

# METODE PEMBOBOTAN TF-IDF UNTUK KLASIFIKASI TEKS QUICK COUNT PEMILIHAN WAKIL PRESIDEN INDONESIA 2024 PADA X TWITTER DENGAN METODE SVM

Ricky Albin Pranata <sup>a,1</sup>, Rudiman <sup>b,2\*</sup>, Naufal Azmi Verdikha <sup>c,3</sup>

<sup>a,b,c</sup> Universitas Muhammadiyah Kalimantan Timur, Jl. Ir. H. Juanda, Samarinda 75124

<sup>1</sup> [2011102441017@umkt.ac.id](mailto:2011102441017@umkt.ac.id) ; <sup>2</sup> [rud959@umkt.ac.id](mailto:rud959@umkt.ac.id) \*; <sup>3</sup> [nav651@umkt.ac.id](mailto:nav651@umkt.ac.id)

\* corresponding author

## ARTICLE INFO

## ABSTRACT

### Keywords

Quick Count, SVM, Text Classification, Twitter, TF-IDF

The Quick Count of Indonesia's 2024 Vice Presidential Election triggered mixed public reactions on Twitter. The large number and variety of opinions expressed make it difficult to identify and classify accurate sentiment. This research aims to develop a text classification model using Support Vector Machine (SVM) to identify sentiment in tweets related to the election Quick Count. Data was obtained through tweet collection, followed by pre-processing, word weighting using TF-IDF, and splitting the data for training and testing the model. The results showed that the developed SVM model achieved 81.13% accuracy in tweet sentiment classification. The model showed varying performance with a precision of 64.29%, recall of 51.92%, and F1-score of 57.41% for the positive class. The implementation of this model is expected to help filter information more effectively and assist stakeholders in understanding public opinion more accurately, although there is still room for improvement, especially in the identification of positive sentiment.

## 1. Pendahuluan

Dalam ranah politik, proses pemilihan umum (pemilu) di Indonesia untuk Calon Wakil Presiden menarik perhatian yang lebih besar dibandingkan dengan pemilihan umum legislatif. Berita tentang hal tersebut menjadi sorotan utama dalam berbagai jenis media, termasuk media cetak, elektronik, maupun online[1]. Kualitas yang tinggi dalam pelaksanaan pemilihan umum menjadi kunci penting dalam menegakkan kedaulatan rakyat dalam sistem pemerintahan demokratis, sejalan dengan prinsip-prinsip Pancasila dan Konstitusi Negara Republik Indonesia Tahun 1945[2]. *Quick Count* merupakan suatu metode pengecekan hasil pemilu yang dilakukan melalui proses menghitung hasil pemilihan secara acak di berbagai TPS di berbagai wilayah[3].

Dengan semakin meluasnya penggunaan media sosial sebagai platform untuk mengekspresikan opini politik, klasifikasi sentimen masyarakat menjadi sangat penting untuk merespons perkembangan politik secara tepat waktu. Menurut data Hootsuite 2024, jumlah pengguna internet di Indonesia mencapai 212,9 juta, sementara pengguna media sosial aktif berjumlah 167 juta. Dari 167 juta pengguna media sosial tersebut, 57,5% menggunakan platform Twitter X. Ketertarikan pengguna media sosial terhadap informasi politik cukup tinggi, banyak masyarakat membagikan dan mengomentari konten hasil *Quick Count*, serta menyampaikan pandangan mereka tentang para calon Wakil Presiden[4].

Klasifikasi teks dokumen adalah masalah yang mendasar dan penting. Dalam teks dokumen, isi yang ada merupakan bahasa alami manusia, yang memiliki struktur kompleks dan jumlah kata yang sangat banyak. Masalah ini menjadi cukup rumit karena penggunaan bahasa alami tersebut. Salah satu algoritma

yang dapat digunakan untuk mengklasifikasikan dokumen dimulai dengan pengelolaan teks dokumen menggunakan *text mining*, yaitu algoritma *Support Vector Machine*(SVM)[5]. Tujuan utama dalam *text mining* untuk menemukan kata-kata yang dapat mempresentasikan isi dokumen tersebut, sehingga memungkinkan analisis hubungan antara dokumen-dokumen tersebut. Setelah tahap *text mining*, algoritma yang digunakan adalah *Term-Frequency Inverse Document Frequency* (TF-IDF), yang merupakan metode untuk meningkatkan bobot relevansi suatu kata (*term*) terhadap suatu dokumen[6].

