

## Perbandingan *Random Forest* dan SVM dalam Analisis Sentimen *Quick Count* Pemilu 2024

Ika Septiana<sup>1</sup>, Debby Alita<sup>1</sup>

<sup>1</sup> Progam Studi Sistem Informasi, Universitas Teknokrat Indonesia,

<sup>1</sup> Jln. ZA. Pagar Alam No.9-11, Labuhan Ratu, Kec. Kedaton, Kota Bandar Lampung, 35132, Indonesia

### Info Artikel

#### Riwayat Artikel:

Received 2024-03-27

Revised 2024-11-28

Accepted 2024-11-28

**Abstract** – The implementation of the 2024 elections is regulated in the General Election Commission Regulation (PKPU) Number 3 of 2022, which also stipulates the election schedule and stages. After the simultaneous general elections that took place on February 14, 2024, problems arose among the public regarding the Quick Count results, especially for the Presidential election. The Quick Count results themselves generated various opinions, both positive and negative. In the post-election Twitter page, there are many conversations in cyberspace related to the Quick Count results on Twitter. Thus, sentiment analysis can be used to classify tweets and comments about the 2024 election quick count results into three categories, namely positive, negative, and neutral. Thus, this analysis is expected to provide some significant benefits related to the quick count results in the 2024 election. Random Forest and Support Vector Machine are two machine learning techniques used to measure how accurate the resulting sentiment analysis is. From the results of the research that has been carried out, there are 2000 data collected during February 2024. After preprocessing and labeling, there are 1,116 positive class data, 730 negative class data and 154 neutral class data. From the results of the comparison of the algorithms evaluated, the accuracy value of the two algorithms was obtained. The Random Forest algorithm produces an accuracy of 78%, while the SVM algorithm produces an accuracy of 80%. This shows that in sentiment analysis on the 2024 election quick count, the SVM method obtained a greater accuracy value compared to Random Forest.

**Keywords:** Machine Learning; Quick Count; Random Forest, Support Vector Machine, Sentiment Analysis.

#### Corresponding Author:

Debby Alita

Email: [Debbyalita@teknokrat.ac.id](mailto:Debbyalita@teknokrat.ac.id)



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

**Abstrak** – Pelaksanaan Pemilu 2024 diatur pada Peraturan Komisi Pemilihan Umum (PKPU) Nomor 3 Tahun 2022, yang juga menetapkan jadwal dan tahapan pemilu. Setelah pemilihan umum (pemilu) serentak yang berlangsung pada tanggal 14 Februari 2024, permasalahan muncul di kalangan masyarakat terkait hasil Quick Count terutama untuk pemilihan Presiden. Hasil Quick Count sendiri timbul berbagai opini baik yang menanggapi secara positif maupun negatif. Dalam laman Twitter pasca-pemilu, banyak perbincangan di dunia maya terkait hasil Quick Count di Twitter. sehingga, analisis sentimen dapat dipergunakan untuk klasifikasi tweet dan komentar tentang hasil quick count pemilu 2024 ke dalam tiga kategori yaitu positif, negatif, dan netral. Dengan demikian, analisis ini diharapkan dapat memberikan beberapa manfaat yang signifikan terkait hasil hitung cepat pada pemilu 2024. Random Forest dan Support Vector Machine adalah dua teknik machine learning yang dilakukan untuk mengukur seberapa akurat analisis sentimen yang dihasilkan. Dari hasil penelitian yang telah dilakukan, terdapat 2000 data yang dikumpulkan selama bulan Februari 2024. Setelah melakukan preprocessing serta pelabelan dihasilkan 1.116 data kelas positif, 730 data kelas negatif dan 154 data kelas netral. Dari hasil komparasi algoritma yang dievaluasi, diperoleh nilai akurasi dari kedua algoritma tersebut. Untuk algoritma Random Forest menghasilkan akurasi sebesar 78%, sedangkan untuk algoritma SVM menghasilkan akurasi sebesar 80%. Hal ini menunjukkan bahwa dalam analisis sentimen pada penghitungan cepat pemilu 2024, metode SVM memperoleh nilai akurasi yang lebih besar dibandingkan dengan Random Forest.

**Kata Kunci:** Analisis Sentimen, Machine Learning, Quick Count, Random Forest, Support Vector Machine.

## I. PENDAHULUAN

Setiap individu dalam negara demokrasi mempunyai hak yang sama untuk dapat berpartisipasi dalam mengambil keputusan yang berdampak pada kehidupan sehari-hari. Dalam demokrasi, warga negara memiliki kesempatan untuk terlibat dalam pembuatan, pengembangan, dan implementasi kebijakan, baik secara langsung maupun melalui perwakilan[1]. Indonesia menjunjung tinggi sistem demokrasi sebagai landasan negaranya. Pelaksanaan Pemilihan umum (Pemilu) secara berkala adalah salah satu aspek penting pada sistem demokrasi, yang memungkinkan warga untuk secara langsung mengekspresikan preferensi dan aspirasi politik mereka[2]. Wujud dari penerapan demokrasi Indonesia yaitu Pemilihan Umum (Pemilu). Melalui Pemilu, proses demokrasi terbuka bagi masyarakat untuk memilih wakil rakyat serta pejabat publik lainnya sesuai dengan aspirasi dan kehendak mereka[3]. Tanggal 14 Februari 2024, Indonesia telah melaksanakan Pemilihan umum (Pemilu) serentak untuk memilih Presiden, DPR, DPRD, dan DPD.

