

Analisis Sentimen Masyarakat Terhadap Hasil Quick Count Pilpres 2024 Pada Media Sosial Instagram Menggunakan Metode Support Vector Machine (Svm)

Yohanes Don Bosco Setiawan¹⁾, Kusno Harianto²⁾, Amelia Yusnita³⁾

Program Studi Sistem Informasi, STMIK Widya Cipta Dharma
Jl. Prof. M. Yamin No 25, Samarinda 75123
E-mail: 2041030@wicida.ac.id¹⁾, kusno@wicida.ac.id²⁾, amelia@wicida.ac.id³⁾

ABSTRAK

Penelitian ini menganalisis sentimen masyarakat terhadap hasil quick count Pemilihan Presiden 2024 di Instagram menggunakan metode Support Vector Machine (SVM). Instagram, sebagai platform media sosial, menyediakan wawasan berharga tentang reaksi publik terhadap isu politik. Data komentar Instagram dikumpulkan dan dianalisis untuk mengidentifikasi sentimen positif, negatif, dan netral. Proses analisis meliputi pengumpulan data, pre-processing, pelatihan model SVM, serta evaluasi menggunakan confusion matrix dan visualisasi. Hasil penelitian menunjukkan bahwa metode SVM mencapai akurasi 81%, presisi 79%, dan recall 80%. Temuan ini diharapkan dapat berkontribusi pada pengembangan metode analisis sentimen yang lebih efektif dan memberikan dasar untuk penelitian lebih lanjut tentang dinamika opini publik dalam konteks politik modern.

Kata Kunci: Analisis Sentimen, Quick Count, Support Vector Machine (SVM)

ANALYSIS OF COMMUNITY SENTIMENT TOWARDS THE RESULTS OF THE 2024 QUICK COUNT PRILPRESS ON INSTAGRAM SOCIAL MEDIA USING THE SUPPORT VECTOR MACHINE (SVM) METHOD

ABSTRACT

This study aims to analyze public sentiment towards the 2024 Presidential Election quick count on Instagram using the Support Vector Machine (SVM) method. Social media, particularly Instagram, serves as a crucial platform for observing public reactions to political issues. Instagram comments data was collected and analyzed to identify positive, negative, and neutral sentiments. The analysis process included data collection, pre-processing, SVM model training, and evaluation using confusion matrix and visualization. The results indicate that the SVM method achieved an accuracy of 81%, precision of 79%, and recall of 80%. This research is expected to contribute to the development of more sophisticated and effective sentiment analysis methods and provide a foundation for further research in understanding public opinion dynamics in the increasingly complex modern political context.

Keywords: *Sentiment Analysis, Quick Count, 2024, Support Vector Machine (SVM)*

1. PENDAHULUAN

Demokrasi, sebagai suatu bentuk pemerintahan, memberikan setiap warga negara kesempatan yang sama untuk berpartisipasi dalam proses pengambilan keputusan yang berpotensi mengubah arah hidup mereka. Dalam konteks Indonesia, negara ini mengadopsi sistem demokrasi yang memungkinkan partisipasi aktif warga negara baik secara langsung maupun melalui perwakilan dalam penyusunan dan pembentukan hukum. Pemilihan umum yang diadakan secara berkala merupakan salah satu indikator penting dari berjalannya sistem demokrasi di Indonesia. Pada tahun 2024, Indonesia menyelenggarakan pemilihan umum serentak yang mencakup pemilihan Presiden, anggota DPR, DPRD, dan

DPD. Pasca dilaksanakannya pemilihan umum serentak pada tanggal 14 Februari 2024, muncul masalah di tengah-tengah masyarakat terkait hasil quick count, terutama untuk pemilihan presiden. Timbul berbagai opini terhadap hasil quick count baik menanggapi hasil quick count secara positif maupun negatif. Media sosial Instagram, dengan pengguna yang terus berkembang, menjadi salah satu sumber utama untuk mengamati reaksi masyarakat terhadap berbagai isu, termasuk politik. Dalam konteks Pilpres 2024, Instagram menjadi platform yang penting untuk menganalisis sentimen masyarakat. Dalam laman Instagram pasca-pemilu, banyak perbincangan di dunia maya terkait hasil quick count di media sosial. Opini-opini yang ada di laman