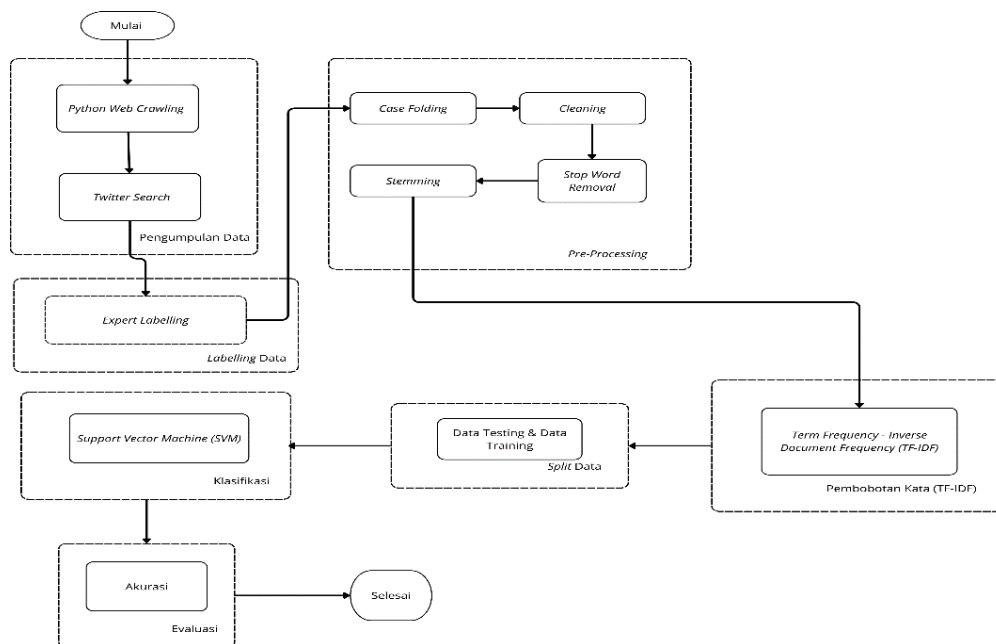
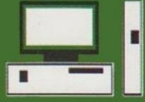
*Support Vector Machine* (SVM) adalah salah satu teknik dalam *machine learning* yang didasarkan pada teori pembelajaran statistik struktural dan memiliki kinerja yang lebih unggul dibandingkan dengan metode *machine learning* lainnya[7]. Algoritma SVM membangun model yang memprediksi apakah data baru termasuk dalam satu kategori atau lainnya. SVM beroperasi dengan cara mencari *hyperplane* optimal yang memaksimalkan jarak antara kelas-kelas[8].

Terdapat penelitian [9] data yang digunakan sebanyak 11.000 tweet yang dibagi menjadi data latih sebanyak 70% dan data uji 30%. Pembobotan dilakukan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil penelitian menunjukkan bahwa metode SVM memiliki akurasi yang lebih tinggi dibandingkan *Naïve Bayes*. Hasil ketepatan klasifikasi menggunakan *Naïve Bayes* mencapai akurasi 59%, sedangkan SVM mencapai akurasi 61%. Selanjutnya penelitian [10] data yang digunakan sebanyak 189919 dataset dengan rasio 70% data train dan 30% data test menghasilkan tingkat akurasi sebesar 68%. Dari penelitian sebelumnya sentimen analisis dengan label positif berjumlah 3000, label netral berjumlah 2500 dan label negatif cukup tinggi yaitu diatas 4000.

Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat di Twitter terhadap hasil *Quick Count* pemilihan Wakil Presiden Indonesia 2024 menggunakan metode *Support Vector Machine* (SVM) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). Tahapan penelitian meliputi pengumpulan data opini publik dari Twitter, penerapan TF-IDF untuk ekstraksi fitur teks, penggunaan SVM untuk klasifikasi sentimen, dan analisis hasil klasifikasi. Melalui pendekatan ini, diharapkan dapat diperoleh hasil klasifikasi sentimen yang lebih akurat, membantu pemangku kebijakan dalam memahami pandangan masyarakat secara lebih menyeluruh, serta mendukung pengambilan keputusan yang lebih informatif terkait hasil pemilihan Wakil Presiden. Penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam klasifikasi teks publik pada konteks politik Indonesia, khususnya dalam peristiwa pemilihan nasional yang krusial.

## 2. Metodologi Penelitian

Penelitian ini dimulai dengan tahap pengumpulan data menggunakan *Python Web Crawling* dan pencarian Twitter. Data yang terkumpul kemudian melalui proses *labelling* data oleh ahli. Selanjutnya, data memasuki tahap *pre-processing* yang terdiri dari *case folding*, *cleaning*, *stop word removal*, dan *stemming*. Setelah *pre-processing*, data diproses menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk pembobotan kata. Data yang telah dibobot kemudian dibagi menjadi data training dan data testing. Proses klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM). Hasil klasifikasi kemudian dievaluasi untuk mengukur akurasi model.



