

## ANALISIS HASIL REKAPITULASI PILKADA DAERAH KHUSUS JAKARTA (DKJ) 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Miftakhul Rahman<sup>1</sup>, Mantri Kromo Fandith Fili<sup>2</sup>, Wardianto<sup>3</sup>

Fakultas Sains dan Teknologi, Universitas Nurul Huda

Email: miftakhulr@student.unuha.ac.id<sup>1</sup>, mantrifandith@student.unuha.ac.id<sup>2</sup>,

wardianto@unuha.ac.id<sup>3</sup>

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap hasil rekapitulasi Pilkada DKJ 2024 menggunakan metode *Support Vector Machine* (SVM). Data penelitian berjumlah 3.500 data yang diperoleh dari komentar Instagram menggunakan "*Turbo IG Comment Export Tool*", menghasilkan 3.100 data bersih setelah *preprocessing* yang meliputi *cleaning*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil analisis menunjukkan distribusi sentimen sebesar 37,8% positif, 32,4% netral, dan 29,8% negatif. Model SVM mencapai akurasi 84% dengan performa seimbang: sentimen positif (presisi 0,92, recall 0,74, f1-score 0,82), negatif (presisi 0,86, recall 0,87, f1-score 0,87), dan netral (presisi 0,75, recall 0,90, f1-score 0,82). Analisis *word frequency* dan *wordcloud* mengungkapkan pola kata dominan yang mencerminkan respons publik terhadap hasil Pilkada, dengan kata-kata positif seperti "alhamdulillah" dan "selamat" kontras dengan kata-kata negatif seperti "kalah". Penelitian ini memberikan wawasan berharga tentang persepsi publik terhadap hasil Pilkada DKJ 2024 dan mendemonstrasikan efektivitas metode SVM dalam analisis sentimen politik.

**Kata kunci:** Analisis Sentimen, Pilkada DKJ, Support Vector Machine.

### 1. PENDAHULUAN

Pemilihan kepala daerah (Pilkada) merupakan manifestasi demokrasi di tingkat lokal yang memiliki peran strategis dalam menentukan arah pembangunan dan kesejahteraan masyarakat. Pilkada Daerah Khusus Jakarta (DKJ) 2024 menjadi sorotan publik karena posisinya sebagai barometer politik nasional dan kompleksitas tantangan yang dihadapi [1]. Dalam era digital saat ini, media sosial telah menjadi platform utama bagi masyarakat untuk menyuarakan aspirasi, kritik, dan dukungan terhadap proses demokrasi ini. Analisis sentimen terhadap opini publik di media sosial menjadi semakin penting untuk memahami persepsi dan respons masyarakat terhadap hasil Pilkada. Tulisan yang dapat dimuat adalah yang mengkaji masalah yang berhubungan dengan teknologi dibidang mesin, elektronika dan ilmu komputer.

Dinamika politik di DKJ memiliki karakteristik unik yang mencerminkan heterogenitas sosial, ekonomi, dan budaya masyarakatnya. Sebagai pusat pemerintahan dan ekonomi nasional, kebijakan dan kepemimpinan di DKJ memiliki dampak yang signifikan tidak hanya bagi warganya tetapi juga bagi pembangunan nasional secara keseluruhan [2]. Oleh karena itu, pemahaman yang mendalam terhadap sentimen publik pasca-Pilkada menjadi crucial untuk mengidentifikasi ekspektasi masyarakat dan potensi tantangan yang mungkin dihadapi oleh kepemimpinan baru.

Media sosial, khususnya Instagram, telah menjadi platform yang sangat berpengaruh dalam membentuk opini publik dan menyebarkan informasi terkait Pilkada. Platform ini memungkinkan interaksi real-time antara masyarakat, kandidat, dan pemangku kepentingan lainnya. Namun, volume besar data yang dihasilkan dari interaksi tersebut membutuhkan pendekatan analitis yang sistematis dan objektif untuk mengekstrak insight yang bermakna. Machine learning, khususnya metode Support Vector Machine (SVM), menawarkan solusi yang efektif untuk menganalisis dan mengklasifikasikan sentimen dari data tekstual dalam skala besar.

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap hasil rekapitulasi Pilkada DKJ 2024 menggunakan metode SVM. Analisis sentimen merupakan cabang dari pembelajaran mesin dan

pemrosesan bahasa alami yang bertujuan untuk mengidentifikasi, mengekstrak, dan menganalisis pendapat, emosi, dan sikap yang terkandung dalam teks terhadap suatu entitas, peristiwa, atau topik tertentu [3]. Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang dikembangkan oleh Vladimir Vapnik, yang bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam kelas-kelas yang berbeda dengan margin maksimal [4]. Analisis ini tidak hanya fokus pada polaritas sentimen (positif, negatif, atau netral) tetapi juga berupaya mengidentifikasi tema-tema utama yang muncul dalam diskusi publik. Melalui pendekatan kuantitatif yang didukung oleh teknologi machine learning, penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang respons masyarakat terhadap hasil Pilkada dan implikasinya bagi stabilitas sosial-politik di DKJ.

