



Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pada Komentar Bitcoin Di Aplikasi X

Yaskur Bearly Fernandes, Elin Haerani*, Fadhilah Syafria, Muhammad Fikry, Lola Oktavia

Fakultas Sains dan Teknologi, Program Studi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹12050111676@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin-suska.ac.id, ³fadhilah.syafria@uin-suska.ac.id

⁴muhammad.fikry@uin-suska.ac.id, ⁵lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Media sosial menjadi sarana utama bagi pengguna dalam menyampaikan opini, termasuk terhadap Bitcoin yang memiliki nilai fluktuatif dan memicu beragam tanggapan. Banyaknya komentar yang tidak terstruktur menyebabkan analisis sentimen secara manual menjadi tidak efisien, sehingga diperlukan pendekatan otomatis berbasis machine learning. Penelitian ini bertujuan mengklasifikasikan sentimen positif dan negatif pada komentar Bitcoin di platform X menggunakan algoritma Support Vector Machine (SVM) dengan pembobotan fitur Term Frequency–Inverse Document Frequency (TF-IDF). Dataset yang digunakan terdiri dari 1.750 komentar berbahasa Indonesia yang dilabeli oleh tiga anotator. Data diproses melalui tahapan case folding, pembersihan teks, tokenisasi, stopword removal, dan stemming. Pengujian dilakukan menggunakan empat rasio pembagian data, yaitu 90:10, 80:20, 70:30, dan 60:40. Hasil penelitian menunjukkan bahwa rasio 90:10 menghasilkan performa terbaik dengan akurasi 72,57%, precision 0,75, recall 0,73, dan F1-score 0,67. Model SVM menunjukkan kinerja yang baik dalam mengidentifikasi sentimen positif, namun kurang optimal dalam mendeteksi sentimen negatif akibat ketidakseimbangan distribusi data. Sebagai percobaan tambahan, dilakukan pengujian menggunakan dataset seimbang hasil proses undersampling serta beberapa jenis kernel SVM sebagai pembanding. Hasil percobaan tambahan menunjukkan bahwa penggunaan dataset seimbang menghasilkan klasifikasi yang lebih merata antar kelas, sementara kernel linear memberikan kinerja yang paling stabil dibandingkan kernel lainnya. Secara keseluruhan, SVM dengan pembobotan TF-IDF terbukti efektif untuk analisis sentimen komentar Bitcoin di media sosial.

Kata Kunci: Analisis Sentimen; Term Frequency-Inverse Document Frequency; Bitcoin; Support Vector Machine; Platform X

Abstract—Social media has become a primary medium for users to express opinions, including those related to Bitcoin, whose fluctuating value often triggers diverse public responses. The large volume of unstructured comments makes manual sentiment analysis inefficient, thereby necessitating an automated approach based on machine learning. This study aims to classify positive and negative sentiments in Bitcoin-related comments on the X platform using the Support Vector Machine (SVM) algorithm with Term Frequency–Inverse Document Frequency (TF-IDF) feature weighting. The dataset consists of 1,750 Indonesian-language comments labeled by three annotators. The data were processed through several preprocessing stages, including case folding, text cleaning, tokenization, stopword removal, and stemming. Model evaluation was conducted using four data split ratios, namely 90:10, 80:20, 70:30, and 60:40. The experimental results indicate that the 90:10 ratio achieved the best performance, with an accuracy of 72.57%, precision of 0.75, recall of 0.73, and an F1-score of 0.67. The SVM model demonstrates strong performance in identifying positive sentiments; however, it is less effective in detecting negative sentiments due to class imbalance in the dataset. As an additional experiment, testing was performed using a balanced dataset obtained through an undersampling process and several SVM kernel types for comparison. The results show that using a balanced dataset leads to more evenly distributed classification performance across sentiment classes, while the linear kernel provides the most stable performance compared to other kernels. Overall, SVM with TF-IDF weighting proves to be an effective approach for sentiment analysis of Bitcoin-related comments on social media.

Keywords: Sentiment Analysis; Term Frequency-Inverse Document Frequency; Bitcoin; Support Vector Machine; Platform X

1. PENDAHULUAN

Perkembangan di bidang teknologi informasi telah mengubah masyarakat untuk mendapatkan, menyebarkan, serta memberikan tanggapan melalui informasi di berbagai bidang kehidupan[1]. Transformasi digital ini didukung oleh kemajuan infrastruktur jaringan dan meningkatnya penggunaan perangkat pintar, khususnya perangkat seluler. Kemudahan akses internet melalui perangkat tersebut mendorong terciptanya berbagai platform media sosial yang berfungsi sebagai sarana komunikasi, berbagi informasi, serta pembentukan opini publik secara luas. Media sosial memungkinkan pengguna untuk berinteraksi secara real-time tanpa batasan geografis, sehingga informasi dapat tersebar dengan cepat dan menjangkau khalayak yang sangat luas.

Salah satu *platform* populer adalah *platform x*[2] yang memungkinkan pengguna membaca, menanggapi, dan mengirim pesan singkat[3]. Platform ini menjadi ruang diskusi publik yang aktif, di mana pengguna bebas menyampaikan pandangan, kritik, maupun dukungan terhadap berbagai isu sosial, ekonomi, dan teknologi. Fenomena tersebut menghasilkan data opini dalam jumlah besar yang tersebar secara masif, sehingga menjadi sumber data yang sangat potensial untuk dianalisis. Namun, besarnya volume data dan keberagaman ekspresi bahasa yang digunakan menjadikan pengolahan data opini ini sebagai tantangan tersendiri.

Salah satu informasi yang banyak dibicarakan melalui *platform x* yaitu bitcoin. Sebagai mata uang digital terdesentralisasi, bitcoin memiliki nilai tukar yang fluktuatif sehingga sering memicu beragam opini positif maupun negatif[4]. Volume komentar yang besar, sifatnya yang tidak terstruktur, serta kompleksitas bahasa yang digunakan membuat analisis manual menjadi tidak efisien. Jadi, dibutuhkan metode otomatis untuk mengidentifikasi dan mengklasifikasikan sentimen publik terhadap bitcoin secara akurat[5].