Instagram menarik untuk diklasifikasikan ke dalam sentimen positif ataupun sentimen negatif. Klasifikasi sebuah opini merupakan bagian dari data mining khususnya text mining atau penambangan teks. Dalam konteks ini, penelitian tentang analisis sentimen masyarakat terhadap hasil quick count pilpres 2024 di Instagram memiliki relevansi yang besar. Menggunakan metode Support Vector Machine (SVM), penelitian ini bertujuan tidak hanya untuk memahami sentimen yang terkandung dalam berbagai unggahan di Instagram, tetapi juga untuk menawarkan solusi konkret bagi berbagai pihak terkait. Metode Support Vector Machine (SVM) dipilih karena keunggulannya dalam mengatasi kompleksitas analisis sentimen pada data teks yang besar dan beragam yang sering ditemukan dalam platform media sosial seperti Instagram. Penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentimen yang lebih canggih dan efektif, serta menjadi landasan bagi penelitian lebih lanjut dalam memahami dinamika opini publik dalam konteks politik modern yang semakin kompleks.

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

2.1 Cakupan permasalahan

Cakupan permasalahan berdasarkan latar belakang pada penelitian ini adalah “Bagaimana menentukan sentimen masyarakat dari komentar pada instagram dengan menggunakan algoritma *Support Vector Machine*?”

2.2 Batasan-batasan penelitian

Agar tidak meluasnya permasalahan maka dibuatlah batasan masalah sebagai berikut :

1. Tidak menangani data yang salah pengetikan.
2. Hanya menggunakan data yang berbahasa Indonesia.
3. Data diambil dari media sosial instagram.

2.3 Rencana hasil yang didapatkan

Penelitian ini bertujuan untuk menganalisis sentimen dari komentar di Instagram dengan memanfaatkan algoritma *Support Vector Machine* dan selanjutnya untuk mengevaluasi kinerja dari *Support Vector Machine* tersebut.

3. BAHAN DAN METODE

Sebagai landasan dilakukannya penelitian ini, maka perlu dikembangkan teori-teori yang berhubungan dengan analisis sentimen, quick count, *Support Vector Machine*.

3.1 Analisis Sentimen

Menurut Liu (Rahmawati, 2023) Analisis sentimen atau opinion mining merupakan salah satu bidang Natural Language Processing (NLP) yang menganalisis pendapat, perilaku, penilaian, dan emosi seseorang terkait dengan produk, topik, layanan, organisasi, individu, atau kegiatan lainnya yang diekspresikan ke dalam bentuk teks

3.2 Quick Count

Menurut Imam Yunianto (Saraswati, 2020) *Quick Count* atau dalam artian bahasa Indonesia disebut hitung

cepat merupakan sebuah metode verifikasi hasil pemilihan umum yang dilakukan Dengan menghitung (satuan persen %) hasil pemilu di tempat berlangsungnya pemungutan suara (TPS) yang dapat dijadikan sampel.

3.3 Support Vector Machine

Menurut (Samsudiney, 2019) *Support Vector Machine* (SVM) merupakan salah satu metode dalam *supervised learning* yang biasanya digunakan untuk klasifikasi (seperti *Support Vector Classification*) dan regresi (*Support Vector Regression*).

3.4 Metode

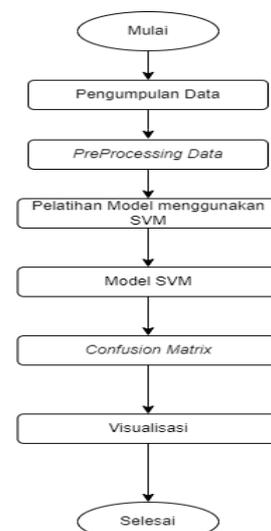
Metode penelitian merupakan sebuah cara yang digunakan untuk mendapatkan dan menganalisa data untuk kegunaan dan tujuan tertentu. Di dalam metode penelitian ini akan ada beberapa hal yang perlu diperhatikan.