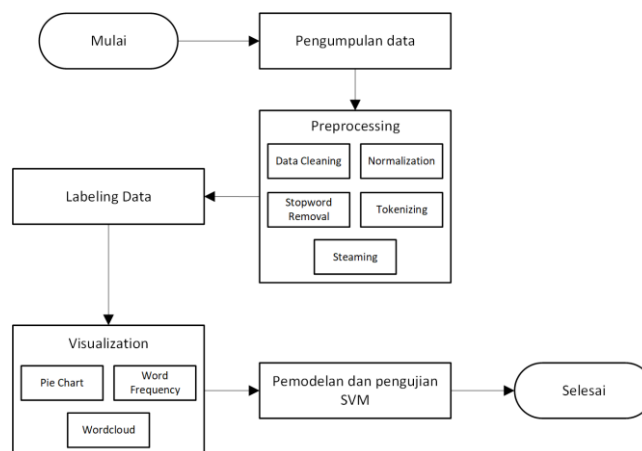
Penggunaan Instagram sebagai sumber data primer didasarkan pada popularitasnya di kalangan masyarakat Indonesia dan karakteristiknya yang memungkinkan interaksi yang lebih terbuka dan dinamis dibandingkan platform media sosial lainnya [5]. Proses preprocessing yang meliputi cleaning, normalization, tokenizing, stopword removal, dan stemming dilakukan untuk memastikan kualitas data yang optimal sebelum analisis.

Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih komprehensif tentang dinamika opini publik pasca-Pilkada DKJ 2024, yang pada gilirannya dapat berkontribusi pada penguatan demokrasi lokal dan peningkatan kualitas tata kelola pemerintahan di DKJ. Hasil analisis juga dapat menjadi basis untuk pengembangan sistem monitoring sentimen publik yang lebih sophisticated untuk keperluan pengambilan keputusan di masa mendatang.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan penelitian merupakan rangkaian langkah-langkah yang dilakukan peneliti secara sistematis dalam melaksanakan proses penelitian. Secara lebih rinci tahapan penelitian meliputi pengumpulan data, preprocessing, labeling data, visualization, serta pemodelan dan pengujian SVM. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

### 2.2 Pengumpulan Data

Pada penelitian ini pengumpulan data komentar di ambil dari platform instagram pada sebuah postingan tentang rekapitulasi hasil pilkada DJK. Data komentar di ambil pada tanggal 12 Desember 2024 dengan metode web scraping menggunakan bantuan sebuah tool yakni Turbo IG Comment Export Tool. Web scraping adalah teknik otomatis untuk mengekstrak informasi dari website melalui serangkaian proses yang melibatkan pengambilan konten HTML, parsing data,

dan penyimpanan informasi dalam format terstruktur [6]. Data yang terkumpul dan digunakan sejumlah 3500 data yang disimpan dalam format csv.

### **2.3 Preprocessing Data**

Preprocessing data merupakan tahapan kritis dalam analisis sentimen yang melibatkan serangkaian proses untuk membersihkan dan mempersiapkan teks agar dapat dianalisis secara efektif [7]. Tahapan preprocessing data akan menjadi proses yang paling penting karena akan menentukan kualitas data yang digunakan. Pada dasarnya preprocessing data diterapkan pada data mentah sebelum data digunakan untuk klasifikasi dan identifikasi. Preprocessing dalam penelitian ini meliputi tahap data cleaning, normalization, tokenizing, Stopword removal, dan stemming.

### **2.4 Labeling Data**

Labeling data merupakan tahapan dimana kita memberikan kategori atau klasifikasi terhadap suatu data berdasarkan sentimen yang terkandung di dalamnya. Dalam proses ini, setiap data akan diberi label atau tanda yang menunjukkan data tersebut termasuk ke dalam kelas sentimen tertentu [8]. Proses pelabelan data dilakukan menggunakan pendekatan berbasis leksikon (lexicon-based approach) dengan memanfaatkan kamus sentimen bahasa Indonesia yang telah dikategorikan sebelumnya. Setiap komentar dianalisis dengan menghitung skor sentimen berdasarkan kata-kata yang cocok dengan kamus sentimen, di mana setiap kata memiliki bobot sentimen tertentu. Skor total kemudian dihitung untuk menentukan label sentimen, dengan aturan jika skor total lebih besar dari 0 diberi label "positive", jika kurang dari 0 diberi label "negative", dan jika sama dengan 0 diberi label "neutral". Proses ini dilengkapi dengan validasi manual terhadap sampel hasil pelabelan untuk memastikan konsistensi dan akurasi.

### **2.5 Visualization**

Visualization data adalah proses representasi informasi secara grafis yang memungkinkan pemahaman yang lebih baik terhadap pola, tren, dan insight yang terkandung dalam data [9]. Visualisasi data dalam penelitian ini dilakukan menggunakan tiga teknik utama untuk memudahkan interpretasi hasil analisis sentimen. Pertama, pie chart digunakan untuk menampilkan distribusi persentase dari ketiga kategori sentimen. Kedua, word frequency chart dalam bentuk diagram batang menampilkan 10 kata dengan frekuensi tertinggi untuk setiap kategori sentimen. Ketiga, wordcloud dibuat untuk setiap kategori sentimen dengan dimensi 800x400 piksel dan background putih, di mana ukuran kata sebanding dengan frekuensi kemunculannya. Semua visualisasi ini diimplementasikan menggunakan library Python matplotlib dan wordcloud untuk memberikan perspektif yang komprehensif dalam memahami pola dan karakteristik sentimen dalam dataset.

### **2.6 Pemodelan dan Pengujian SVM**

Pada tahap ini data yang telah diproses diubah menjadi representasi vektor menggunakan TF-IDF. Untuk menangani ketidakseimbangan data, digunakan teknik SMOTE (Synthetic Minority Oversampling Technique). Data kemudian dibagi menjadi data latih (90%) dan data uji (10%). Model SVM dilatih menggunakan data latih dan dievaluasi menggunakan data uji.