Analisis sentimen merupakan pendekatan yang bertujuan mengidentifikasi kecenderungan opini masyarakat terhadap suatu topik tertentu. Dalam konteks big data dan media sosial, analisis sentimen berbasis machine learning menjadi penting karena mampu mengenali pola dan mengklasifikasikan teks dengan tingkat akurasi yang tinggi. Salah satu algoritma yang paling populer untuk klasifikasi teks adalah algoritma support vector machine, yang mampu menangani data besar dan mampu memisahkan kelas dengan margin maksimal[6]. Dengan perkembangan metode pembelajaran mesin yang semakin pesat, memilih algoritma yang baik menjadi faktor-faktor yang sangat penting untuk meningkatkan akurasi[7]. Oleh karena itu, berbagai penelitian sebelumnya telah membandingkan kinerja beberapa algoritma untuk menemukan model yang paling efektif dalam mengklasifikasikan opini pengguna.

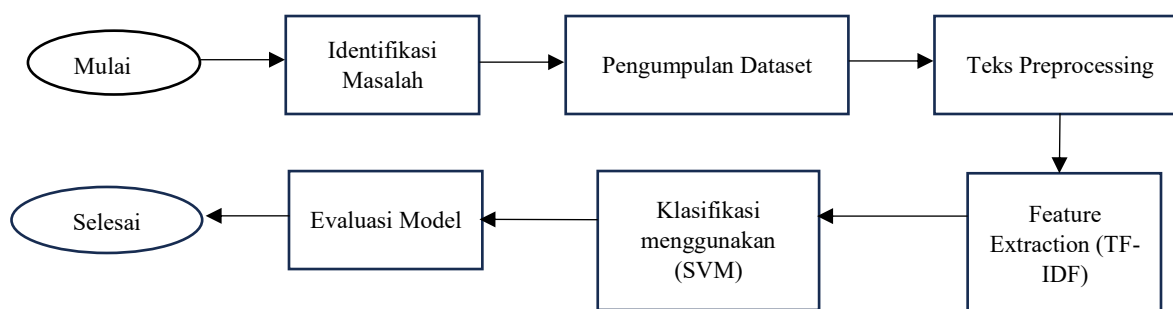
Penelitian yang dilakukan oleh Raja Allifin Muhammad (2025)[3] menggunakan metode *naive bayes* dioptimasi dengan *particle swarm optimization* untuk analisis sentimen pengguna aplikasi x terhadap bitcoin menghasilkan akurasi tertinggi 69,14 %, yang mencakup kelas sedang. Hasil menunjukkan bahwa meskipun integrasi PSO dapat meningkatkan performa *naive bayes*, model ini masih memiliki keterbatasan dalam mengenali variasi sentimen, terutama dalam kelas negatif. Sementara, beberapa penelitian lain[8] yang membandingkan *naive bayes* dan *support vector machine* menunjukkan SVM mampu menghasilkan tingkat akurasi yang lebih baik karena kemampuannya memisahkan data secara optimal melalui *hyperplane* di ruang fitur berdimensi tinggi. Selain itu, penelitian lain[9] yang menerapkan algoritma SVM dalam analisis sentimen dengan membandingkan beberapa rasio data latih dan data uji menunjukkan tingkat akurasi yang tinggi, yaitu di atas 90%. Berdasarkan kesenjangan hasil tersebut, penelitian ini menerapkan algoritma SVM untuk analisis sentimen terkait bitcoin, dengan harapan dapat menghasilkan tingkat akurasi yang lebih baik dan meningkatkan ketepatan klasifikasi dibandingkan dengan metode *naive bayes* yang digunakan dalam studi sebelumnya.

Dalam penelitian ini, metode SVM digunakan untuk menganalisis sentimen komentar terkait bitcoin di platform x. Proses penelitian diawali dengan pengumpulan data komentar, dilanjutkan dengan tahapan text preprocessing untuk membersihkan dan menormalisasi teks agar siap diolah. Selanjutnya, proses ekstraksi fitur dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) untuk merepresentasikan teks ke dalam bentuk numerik. Dengan penerapan metode tersebut, diharapkan penelitian ini mampu menghasilkan tingkat akurasi yang lebih tinggi, ketepatan klasifikasi yang lebih baik, serta memberikan kontribusi dalam pengembangan sistem analisis sentimen opini publik berbasis media sosial.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahap penelitian pada studi ini disusun secara sistematis sehingga proses analisis dapat berjalan dengan lancar. Gambar 1 menunjukkan alur penelitian yang digunakan, yang mencakup identifikasi masalah, pengumpulan dataset, text preprocessing, fitur extraction, klasifikasi menggunakan SVM dan evaluasi model.



Gambar 1. Tahapan Penelitian

Langkah - langkah penting dalam tahapan penelitian untuk mencapai tujuan penelitian.

- Identifikasi Masalah: Tahap awal penelitian yang bertujuan untuk merumuskan permasalahan yang akan diteliti.
- Pengumpulan Dataset: Tahap ini bertujuan untuk memperoleh data yang mendukung proses penelitian.
- Teks Preprocessing: Membersihkan dan menyiapkan data teks agar dapat diolah dengan baik oleh algoritma machine learning
- Feature Extraction: Mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi
- Klasifikasi Menggunakan SVM: Tahap klasifikasi dilakukan menggunakan algoritma Support Vector Machine (SVM). Algoritma ini bekerja dengan menentukan batas pemisah terbaik untuk membedakan sentimen positif dan negatif.
- Evaluasi Model: Evaluasi model dilakukan untuk mengetahui kinerja algoritma SVM dalam mengklasifikasikan sentimen komentar terkait bitcoin.

2.2 Identifikasi Masalah

Penelitian ini akan mengidentifikasi komentar pengguna terhadap bitcoin menjadi sentimen positif atau negatif secara otomatis dan akurat menggunakan algoritma *machine learning*, khususnya dengan menggunakan metode SVM.



2.3 Pengumpulan Dataset

a. Pengumpulan data

Data ini diperoleh melalui tahap *webscraping* pada platform x[10]. Metode ini memungkinkan pengambilan data publik berdasarkan kata kunci pencarian "Bitcoin" secara otomatis dan sudah melalui tahap pembersihan serta *filtering* bahasa, sehingga dapat digunakan dalam penelitian analisis sentimen dengan algoritma SVM untuk mengklasifikasikan sentiment antara positif atau negatif terhadap bitcoin[11]. Tabel 1 adalah hasil yang didapat setelah *webscraping*.