3.1 Tempat dan Waktu Penelitian

Tempat penelitian dilakukan di Komisi Pemilihan Umum Kota Samarinda. Penelitian ini dilakukan selama kurang lebih satu bulan yang dimulai pada bulan mei sampai juni 2024.

3.2 Tahapan Penelitian

Diagram Tahapan Penelitian yang dibuat adalah rencana yang akan dilakukan oleh peneliti dalam melakukan analisis sentimen. Pada diagram tahapan penelitian terdapat rincian dalam proses pencarian awal data sampai dengan akhir, pada masing-masing proses yang ditulis pada diagram tahapan akan dijelaskan satu-persatu guna memperjelas diagram yang diolah tersebut. Bisa dilihat pada gambar. Penjelasan rinci ini mencakup semua langkah penting, mulai dari pengumpulan data mentah hingga tahap analisis dan interpretasi hasil, memastikan bahwa setiap tahapan penelitian diuraikan dengan jelas untuk memudahkan pemahaman keseluruhan proses penelitian. Langkah-langkah ini meliputi pengumpulan data, pra-pemrosesan data, metode analisis, dan lain – lain secara terstruktur dalam diagram tersebut.



Gambar 3. 1 Tahapan Penelitian

3.2.1 Pengumpulan Data

Penelitian ini menggunakan data kuantitatif untuk menganalisis data dari teks berbahasa Indonesia berupa komentar dari postingan narasineersroom yang diperoleh dari instagram terkait hasil rekap quick count 2024. Metode yang digunakan dalam pengumpulan data selama penelitian adalah :

1. Observasi

Observasi dilakukan sebagai studi pendahuluan, peneliti mengamati secara langsung postingan hasil rekap quick count 2024 dari narasineersroom dengan melihat pendapat atau komentar yang ada pada media sosial Instagram. Observasi dilakukan peneliti pada bulan Mei 2024. Beberapa hal yang diperoleh peneliti dari observasi yang telah dilakukan adalah memperoleh informasi bahwa terdapat tanggapan atau komentar yang beragam dari postingan narasineersroom terkait hasil rekap quick count 2024 pada Instagram.

2. Studi Pustaka

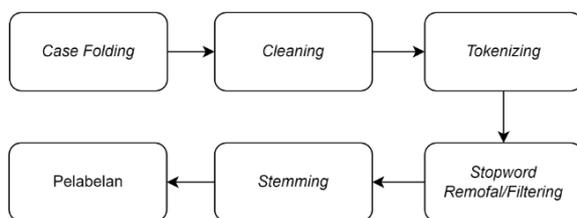
Metode pengumpulan studi pustaka ini dilakukan untuk mencari, mempelajari, dan menggunakan berbagai literatur seperti, buku, jurnal, paper, e- book, atau literatur lain yang berhubungan dengan tema penelitian ini. Informasi yang terkumpul akan dijadikan data pendukung dan pembanding untuk penelitian ini.

3. Crawling Data

Pada tahapan ini dilakukan pengumpulan dan pencarian dataset Instagram Bahasa Indonesia. Data yang dikumpulkan adalah data komentar masyarakat dari Instagram yang merupakan pesan ataupun opini dari masyarakat terkait hasil rekap quick count 2024.

3.2.2 Pre-Processing Data

Pre-processing merupakan tahapan awal yang akan dilalui dalam memproses teks. Pada penelitian ini akan dilakukan tahapan Pre-processing dengan tahapan case folding, data cleaning, tokenisasi, stopword removal/filtering, stemming dan pelabelan. Alur proses pada tahapan ini ditunjukkan pada Gambar 3.2



Gambar 3. 2 Tahapan Text Processing

1. Case Folding

Case folding adalah proses mengubah semua karakter dalam sebuah dokumen menjadi huruf kecil atau huruf besar yang sama, untuk mempercepat perbandingan selama proses pemrosesan data.

