan 1.00, yang mengindikasikan kinerja model yang sempurna.

#### E. Evaluasi Model

Tahap terakhir dari analisis ini adalah melakukan pengujian evaluasi yaitu untuk mengetahui akurasi, *precision*, *recall* dan *f1-score* pada model SVM dan seleksi fitur RFE (*Recursive Feature Elimination*). Dari hasil pengujian, SVM dengan seleksi fitur RFE mendapatkan hasil akurasi yang tinggi dalam evaluasi model ini. Hasil evaluasi pada fitur SVM RFE mendapatkan akurasi sebesar 0.95. Sedangkan hasil evaluasi pada SVM biasa mendapatkan hasil akurasi sebesar 0.93.

TABEL VIII  
 HASIL PENGUJIAN EVALUASI

Test Accuracy	Accuracy
SVM	0.9381625441696113
SVM + RFE	0.9572222222222221

Setelah melakukan evaluasi model, bisa ditinjau bahwa model SVM dengan RFE mempunyai akurasi yang lebih tinggi daripada model SVM biasa. Hal ini menunjukkan bahwa RFE dapat membantu meningkatkan akurasi pada model SVM dan meningkatkan performa model. Hasil klasifikasi model SVM dan RFE dapat dilihat pada Tabel IX dan Tabel X.

TABEL IX  
 HASIL KLASIFIKASI MODEL SVM BIASA

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>support</i>
<i>Positive</i>	0.80	0.09	0.15	35
<i>Negative</i>	0.95	1.00	0.97	531
<i>Accuracy</i>			0.93	566
<i>Macro avg</i>	0.47	0.50	0.48	566
<i>Weighted avg</i>	0.88	0.94	0.91	566

TABEL X  
 HASIL KLASIFIKASI MODEL SVM RFE

	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>support</i>
<i>Positive</i>	1.00	0.02	0.04	47
<i>Negative</i>	0.95	1.00	0.97	801
<i>Accuracy</i>			0.95	848
<i>Macro avg</i>	0.64	0.51	0.51	848
<i>Weighted avg</i>	0.95	0.95	0.92	848

Selanjutnya adalah menghitung nilai *presisi*, *recall* dan *f1-score* pada model SVM dan RFE berdasarkan hasil dari *Confusion Matrix* Saat melakukan pengujian tanpa menggunakan metode SVM RFE, tingkat akurasi yang diperoleh adalah sebesar 93%, dengan *precision* mencapai 85%, *recall* 94% dan *f1-score* 88%. Namun, ketika menerapkan SVM dengan fitur RFE, berhasil mencapai akurasi sebesar 95%, dengan presisi mencapai 99%, *recall* 94% dan *f1-score* 97%. Hal ini menunjukkan bahwa pemilihan fitur yang lebih cermat dapat memberikan

kontribusi besar terhadap peningkatan kinerja model. Dengan demikian, kami merekomendasikan penggunaan fitur RFE (*Recursive Feature Elimination*) sebagai cara yang efektif dalam meningkatkan akurasi dan presisi pada evaluasi model SVM.

Pada penelitian serupa yang dilakukan oleh [5] yang juga menggunakan metode SVM dan RFE untuk menganalisis sentiment pada Aplikasi E-Wallet dengan menggunakan dataset sebanyak 5.309 yang didapatkan dari twitter. Berdasarkan hasil penelitian tersebut dengan menggunakan split data 80:20, SVM menunjukkan hasil akurasi sebesar 74% dalam mengklasifikasikan sentimen positif dan negatif. Sementara itu, hasil yang menggabungkan SVM dengan seleksi fitur RFE berhasil meningkatkan akurasi model SVM dengan nilai akurasi sebesar 81%. Perbandingan ini menunjukkan bahwa meskipun terdapat berbagai metode dalam analisis sentimen, SVM tetap relevan dan efektif untuk melakukan analisis sentiment pada data teks yang besar, terutama dalam meningkatkan akurasi pada model. Dengan demikian, penelitian ini tidak hanya memperkuat validitas penggunaan SVM dalam analisis sentimen, tetapi juga memberikan bukti tambahan tentang kemampuannya untuk digunakan pada berbagai jenis data teks.

#### F. Visualisasi Wordcloud

Visualisasi *Wordcloud* merupakan representasi visual dari kata-kata yang paling sering digunakan dalam teks atau kumpulan teks tertentu. Kata-kata seperti "Pajak", "Hiburan", dan "Naik" mendominasi visualisasi yang menunjukkan bahwa fokus utama pada penelitian ini. Ukuran font yang lebih besar pada kata-kata ini menunjukkan frekuensinya yang lebih tinggi dalam teks. Hasil visualisasi *Wordcloud* bisa ditinjau pada Gambar 7.



Gambar 7. Visualisasi Wordcloud

#### IV. KESIMPULAN

Berdasarkan hasil analisis sentimen dan pengujian yang telah dilakukan, dapat disimpulkan bahwa hasil analisis sentimen masyarakat terhadap kebijakan kenaikan pajak hiburan pada umumnya negatif. Hal ini ditunjukkan dengan persentase sentimen negatif yang mencapai 95% dengan total 2.682 sentimen. Sedangkan sentimen positif hanya mencapai 5% dengan total 144 sentimen. Ini menunjukkan bahwa masyarakat mayoritas tidak setuju dengan kebijakan tersebut. Selain itu, hasil pengujian model klasifikasi SVM menggunakan fitur RFE berhasil meningkatkan akurasi model SVM. Hal ini dibuktikan dengan perbandingan akurasi model SVM tanpa seleksi fitur RFE yang hanya mencapai 93% dengan *precision* mencapai 85%, *recall* 95% dan *f1-score* 97%, sedangkan ketika menggunakan model SVM RFE akurasi meningkat mencapai 95% dengan presisi mencapai 95%, *recall* 99% dan *f1-score* 97%. Dengan demikian, metode SVM ketika menggunakan seleksi fitur RFE ini sangat efektif digunakan untuk melakukan analisis sentimen.