Peraturan Komisi Pemilihan Umum (PKPU) Nomor 3 Tahun 2022 yang mengatur tentang Tahapan serta Jadwal Pemilu 2024 mengatur pelaksanaan Pemilu 2024. Setelah pemilihan umum (pemilu) serentak dilakukan, yaitu pada

tanggal 14 Februari 2024, Permasalahan muncul di kalangan masyarakat terkait hasil Quick Count terutama untuk pemilihan Presiden. Hasil Quick Count sendiri timbul berbagai opini baik yang menanggapi secara positif maupun negatif. Metode Quick Count adalah cara cepat untuk menghitung hasil pemilu. Komisi Pemilihan Umum (KPU) mendefinisikan Quick Count sebagai proses penghitungan suara secara cepat dengan bantuan teknologi informasi[4].

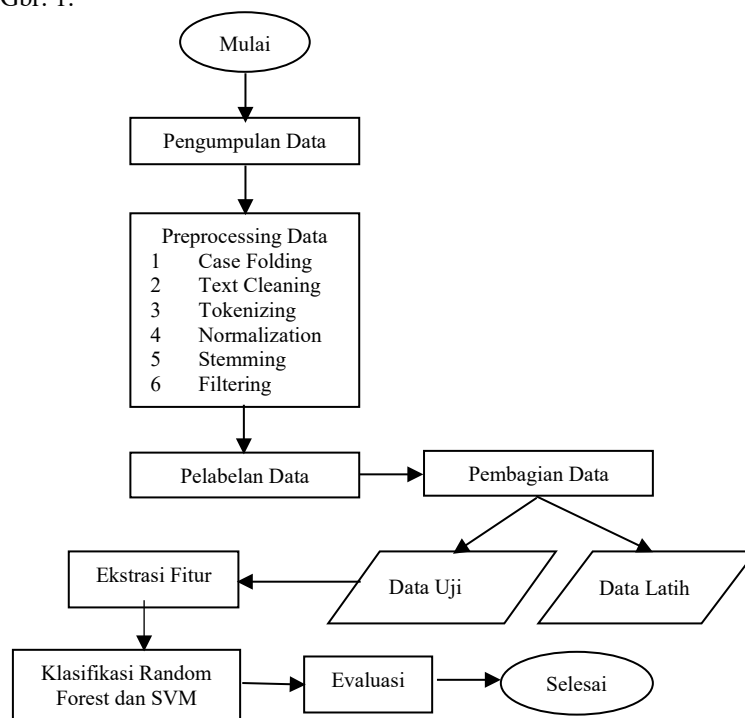
Pada saat ini, Media sosial adalah platform yang umum digunakan oleh orang-orang untuk menyuarakan pendapat mereka. Twitter merupakan salah satu platform yang sering dipergunakan dalam berbagai sudut pandang[5]. Pada bulan Januari 2022 Indonesia mencatatkan sebanyak 191 juta pengguna aktif media sosial selanjutnya meningkat dari tahun sebelumnya sebesar 12,35%. Twitter, menjadi salah satu platform terkemuka, mempunyai 18,45 juta pengguna di Indonesia pada tahun 2022[6]. Twitter menjadi tempat di mana pengguna dapat dengan bebas menyampaikan pendapat pribadi maupun publik mengenai berbagai topik. Pengguna Twitter memiliki kemampuan untuk membagikan opini mereka pada fitur yang telah disediakan dan tulisan yang mereka unggah bisa diakses oleh pengguna Twitter lainnya[5].

Setelah pemilihan umum berlangsung, laman Twitter menjadi tempat banyaknya perbincangan di dunia maya mengenai hasil Quick Count. Opini-opini yang muncul di Twitter memiliki nilai penting untuk diklasifikasikan ke dalam sentimen positif, negatif, maupun netral[7]. Maka dari itu, peneliti mengumpulkan data-data mengenai opini masyarakat untuk dianalisis. Dari data-data yang peneliti kumpulkan akan digunakan untuk melihat sentimen masyarakat terhadap keputusan Quick Count khususnya pada Pemilu 2024. Untuk mengolah opini pada setiap tweet, peneliti menggunakan model Machine Learning bernama Random Forest Classifier dan Support Vector Machine (SVM).

Tujuan penelitian ini yaitu membandingkan metode Support Vector Machine (SVM) dan Random Forest Classifier untuk menilai opini publik terkait hasil Quick Count Pemilu 2024. Tidak seperti penelitian lain yang hanya menerapkan algoritma Naive Bayes, Penelitian ini memperluas pengetahuan dengan membandingkan efektivitas kinerja antara Random Forest dan SVM dalam melakukan analisis sentiment terhadap hasil Quick Count Pemilu 2024. Selain itu, penelitian ini diharapkan dapat memberikan perspektif baru yang penting bagi para pemangku kepentingan tentang bagaimana masyarakat bereaksi dan merasakan hasil pemilu yang diumumkan melalui Quick Count.

## II. METODE

Salah satu pendekatan pembelajaran mesin yang paling populer adalah analisis sentimen. Untuk menafsirkan data teks dan menghasilkan informasi dari data tersebut, komputasi linguistik, analisis teks, dan pemrosesan bahasa alami digunakan[8]. Pada penelitian ini, sentimen dikelompokkan kedalam tiga kategori, yaitu positif, negatif, serta netral. Penganalisaan sentimen dilaksanakan dengan menerapkan dua algoritma Machine Learning, yaitu Random Forest dan Support Vector Machine. Dua algoritma ini dikatakan efisien dalam hal penganalisaan sentimen. Alur proses penelitian digambarkan dapat dilihat di Gbr. 1.