Gambar 1 Alur Penelitian

### 2.1. Pengumpulan Data

Dalam studi ini pengambilan data dilakukan dengan cara *crawling* data pada media sosial X Twitter dengan menggunakan salah satu library *tweet-harvest* yang dibuat dengan menggunakan Node.js yang ada dalam pemrograman Python. Library *tweet-harvest* digunakan sebagai alat untuk mengumpulkan data dan menganalisis informasi dari Twitter dengan keyword “Quick Count”. Data yang berhasil diekstrak akan disimpan melalui format file CSV atau Excel.

### 2.2. Labelling Data

Dalam proses melakukan klasifikasi teks pada data *tweet* dilakukan pelabelan data oleh ahli yang memiliki pengalaman dalam mengategorikan teks berdasarkan opini publik. Tujuan utama pelibatan ahli ini adalah untuk memastikan akurasi dan konsistensi dalam proses pelabelan data, yang merupakan tahap kritis dalam penelitian klasifikasi teks ini. Dengan pendekatan ini, diharapkan dapat diperoleh dataset yang berkualitas tinggi untuk mendukung klasifikasi teks yang lebih akurat dan relevan.

### 2.3. Pre-Processing

Proses pre-processing ini bertujuan untuk membersihkan atau menghapus bagian teks yang tidak relevan dalam dokumen yang tidak relevan[11]. Langkah – langkah yang dilakukan dalam tahap *pre-processing* meliputi 4 (empat) langkah pada sebagai berikut:

- a. *Case Folding* adalah proses merubah seluruh huruf dalam teks menjadi huruf kecil atau besar.
- b. *Cleaning* adalah proses pembersihan teks yang melibatkan penghapusan elemen-elemen yang tidak relevan atau mengganggu dalam teks, seperti tanda baca, karakter khusus, atau karakter yang tidak diinginkan lainnya.

- c. *Stop Word Removal* adalah proses menghapus kata-kata penghenti dari teks. Kata-kata penghenti adalah kata-kata yang sangat umum dalam bahasa tertentu dan biasanya tidak menyediakan banyak informasi yang signifikan dalam analisis teks.
- d. *Stemming* adalah proses untuk melibatkan pemangkasan akhiran atau awalan kata untuk menghasilkan akar kata atau bentuk dasarnya.

#### 2.4. Pembobotan Kata (TF-IDF)

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik dalam pemrosesan teks yang digunakan untuk menilai pentingnya sebuah kata dalam suatu dokumen. Teknik ini umumnya digunakan untuk klasifikasi dokumen, dan juga mengurangi bobot kata-kata umum yang muncul dalam banyak dokumen, seperti kata penghenti, yang mungkin tidak memberikan banyak informasi penting[12]. Menggunakan *TfidfVectorizer* untuk transformasi data teks adalah langkah penting dalam klasifikasi teks. *TfidfVectorizer* mengonversi teks menjadi bentuk numerik yang bisa digunakan oleh algoritma data mining[13].

Berdasarkan penggunaan TF-IDF dapat menggunakan rumus persamaan 1 di bawah ini:

$$tf_{t,d} = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d} \quad (1)$$

*Term Frequency* (TF) menghitung frekuensi kemunculan sebuah kata muncul dalam sebuah dokumen. Dan terkait rumus IDF persamaan 2 yaitu:

$$idf_d = \log \frac{N}{n_t} \quad (2)$$

*Inverse Document Frequency* (IDF) mengukur seberapa penting sebuah kata dengan memperhitungkan seberapa sering kata tersebut muncul dalam semua dokumen. N melambangkan jumlah total dokumen dalam kumpulan teks. Mengenai rumus TF-IDF score persamaan 3 yakni:

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

Keterangan :

- t = Kata kunci, term
- d = Dokumen
- t.d = nilai TF-IDF untuk kata t dalam dokumen d
- Tf = Banyaknya t (kata) yang dicari dalam dokumen
- Idf = Banyak t kebalikan dari kata yang dicari

#### 2.5. Split Data

Pada langkah ini, pemisahan dataset yang dilakukan dalam rangka penelitian terbagi menjadi beberapa bagian, umumnya untuk mengevaluasi keberhasilan model atau algoritma yang diusulkan. Dataset awal biasanya terpisah menjadi dua bagian utama, yaitu data (*training*) dan data uji (*testing*). Data training digunakan untuk melatih algoritma yang diusulkan, sedangkan data uji (*testing*) dimanfaatkan untuk mengevaluasi kinerja algoritma[14]. Data awal dipecah menjadi dua bagian di mana 90% dari total data digunakan untuk melatih model, dan 10% sisanya dialokasikan sebagai set pengujian. Pembagian data menggunakan random state 42 untuk menjamin konsistensi dan kemampuan mengulang eksperimen.