Tabel 1. Hasil Webscraping

No	Full text	Id_str	Created_at	Username	...	Source
1	Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a	1609461728 739290000	2023-01-01 08:09:13+00 :00	cintabitcoin	...	cintabitcoin
2	@cryptondo Kapan bitcoin turun ke bawah lagi testing ke harga 10	1609743567 312020000	2023-01-02 02:49:08+00 :00	miswar_riki	...	miswar_riki
...
1750	1. #BTC Cut Loss - Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	1804303293 684020000	2024-06-22 00:00:02+00 :00	essentscars	...	essentscars

Tabel 1 menunjukkan struktur awal dataset yang digunakan dalam penelitian ini, yang terdiri dari beberapa atribut seperti *full text*, *id_str*, *created_at*, *username*, dan *source*. Isi komentar yang digunakan untuk analisis sentimen disimpan dalam kolom *full text*, sedangkan *id_str* menunjukkan identitas unik dari setiap komentar, *created_at* menunjukkan waktu unggahan, *username* yang memposting komentar tersebut ditunjukkan dalam kolom *username*. Dataset ini kemudian digunakan sebagai dasar untuk tahap *text preprocessing* dan analisis sentimen menggunakan metode TF-IDF dan algoritma svm.

b. Memberikan label pada data

Setelah data dikumpulkan dan dibersihkan, selanjutnya memberikan label sentimen pada masing-masing data. Proses pelabelan dilakukan secara manual oleh tiga anotator. Anotator melakukan pelabelan ini secara manual dengan memperhatikan setiap entri yang mengandung kata kunci "Bitcoin". Setiap data dikategorikan ke dalam kategori positif dan negatif. Untuk menjaga akurasi model, data netral dieliminasi dari dataset. Apabila terjadi perbedaan pendapat antara anotator dalam menentukan label sentimen, maka dilakukan diskusi bersama untuk mencapai kesepakatan akhir. Tujuan dari proses ini adalah untuk menghasilkan data latih berkualitas tinggi untuk membantu algoritma support vector machine dalam proses klasifikasi sentimen.

c. Membagi data

Sebanyak 1.750 data yang telah melalui proses pelabelan digunakan sebagai dasar dalam penerapan metode *Support Vector Machine*. Dataset tersebut dibagi ke dalam data latih dan data uji menggunakan empat skenario rasio pembagian, yaitu 90:10, 80:20, 70:30, dan 60:40.

Pembagian data ini bertujuan untuk mengamati pengaruh variasi rasio data latih dan data uji terhadap kinerja model dalam mengklasifikasikan sentimen. Pada setiap skenario, data latih digunakan untuk membangun model SVM, sedangkan data uji digunakan untuk mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya. Hasil evaluasi dari masing-masing rasio kemudian dibandingkan untuk menentukan rasio pembagian data yang menghasilkan performa terbaik.

2.4 Text Preprocessing

Proses text preprocessing dilakukan untuk melakukan pembersihan dan penyiapan data teks mentah agar dapat diolah secara optimal dalam analisis data[12]. Adapun proses dari text preprocessing.

a. Case folding

Untuk menjaga konsistensi dalam penulisan, proses ini mengubah setiap karakter dalam teks menjadi huruf kecil.

b. Hapus URL

Pola pencocokan berbasis ekspresi reguler (Regex) digunakan untuk menghapus semua tautan yang ada dalam teks.

c. Hapus karakter khusus

Untuk membersihkan teks dari tanda baca, mention akun (@nama), tagar (#), emoticon, dan tautan web.

d. Hapus angka

Dengan menggunakan pola pencocokan menggunakan ekspresi reguler, untuk menghapus semua angka dari teks.

e. Tokenisasi



Menggunakan fungsi tokenisasi kata yang tersedia di pustaka NLTK, akan membagi teks menjadi bagian-bagian kata, yang juga dikenal sebagai token.

f. Stopword removal

Menghapus kata-kata lazim yang tidak signifikan, seperti "dan", "di", "untuk", dan lainnya.

g. Stemming

Stemming meningkatkan konsistensi data dengan mengubah kata ke bentuk dasar sehingga model tidak dapat membedakan variasi kata dengan makna yang sama.

2.5 Feature Extraction

TF-IDF menghitung bobot kata dari frekuensi kemunculannya dalam suatu dokumen TF dan seberapa jarang kata tersebut muncul di seluruh dokumen IDF. Nilai akhir dihitung dengan perkalian TF dan IDF, menghasilkan vektor fitur yang merepresentasikan pentingnya kata dalam teks[13] (Persamaan 1,2,3).

$$TF\ IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$TF(t, d) = \frac{f(t,d)}{\sum f(t,d)} \quad (2)$$

$$IDF(t) = \log \frac{N}{df(t)} \quad (3)$$

Pada perhitungan Term Frequency (TF), digunakan beberapa notasi untuk merepresentasikan komponen pembentuknya. Simbol t merepresentasikan sebuah term atau kata, sedangkan d menunjukkan dokumen tempat term tersebut muncul. Nilai $f(t, d)$ menyatakan jumlah kemunculan t dalam d . Sementara itu, $\sum f(t, d)$ menunjukkan total seluruh kata yang terdapat dalam d . Berdasarkan notasi tersebut, nilai *term frequency* dihitung untuk menggambarkan tingkat kepentingan suatu kata dalam sebuah dokumen, di mana semakin sering sebuah kata muncul dalam dokumen tertentu, maka semakin besar pula kontribusinya terhadap representasi dokumen tersebut.

Setiap kata dinilai berdasarkan signifikansinya dalam dokumen dengan rumus ini, sehingga kata yang memiliki keterkaitan kuat dengan isi teks memperoleh nilai bobot lebih tinggi dan dapat digunakan secara efektif oleh algoritma SVM dalam klasifikasi sentimen.

2.6 Klasifikasi SVM

Algoritma *support vector machine*[14]dipergunakan untuk mempelajari pola antar fitur dan label, sehingga dapat mengklasifikasikan data baru secara akurat dengan menemukan *hyperplane* terbaik yang dipisahkan oleh tiap kelas sentimen, dengan langkah – langkah berikut.