Gambar 1. Tahapan Penelitian

### A. Pengumpulan Data

Peneliti menggunakan metode crawling dalam melakukan proses pengumpulan data. Teknik crawling data adalah metode otomatis untuk mengambil informasi dari sebuah situs web dengan bantuan program komputer. Dengan pendekatan ini, pengumpulan data dari berbagai sumber menjadi lebih efisien dan terstruktur, maka bisa dimanfaatkan untuk berbagai tujuan seperti penelitian, analisis, dan pengembangan aplikasi[9]. Penelitian ini penulis mengumpulkan data dari media sosial twitter dengan mempergunakan keyword tentang Quick Count Pemilihan Umum (Pemilu) 2024. Data yang dikumpulkan pada periode Februari 2023 dan berhasil mengumpulkan 2000 data. Data tersebut disimpan dalam format CSV sebelum dipakai untuk analisis sentimen. Google Colab digunakan pada proses crawling data dengan bahasa pemrograman Python dan library pandas sebagai dasar perintah saat penambangan datanya. Analisis sentiment yang dilakukan ini memakai variabel full text yang terdapat dalam data Quick Count Pemilihan Umum (Pemilu) 2024[6]. Berikut adalah dataset yang dipergunakan dalam pemodelan, dapat dilihat pada tabel I.

TABEL 1  
 SAMPEL DATASET AWAL

Username	Full text
belokchain	@BurhanMuhtadi Apakah bisa jadi sampling dalam quick count bisa saja salah terhadap PSI sebagai partai baru. Secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara PSI signifikan sehingga MOE jadi kisaran mendekati 2%. Atau bisa jadi sebaliknya.
Simanjunta 9Nico	Anies Akui Quick Count Ibarat Tes Darah di Lab, tapi Pasien Harus Puasa 12 Jam Sebelumnya bukan Dicekoki Berbagai Makanan. Hasilnya diragukan bahkan tdk dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang. <a href="https://t.co/08tvNAa70X">https://t.co/08tvNAa70X</a> @choymarkochoy

### B. Preprocessing Data

Tahapan preprocessing merupakan tahapan yang dilakukan sebelum proses klasifikasi yaitu untuk mengelola teks pada data tweet (Negara, Muhardi and Putri, 2020)[10]. Pada proses penelitian ini data yang digunakan yaitu tweet dari platform twitter dimana terdapat 2000 data berbentuk teks yang sifatnya tak terstruktur, itulah sebabnya membutuhkan tahap preprocessing data. Urutan tahapan preprocessing data yang dilakukan yaitu Text Cleaning, Case Folding, Tokenizing, Normalization, Stemming, dan Filtering. tahapan-tahapan ini bertujuan untuk membersihkan, mengubah format, dan mengelompokkan kata-kata sehingga data menjadi lebih siap untuk diproses dan dianalisis[11].

### C. Case Folding

Case folding dilakukan dengan maksud merubah teks atau kalimat yang awalnya menggunakan huruf kapital menjadi huruf kecil. Tujuannya untuk memudahkan proses pemrosesan data yang lebih lanjut serta menjaga konsistensi pada analisis teks. Sampel hasil case folding dapat dilihat pada tabel 2.

TABEL 2  
 HASIL CASE FOLDING

Full text	text case folding
@BurhanMuhtadi Apakah bisa jadi sampling dalam quick count bisa saja salah terhadap PSI sebagai partai baru. Secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara PSI signifikan sehingga MOE jadi kisaran mendekati 2%. Atau bisa jadi sebaliknya.	@BurhanMuhtadi apakah bisa jadi sampling dalam quick count bisa saja salah terhadap psi sebagai partai baru. secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara psi signifikan sehingga moe jadi kisaran mendekati 2%. atau bisa jadi sebaliknya.
Anies Akui Quick Count Ibarat Tes Darah di Lab, tapi Pasien Harus Puasa 12 Jam Sebelumnya bukan Dicekoki Berbagai Makanan. Hasilnya diragukan bahkan tdk	anies akui quick count ibarat tes darah di lab, tapi pasien harus puasa 12 jam sebelumnya bukan dicekoki berbagai makanan. hasilnya diragukan bahkan tdk

dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang. <a href="https://t.co/08tvNAa70X">https://t.co/08tvNAa70X</a> @choymarkochoy	dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang. <a href="https://t.co/08tvNAa70X">https://t.co/08tvNAa70X</a> @choymarkochoy
---	---

#### D. Text Cleaning

Dalam tahap cleaning, karakter yang dianggap tidak penting akan dihilangkan. Hal terdiri dari tanda baca, simbol, emotikon, URL, hashtag, karakter khusus, dan angka yang terdapat dalam teks. Proses pembersihan ini bertujuan untuk membersihkan teks dari elemen-elemen yang tak relevan sehingga meningkatkan kualitas analisis data. Sampel hasil text cleaning dapat dilihat pada tabel 3.