Metode ini memastikan pemilihan sampel acak yang sama setiap kali eksperimen dijalankan, sehingga mendukung reproduktibilitas hasil penelitian.

## 2.6. Evaluasi

Pada proses klasifikasi peneliti menggunakan metode *Support Vector Machine* (SVM), yang merupakan metode pembelajaran *supervised* yang menganalisis data serta mengidentifikasi pola. SVM berperan sebagai pengklasifikasi yang memerlukan data uji (*testing*) di mana setiap data ditandai sebagai anggota dari satu kategori tertentu. Melalui algoritma pelatihan SVM, sebuah model dibangun untuk mengidentifikasi apakah data yang telah diolah termasuk dalam kategori tersebut atau tidak [15]. Kelebihan metode SVM terletak pada kemampuannya dalam menemukan *hyperplane* yang dapat memisahkan kelas-kelas dengan maksimal, sehingga meningkatkan margin antara kelas yang berbeda [16]. Pada prinsipnya, SVM adalah metode klasifikasi linear. Meski begitu, SVM dapat dimodifikasi untuk menangani klasifikasi nonlinear. Untuk menghadapi kasus nonlinear, SVM memanfaatkan teknik yang disebut *kernel trick*, yang memetakan data ke ruang dimensi yang lebih tinggi. Fungsi *kernel trick* pada SVM yang digunakan yaitu *kernel linear*. Fungsi *kernel* ini berfungsi untuk mengukur kesamaan antara dua data titik dalam ruang fitur, membantu memetakan data ke dalam dimensi yang lebih tinggi agar dapat diolah lebih efektif oleh algoritma pembelajaran [17].

Berikut rumus *Support Vector Machine* (SVM) pada persamaan 4:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \quad (4)$$

Keterangan :

- f(x) = nilai prediksi untuk titik data x
- $\alpha_i$  = variabel pendukung yang menentukan titik data
- $y_i$  = label kelas dari titik data ke i
- K ( $x_i, x$ ) = pengukuran kesamaan antara titik data  $x_i$  dan x
- b = konstanta yang mempengaruhi posisi hyperplane

Evaluasi dilakukan dengan memanfaatkan metode *Confusion Matrix*, serta mengukur performa menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Akurasi digunakan untuk menilai seberapa efektif metode yang diterapkan dalam proses klasifikasi. *Recall* akan menunjukkan rasio antara prediksi benar positif dan jumlah keseluruhan data yang benar positif. *Precision* menggambarkan rasio antara prediksi benar positif dan seluruh hasil yang diprediksi positif. *F1-Score* kemudian dihitung dengan mengombinasikan rata-rata tertimbang dari nilai *Recall* dan *Precision* [18]. Berikut persamaanya 5,6,7,8 :

$$\text{Akurasi} = \left( \frac{TP + TN}{(TP + TN + FP + FN)} \right) \times 100\% \quad (5)$$

$$Precision = \left( \frac{TP}{TP + FP} \right) \times 100\% \tag{6}$$

$$Recall = \left( \frac{TP}{TP + FN} \right) \times 100\% \tag{7}$$

$$F1 - Score = \left( 2 \times \frac{Precision \times Recall}{Precision + Recall} \right) \times 100\% \tag{8}$$

Terdapat empat istilah dalam *confusion matrix* yaitu, *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *False Positive* (FP) mengacu pada situasi di mana data yang seharusnya negatif dikelirukan sebagai positif. Di sisi lain, *True Negative* (TN) adalah saat data yang sebenarnya negatif teridentifikasi dengan benar sebagai negatif[19]. Tabel *confusion matrix* akan ditampilkan pada Tabel 1.

Tabel 1 *Confusion Matrix*

		<i>Predicted Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Actual Class</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

### 3. Hasil dan Pembahasan

#### 3.1 Pengumpulan Data

Pada penelitian ini, digunakan data ulasan terkait hasil *Quick Count* yang diambil dari media sosial X Twitter. Data yang berhasil dikumpulkan sebanyak 2113 ulasan, dan disimpan dalam format CSV. Proses pengumpulan data dilakukan dengan metode *crawling* selama 29 menit pada tanggal 21 April 2024, dari pukul 21:11 WITA hingga pukul 21:40 WITA. Hasil pengumpulan data akan ditampilkan pada tabel 2.