## **3. HASIL DAN PEMBAHASAN**

### **3.1 Pengumpulan Data**

Data komentar instagram tentang rekapitulasi pilkada daerah khusus jakarta dikumpulkan dengan teknik scraping menggunakan Turbo IG Comment Export Tool, data yang diperoleh yaitu 3500 data dengan beberapa atribut yakni profile, avatar, user\_id, user\_name, comment, dan date. Selanjutnya

dilakukan seleksi, atribut yang digunakan yaitu hanya comment. Dataset setelah dilakukan seleksi data setelah di dapat dilihat pada tabel 1.

**Tabel 1. Dataset Awal**

No	Comment
1	selamat warga jakarta! 🙌🙌
2	Biasa bro pas pilpres dia milih abah anies 🙌🙌
3	Tolong pak @aniesbaswedan dibantu pantau pak. Banyak yg memilih mereka krn bapak
...	...
35	Mas Pram dan Bang Doel semoga amanah dalam menjalankan amanat baru. Oh iya semoga
00	Pembangunan Sirkuit F1 di PIK 2 jadi ya, agar bisa menjadi daya tarik dunia, semoga sukses

### 3.2 Preprocessing Data

#### 3.2.1 Cleaning

Dalam proses pembersihan data, dilakukan penghapusan karakter-karakter selain huruf, seperti menghilangkan simbol pagar (#), tautan, dan tanda baca seperti garis miring (/), titik koma (;), titik dua (:), dan titik (.). ("). Seluruh teks juga dikonversi menjadi huruf kecil, menghapus mention akun (@...), menstandarisasi kata dengan menghilangkan apostrof (') dan karakter alfanumerik yang menempel, serta menghapus spasi ganda agar diperoleh data teks yang lebih bersih untuk analisis selanjutnya setelah proses cleaning data berkurang menjadi 3100 data. Hasil pada tahap cleaning dapat dilihat pada tabel 2.

**Tabel 2. Hasil Cleaning**

No	Comment	Clean Text
1	selamat warga jakarta! 🙌🙌	selamat warga jakarta
2	Biasa bro pas pilpres dia milih abah anies 🙌🙌	Biasa bro pas pilpres dia milih abah anies
3	Tolong pak @aniesbaswedan dibantu pantau pak. Banyak yg memilih mereka krn bapak	tolong pak aniesbaswedan dibantu pantau pak banyak yg memilih mereka krn bapak
...	...	...
3100	Mas Pram dan Bang Doel semoga amanah dalam menjalankan amanat baru. Oh iya semoga Pembangunan Sirkuit F1 di PIK 2 jadi ya, agar bisa menjadi daya tarik dunia, semoga sukses	mas pram dan bang doel semoga amanah dalam menjalankan amanat baru oh iya semoga pembangunan sirkuit f1 di pik 2 jadi ya agar bisa menjadi daya tarik dunia semoga sukses

#### 3.2.2 Normalization

Dalam proses normalisasi data, dilakukan standarisasi ejaan kata-kata yang sering disingkat atau ditulis dengan ejaan tidak baku menjadi bentuk baku. Sebagai contoh, kata "gue" dinormalisasi menjadi "saya", "dgn" menjadi "dengan", "bgmn" menjadi "bagaimana", dan seterusnya. Tujuan normalisasi ini adalah untuk menyeragamkan penulisan kata agar lebih baku sehingga dapat dianalisis dengan lebih akurat. Hasil pada tahap normalization dapat dilihat pada tabel 3.

**Tabel 3. Hasil Normalization**

No	Clean Text	Normalization
1	selamat warga jakarta	selamat warga jakarta
2	Biasa bro pas pilpres dia milih abah anies	biasa bro pas pilpres dia memilih abah anies
3	tolong pak aniesbaswedan dibantu pantau pak banyak yg memilih mereka krn bapak	tolong pak dibantu pantau pak banyak yang memilih mereka karena bapak
...	...	...
3100	mas pram dan bang doel semoga amanah dalam menjalankan amanat baru oh iya semoga pembangunan sirkuit f1 di pik 2 jadi ya agar bisa menjadi daya tarik dunia semoga sukses	mas pram dan bang doel semoga amanah dalam menjalankan amanat baru oh iya semoga pembangunan sirkuit f di pik jadi ya agar bisa menjadi daya tarik dunia semoga sukses

### 3.2.3 Tokenizing

Dalam proses tokenizing, teks dipotong menjadi token berupa kata-kata individual dengan menggunakan RegexpTokenizer. Teks dipotong pada setiap kata yang terdiri dari karakter alfanumerik (r<sup>w</sup>+). Dengan tokenizing ini, teks yang awalnya berupa kalimat utuh dapat dipecah menjadi kata per kata sehingga lebih mudah dianalisis secara leksikal. Hasil tokenizing ditambahkan ke dalam kolom baru 'text\_token' pada data frame untuk digunakan pada proses analisis selanjutnya. Hasil pada tahap cleaning dapat dilihat pada tabel 4.