2. Cleansing

Cleansing merupakan tahapan yang bertujuan menghilangkan karakter atau symbol link url (<http://situs.com>), username atau mention (@username),

hashtag(#), retweet, dan emoticon. Selain itu, pada tahap ini juga membersihkan duplikasi data.

3. Tokenizing

Pada proses *tokenizing* dilakukan untuk memisahkan teks menjadi bagian- bagian yang lebih terstruktur sehingga mempermudah proses analisis selanjutnya, seperti penghapusan karakter khusus, menghitung frekuensi kata, atau membangun representasi vektor dari teks.

4. Stopword Removal/Filtering

Tahap *stopword removal* membuang kata-kata tidak bermakna yang tidak mempengaruhi klasifikasi sentimen, seperti kata penghubung dan kata depan, berdasarkan kamus stopword. Hal ini akan membantu masyarakat memahami dinamika dan keragaman pendapat yang ada di tengah-tengah mereka.

5. Stemming

Tahap ini bertujuan untuk mengembalikan suatu kata ke dalam bentuk dasarnya. Dengan menghilangkan kata imbuhan awalan, akhiran, sisipan atau pun kombinasi imbuhan. Pada penelitian ini akan digunakan library sastrawi untuk melakukan proses *stemming*. Sastrawi merupakan library pada Bahasa pemrograman python yang dibangun dengan algoritma NA. Algoritma tersebut berdasarkan aturan Bahasa Indonesia.

6. Pelabelan

Pelabelan adalah memberikan kategori atau tanda pada data agar atribut tertentu dapat diidentifikasi. Dengan pelabelan ini memudahkan menentukan kalimat sentimen yang terdiri dari negative positif ataupun netral. Pelabelan akan dibantu dengan kamus lexicon, berupa kamus kata positif dan negative.

3.2.3 Pelatihan Model Menggunakan SVM

1. Pembobotan TF-IDF

Setelah tahap preprocessing berikutnya adalah tahap pembobotan kata. Pembobotan bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. TF-IDF (Term Frequency-Inverse Document Frequency) memberi nilai penting pada kata dalam satu dokumen dengan mempertimbangkan seberapa sering kata tersebut muncul di dokumen itu dan seberapa jarang kata itu muncul di seluruh kumpulan dokumen. Rumus pada TF-IDF seperti pada dibawah berikut ini :

$$TF_{(t,d)} = \frac{\text{Number of times } t \text{ appear in document } d}{\text{Total number of terms in document } d}$$

$$IDF_{(t,d)} = \log \left(\frac{\text{Total number of documents } D}{\text{Number of document the term in it}} \right)$$

$$TFIDF(t, d, D) = TF(t,d) * IDF(t,d)$$

dimana : t_i = Dokumen ke - i ,

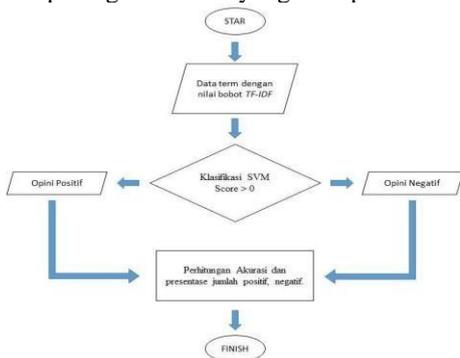
df = Document frequency,

n = Banyaknya data

idf = Inverse document frequency.