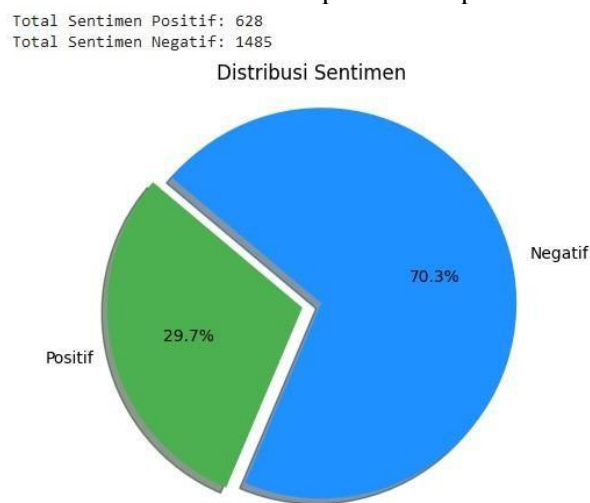
Tabel 2 Hasil Pengumpulan Data

No	created_at	full_text	...	username
1	Wed Feb 14 23:59:58 +0000 2024	@m4retha Bilas muka gosok gigi liat quick count lagi	...	Irfanalawiii
2	Wed Feb 14 23:59:49 +0000 2024	@nvtamarsaoly @convomfs Bener bgt. Ga jauh2 hasil real count nanti sama quick count. Udah nder capee ngomong sama yang ga mau kalah sipaling dicurangi dan tersakiti .	...	jisookimjichuu
...	...	...	...	...
2112	Wed Feb 14 16:29:25 +0000 2024	@tanyarfes Perasaan itu bukan pidato kemenangan deh tadi gw nonton fyi aja itu tuh emang udah direncanain kali acaranya buat ngelihat hasil quick count bareng timsesnya plis deh jangan kemakan judul berita minimal tonton sampe habis	...	ourrbee
2113			...	msmrrleia

No	created_at	full_text	...	username
	Wed Feb 14 16:29:25 +0000 2024	@m4retha Update 23.29: 02 unggul dengan suara diatas 50% pada 29 provinsi. Total quickcount di 38 provinsi + Luar Negeri: 01 Amin 32 50% 02 Pagi 51 21% 03 Gama 16 29%		

### 3.2 Labelling Data

Pelabelan data dilakukan oleh ahli bahasa (*expert*) untuk 2113 data ulasan pada media sosial X Twitter. Data ulasan pada penelitian ini akan dikategorikan ke dalam dua kelompok utama yaitu positif, dan negatif. Terdapat 1485 data dengan sentimen negatif dan 628 data dengan sentimen positif. Data negatif mendominasi dengan 70,3% dari keseluruhan data, sementara data positif berjumlah 29,7%. Berikut terdapat dua kelompok ulasan setelah diberi label diperlihatkan pada Gambar 2.



Gambar 2 Labelling Data

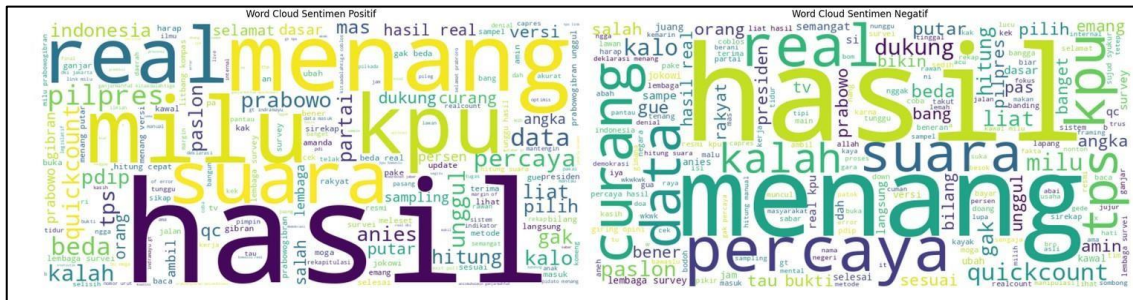
### 3.3 Pre-Processing

Gambar 3 menampilkan *wordcloud* dari teks mentah sebelum dilakukan *pre-processing*. Terlihat berbagai kata dalam bahasa Indonesia dengan frekuensi dan ukuran yang bervariasi. Kata-kata seperti "count", "quick", "hasil", "dan", serta "ini" tampak dominan. Namun, masih terdapat banyak kata-kata umum (*stopwords*) dan variasi bentuk kata yang sama, seperti "quickcount" dan "quick count". Wordcloud ini mencerminkan data teks asli dengan segala kata-kata yang tidak relevan dan ketidakseragaman penulisannya.





isu-isu yang menjadi perhatian utama selama periode *quick count*. Hasil tersebut ditampilkan pada Gambar 5.