**Tabel 4. Hasil Tokenizing**

No	Normalization	Text Token
1	selamat warga jakarta	['selamat', 'warga', 'jakarta']
2	biasa bro pas pilpres dia memilih abah anies	['biasa', 'bro', 'pas', 'pilpres', 'dia', 'memilih', 'abah', 'anies']
3	tolong pak dibantu pantau pak banyak yang memilih mereka karena bapak	['tolong', 'pak', 'dibantu', 'pantau', 'pak', 'banyak', 'yang', 'memilih', 'mereka', 'karena', 'bapak']
...	...	...
3100	mas pram dan bang doel semoga amanah dalam menjalankan amanat baru oh iya semoga pembangunan sirkuit f di pik jadi ya agar bisa menjadi daya tarik dunia semoga sukses	['mas', 'pram', 'dan', 'bang', 'doel', 'semoga', 'amanah', 'dalam', 'menjalankan', 'amanat', 'baru', 'oh', 'iya', 'semoga', 'pembangunan', 'sirkuit', 'f', 'di', 'pik', 'jadi', 'ya', 'agar', 'bisa', 'menjadi', 'daya', 'tarik', 'dunia', 'semoga', 'sukses']

### 3.2.4 Stopword removal

Dalam proses Stopword removal, kata-kata yang termasuk Stopword dalam bahasa Indonesia dihapus dari hasil tokenisasi agar tidak terbawa ke proses analisis selanjutnya. Stopword adalah kata-kata yang sangat umum seperti 'yang', 'dan', 'di', dan lain-lain yang kurang memiliki makna penting secara leksikal. Dengan menghapus Stopword, diharapkan kata-kata yang tersisa lebih merepresentasikan substansi dan topik teks sebenarnya. Stopword bahasa Indonesia diunduh dari corpus NLTK kemudian daftar Stopword ini digunakan untuk menyaring kata hasil tokenisasi dengan menghapus kata yang termasuk Stopword. Hasilnya adalah kumpulan kata tanpa Stopword yang lebih bermakna untuk analisis. Hasil pada tahap Stopword removal dapat dilihat pada tabel 5.

**Tabel 5. Hasil Stopword Removal**

No	Text Token	Text Token
1	['selamat', 'warga', 'jakarta']	['selamat', 'warga', 'jakarta']
2	['biasa', 'bro', 'pas', 'pilpres', 'dia', 'memilih', 'abah', 'anies']	['bro', 'pas', 'pilpres', 'memilih', 'abah', 'anies']
3	['tolong', 'pak', 'dibantu', 'pantau', 'pak', 'banyak', 'yang', 'memilih', 'mereka', 'karena', 'bapak']	['tolong', 'dibantu', 'pantau', 'memilih']
...	...	...
3100	['mas', 'pram', 'dan', 'bang', 'doel', 'semoga', 'amanah', 'dalam', 'menjalankan', 'amanat', 'baru', 'oh', 'iya', 'semoga', 'pembangunan', 'sirkuit', 'f', 'di', 'pik', 'jadi', 'ya', 'agar', 'bisa', 'menjadi', 'daya', 'tarik', 'dunia', 'semoga', 'sukses']	['mas', 'pram', 'bang', 'doel', 'semoga', 'amanah', 'menjalankan', 'amanat', 'oh', 'iya', 'semoga', 'pembangunan', 'sirkuit', 'f', 'pik', 'ya', 'daya', 'tarik', 'dunia', 'semoga', 'sukses']

### 3.2.5 Steaming

Dalam proses steaming, kata-kata hasil tokenisasi diubah ke bentuk kata dasarnya menggunakan steamer bahasa Indonesia dari library Sastrawi. Proses steaming dilakukan pada setiap kata untuk mendapatkan akar katanya dengan menghilangkan imbuhan seperti awalan, akhiran, dan sisipan. Misalnya kata "bermain" di-steaming menjadi "main". Dengan steaming ini, kata-kata yang memiliki akar kata yang sama dapat dianggap sebagai kata yang sama meskipun bentuk infleksinya berbeda. Hasil steaming kemudian diubah kembali dari kumpulan token menjadi kalimat utuh agar siap digunakan pada proses analisis selanjutnya yaitu pembentukan model. Hasil pada tahap steaming dapat dilihat pada tabel 6.

**Tabel 6. Hasil Steaming**

No	Text Token	Steamed
1	['selamat', 'warga', 'jakarta']	['selamat', 'warga', 'jakarta']
2	['bro', 'pas', 'pilpres', 'memilih', 'abah', 'anies']	['bro', 'pas', 'pilpres', 'pilih', 'abah', 'anies']
3	['tolong', 'dibantu', 'pantau', 'memilih']	['tolong', 'bantu', 'pantau', 'pilih']
...	...	...
3100	['mas', 'pram', 'bang', 'doel', 'semoga', 'amanah', 'menjalankan', 'amanat', 'oh', 'iya', 'semoga', 'pembangunan', 'sirkuit', 'f', 'pik', 'ya', 'daya', 'tarik', 'dunia', 'semoga', 'sukses']	['mas', 'pram', 'bang', 'doel', 'moga', 'amanah', 'jalan', 'amanat', 'oh', 'iya', 'moga', 'bangun', 'sirkuit', 'f', 'pik', 'ya', 'daya', 'tarik', 'dunia', 'moga', 'sukses']