- Input Data Train: Memasukkan data *train* ke dalam sistem. Data ini biasanya sudah melalui tahap *preprocessing* (pembersihan data, tokenisasi, dan TF-IDF). Tujuannya adalah agar model bisa belajar dari pola-pola dalam data[15].
- Proses Penerapan Fungsi: Menentukan dan menerapkan fungsi kernel. Fungsi kernel berperan dalam mengkonversi data ke ruang fitur berdimensi tinggi, supaya data lebih mudah dipisah secara *linear* oleh SVM[16]. Rumus *kernel linear* (Persamaan 4).

$$(k(x_i x) = x^t x) \quad (4)$$

- Penentuan parameter SVM: Selain fungsi kernel, SVM memiliki parameter penting yaitu parameter penalti (C) yang berfungsi untuk mengatur tingkat toleransi kesalahan klasifikasi pada data latih. Nilai parameter C yang digunakan dalam penelitian ini adalah $C = 1$, yang merupakan nilai standar dan umum digunakan untuk menjaga keseimbangan antara margin maksimum dan kesalahan klasifikasi. Pemilihan nilai C tersebut bertujuan untuk menghindari overfitting serta tetap mempertahankan kemampuan generalisasi model.
- Proses Perhitungan *Train* SVM: Di tahap ini, model dilatih menggunakan algoritma SVM. SVM menghitung *hyperplane* terbaik yang membedakan kelas dengan margin maksimal[17]. Proses juga disebut sebagai model training (Persamaan 5,6,7,8).

$$(w^t x + b = 0) \quad (5)$$

Kelas positif

$$(w^t x + b \geq +1) \quad (6)$$

Kelas negatif

$$(w^t x + b \leq -1) \quad (7)$$

Dalam persamaan SVM, simbol w merepresentasikan vektor bobot (weight vector) yang berfungsi untuk menentukan arah dan orientasi *hyperplane* pemisah antar kelas. Simbol x menyatakan vektor fitur yang berasal dari hasil ekstraksi fitur menggunakan metode TF-IDF pada data teks. Sementara itu, b merupakan bias (intercept) yang berperan dalam mengatur posisi *hyperplane* terhadap titik asal ruang fitur. Kombinasi antara vektor bobot w dan bias b memungkinkan SVM membentuk *hyperplane* optimal yang mampu memisahkan data ke dalam kelas yang berbeda dengan margin maksimum.



$$\text{Margin} = \frac{2}{\|w\|} \quad (8)$$

Keterangan.

$\|w\| = \sqrt{w_1^2 + w_2^2 + \dots + w_n^2}$ adalah panjang (normal) *vector* bobot w .

- e. Proses Testing SVM: Data uji (testing data) dimasukkan untuk melihat performa model yang sudah dilatih. Model memprediksi label data berdasarkan *hyperplane* yang sudah dibuat[18].
- f. Hasil Klasifikasi: Model menghasilkan output berupa, Label sentimen (misalnya: sentimen positif, atau negatif). Tahap ini digunakan untuk evaluasi[19].

2.7 Evaluasi

Evaluasi dilakukan untuk mengevaluasi kemampuan model SVM untuk mengkategorikan sentimen positif dan negatif terhadap data komentar bitcoin di platform x[20]. Proses evaluasi akan membagi data rasio 90% sebagai data train dan 10% sebagai data test. Model akan di uji oleh empat metrik utama, akurasi, presisi, *recall*, dan *f1-score*. Empat metrik digunakan untuk menilai seberapa model bisa mengidentifikasi kelas sentimen dengan benar[21] (Persamaan 9,10,11,12).

$$a. \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$b. \text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$c. \text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

$$d. \text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

Dalam menilai hasil kerja model klasifikasi, terdapat beberapa istilah dasar yang digunakan. *True Positive* (TP) adalah kondisi ketika data yang sebenarnya bersentimen positif berhasil diprediksi sebagai positif oleh sistem. *True Negative* (TN) terjadi saat data yang sebenarnya negatif juga diprediksi sebagai negatif. *False Positive* (FP) merupakan kesalahan ketika data yang seharusnya negatif justru diprediksi sebagai positif. Sebaliknya, *False Negative* (FN) adalah kondisi ketika data yang sebenarnya positif malah diprediksi sebagai negatif sehingga tidak terdeteksi oleh sistem. Dari hasil tersebut, dihitung nilai *Precision* untuk melihat seberapa tepat prediksi positif yang dihasilkan, *Recall* untuk mengetahui seberapa banyak data positif yang berhasil ditemukan oleh sistem, serta *F1-Score* yang digunakan sebagai ukuran keseimbangan antara ketepatan dan kelengkapan hasil prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Data yang digunakan dalam penelitian ini terdiri dari 1.750 komentar yang diperoleh dari tahap pengumpulan dataset, yang masing-masing dikategorikan sebagai positif atau negatif. Kategorisasi tersebut divalidasi oleh tiga guru bahasa Indonesia untuk memastikan konsistensi makna dengan kaidah KBBI. Dataset tersebut berisi 1.167 komentar positif dan 583 komentar negatif. Dataset kemudian dibagi dengan rasio 90:10, menghasilkan 1.575 titik data latihan dan 175 titik data uji. Tabel 2 menunjukkan hasil dari dataset tersebut.

Tabel 2. Tampilan hasil dataset

komentar	label
Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a	negatif
@cryptondo Kapan bitcoin turun ke bawah lagi testing ke harga 10	negatif
...	...
Liquid restaking mengubah Bitcoin dari sekedar menyimpan nilai menjadi aset aktif yang dapat digunakan di berbagai kasus penggunaan DeFi. Pelajari lebih lanjut https://t.co/4FC76jbWeh	positif
1. #BTC Cut Loss – Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	negatif

Tabel 2 menunjukkan beberapa contoh data komentar yang digunakan dalam studi ini, yang dikategorikan sebagai positif atau negatif. Setiap komentar berisi opini atau informasi tentang Bitcoin, yang ditentukan berdasarkan isinya. Komentar optimis atau menyetujui dikategorikan sebagai positif, sementara komentar pesimis, kritis, atau khawatir tentang Bitcoin dikategorikan sebagai negatif. Data tersebut kemudian digunakan sebagai dasar untuk prapemrosesan teks dan pelatihan model klasifikasi.



3.2 Text Preprocessing

Tahap ini melibatkan text preprocessing berupa pembersihan data dengan menghapus URL, karakter khusus, dan angka, kemudian dilakukan *case folding*, *tokenization*, *stopword removal*, serta *stemming* untuk menyamakan bentuk dasar kata. Tabel 3 menunjukkan hasil sebelum dan sesudah *text preprocessing*.