Gambar 5 Tampilan WordCloud

### 3.5 Pembobotan Kata (TF-IDF)

Berdasarkan hasil pembobotan TF-IDF yang ditampilkan pada Gambar 6, setiap term dalam dataset telah diubah menjadi *vektor numerik*. Total kosakata yang dianalisis adalah 4999 term. Pada output (0, 838) yang menunjukkan bahwa pada dokumen ke-0 terdapat term pada *index* ke-838 dengan nilai 0,0688. Untuk output (2112, 146) menunjukkan bahwa pada dokumen ke-2112 terdapat term pada *index* ke-146 dengan nilai 0,2451. Proses ini menghasilkan matriks TF-IDF yang mencakup 2112 dokumen dan 4999 term.

Total Vocabulary: 4999

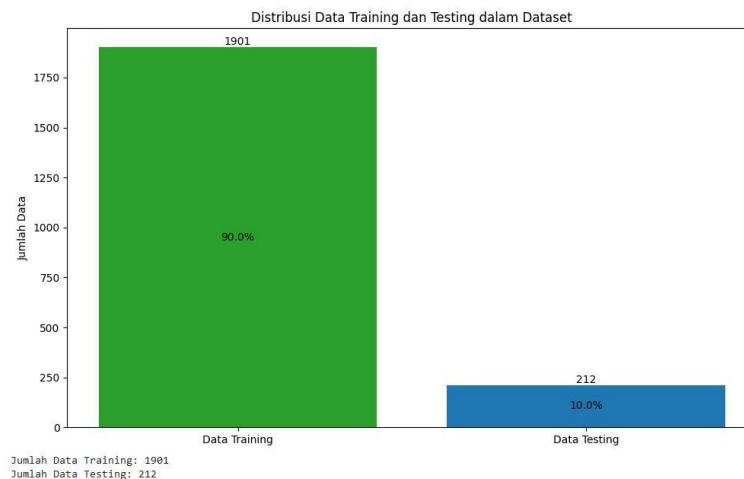
Term	Frequency-Inverse Document Frequency (TF-IDF)
(0, 838)	0,06886996842891753
(0, 3669)	0,06886996842891753
(0, 2466)	0,23852635689019155
(0, 1458)	0,48435830847651
(0, 1494)	0,48435830847651
(0, 2868)	0,4516208951228151
(0, 536)	0,5103440335708471
(1, 3861)	0,37087090191925937
(1, 880)	0,22228298442708463
(1, 4169)	0,46581517195266436
(1, 2001)	0,225945060852625
(1, 3067)	0,39180913667826514
(1, 716)	0,4908057348679893
(1, 3720)	0,1733645454710436
(1, 1583)	0,13154268294619828
(1, 449)	0,28542760295740477
(1, 830)	0,13246662346017476
(1, 3669)	0,0662331173000738
(2, 1841)	0,3231369472829371
(2, 146)	0,2701786628991291
(2, 1958)	0,29363604743878893
(2, 4029)	0,285154569615561
(2, 295)	0,486515590556142
(2, 3411)	0,19556787121704372
(2, 4828)	0,3676290869085403
(2, 1054)	0,37704700921422624
(2, 3720)	0,17184915083145802
(2, 838)	0,1313087211308153
(2, 3669)	0,06565436056515765
(2, 2466)	0,2273893219790304
(3, 1129)	0,3293850799960921
(3, 987)	0,3621001212629606
(3, 2414)	0,36264456799210665
(3, 4851)	0,30992952672523816
(3, 3461)	0,24906216927890243
(3, 350)	0,22395499746235517
(3, 1432)	0,3821001212629606
(3, 718)	0,25378564866666207
(3, 3463)	0,282321657274917
(3, 3411)	0,3071906166874691
...	...
(2112, 3670)	0,1995555639791592
(2112, 4317)	0,18476864059004822
(2112, 146)	0,2451273162975515

Gambar 6 Pembobotan TF-IDF

Matriks TF-IDF yang dihasilkan menggambarkan bobot setiap term dalam setiap dokumen. Nilai-nilai dalam matriks ini mencerminkan tingkat kepentingan suatu term dalam dokumen tertentu relatif terhadap seluruh kumpulan dokumen. Term dengan nilai TF-IDF yang lebih tinggi cenderung lebih signifikan dalam konteks dokumen tersebut. Nilai 0,2451 untuk term pada *index* 146 di dokumen 2112 menunjukkan bahwa term ini memiliki peran penting dalam dokumen tersebut. Nilai yang lebih rendah seperti 0,0688 untuk term pada *index* 838 di dokumen 0 menandakan bahwa term tersebut mungkin kurang spesifik atau kurang penting dalam konteks dokumen. Visualisasi matriks ini memungkinkan untuk melihat penyebaran bobot term di seluruh kumpulan dokumen, membantu dalam identifikasi pola dan tema utama dalam kumpulan dokumen.