### 3.3 Labeling Data

Proses pemberian label dilakukan dengan menggunakan pendekatan berbasis kamus (lexicon). Metode Lexicon merupakan suatu pendekatan dalam analisis sentimen yang menggunakan basis kamus atau kumpulan kata sebagai acuannya [10]. Teks yang dianalisis diperiksa kata per kata untuk mencocokkannya dengan daftar kata positif dan negatif dalam kamus sentimen, di mana setiap kata memiliki skor tertentu. Jika kata ditemukan dalam kamus positif, skornya ditambahkan, sedangkan jika ditemukan dalam kamus negatif, skornya juga ditambahkan tapi karena nilainya negatif maka akan berkurang. Total skor dihitung untuk menentukan polaritas sentimen yakni "positive" jika skor > 0, "negative" jika skor < 0, dan "neutral" jika skor = 0. Dengan demikian, sentimen teks ditentukan berdasarkan akumulasi skor dari kata-kata yang dikenali dalam kamus. Proses analisis ini memungkinkan setiap teks mendapatkan label berdasarkan akumulasi skor dari kata-kata yang teridentifikasi dalam kamus sentimen. Distribusi hasil pelabelan data yang dihasilkan dari pendekatan ini dirangkum dalam tabel 7.

**Tabel 7. Distribusi Label Sentimen**

No	Polarity	Jumlah Data
1	Positive	1172
2	Negative	1004
3	Neutral	924

Setelah proses pelabelan selesai, data yang telah terklasifikasi dianalisis lebih lanjut untuk mendapatkan gambaran yang lebih mendalam. Hasil analisis distribusi sentimen dari data yang telah dilabeli secara rinci dapat dilihat pada tabel 3 berikut.

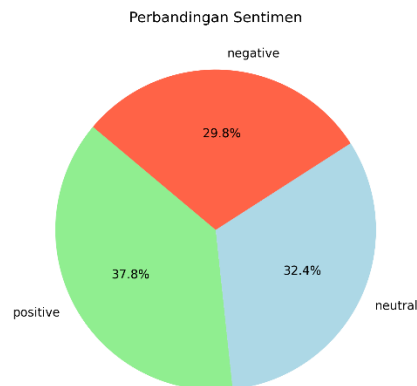
**Tabel 8. Hasil Distribusi Sentimen**

No	Steamed	Polarity Score	Polarity
1	['selamat', 'warga', 'jakarta']	5	Positive
2	['bro', 'pas', 'pilpres', 'pilih', 'abah', 'anies']	4	Positive
3	['tolong', 'bantu', 'pantau', 'pilih']	0	Neutral
...			
3100	['mas', 'pram', 'bang', 'doel', 'moga', 'amanah', 'jalan', 'amanat', 'oh', 'iya', 'moga', 'bangun', 'sirkuit', 'f', 'pik', 'ya', 'daya', 'tarik', 'dunia', 'moga', 'sukses']	16	Positive

### 3.4 Visualization

#### 3.4.1 Pie chart

Dalam visualisasi diagram pie sentimen, data label sentimen dihitung jumlah kemunculannya untuk masing-masing kategori positif, positifve, negatif menggunakan fungsi `value_counts()`. Dari data ulasan yang ada, terdapat 1172 ulasan dengan sentimen positif, 1004 ulasan dengan sentimen neutral, dan 924 ulasan dengan sentimen negative. Persentase ulasan yang telah diberi label sebagai sentimen positif, negatif, dan netral dapat dilihat pada Gambar 2.

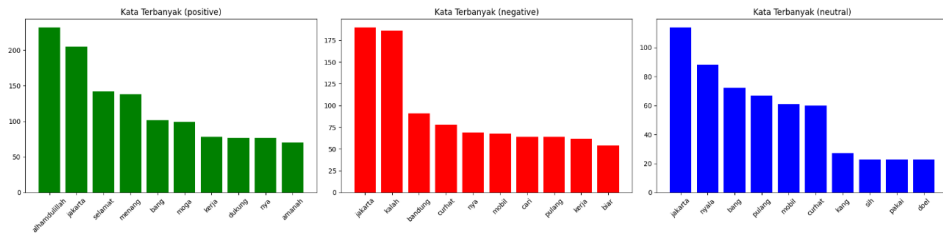


**Gambar 2. Pie Chart Sentimen**

Berdasarkan analisis pada Gambar 2, dapat disimpulkan bahwa persentase data tertinggi berada pada sentimen sangat positive, yaitu mencapai 37,8%. Untuk sentimen dengan tingkat menengah, persentase terbanyak ada pada sentimen neutral sebesar 32,4%, sedangkan sentimen negative memiliki persentase paling rendah yaitu 29,8%.

#### 3.4.2 Word Frequency

Visualisasi frekuensi kata dilakukan untuk menganalisis kata-kata yang paling sering muncul pada setiap kategori sentimen (positif, negatif, dan netral). Teks yang telah diproses dipisahkan berdasarkan polaritas sentimen. Untuk setiap kategori, kata-kata dalam teks dihitung frekuensinya menggunakan metode berbasis penghitung (counter). Frekuensi kata dihitung dengan mengidentifikasi kata-kata yang muncul di dalam teks, kemudian menghitung jumlah kemunculannya. Selanjutnya, 10 kata yang paling sering muncul dari masing-masing kategori dipilih dan divisualisasikan dalam bentuk diagram batang. Diagram ini memberikan gambaran tentang distribusi kata dominan pada setiap sentimen. Warna batang digunakan untuk membedakan polaritas sentimen, yaitu hijau untuk positif, merah untuk negatif, dan biru untuk netral. Pada Gambar 3, diagram batang menunjukkan kata-kata yang paling sering digunakan dalam teks positive, Negative dan Neutral.