Tabel 3. Sebelum dan sesudah text preprocessing

Sebelum text preprocessing	Sesudah text preprocessing
Berita Bitcoin: 3 Alasan Kenapa BTC Masih Akan Tetap Turun di Bawah US\$16.000 #bitcoin #kripto https://t.co/HU0DIFiV0a @cryptondo Kapan bitcoin turun ke bawah lagi testing ke harga 10	berita bitcoin alas btc turun us bitcoin kripto bitcoin turun testing harga
...	...
Liquid restaking mengubah Bitcoin dari sekedar penyimpan nilai menjadi aset aktif yang dapat digunakan di berbagai kasus penggunaan DeFi. Pelajari lebih lanjut https://t.co/4FC76jbWeh	liquid restaking ubah bitcoin dari simpan nilai jadi aktif guna defi ajar
1. #BTC Cut Loss – Minggu ini Bitcoin mengalami volatilitas tinggi yang mengharuskan kita untuk memotong kerugian lebih awal demi menjaga modal.	Btc cut loss minggu bitcoin alami volatilitas harus potong rugi jaga modal

Tabel 3 memperlihatkan hasil perbandingan teks sebelum dan sesudah dilakukan *text preprocessing*. Terlihat bahwa kalimat asli yang masih mengandung elemen seperti *URL*, angka, simbol, dan tanda baca telah dibersihkan menjadi teks yang lebih sederhana dan relevan. Selain itu, proses seperti *stopword removal*, *case folding*, dan *stemming* berhasil mengubah sentimen ke bentuk dasarnya, sehingga menjadi terstruktur dan siap digunakan pada proses sentimen menggunakan metode TF-IDF dan SVM.

3.3 TF - IDF

Tahap TF-IDF, teks akan dikonversikan membentuk numerik berupa nilai TF-IDF untuk setiap kata yang menunjukkan tingkat kepentingannya dalam suatu dokumen. Nilai bobot tersebut selanjutnya dimanfaatkan sebagai fitur masukan bagi model. Hasil vektorisasi ditunjukkan pada Tabel 4.

Tabel 4. Hasil vektorisasi

No	Kata	Nilai TF-IDF
1	bitcoin	0.035986
2	turun	0.027491
3	harga	0.020700
4	bitcoin turun	0.015815
5	btc	0.013659
6	harga bitcoin	0.012866
7	kripto	0.011246
8	us	0.009310
9	yg	0.009173
10	pasar	0.008364

Tabel 4 menunjukkan istilah yang paling berpengaruh dalam kumpulan data berdasarkan bobot TF-IDF. Terlihat bahwa kata bitcoin, turun, dan harga memiliki nilai tertinggi, yang menandakan bahwa topik utama pembahasan dalam data berkaitan erat dengan pergerakan harga dan kondisi pasar Bitcoin. Metode TF-IDF digunakan untuk mengidentifikasi kata-kata yang paling berpengaruh dalam data pelatihan. Pendekatan ini menyoroti istilah-istilah yang tidak hanya sering muncul tetapi juga sangat signifikan karena distribusinya yang tidak merata di seluruh dokumen. Secara umum, kata-kata dengan skor TF-IDF tertinggi mencerminkan tema utama dari kumpulan data. Berdasarkan hasil analisis, sepuluh kata dengan skor TF-IDF total tertinggi ditunjukkan pada Tabel 5.

Tabel 5. Hasil total skor TF – IDF tertinggi

No	Kata	Total TF-IDF
1	bitcoin	55.939798
2	turun	43.316927
3	harga	32.109944
4	bitcoin turun	25.071156
5	btc	21.614785
6	harga bitcoin	20.370698
7	kripto	17.586678
8	us	14.188759
9	beli	13.817385



No	Kata	Total TF-IDF
10	yg	13.412041

Tabel 5 menunjukkan sepuluh kata dengan nilai total TF-IDF tertinggi yang muncul dalam data latih (training data). Kata “bitcoin”, “turun”, dan “harga” memiliki bobot TF-IDF paling tinggi, menandakan bahwa istilah tersebut paling dominan dan relevan dalam menggambarkan konteks sentimen yang dibahas dalam dataset mengenai bitcoin.

3.4 Hasil Klasifikasi SVM

SVM adalah metode klasifikasi yang untuk mencari *hyperplane* terbaik untuk membedakan Data dalam dua kategori, yaitu sentimen positif dan sentimen negatif. Pada penelitian ini digunakan kernel linear karena hasil pengolahan teks dengan TF-IDF bersifat linier, sehingga model dapat memisahkan data secara langsung berdasarkan fitur yang terbentuk. Selain itu, penggunaan kernel linear juga memudahkan peneliti dalam melihat bobot setiap fitur yang berkontribusi terhadap proses klasifikasi.

Setelah model berhasil dilatih, peneliti melakukan analisis terhadap fitur-fitur yang memiliki pengaruh terbesar dalam menentukan hasil klasifikasi. Fitur dengan bobot positif menunjukkan kecenderungan terhadap sentimen positif, sedangkan fitur dengan bobot negatif lebih mengarah ke sentimen negatif. Berikut akan ditampilkan hasil fitur yang paling berpengaruh terhadap hasil klasifikasi SVM dapat dilihat pada tabel 6.

Tabel 6. Fitur yang paling mempengaruhi klasifikasi

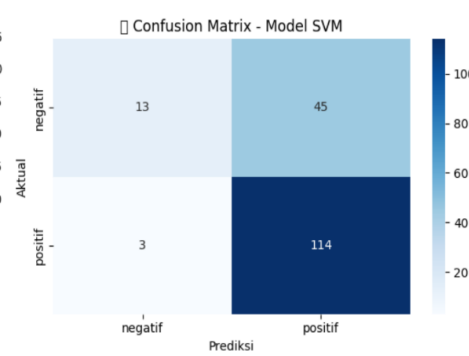
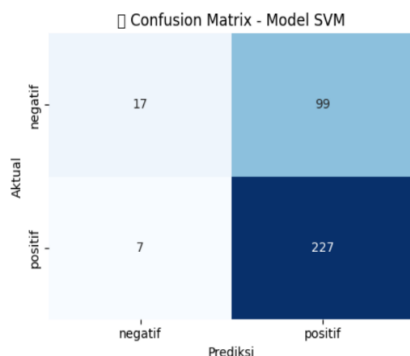
No	Fitur Positif	Bobot	Fitur Negatif	Bobot
1	bitcoin turun altcoin	1.1320	turun	-3.2843
2	beli	1.0500	bitcoin turun	-2.1908
3	turun altcoin	0.8882	gara	-1.4345
4	altcoin	0.8709	bitcoin halving turun	-1.2267
5	coin	0.8678	bitcoin nya	-1.1622
6	jt	0.8519	kayak	-1.1415
7	data	0.8475	turun persen	-1.0725
8	bitcoin turun beli	0.8080	rugi	-1.0410
9	tren	0.7859	turun us	-1.0217
10	gt	0.7392	matauangkripto	-0.9888