### 3.6 Split Data

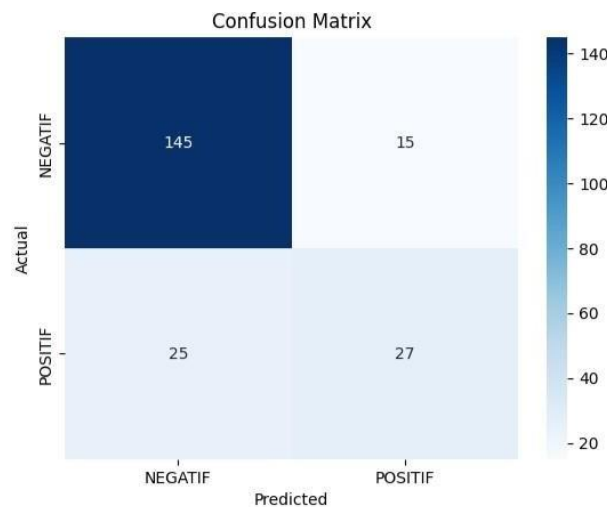
Pembagian data dilakukan dengan proporsi 90 berbanding 10. Sebanyak 1901 data, yang mewakili 90% dari total, digunakan untuk melatih model (data training). Sementara itu, 212 data sisanya, yang merupakan 10% dari keseluruhan, digunakan untuk menguji model (data testing). Pembagian data dengan rasio 90:10 akan ditampilkan pada Gambar 7.



Gambar 7 Pembagian Data

### 3.7 Klasifikasi dan Evaluasi

Hasil klasifikasi menggunakan metode *Support Vector Machine* (SVM) ditampilkan dalam *confusion matrix*. *Matrix* memperlihatkan 145 *true negative* dan 27 *true positive*, menunjukkan kemampuan model dalam mengenali kedua kelas dengan benar. Namun, terdapat 15 *false positive* dan 25 *false negative*, yang mengungkapkan adanya kesulitan dalam klasifikasi, terutama untuk kelas positif. Kinerja SVM cukup baik untuk kelas negatif, tetapi cenderung lebih berat ke kelas tersebut. Hasil *confusion matrix* ditampilkan pada Gambar 8.



Gambar 8 *Confusion Matrix SVM*

Hasil evaluasi kinerja model klasifikasi menunjukkan bahwa model memiliki akurasi sebesar 81.13%, yang berarti model ini mampu mengklasifikasikan 81.13% dari total data dengan benar. *Precision* untuk kelas "Positif" tercatat sebesar 64.29%, menunjukkan bahwa dari semua prediksi positif yang dibuat oleh model, sekitar 64.29% di antaranya benar-benar positif. Sementara itu, *recall* untuk kelas "Positif" adalah 51.92%, yang menunjukkan bahwa model ini berhasil mengidentifikasi 51.92% dari seluruh data yang sebenarnya positif. *F1-Score*, yang merupakan rata-rata keseimbangan antara *precision* dan *recall*, tercatat sebesar 57.41%, yang berarti model memiliki keseimbangan kinerja antara *precision* dan *recall*. Detail hasil evaluasi kinerja model dapat ditemukan dalam Gambar 9.

```

Accuracy: 0.8113207547169812

Classification Report:

```

	precision	recall	f1-score	support
Negatif	0.85	0.91	0.88	160
Positif	0.64	0.52	0.57	52
accuracy			0.81	212
macro avg	0.75	0.71	0.73	212
weighted avg	0.80	0.81	0.80	212

Gambar 9 Evaluasi Kinerja Model3

#### 4. Kesimpulan

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa metode *Support Vector Machine* (SVM) dengan pembobotan TF-IDF efektif dalam mengklasifikasikan sentimen masyarakat terhadap hasil *Quick Count* pemilihan Wakil Presiden Indonesia 2024 pada platform X Twitter. Model yang dikembangkan mencapai akurasi 81.13%, menunjukkan kemampuan yang baik dalam membedakan sentimen positif dan negatif. Namun, model menunjukkan kinerja yang lebih baik dalam mengidentifikasi sentimen negatif dibandingkan positif, dengan *precision* 64.29%, *recall* 51.92%, dan *F1-score* 57.41% untuk kelas positif. *F1-score* ini menggambarkan keseimbangan antara *precision* dan *recall*, memperlihatkan bahwa meski model cukup akurat secara keseluruhan, masih ada ruang untuk

peningkatan, terutama dalam identifikasi sentimen positif. Penelitian ini memberikan wawasan berharga tentang opini publik terkait *Quick Count* dan dapat menjadi dasar untuk pengembangan model analisis sentimen yang lebih canggih di masa depan.