TABEL 3  
 SAMPEL HASIL TEXT CLEANING

text case folding	text cleaning
@BurhanMuhtadi apakah bisa jadi sampling dalam quick count bisa saja salah terhadap psi sebagai partai baru. secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara psi signifikan sehingga moe jadi kisaran mendekati 2%. atau bisa jadi sebaliknya.	apakah bisa jadi sampling dalam quick count bisa saja salah terhadap psi sebagai partai baru secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara psi signifikan sehingga moe jadi kisaran mendekati 2% atau bisa jadi sebaliknya
anies akui quick count ibarat tes darah di lab, tapi pasien harus puasa 12 jam sebelumnya bukan dicekoki berbagai makanan. hasilnya diragukan bahkan tdk dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang. <a href="https://t.co/08tvNAa70X">https://t.co/08tvNAa70X</a> @choymarkochoy	anies akui quick count ibarat tes darah di lab tapi pasien harus puasa 12 jam sebelumnya bukan dicekoki berbagai makanan hasilnya diragukan bahkan tdk dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang

#### E. Tokenizing

Tokenizing adalah tindakan memisahkan teks atau kalimat menjadi sebuah kata-kata individual. Dengan dilakukannya tokenizing, kita dapat lebih mudah menghitung frekuensi kemunculan setiap kata pada teks tersebut. Ini memungkinkan analisis yang lebih efektif terhadap teks. Sampel hasil tokenizing dapat dilihat pada tabel 4.

TABEL 4  
 SAMPEL HASIL TOKENIZING

text cleaning	text tokenizing
apakah bisa jadi sampling dalam quick count bisa saja salah terhadap psi sebagai partai baru secara mungkin saja kebenaran di tps yg tidak diambil sampling justru suara psi signifikan sehingga moe jadi kisaran mendekati 2% atau bisa jadi sebaliknya	apakah,bisa,jadi,sampling,dalam ,quick,count,bisa,saja,salah,terhadap,psi,sebagai,partai,baru,secara,mungkin,saja,kebenaran,di,tps,yg,tidak,diambil,sampling,justru ,suara,psi,signifikan,sehingga,moe,jadi,kisaran,mendekati,2%,atau,bisa,jadi,sebaliknya.
anies akui quick count ibarat tes darah di lab tapi pasien harus puasa 12 jam sebelumnya bukan dicekoki berbagai makanan hasilnya diragukan bahkan tdk dipercaya kalau rakyat sebelum pemilihan diintimidasi dan diserang dgn politik uang	anies,akui,quick,count,ibarat,tes,darah,di,lab,,tapi,pasien,harus,puasa,12,jam,sebelumnya,bukan,dicekoki,berbagai,makanan,,hasilnya,diragukan,bahkan,tdk,dipercaya,kalau,rakyat,sebelum,pemilihan,diintimidasi,dan,diserang,dgn,politik,uang,.

#### F. Normalization

Pada proses ini kata yang tidak baku dinormalisasi dengan mengubahnya menjadi bentuk kata baku atau standar. Kata-kata tidak standar ini bisa berupa singkatan, kata-kata slang, atau kata-kata yang memiliki variasi ejaan. Dengan

normalization, kita dapat memastikan konsistensi dalam analisis teks dan mengurangi ambiguitas dalam pemahaman makna kata. Sampel hasil normalization dapat dilihat pada tabel 5.

TABEL 5  
 SAMPEL HASIL NORMALIZATION

text tokenizing	text normalization
apakah,bisa,jadi,sampling,dalam,q uick,count,bisa,saja,salah,terhadap ,psi,sebagai,partai,baru,secara,mun gkin,saja,kebenaran,di,tps,yg,tidak ,diambil,sampling,justru,suara,psi, signifikan,sehingga,moe,jadi,kisar an,mendekati,2,%,atau,bisa,jadi,se baliknya.	apakah,bisa,jadi,sampling,dalam ,quick,count,bisa,saja,salah,terha dap,psi,sebagai,partai,baru,secar a,mungkin,saja,kebenaran,di,tps, yg,tidak,diambil,sampling,justru ,suara,psi,signifikan,sehingga,m oe,jadi,kisaran,mendekati,2,%,at au,bisa,jadi,sebaliknya.
anies,aku,quick,count,ibarat,tes,d arah,di,lab,,tapi,pasien,harus,puas a,12,jam,sebelumnya,bukan,dicek oki,berbagai,makanan,,hasilnya,di ragukan,bahkan,tdk,dipercaya,kala u,rakyat,sebelum,pemilihan,diinti midasi,dan,diserang,dgn,politik,ua ng,.	anies,aku,quick,count,ibarat,tes, darah,di,lab,,tapi,pasien,harus,p uasa,12,jam,sebelumnya,bukan, dicekoki,berbagai,makanan,,has ilnya,diragukan,bahkan,tdk,dipe rcaya,kalau,rakyat,sebelum,pemi lihan,diintimidasi,dan,diserang,d gn,politik,uang,.

### G. Stemming

Stemming merupakan proses dalam pengolahan teks yang digunakan dalam penghapusan imbuhan pada sebuah kata maka diperoleh kata dasar. Variasi kata dengan akar kata yang sama dapat direpresentasikan sebagai satu kesatuan dengan menggunakan stemming, sehingga mempermudah analisis teks dan pengambilan keputusan. Sampel hasil stemming dapat dilihat pada tabel 6.

TABEL 6  
 SAMPEL HASIL STEMMING

text normalization	text stemming
apakah,bisa,jadi,sampling,dalam,q uick,count,bisa,saja,salah,terhadap ,psi,sebagai,partai,baru,secara,mun gkin,saja,kebenaran,di,tps,yg,tidak ,diambil,sampling,justru,suara,psi, signifikan,sehingga,moe,jadi,kisar an,mendekati,2,%,atau,bisa,jadi,se baliknya.	apakah,bisa,jadi,sampling,dalam ,quick,count,bisa,saja,salah,hada p,psi,bagai,partai,baru,cara,mun gkin,saja,benar,di,tps,yg,tidak,a mbil,sampling,justru,suara,psi,si gnifikan,sehingga,moe,jadi,kisar ,dekat,2,atau,bisa,jadi,balik,
anies,aku,quick,count,ibarat,tes,d arah,di,lab,,tapi,pasien,harus,puas a,12,jam,sebelumnya,bukan,dicek oki,berbagai,makanan,,hasilnya,di ragukan,bahkan,tdk,dipercaya,kala u,rakyat,sebelum,pemilihan,diinti midasi,dan,diserang,dgn,politik,ua ng,.	anies,aku,quick,count,ibarat,tes, darah,di,lab,,tapi,pasien,harus,p uasa,12,jam,belum,bukan,cekok, bagai,makan,,hasil,ragu,bahkan,t dk,percaya,kalau,rakyat,belum,p ilih,intimidasi,dan,serang,dgn,po litik,uang,