Tabel 6 menunjukkan sepuluh fitur (kata atau frasa) dengan bobot TF-IDF tertinggi yang paling memengaruhi hasil klasifikasi SVM untuk setiap kelas sentimen. Fitur pada kelas positif seperti altcoin, beli, dan tren menggambarkan konteks optimis terhadap kondisi pasar Bitcoin. Sebaliknya, fitur pada kelas negatif seperti turun, rugi, dan bitcoin halving turun mencerminkan persepsi pesimis atau kekhawatiran terhadap penurunan nilai Bitcoin. Nilai bobot yang lebih tinggi menunjukkan pengaruh yang lebih kuat dalam menentukan arah sentimen. Bobot yang tinggi—misalnya nilai 1.1320 pada fitur “bitcoin turun altcoin”—dapat muncul karena frasa tersebut lebih sering muncul pada komentar yang berlabel positif dibandingkan komentar negatif. Selain karena frekuensi kemunculannya yang dominan pada satu kelas, frasa tersebut juga dianggap unik oleh perhitungan TF-IDF dan terbukti membantu SVM dalam membedakan sentimen positif dan negatif. Kombinasi kemunculan yang sering pada kelas positif, jarang muncul pada kelas negatif, serta kontribusinya dalam memisahkan dua kelas inilah yang membuat bobot fitur tersebut menjadi tinggi.

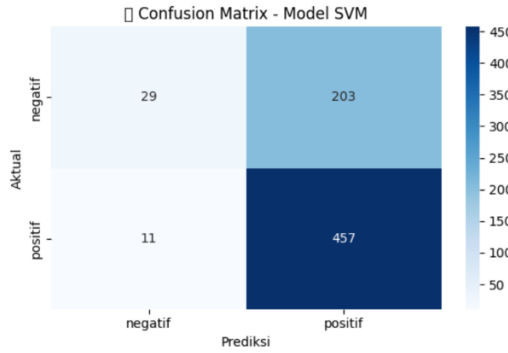
3.5 Evaluasi

3.5.1 Hasil Klasifikasi SVM

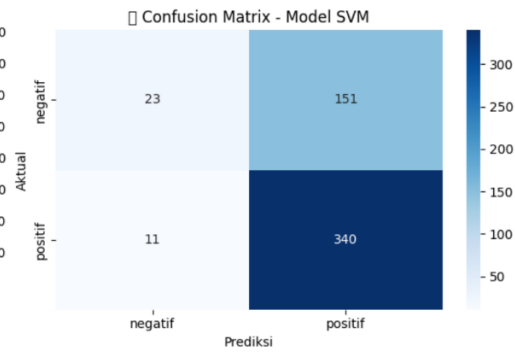
Pada tahap ini, para peneliti menganalisis kinerja model SVM untuk menentukan rasio pemisahan data yang optimal. Pengujian dilakukan dengan empat rasio 90:10, 80:20, 70:30, dan 60:40. Untuk menilai kemampuan model dalam mendeteksi sentimen positif dan negatif, digunakan visualisasi *confusion matrix* pada setiap rasio dapat dilihat pada Gambar 2 hingga Gambar 5.



Gambar 2. Confusion matrix SVM 90:10 **Gambar 3.** Confusion matrix SVM 80:20



Gambar 4. Confusion matrix SVM 70:30



Gambar 5. Confusion matrix SVM 60:40

Secara keseluruhan, keempat confusion matrix menunjukkan bahwa model SVM memiliki kemampuan yang sangat baik dalam menunjukkan bahwa model SVM unggul dalam mengidentifikasi sentimen positif tetapi kurang sensitif terhadap sentimen negatif. Hal ini terbukti dari tingkat keberhasilan yang lebih tinggi pada kelas positif dibandingkan dengan kelas negatif, yang menunjukkan bahwa model lebih mampu mempelajari pola kata yang mengarah pada sentimen positif. Lebih lanjut, rasio 90:10 terbukti optimal, karena menyediakan lebih banyak data pelatihan, yang memungkinkan model untuk mempelajari pola sentimen secara lebih efektif dan mencapai batas keputusan yang lebih stabil dibandingkan rasio lainnya. Kondisi ini menjadi dasar untuk fase evaluasi, yang tujuannya adalah untuk menunjukkan kemampuan model dalam mengidentifikasi pola hubungan antara fitur teks dan label sentimen, sehingga secara akurat mengkategorikan opini pengguna tentang Bitcoin sebagai positif atau negatif. Evaluasi ini juga berfungsi untuk mengukur generalisasi model terhadap data baru yang sebelumnya tidak terlihat. Dengan begitu, hasil yang diperoleh dapat menjadi acuan dalam menentukan efektivitas metode yang digunakan. Berikut ditampilkan hasil evaluasi SVM pada tabel 7.

Tabel 7 Hasil evaluasi SVM

rasio		precision	recall	f1-score	support
90:10	negatif	0.81	0.22	0.35	58
	positif	0.72	0.97	0.83	117
	accuracy			0.73	175
	macro avg	0.76	0.60	0.59	175
	weighted avg	0.75	0.73	0.67	175
80:20	negatif	0.71	0.15	0.24	116
	positif	0.70	0.97	0.81	234
	accuracy			0.70	350
	macro avg	0.70	0.56	0.53	350
	weighted avg	0.70	0.70	0.62	350
70:30	negatif	0.68	0.13	0.22	174
	positif	0.69	0.97	0.81	351
	accuracy			0.69	525
	macro avg	0.68	0.55	0.51	525
	weighted avg	0.69	0.69	0.61	525
60:40	negatif	0.72	0.12	0.21	232
	positif	0.69	0.98	0.81	468
	accuracy			0.69	700
	macro avg	0.71	0.55	0.51	700
	weighted avg	0.70	0.69	0.61	700

Tabel 7 menampilkan hasil pengujian untuk empat rasio pembagian data 90:10, 80:20, 70:30, dan 60:40. rasio 90:10 menunjukkan kinerja terbaik. Dalam pengujian ini, model SVM mencapai skor akurasi 0,73, dengan presisi 0,81 untuk kelas negatif dan 0,72 untuk kelas positif. Selain itu, perolehan recall pada kelas positif mencapai 0,97 dengan F1-score 0,83, yang menunjukkan kemampuan model yang kuat dalam mendeteksi sentimen positif.

Meskipun nilai recall untuk kelas negatif tetap rendah, yaitu 0,22. Hal ini menunjukkan bahwa model tidak dapat mendeteksi pola sentimen negatif secara andal. Hal ini kemungkinan disebabkan oleh ketidakseimbangan dalam kumpulan data antara kelas positif dan negatif, dengan data positif lebih banyak daripada data negatif.