Gambar 3. Word Frequency

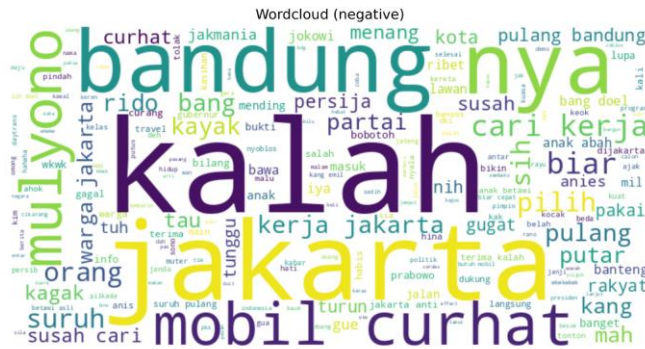
### 3.4.3 Wordcloud

Wordcloud digunakan untuk memvisualisasikan kata-kata yang sering muncul dalam data teks berdasarkan sentimen positif, negatif, dan netral. Proses pembentukan wordcloud diawali dengan menggabungkan semua teks dalam kategori sentimen tertentu menjadi satu string panjang. Teks yang dihasilkan kemudian digunakan untuk menghasilkan wordcloud menggunakan library wordcloud. Wordcloud dibuat dengan parameter dimensi lebar 800 piksel dan tinggi 400 piksel, serta latar belakang berwarna putih. Kata-kata yang lebih sering muncul dalam teks ditampilkan dengan ukuran huruf yang lebih besar, sehingga mempermudah identifikasi kata-kata dominan pada setiap kategori sentimen. Untuk setiap kategori sentimen (positif, negatif, dan netral), wordcloud divisualisasikan secara terpisah untuk memberikan perbandingan yang jelas.

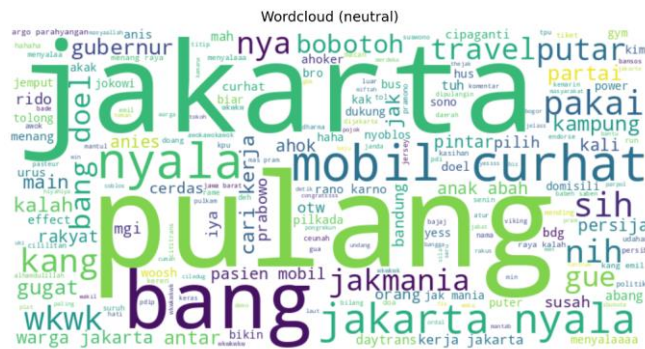
Pada Gambar 4, visualisasi wordcloud untuk sentimen positive menunjukkan kata-kata yang mendominasi teks dengan sentimen positive. Gambar 5 memperlihatkan wordcloud untuk sentimen negative, sementara Gambar 6 menampilkan wordcloud untuk sentimen netral. Visualisasi ini memberikan wawasan mendalam mengenai distribusi dan pola kata yang mendasari sentimen dalam data teks.



Gambar 4. Wordcloud Positive



Gambar 5. Wordcloud Negative



Gambar 6. Wordcloud Neutral

Berdasarkan gambar 4, 5 dan 6, kata-kata seperti "alhamdulillah", "menang", "dukung", "selamat", dan "mantap" banyak muncul pada komentar positive. Sedangkan kata-kata seperti "kalah", "jakarta", "mulyono", dan "curhat" sering muncul pada komentar negative. Sementara itu, kata-kata seperti "pulang", "nyala", dan "jakarta" banyak ditemukan pada komentar Neutral.

### 3.5 Pemodelan dan Pengujian SVM

Pada tahap analisis, data teks komentar diubah menjadi representasi numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Representasi TF-IDF membantu menyoroti kata-kata yang signifikan dalam teks berdasarkan frekuensi kemunculannya, dengan mempertimbangkan pengaruh dari kata-kata yang sering muncul di seluruh dokumen [11]. Karena data sentimen memiliki distribusi yang tidak seimbang, teknik oversampling dilakukan menggunakan metode SMOTE (Synthetic Minority Oversampling Technique). Metode ini bertujuan untuk meningkatkan jumlah data pada kelas minoritas agar lebih proporsional dengan kelas mayoritas. Setelah proses ini, data dibagi menjadi data latih dan data uji menggunakan fungsi `train_test_split` dengan rasio 90:10.

Model Support Vector Machine (SVM) digunakan untuk melatih data latih dan memprediksi data uji. SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan hyperplane yang optimal untuk klasifikasi. Evaluasi performa model dilakukan dengan menggunakan metrik accuracy, precision, recall, dan f1-score yang dirangkum dalam classification report. Hasil evaluasi performa model SVM dapat dilihat pada Gambar 7.

	precision	recall	f1-score	support
negative	0.86	0.87	0.87	111
neutral	0.75	0.90	0.82	115
positive	0.92	0.74	0.82	126
accuracy			0.84	352
macro avg	0.84	0.84	0.84	352
weighted avg	0.85	0.84	0.84	352