### H. Filtering

Pada proses filtering, kita menentukan apakah sebuah kata akan dipakai atau dibuang dalam analisis teks. Sebelum melakukan proses penghapusan stopword, daftar stoplist akan disusun terlebih dahulu. Sebuah kata akan dihilangkan dari teks jika kata tersebut ada dalam daftar stoplist. Akibatnya, istilah yang masih ada setelah prosedur ini dianggap mendefinisikan substansi dokumen. Sampel hasil filtering dapat dilihat pada tabel 7.

TABEL 7  
 SAMPEL HASIL FILTERING

text stemming	text filtering
apakah,bisa,jadi,sampling,dalam,q uick,count,bisa,saja,salah,hadap,ps i,bagai,partai,baru,cara,mungkin,s aja,benar,di,tps,yg,tidak,ambil,sam pling,justru,suara,psi,signifikan,se	sampling,quick,count,salah,hada p,psi,partai,tps,ambil,sampling,s uara,psi,signifikan,moe,kisar,2,

hingga,moe,jadi,kisar,dekat,2,atau, bisa,jadi,balik,	anies,aku,quick,count,ibarat,tes,da rah,di,lab,,tapi,pasien,harus,puasa, 12,jam,belum,bukan,cekok,bagai, makan,,hasil,ragu,bahkan,tdk,perc aya,kalau,rakyat,belum,pilih,intim idasi,dan,serang,dgn,politik,uang,
---	---

I. Pelabelan Data

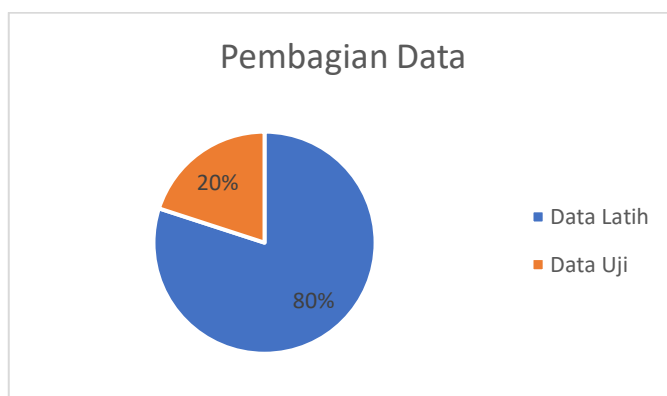
Setelah preprocessing data dilakukan, langkah selanjutnya adalah melakukan pelabelan pada setiap data tweet yang berbasis leksikon[12]. Pelabelan adalah proses mengklasifikasikan data yang telah dikumpulkan untuk menetapkan apakah tiap kalimat pada dataset itu mempunyai makna yang positif, negatif atau netral[13]. Label positif menampilkan hasil tweet tersebut sifatnya setuju, mendukung, atau menerima terkait hasil Quick Count. Sementara itu label negatif menampilkan hasil tweet tersebut sifatnya menyangkal, menolak atau membantah terkait hasil Quick Count[12]. Adapun sampel hasil pelabelan data quick count pada penelitian ini dapat dilihat pada tabel 8.

TABEL 8  
 SAMPEL HASIL PELABELAN DATA

text filtering	polarity score	sentimen
sampling,quick,count,salah,hadap,psi,part ai,,tps,ambil,sampling,suara,psi,signifikan, moe,kisar,2,,,	-8	Negative
atur,mudah,1,putar,syarat,20,,20,provinsi,t p,fix,1,putar,lihat,kepala,negara2,selamat, prabowo,bahasa,inggris,asli,selamat,quick ,count,,bgmn,tanggap,	17	Positif
id,kasih,deklarasi,percaya,hasil,quick,co unt,juluk,presiden,quick,count,,malu,sema put,	0	Neutral

J. Pembagian data

Tweet yang sudah diberikan label lalu membangnya menjadi dua kategori yaitu data uji dan data latih selama langkah pembagian data. Sebelum pembagian data dilakukan, terlebih dahulu melakukan proses label encode untuk variabel target, yaitu "Indonesia Sentimen". Diikuti dengan menggunakan nilai random state sebesar 42, 80% dari dataset terdiri dari data latih dan 20% terdiri dari data uji. Hasilnya, setiap dataset memiliki 1.600 untuk data latih dan 400 data untuk data uji dari total dataset. Pada bagian ini, sebagian besar data akan dipergunakan untuk melatih model, sementara sebagian data lainnya akan dipergunakan untuk memverifikasi keakuratan model sebelum digunakan[12].