Meskipun demikian, rasio 90:10 menunjukkan keseimbangan yang cukup baik antara keduanya, terutama dalam hal akurasi dan presisi, terutama dalam kategori positif. Secara keseluruhan, hasil ini menunjukkan bahwa kemampuan model untuk mempelajari pola sentimen meningkat dengan proporsi data pelatihan yang lebih besar. Jumlah data pelatihan yang lebih besar memungkinkan model untuk merepresentasikan karakteristik data secara lebih akurat.

Dengan mempertimbangkan seluruh metrik evaluasi, rasio 90:10 dapat dikatakan sebagai konfigurasi paling efektif pada penelitian ini. Rasio tersebut tidak hanya menampilkan rasio tertinggi, tetapi menghasilkan performa yang



lebih stabil dibandingkan rasio lainnya. Oleh sebab itu, rasio 90:10 dianggap paling sesuai untuk merepresentasikan performa terbaik model SVM dalam analisis sentimen komentar Bitcoin pada platform X.

3.5.2 Percobaan Tambahan Pengaruh Kernel dan Distribusi Data

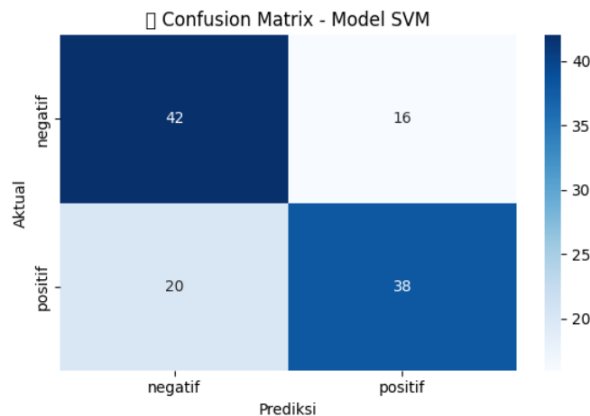
Untuk melihat pengaruh kombinasi jenis kernel dan distribusi data terhadap performa model, dilakukan percobaan tambahan dengan mengombinasikan kernel SVM dan kondisi distribusi data. Percobaan ini dilakukan pada dataset tidak seimbang dan dataset seimbang hasil proses undersampling, dengan tujuan untuk memberikan gambaran tambahan mengenai stabilitas dan konsistensi kinerja model. Hasil evaluasi dari masing-masing kombinasi kernel dan distribusi data disajikan dalam Tabel 8.

Tabel 8. Tabel kombinasi kernel dan distribusi data

Rasio	Dataset	Kernel	Akurasi	Presisi	Recall	F1-Score
90:10	Tidak seimbang	Linear	72.57%	0.75	0.73	0.66
	Tidak seimbang	RBF	69.14%	0.73	0.69	0.59
	Seimbang	Linear	68.96%	0.69	0.68	0.68
	Seimbang	RBF	59.48%	0.61	0.59	0.57
80:20	Tidak seimbang	Linear	69.71%	0.70	0.69	0.62
	Tidak seimbang	RBF	69.14%	0.75	0.69	0.59
	Seimbang	Linear	63.63%	0.63	0.63	0.63
	Seimbang	RBF	59.48%	0.62	0.59	0.56
70:30	Tidak seimbang	Linear	69.14%	0.68	0.69	0.61
	Tidak seimbang	RBF	68.57%	0.75	0.68	0.57
	Seimbang	Linear	64.65%	0.64	0.64	0.64
	Seimbang	RBF	59.19%	0.62	0.59	0.56
60:40	Tidak seimbang	Linear	69.43%	0.70	0.69	0.61
	Tidak seimbang	RBF	68.29%	0.73	0.68	0.57
	Seimbang	Linear	61.42%	0.61	0.61	0.61
	Seimbang	RBF	59.48%	0.63	0.59	0.56

Berdasarkan Tabel 8, dapat dilihat bahwa penggunaan kernel linear secara umum memberikan performa yang lebih baik dibandingkan kernel RBF, baik pada dataset tidak seimbang maupun dataset seimbang. Kernel linear menunjukkan nilai akurasi dan F1-score yang lebih tinggi serta relatif stabil, sementara kernel RBF menghasilkan performa yang lebih rendah pada kedua kondisi distribusi data. Hasil ini mengindikasikan bahwa kernel linear lebih sesuai digunakan pada data teks yang direpresentasikan menggunakan fitur TF-IDF. Selain itu, penggunaan dataset seimbang melalui undersampling menghasilkan performa klasifikasi yang lebih merata antar kelas, meskipun nilai akurasi keseluruhan cenderung sedikit menurun akibat berkurangnya jumlah data latih.

Untuk memperoleh gambaran yang lebih rinci mengenai hasil klasifikasi pada dataset seimbang, dilakukan analisis lanjutan menggunakan confusion matrix. Confusion matrix digunakan untuk menunjukkan jumlah data yang diklasifikasikan dengan benar serta kesalahan klasifikasi yang terjadi pada masing-masing kelas sentimen. Pada tahap ini, confusion matrix ditampilkan untuk model SVM dengan kernel linear pada dataset seimbang hasil undersampling dengan rasio pembagian data 90:10, karena konfigurasi tersebut menunjukkan performa yang paling stabil pada percobaan tambahan sebelumnya.



Gambar 6. Confusion matrix SVM 90:10 data seimbang dan kernel linear

Berdasarkan confusion matrix yang ditampilkan pada Gambar 6, model SVM dengan kernel linear berhasil mengklasifikasikan 42 data sentimen negatif dan 38 data sentimen positif dengan benar. Sementara itu, terdapat 16 data sentimen negatif yang salah diklasifikasikan sebagai positif dan 20 data sentimen positif yang salah diklasifikasikan sebagai negatif. Hasil ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang relatif seimbang pada kedua



kelas sentimen, meskipun masih terdapat kesalahan prediksi yang perlu diperhatikan, khususnya pada data dengan karakteristik sentimen yang ambigu.