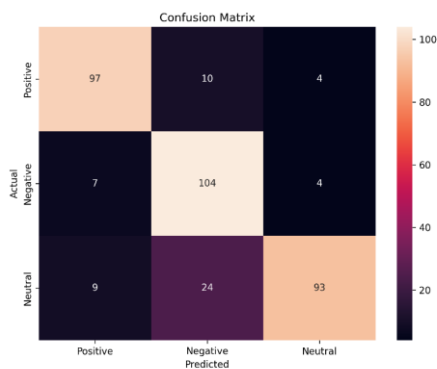
**Gambar 7. Classification Report**

Hasil klasifikasi pada gambar 7 menunjukkan performa yang baik dengan tingkat akurasi keseluruhan mencapai 84%. Model berhasil mengklasifikasikan data ke dalam tiga kategori sentimen: positif, negatif, dan netral. Berdasarkan hasil classification report, model menunjukkan performa yang bervariasi untuk setiap kelas sentimen. Untuk sentimen positif, model mencapai presisi yang tinggi sebesar 0,92, namun dengan recall yang lebih rendah yaitu 0,74, menghasilkan f1-score 0,82. Hal ini mengindikasikan bahwa model sangat selektif dalam mengklasifikasikan sentimen positif, dengan tingkat false positive yang rendah namun cenderung melewatkan beberapa kasus positif yang seharusnya teridentifikasi.

Pada klasifikasi sentimen negatif, model menunjukkan keseimbangan yang baik antara presisi (0,86) dan recall (0,87), menghasilkan f1-score tertinggi di antara ketiga kelas yaitu 0,87. Sementara untuk sentimen netral, model mencapai recall yang tinggi sebesar 0,90 dengan presisi 0,75, menghasilkan f1-score 0,82. Distribusi data dalam dataset cukup seimbang, dengan jumlah sampel untuk setiap kelas berkisar antara 111 hingga 126 data (total 352 sampel). Performa model secara keseluruhan konsisten, ditunjukkan dengan nilai macro average dan weighted average yang serupa (0,84), mengindikasikan bahwa model dapat menangani ketiga kelas dengan baik tanpa bias yang signifikan terhadap kelas tertentu.

Hasil evaluasi menunjukkan bahwa model SVM yang diimplementasikan mampu mengklasifikasikan sentimen dengan tingkat kehandalan yang memadai. Model menunjukkan kekuatan dalam meminimalkan false positive untuk sentimen positif, meskipun hal ini mengakibatkan beberapa missed detection. Keseimbangan performa yang ditunjukkan oleh nilai macro dan weighted average yang identik (0,84) mengkonfirmasi bahwa model ini dapat diandalkan untuk analisis sentimen pada dataset yang dianalisis.

Selanjutnya Confusion matrix digunakan untuk mengevaluasi performa model klasifikasi dengan memberikan gambaran tentang prediksi yang benar dan salah untuk setiap kelas sentimen. Sumbu horizontal merepresentasikan prediksi model, sedangkan sumbu vertikal menunjukkan kelas aktual. Setiap elemen dalam matriks menunjukkan jumlah data yang diklasifikasikan berdasarkan kombinasi prediksi dan kelas sebenarnya. Elemen pada diagonal utama (dari kiri atas ke kanan bawah) menunjukkan prediksi yang benar untuk setiap kelas sentimen, yaitu positif, negatif, dan netral. Nilai tinggi pada diagonal utama mengindikasikan bahwa model mampu mengklasifikasikan data dengan akurasi yang baik. Sebaliknya, elemen di luar diagonal utama menunjukkan kesalahan klasifikasi, yaitu data yang diklasifikasikan ke dalam kelas yang salah. Misalnya, beberapa data sentimen netral diklasifikasikan sebagai negatif, atau data positif diklasifikasikan sebagai netral. Hasil visualisasi Confusion Matrix dapat dilihat pada gambar 8.



**Gambar 8. Confusion Matrix**

Analisis confusion matrix menunjukkan performa detail model SVM dalam mengklasifikasikan setiap kelas sentimen. Dari total 111 data sentimen positif aktual, model berhasil mengklasifikasikan dengan benar 97 data sebagai positif, sementara 10 data salah diklasifikasikan sebagai negatif dan 4 data sebagai netral. Hal ini menunjukkan kemampuan model yang kuat dalam mengenali sentimen positif dengan tingkat kesalahan klasifikasi yang relatif rendah.

Untuk kelas sentimen negatif, dari 115 data aktual, model berhasil mengidentifikasi dengan tepat 104 data sebagai negatif, dengan hanya 7 data yang salah diklasifikasikan sebagai positif dan 4 data sebagai netral. Performa ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali sentimen negatif, dengan tingkat akurasi klasifikasi yang tinggi. Namun, perlu dicatat bahwa model juga cenderung over-classified beberapa data sebagai negatif, terlihat dari 10 data positif dan 24 data netral yang salah diklasifikasikan sebagai negatif.