Gambar 2. Diagram Pembagian Data

K. Ekstraksi fitur

Untuk meningkatkan akurasi dan efektivitas dalam proses klasifikasi dalam penelitian ini, seleksi fitur dilakukan dengan tujuan mengurangi dimensi dari kumpulan teks dengan menghilangkan kata-kata yang dinilai tidak penting atau tidak relevan dengan isi dokumen. Metode ekstraksi fitur yang dilakukan yaitu Count-Vectorizer. Cara kerja Count-Vectorizer adalah menghitung jumlah frekuensi kemunculan kata (fitur) pada dokumen kemudian direpresentasikan dalam bentuk vector[14]. Pada penelian ini menghasilkan 1.116 data untuk kelas positif, 730 data untuk kelas negatif,

dan 154 data untuk kelas netral membentuk distribusi kelas hasil ekstraksi fitur. Adapun hasil ekstraksi fitur quick count pada penelitian ini dapat dilihat pada tabel 9.

TABEL 9  
SAMPEL HASIL EKSTRAKSI FITUR

Label	Jumlah Data
Positif	1.116
Negative	730
Netral	154

#### L. Klasifikasi

Klasifikasi dilakukan dengan menggunakan pendekatan algoritma Random Forest dan Support Vector Machine (SVM) memanfaatkan data training dan mengujinya pada data testing untuk menetapkan polaritas sentiment. Tahapan ini memakai bahasa pemrograman python dengan library sklearn[15].

Random Forest adalah metode yang efektif dalam meningkatkan akurasi dengan cara membangun banyak pohon keputusan secara acak. Setiap pohon keputusan dalam Random Forest menghasilkan atribut untuk setiap node secara acak. sekumpulan pohon keputusan ini kemudian dipergunakan untuk mengklasifikasikan data ke dalam kelas yang sesuai. Dengan menggunakan Random Forest, kita dapat mengambil keputusan berdasarkan mayoritas suara dari semua pohon keputusan yang terbentuk, sehingga menghasilkan model yang lebih stabil dan memiliki kinerja yang lebih baik dalam klasifikasi data[16]. Pemilihan kelas pada Random Forest yang paling populer[17], dapat menggunakan persamaan berikut.

$$f(x) = \text{Average}(f_1(x), f_2(x), \dots, f_n(x)) \quad (1)$$

Keterangan :

- $f(x)$  : Hasil prediksi
- $f_1(x), f_2(x), \dots, f_n(x)$  : Hasil prediksi dari setiap pohon keputusan
- $x$  : Inputan data

Structural Risk Minimization (SRM) adalah fondasi dari metode pembelajaran mesin yang terkenal sebagai Support Vector Machine (SVM). Metode Support Vector Machine (SVM) termasuk dalam kategori supervised learning, dataset yang akan dianalisis diberikan label terlebih dahulu untuk melatih model[18]. Memperoleh hyperplane yang optimal untuk membagi dua kelas dalam ruang input adalah tujuan utamanya. SVM telah menunjukkan kinerja yang kuat dalam klasifikasi data, terutama dalam pekerjaan yang melibatkan kumpulan data yang sangat besar untuk kategorisasi teks[19]. Penyelesaian klasifikasi SVM[20], dapat menggunakan persamaan berikut.

$$(w \cdot xi) + b = 0 \quad (2)$$

Data xi yang termasuk pada kelas -1 dapat direpresentasikan dengan persamaan (3).

$$(w \cdot xi + b) \leq yi = -1 \quad (3)$$

Untuk data xi yang termasuk pada kelas +1, rumusnya dapat dilihat pada persamaan (4).

$$(w \cdot xi + b) \geq yi = 1 \quad (4)$$

Keterangan :

- $w \cdot xi$  : Data ke - i
- $xi$  : Nilai bobot untuk kelas data ke - i
- $b$  : Nilai bias
- $yi$  : Kelas data ke - i

#### M. Evaluasi

Penelitian ini menggunakan metode evaluasi confusion matrix dengan menggunakan data testing. Confusion matrix adalah alat yang sederhana dan berguna untuk mengevaluasi keefektifan algoritma kategorisasi. Tujuannya adalah untuk mengevaluasi kinerja atau akurasi sistem klasifikasi dalam mengkategorikan data uji[[21]. Terdapat tiga kelas sentimen yang digunakan untuk penelitian ini yaitu kelas positif, kelas negatif dan kelas netral. Tujuan menggunakan Confusioan matrix adalah untuk memperhitungkan nilai accuracy, precision, recall dan f1-score[22].

### III. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan opini atau komentar dalam media sosial twitter dengan menggunakan kata kunci quick count dalam bahasa indonesia sebanyak 2000 data. Dalam penelitian ini data dibagi menjadi 80:20 yaitu, 80% data training dan 20% data testing. Data tersebut digunakan sebagai data machine learning dan data uji untuk mengevaluasi kinerja sistem. Adapun Langkah langkah untuk melakukan analisis sentimen sebagai berikut.

#### A. Random Forest

Hasil evaluasi model Random Forest pada data tweet mengenai quick count pemilu 2024 periode Februari 2024 menunjukkan hasil yang menonjol. Presentasi precision tertinggi, mencapai 79%, terdapat pada kelas positif, menandakan bahwa sebagian besar dari hasil yang diprediksi benar-benar relevan. Sementara itu, presentasi recall mencapai 88% untuk kelas positif, menunjukkan bahwa model mampu mengidentifikasi sebagian besar dari semua sampel yang sebenarnya adalah kelas positif. F1-Score, sebagai rata-rata harmonis precision dan recall, juga mencapai nilai tertinggi pada kelas positif, yaitu 84%. Dari segi akurasi keseluruhan, model Random Forest mencapai 78%. Keseluruhan, model ini memberikan kinerja yang baik dalam mengklasifikasikan tweet terkait quick count pemilu 2024 pada bulan Februari 2024, terutama dalam mengidentifikasi tweet yang relevan atau mendukung. Gbr. 5 menunjukkan hasil evaluasi model Random Forest.