#### Daftar Pustaka

- [1] A. Fadly, "Realisasi Tindak Pengancam Muka (Face-Threatening Act) Pada Deklarasi Kemenangan Pilpres 2019 Berdasarkan Hasil Hitung Cepat Lembaga Survei," *J. Ilm. Bhs. Sastra dan Pembelajarannya*, vol. 9, no. 2, pp. 184–193, 2022, doi: 10.26740/paramasastra.v9n2.p184-193.
- [2] S. Fachrian and A. Adriadi, "Analisis Kinerja Komisi Pemilihan Umum Dalam Meningkatkan Partisipasi Politik Masyarakat Pada Pilkada Tahun 2020 Di Kabupaten Musi Rawas Utara," *J. Ilmu Sos. dan Ilmu Polit.*, vol. 12, no. 1, pp. 1–8, 2023, doi: 10.33366/jisip.v12i1.2540.
- [3] R. I. Saraswati and A. Prasakti, "Kepercayaan Masyarakat Terhadap Hasil Quick Count Pada Pemilihan Presiden Dan Legislatif Tahun 2019," *J. Kewarganegaraan*, vol. 4, no. 1, pp. 25–30, 2020.
- [4] R. Nasywa Kirana and S. Ardyan, "Analisis Pendukung Paslon 02 di Pemilu 2024 Pada Platform Media Sosial Twitter 'X,'" *J. Ilm. Multidisiplin*, vol. 3, no. 5, pp. 28–35, 2024.
- [5] D. Ariyanti and K. Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes," *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- [6] A. R. Purnajaya, V. Lieputra, V. Tayanto, and J. G. Salim, "Implementasi Text Mining untuk Mengetahui Opini Masyarakat Tentang Climate Change," *J. Inf. Syst. Technol.*, vol. 3, no. 3, p. 36, 2022, doi: 10.37253/joint.v3i3.7337.
- [7] D. Oktavia, Y. R. Ramadahan, and M. Minarto, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [8] Sarimole Francis Matheos and Septian Wahyu, "Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *J. Sains dan Teknol.*, vol. 5, no. 3, p. 2024, 2024, [Online]. Available: <https://doi.org/10.55338/saintek.v5i1.2789>
- [9] F. Nurwanda and J. R. Rizkiani, "Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik Lifestyle," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 21, pp. 314–323, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.10077023>
- [10] M. Muadin and H. Asnal, "Implementasi Metode Support Vector Machine Pada Opinion Mining Masyarakat Terkait Chatgpt," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 7, no. 1, pp. 78–84, 2023.
- [11] R. Kurniawan R and I. Zufria, "Penerapan Text Mining Pada Sistem Penyeleksian Judul Skripsi Menggunakan Algoritma Latent Dirichlet Allocation(LDA)," *Indones. J. Comput. Sci.*, vol. 11, no. 3, pp. 1036–1052, 2022, doi: 10.33022/ijcs.v11i3.3120.



- [12] Nuzul Hikmah, Dyah Ariyanti, and Ferry Agus Pratama, "Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 2, pp. 133–148, 2022, doi: 10.35746/jtim.v4i2.225.
- [13] M. R. A. Surya, Martanto, and U. Hayati, "Analisis Sentimen Ulasan Pengguna Ovo Menggunakan Algoritma Naive Bayes Pada Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 3, pp. 2780–2786, 2024.
- [14] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [15] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 17–23, 2021, doi: 10.29207/resti.v5i1.2700.
- [16] F. Bei and S. Saepudin, "Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm)," *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [17] Ade Dwi Dayani, Yuhandri, and G. Widi Nurcahyo, "Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine," *J. KomtekInfo*, vol. 11, pp. 1–10, 2024, doi: 10.35134/komtekinfo.v11i1.439.
- [18] A. Muhaimin, M. Amin Hariyadi, and M. I. Imamudin, "Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 193–202, 2024, doi: 10.55338/jikomsi.v7i1.2865.
- [19] J. Khatib Sulaiman, M. Dzakwan Ar Rosyid, and I. Artikel Abstrak, "Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vectore Machine Universitas AMIKOM Yogyakarta," *Indones. J. Comput. Sci. Attrib.*, vol. 12, no. 5, pp. 2023–2798, 2023.