Pada klasifikasi sentimen netral, dari 126 data aktual, model berhasil mengidentifikasi 93 data dengan benar sebagai netral, sementara 9 data salah diklasifikasikan sebagai positif dan 24 data sebagai negatif. Meskipun tingkat akurasi untuk kelas netral cukup baik, jumlah misklasifikasi yang relatif tinggi ke kelas negatif (24 kasus) mengindikasikan adanya tantangan bagi model dalam membedakan sentimen netral dari sentimen negatif. Hal ini mungkin disebabkan oleh kesamaan karakteristik atau fitur antara kedua kelas tersebut dalam dataset yang digunakan. Secara umum performa model SVM dalam klasifikasi sentimen teks sudah cukup baik ditinjau dari nilai akurasi, presisi, recall dan F1-score untuk semua kelas. Model ini layak digunakan pada kasus nyata untuk prediksi sentimen teks yang baru. Namun demikian, performa masih dapat ditingkatkan lebih lanjut dengan penyesuaian hiperparameter dan algoritma lainnya.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen terhadap hasil rekapitulasi Pilkada DKJ 2024 menggunakan metode Support Vector Machine (SVM), dapat disimpulkan bahwa dari total 3.500 komentar Instagram yang dianalisis, setelah melalui tahapan preprocessing yang meliputi cleaning, normalization, tokenizing, stopword removal, dan stemming, diperoleh 3.100 data bersih dengan distribusi sentimen 37,8% positif, 32,4% netral, dan 29,8% negatif. Model SVM yang diimplementasikan menunjukkan performa yang baik dengan akurasi keseluruhan mencapai 84%, dimana sentimen positif mencapai presisi 0,92 dan recall 0,74 (f1-score 0,82), sentimen negatif dengan presisi 0,86 dan recall 0,87 (f1-score 0,87), serta sentimen netral dengan presisi 0,75 dan recall 0,90 (f1-score 0,82). Analisis word frequency dan wordcloud mengungkapkan bahwa sentimen positif didominasi kata-kata seperti "alhamdulillah", "menang", "dukung", dan "selamat", sementara sentimen negatif sering mengandung kata-kata seperti "kalah" dan "mulyono", sedangkan sentimen netral ditandai dengan kata-kata seperti "pulang" dan "nyala", yang mencerminkan dinamika opini publik terhadap hasil Pilkada DKJ 2024.

#### DAFTAR PUSTAKA

- [1] L. Romli, "Daerah Khusus Jakarta dan Pilkada 2024: Masihkah Menjadi Barometer Politik Nasional?," *J. Demokr.*, vol. 01, pp. 1–88, 2024, [Online]. Available: <https://bakesbangpol.jakarta.go.id/v2/file/download/6bb2e067d9361d1ae68ada4f74f29779.pdf>
- [2] M. S. Anam, "Analisis Framing Terhadap Branding Politik Anies Baswedan Sebagai Gubernur Daerah Khusus Ibu Kota (DKI) Jakarta," *Polit. Islam*, vol. 3, no. 1, pp. 75–94, Jun. 2024, doi: 10.31958/pi.v3i1.12507.
- [3] G. A. Ramadhan, "Analisis sentimen ulasan aplikasi ruangguru dengan algoritma long short term memory," Universitas Islam Negeri Syarif Hidayatullah jakarta, 2023. [Online]. Available: [https://repository.uinjkt.ac.id/dspace/handle/123456789/71510/1?https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71510/1/GILANG\\_AMBANG\\_RAMADHAN-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/71510/1?https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71510/1/GILANG_AMBANG_RAMADHAN-FST.pdf)
- [4] F. Ramadhan, "Analisis Sentimen Tweet Pelayanan Dompot Digital Menggunakan Algoritma Support Vector Machine Dan Random Forest," Universitas Islam Negeri Syarif

- Hidayatullah jakarta, 2022. [Online]. Available:  
<https://repository.uinjkt.ac.id/dspace/handle/123456789/76619>
- [5] J. Sarangih, M. Purba, M. Manik, N. D. Aulia, W. Wulandari, and O. A. Sihaloha, "Peran Influencer Instagram Dalam Membentuk Opini Publik Dan Partisipasi Politik," *J. Educ. Gov. Wiyata*, vol. 2, no. 4, pp. 396–406, 2024, doi: <https://doi.org/10.71128/e-gov.v2i4.157>.
- [6] T. M. Fahrudin, P. A. Riyantoko, and K. M. Hindrayani, "Implementation of Web Scraping on Google Search Engine for Text Collection Into Structured 2D List," *Telemat. J. Inform. dan Teknol. Inf.*, vol. 20, no. 2, p. 139, Jun. 2023, doi: [10.31315/telematika.v20i2.9575](https://doi.org/10.31315/telematika.v20i2.9575).
- [7] R. Saputra and F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 411–419, Jul. 2024, doi: [10.47233/jteksis.v6i3.1378](https://doi.org/10.47233/jteksis.v6i3.1378).
- [8] Muhammad Fernanda Naufal Fathoni, Eva Yulia Puspaningrum, and Andreas Nugroho Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem J. Inform. dan Sains Teknol.*, vol. 2, no. 3, pp. 62–76, Jul. 2024, doi: [10.62951/modem.v2i3.112](https://doi.org/10.62951/modem.v2i3.112).
- [9] H. Raxy Sragih and A. Ayu Eilasihtta, "Implementasi Tableau Dalam Memvisualisasikan Data Untuk Optimalisasi Keputusan Keuangan Dalam Praktik Akuntansi," *J. Akuntansi dan Keuang. Methodist*, vol. 8, no. 2, pp. 140–152, 2024, doi: <https://doi.org/10.46880/jsika.Vol8No2.pp140-152>.
- [10] F. S. Yerzi and Y. Sibaroni, "Analisis Sentimen terhadap Kebijakan Pemerintah dalam Menangani Covid-19 dengan Pendekatan Lexicon Based," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 11354–11366, 2021, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/15616/15329>
- [11] S. Bila Rahmania Sharafi, "Identifikasi Pola Diskusi Publik mengenai Pemindahan Ibu Kota Negara Menggunakan Analisis TF-IDF dan K-Means Clustering," *Semin. Nas. Sist. Inf.*, vol. 08, pp. 4654–4663, 2024, [Online]. Available: <https://jurnalfti.unmer.ac.id/index.php/senasif/article/view/557>