Classification Report Random Forest Model:				
	precision	recall	f1-score	support
Negative	0.78	0.72	0.75	150
Neutral	0.67	0.39	0.49	31
Positive	0.79	0.88	0.84	219
accuracy			0.78	400
macro avg	0.75	0.66	0.69	400
weighted avg	0.78	0.78	0.78	400

Gambar 3. Hasil Evaluasi Model Random Forest

#### B. Support Vector Machine

Hasil evaluasi model Support Vector Machine (SVM) pada data tweet mengenai quick count pemilu 2024 periode Februari 2024 menunjukkan beberapa metrik evaluasi yang penting. Precision tertinggi, mencapai 81%, terdapat pada kelas negatif, menunjukkan bahwa sebagian besar tweet yang diprediksi sebagai negatif oleh model SVM memang benar-benar negatif. Sementara itu, recall tertinggi, sebesar 93%, terdapat pada kelas positif, menandakan bahwa model dapat mengidentifikasi sebagian besar tweet yang sebenarnya positif. F1-score, yang menggabungkan precision dan recall, mencapai nilai tertinggi, yaitu 86%, untuk kelas positif, menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall dalam mengenali tweet yang positif. Akurasi keseluruhan dari model SVM adalah 80%, menandakan bahwa sebagian besar tweet telah diklasifikasikan dengan benar oleh model. Gbr. 6 menampilkan hasil dari evaluasi model SVM.

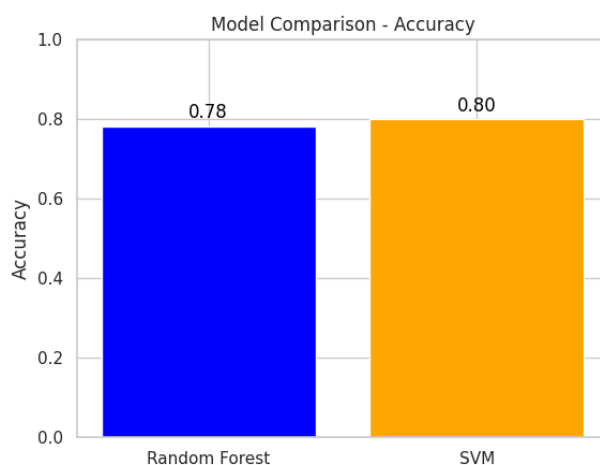
Classification Report SVM Model:				
	precision	recall	f1-score	support
Negative	0.81	0.77	0.79	150
Neutral	0.33	0.03	0.06	31
Positive	0.80	0.93	0.86	219
accuracy			0.80	400
macro avg	0.65	0.58	0.57	400
weighted avg	0.77	0.80	0.77	400

Gambar 4. Hasil Evaluasi Model SVM

#### C. Visualisasi Perbandingan Model

Untuk penelitian ini, visualisasi perbandingan model dilakukan untuk menunjukkan perbandingan akurasi antara model Random Forest dan SVM pada data tweet mengenai quick count pemilu 2024 periode Februari 2024. Nilai performa untuk akurasi dengan pembagian data 80% data training dan 20% data testing dengan model algoritma Support Vector Machine memiliki tingkat akurasi yang lebih tinggi dengan nilai akurasi sebesar 80% dengan selisih 2%

jika dibandingkan dengan model algoritma Random Forest dengan nilai yang lebih rendah hanya sebesar 78% pada konfigurasi dataset sebanyak 2000 data. Dengan demikian dapat ditarik kerimpulan model algoritma Support Vector Machine pada penelitian ini memiliki tingkat performa yang lebih baik dari pada model Random Forest. Gbr. 7 menunjukkan perbandingan akurasi model.



Gambar 5. Visualisasi Perbandingan Akurasi Model

#### IV. SIMPULAN

Pengumpulan data tweet mengenai quick count pemilu 2024 pada periode Februari 2024 memakai metode crawling berhasil menghasilkan 2.000 data. Setelah melakukan proses preprocessing dan pelabelan, diperoleh 1.116 data untuk kelas positif, 730 data kelas negatif, dan kelas netral terdapat 154 data. Berdasarkan penelitian diatas terdapat beberapa kesimpulan, Adapun kesimpulannya sebagai berikut yaitu Data tweet mengenai quick count pemilu 2024 pada proses crawling data menggunakan Google Colab dengan bahasa pemrograman Python dan library pandas sebagai dasar perintah saat penambahan datanya. Algoritma Random Forrest dan support vector machine telah berhasil mengklasifikasikan data tweet mengenai quick count pemilu 2024.

Hasil perhitungan evaluasi yang telah dilakukan dengan confusion matrix bahwa tingkat akurasi yang didapat hasil klasifikasi untuk algoritma Random Forrest memperoleh akurasi sebesar 78% sedangkan algoritma Support Vector Machine memperoleh akurasi 80%. Dapat disimpulkan bahwa algoritma support vector machine lebih unggul dalam mengklasifikasikan tweet yang berkaitan dengan quick count pemilu 2024 dibandingkan algoritma Random Forrest terdapat perbedaan sebesar 2% dengan algoritma Support Vector Machine.

Adapun saran untuk penelitian selanjutnya, metode klasifikasi yang telah digunakan dapat diujikan pada dataset yang berbeda guna membandingkan nilai akurasi dari dua algoritma yang diterapkan. Disamping itu, pendekatan Deep Learning juga dapat dieksplorasi untuk memproses data, agar nilai akurasinya dapat dibandingkan dengan teknik yang digunakan sebelumnya.