Secara keseluruhan, percobaan tambahan ini menunjukkan bahwa jenis kernel dan distribusi data memiliki pengaruh terhadap kinerja model SVM dalam klasifikasi sentimen. Kernel linear terbukti memberikan performa yang lebih stabil dibandingkan kernel RBF pada berbagai kondisi distribusi data. Selain itu, penggunaan dataset seimbang melalui proses undersampling mampu meningkatkan keseimbangan hasil klasifikasi antar kelas, meskipun berdampak pada penurunan jumlah data latih. Namun demikian, percobaan ini hanya digunakan sebagai analisis pendukung, sedangkan kesimpulan utama penelitian tetap mengacu pada hasil pengujian pada dataset asli sesuai dengan metodologi yang telah ditetapkan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma Support Vector Machine (SVM) dengan representasi fitur Term Frequency–Inverse Document Frequency (TF-IDF) mampu mengklasifikasikan sentimen pengguna terhadap Bitcoin secara cukup akurat dengan tingkat akurasi sebesar 73%. Hasil ini menunjukkan bahwa kombinasi SVM dan TF-IDF efektif dalam menangani data teks tidak terstruktur yang berasal dari media sosial. Model yang dibangun menunjukkan kemampuan yang baik dalam mengenali komentar dengan sentimen positif, namun masih memiliki keterbatasan yang signifikan dalam mendeteksi sentimen negatif, yang tercermin dari nilai recall kelas negatif yang relatif rendah. Kondisi ini mengindikasikan bahwa sebagian besar komentar negatif belum dapat teridentifikasi dengan optimal, terutama akibat ketidakseimbangan distribusi data, di mana jumlah komentar positif lebih dominan dibandingkan komentar negatif. Jika dibandingkan dengan penelitian Raja Allifin Muhammad (2025) yang menggunakan algoritma Naïve Bayes dengan optimasi Particle Swarm Optimization (PSO) dan menghasilkan akurasi pada kisaran 68–69%, SVM pada penelitian ini menunjukkan peningkatan akurasi secara keseluruhan. Namun demikian, peningkatan akurasi tersebut tidak serta-merta menunjukkan keunggulan mutlak dari sisi praktis, mengingat rendahnya kemampuan model dalam mendeteksi kelas negatif. Oleh karena itu, efektivitas SVM dalam penelitian ini lebih tepat dipandang sebagai keunggulan dalam klasifikasi kelas dominan, bukan sebagai solusi yang sepenuhnya optimal untuk seluruh kelas sentimen. Sebagai pengembangan penelitian selanjutnya, disarankan untuk meningkatkan jumlah data, menerapkan teknik penyeimbangan kelas sentimen, serta melakukan optimasi parameter SVM seperti nilai C dan pemilihan kernel. Selain itu, eksplorasi metode lanjutan seperti LSTM atau BERT diharapkan dapat meningkatkan kemampuan model dalam mengenali sentimen minoritas secara lebih akurat.

REFERENCES

- [1] V. Albarysah, "Perkembangan Ekonomi Digital Mengenai Perilaku Pengguna Media Sosial dalam Melakukan Transaksi," *JMEB Jurnal Manajemen Ekonomi & Bisnis*, vol. 1, no. 2, 2023.
- [2] Radhika Chabra, "Twitter Diplomacy: A Brief Analysis," *Orf Issue Brief*, no. 335, 2020.
- [3] E. H. F. W. L. O. Raja Allifin Muhammad, "Penerapan Metode Naive Bayes Dengan PSO Pada Analisis Sentimen Pengguna Aplikasi X Terhadap Bitcoin," *Jurnal Sistem Informasi*, vol. 14, no. 4, 2025.
- [4] R. Parlika, S. Ilham Pradika, A. M. Hakim, and R. N. M. Kholilul, "Analisis Sentimen Twitter Terhadap Bitcoin dan Cryptocurrency Berbasis Python TextBlob," *JIFTI-Jurnal Ilmiah Teknologi Informasi dan Robotika*, vol. 2, no. 2, 2020.
- [5] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conf Ser Mater Sci Eng*, vol. 1090, no. 1, p. 012053, Mar. 2021, doi: 10.1088/1757-899x/1090/1/012053.
- [6] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 67–75, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.5.
- [7] D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2780–2788, Oct. 2021, doi: 10.11591/eei.v10i5.3157.
- [8] T. Setiawan, S. Liem, and D. M. R. Pribadi, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare," *Applied Information Technology and Computer Science*, vol. 3, no. 2, pp. 28–32, 2024.
- [9] A. Septini, Susanto, and Elmayati, "Analisis Sentimen Masyarakat di Twitter Mengenai Open AI CHATGPT Menggunakan Metode Support Vector Machine (SVM)," *Bulletin of Computer Science Research*, vol. 5, no. 2, pp. 138–149, Feb. 2025, doi: 10.47065/bulletincsr.v5i2.475.
- [10] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia Di Twitter," *Jurnal TEKNOKOMPAK*, vol. 14, no. 2, p. 86, 2020.
- [11] M. Dogucu and M. Çetinkaya-Rundel, "Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities," *Journal of Statistics and Data Science Education*, vol. 29, no. S1, pp. S112–S122, 2021, doi: 10.1080/10691898.2020.1787116.
- [12] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, Jul. 2020, doi: 10.1088/1757-899X/874/1/012017.
- [13] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 2, no. 1, 2022.
- [14] N. Jasmin, R. K. Dinata, and I. Sahputra, "Implementation of Data Mining for Vertigo Disease Classification Using the Support Vector Machine (SVM) Method," *JACKA-Journal Of Computer Knowledge and Algorithms*, vol. 1, no. 4, pp. 103–108, 2024.



- [15] Sulisetyo Puji Widodo and Adila Alfa Krisnadhi, "Analyzing Twitter Sentiments on Booster Vaccination with Support Vector Machine (SVM) Method," *International Journal of Innovation in Enterprise System*, vol. 8, no. 2, pp. 11–21, Oct. 2024, doi: 10.25124/ijies.v8i02.236.
- [16] P. Fremmuzar and A. Baita, "Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter," *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 57–66, Sep. 2023, doi: 10.34010/komputika.v12i2.9460.
- [17] D. Alita and Y. Fernando, "Multiclass SVM Algorithm For Sarcasm Text In Twitter," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, 2021.
- [18] M. Farhan Saleh and R. Imanda, "Public Sentiment Analysis of the Free Meal Program: A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 1, pp. 2548–6861, 2025.
- [19] H. Zhou, "Research of Text Classification Based on TF-IDF and CNN-LSTM," *J Phys Conf Ser*, vol. 2171, no. 1, Jan. 2022, doi: 10.1088/1742-6596/2171/1/012021.
- [20] S. D. Fitri, D. Lestari, R. Raaiqa Bintana, R. Aryani, M. Ilhami, and Y. Noverina, "Implementasi Model Support Vector Machine Dalam Analisa Sentimen Masyarakat Mengenai Kebijakan Penerapan Aplikasi Mypertamina," *BRIDGE*, vol. 2, no. 2, pp. 176–193, 2024, doi: 10.62951/bridge.v2i2.180.
- [21] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.