#### DAFTAR PUSTAKA

- [1] C. Of, E. Tax, R. To, B. City, and R. Original, "TERHADAP PENDAPATAN ASLI DAERAH KOTA BANDUNG EFFECTIVENESS AND CONTRIBUTION OF ENTERTAINMENT TAX," vol. 4, no. 1, pp. 26–31, 2024.
- [2] A. Handayani and I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.
- [3] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [4] A. Agustin, S. Andrean, S. Susanti, R. Rahmiati, and H. Hamdani, "Review Aplikasi Kredivo Menggunakan Analisis Sentimen Dengan Algoritma Support Vector Machine," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 9, no. 1, pp. 39–49, 2023, doi: 10.36341/rabit.v9i1.4107.
- [5] E. F. Saraswita, D. P. Rini, and A. Abdiansah, "Analisis Sentimen E-Wallet di Twitter Menggunakan Support Vector Machine dan Recursive Feature Elimination," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1195, 2021, doi: 10.30865/mib.v5i4.3118.
- [6] E. R. Kaburuan and N. R. Setiawan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 1, pp. 105–116, 2023, doi: 10.32736/sisfokom.v12i1.1614.
- [7] A. R. Fitriansyah and Y. Sibaroni, "Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding," *e-Proceeding Eng.*, vol. 10, no. 2, p. 1713, 2023.

- [8] B. Laurensz and E. Sedyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19 (Analysis of Public Sentiment on Vaccination in Efforts to Overcome the Covid-19 Pandemic)," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021.
- [9] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecom.v5i1.3708.
- [10] P. Fremmuzar and A. Baita, "Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter SVM Kernel Test in Sentiment Analysis of Telkomsel Services on Twitter Social Media," vol. 12, no. 148, 2023, doi: 10.34010/komputika.v12i2.9460.
- [11] B. Dan, C. Untuk, K. Sms, and R. Sari, "Komparasi Algoritma Support Vector Machine , Naïve," vol. 2, no. 2, pp. 7–13, 2017.
- [12] P. G. Aryanti and I. Santoso, "Analisis Sentimen Pada Twitter Terhadap Mobil Listrik Menggunakan Algoritma Naive Bayes," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 7, no. 2, pp. 133–137, 2023, [Online]. Available: <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2821>
- [13] S. Ariqoh, M. A. Sunandar, and Y. Muhyidin, "Analisis Sentimen Pada Produk Cushion Di Website Female Daily Menggunakan Metode Support Vector Machine (Svm)," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 137–142, 2023, doi: 10.55123/storage.v2i3.2345.
- [14] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
- [15] D. D. Nada, S. Soehardjoepri, and R. M. Atok, "c," *J. Sains dan Seni ITS*, vol. 11, no. 6, 2023, doi: 10.12962/j23373520.v11i6.96330.
- [16] D. Krisnandi, R. N. Ambarwati, A. Y. Asih, and ..., "Analisis Komentar Cyberbullying Terhadap Kata Yang Mengandung Toksisitas Dan Agresi Menggunakan Bag of Words dan TF-IDF Dengan Klasifikasi SVM," *J. Linguist. ...*, vol. 6, no. 2, pp. 36–41, 2023, [Online]. Available: <https://www.inacl.id/journal/index.php/jlk/article/view/85%0Ahttps://www.inacl.id/journal/index.php/jlk/article/download/85/64>
- [17] A. Dewan, D. Wibiyanto, and A. Wibowo, "Penerapan Algoritma Multiclass Support Vector Machine dan TF-IDF Untuk Klasifikasi Topik Tugas Akhir," vol. 6, pp. 42–51, 2023.
- [18] I. M. Karo Karo, J. A. Karo Karo, Y. Yunianto, H. Hariyanto, M. Falah, and M. Ginting, "Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1423–1430, 2023, doi: 10.47065/josh.v4i4.3943.
- [19] P. A. N. Aryanti and I. B. M. Mahendra, "Analisis Sentimen Opini Berbahasa Indonesia Pada Sosial Media Menggunakan TF-IDF dan Support Vector Machine," *J. Elektron. Ilmu Komput. Udayana*, vol. 12, no. 1, pp. 2654–5101, 2023.
- [20] V. No, O. Hal, A. Devia, and B. Soewito, "Analisis Perbandingan Metode Seleksi Fitur untuk Mendeteksi Anomali pada Dataset CIC-IDS-2018," vol. 5, no. 4, pp. 572–578, 2023.
- [21] T. B. Sasongko, "Komparasi dan Analisis Kinerja Model Algoritma SVM dan PSO-SVM (Studi Kasus Klasifikasi Jalur Minat SMA)," *J. Tek. Inform. dan Sist. Inf.*, vol. 2, no. 2, pp. 244–253, 2016, doi: 10.28932/jutisi.v2i2.476.