#### DAFTAR PUSTAKA

- [1] Id.wikipedia.org, "Demokrasi," id.wikipedia.org. Accessed: Feb. 25, 2024. [Online]. Available: <https://id.wikipedia.org/wiki/Demokrasi>
- [2] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.
- [3] A. E. Subiyanto, "Pemilihan Umum Serentak yang Berintegritas sebagai Pembaruan Demokrasi Indonesia," *J. Konstitusi*, vol. 17, no. 2, p. 355, Aug. 2020, doi: 10.31078/jk1726.
- [4] D. Savitri, "Metode dan Cara Kerja Quick Count Pilpres 2024, Ternyata Pakai Langkah Ilmiah!," detikEdu. Accessed: Feb. 25, 2024. [Online]. Available: <https://www.detik.com/edu/detikpedia/d-7194893/metode-dan-cara-kerja-quick-count-pilpres-2024-ternyata-pakai-langkah-ilmiah%0A>
- [5] O. Manullang and C. Prianto, "Analisis Sentimen dalam Memprediksi Hasil Pemilu Presiden dan Wakil Presiden : Systematic Literature Review," *J. J-COM (Jurnal Inform. dan Teknol. Komputer)*, vol. 04, pp. 104–113, 2023, [Online]. Available: <https://ejournalunsam.id/index.php/jicom/>
- [6] P. Elisa and A. Rahman Isnain, "COMPARISON OF RANDOM FOREST, SUPPORT VECTOR MACHINE AND NAIVE BAYES ALGORITHMS TO ANALYZE SENTIMENT TOWARDS MENTAL HEALTH STIGMA," *J. Tek. Inform.*, vol. 5, no. 1, pp. 321–329, 2024, doi: 10.52436/1.jutif.2024.5.1.1817.
- [7] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, 2019.
- [8] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," *J. Tek. Inform.*, vol. 8, no. 3, 2023.
- [9] S. Algifari Rismawan, Y. Syahidin, P. Piksi Ganesha Bandung, J. Gatot Subroto No, and K. Batununggal, "Implementasi Website Berita

- Online Menggunakan Metode Crawling Data Dengan Bahasa Pemrograman Python,” *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 3, pp. 167–178, 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [10] D. Alita, Y. Fernando, and H. Sulistiani, “Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter,” *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.
- [11] A. Nofandi, N. Y. Setiawan, and D. W. Brata, “Analisis Sentimen Ulasan Pelanggan dengan Metode Support Vector Machine (SVM) untuk Peningkatan Kualitas Layanan pada Restoran Warung Wareg,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 458–466, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [12] A. Ilham and W. Pramusinto, “ANALISIS SENTIMEN MASYARAKAT TERHADAP KESEHATAN MENTAL PADA TWITTER MENGGUNAKAN ALGORITME K-NEAREST NEIGHBOR,” *Indones. J. Appl. Stat.*, vol. 2, no. 2, 2023.
- [13] V. Fitriyana *et al.*, “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine,” 2023.
- [14] R. Vincent, I. Maulana, and O. Komarudin, “PERBANDINGAN KLASIFIKASI NAIVE BAYES DAN SUPPORT VECTOR MACHINE DALAM ANALISIS SENTIMEN DENGAN MULTICLASS DI TWITTER,” 2023.
- [15] M. Fachriza and Munawar, “Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest,” *J. Tek. Univ. Muhammadiyah Ponorogo*, pp. 49–58, 2023, [Online]. Available: <http://studentjournal.umpo.ac.id/index.php/komputek>
- [16] D. Alita and A. R. Isnain, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *J. Komputasi*, vol. 8, no. 2, pp. 50–58, 2020, doi: 10.23960/komputasi.v8i2.2615.
- [17] Intan Permata and Esther Sorta Mauli Nababan, “Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit,” *J. Ris. RUMPUN Mat. DAN ILMU Pengetah. ALAM*, vol. 2, no. 2, pp. 65–71, Jul. 2023, doi: 10.55606/jurrimipa.v2i2.1336.
- [18] V. D. Yunanda and N. Hendrastuty, “Perbandingan Kernel Polynomial dan RBF Pada Algoritma SVM Untuk Analisis Sentimen Skincare di Indonesia,” vol. 8, no. April, pp. 726–735, 2024, doi: 10.30865/mib.v8i2.7425.
- [19] N. Fathirachman Mahing *et al.*, “KLASIFIKASI TINGKAT STRES DARI DATA BERBENTUK TEKS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN RANDOM FOREST,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 7, pp. 1527–1536, 2023, doi: 10.25126/jtiik2023108010.
- [20] A. Saepudin, A. Faqih, and G. Dwilestari, “Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee,” *J. TEKNO KOMPAK*, vol. 18, no. 1, pp. 178–192, 2024.
- [21] R. Nurhidayat and K. E. Dewi, “PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN FITUR EKSTRAKSI N-GRAM DALAM ANALISIS SENTIMEN BERBASIS ASPEK,” *KOMPUTA J. Ilm. Komput. dan Inform.*, vol. 12, no. 1, pp. 91–100, 2023, [Online]. Available: <https://www.kaggle.com/datasets/hafidhmusthaanah/skincare-review?select=00.+Review.csv>.
- [22] A. C. Khotimah and E. Utami, “Comparison Naive Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine in the classification of individual on twitter account,” *J. Tek. Inform.*, vol. 3, no. 3, pp. 673–680, 2022, [Online]. Available: <https://doi.org/10.20884/1.jutif.2022.3.3.254>