3.2.4 Pemodelan Dengan Metode Support Vector Machine

Metode *Support Vector Machine* (SVM) digunakan dalam penelitian ini untuk analisis sentimen. Hasilnya adalah klasifikasi teks menjadi kelas positif, negatif dan netral berdasarkan bobot fitur masing-masing dokumen. Proses ini melibatkan pembobotan fitur dengan SVM, menghasilkan pola yang digunakan dalam pengujian. Evaluasi dilakukan dengan memberi skor pada dokumen, dan analisis keputusan didasarkan pada tingkat akurasi serta jumlah dokumen di kelas positif dan negatif. Untuk memastikan hasil yang akurat, data dibagi menjadi dua bagian, yaitu data training dan data testing, menggunakan rasio 0,1. Ini berarti bahwa 90% dari data digunakan untuk proses training, di mana model SVM dilatih untuk mengenali pola dalam teks, sedangkan 10% sisanya digunakan untuk data testing, yang berfungsi untuk menguji seberapa baik model tersebut bekerja pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk menghindari overfitting dan memastikan bahwa model yang dihasilkan memiliki kemampuan generalisasi yang baik pada data baru.



Gambar 3. 3 Tahap Proses Metode SVM Sumber : (Herwinsyah, 2022)

<https://ejournal.lppmunbaja.ac.id/index.php/jsii/article/view/1411/926>

3.2.5 Confusion Matrix

Setelah hasil pengujian menggunakan Support Vector Machine selesai, akan dilakukan confusion matrix untuk menentukan 3 macam hasil performa machine learning yakni akurasi, presisi, dan recall. Hasil klasifikasi yaitu berupa jumlah sentimen positif dan negatif komentar masyarakat terhadap postingan hasil rekap quick count 2024 di instagram, sehingga akan diperoleh kesimpulan mengenai tanggapan masyarakat Indonesia terhadap hasil rekap quick count 2024 di Indonesia.

3.2.6 Visualisasi

Tahap ini merupakan tahap memvisualisasikan kata-kata yang sering muncul pada data komentar ke dalam bentuk wordcloud. Hal ini untuk mengetahui kata-kata apa saja yang paling banyak terdapat pada data komentar dan berapa frekuensi kata tersebut muncul

dalam sebuah komentar. Visualisasi juga digunakan untuk menampilkan jumlah emosi pada data komentar dalam bentuk barplot yang terdiri atas beberapa parameter emosi.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada penelitian ini dataset yang digunakan ialah data yang berasal dari media sosial instagram, objek penelitiannya sendiri ialah postingan dari akun narasinewsroom. Adapun tahapan proses pengolahannya hingga menjadi dataset yang nantinya akan digunakan untuk pengujian penelitian ini, berikut tahapan proses pengambilan dataset :



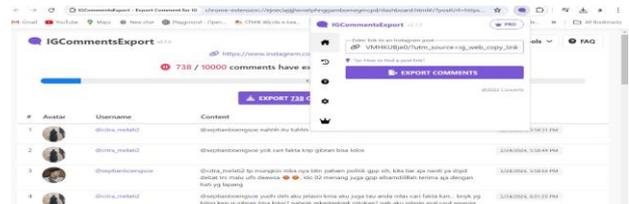
Gambar 4. 1 Instagram narasinewsroom

4.2 Pembahasan

4.2.1 Pengumpulan Data

1. Crawling Data

Proses crawling dilakukan menggunakan Export Comment dari Google Chrome yang secara otomatis melakukan pengumpulan data. Data yang terkumpul kemudian disimpan dalam format yang terstruktur untuk memudahkan analisis



Gambar 4. 2 Proses Pengumpulan Data

Tahap ini melakukan pengambilan data komentar pada salah satu postingan pada akun sosial narasinewsroom Indonesia. Gambar di atas adalah contoh mengambil data nama pada komentar yang akan di ambil. Pada gambar 4.2 menunjukkan jumlah nama yang di dapat pada komentar pada suatu postingan.

User Id	Username	Comment Id	Comment Text	Profile URL	Avatar URL	Date
9	5870918800	citra_meia2	18014354803278791	@septiarboengoe nahhh bu bunn	https://www.instagram.com/citra_meia2	2/24/2024, 5:58:31 PM
1	5870918800	citra_meia2	1788555326070519	@septiarboengoe yok cari fakta krip gilaan ba...	https://www.instagram.com/citra_meia2	2/24/2024, 5:58:44 PM
2	2158929361	septiarboengoe	18041098867712513	@citra_meia2 lo mungkin rba nya bin patam po...	https://www.instagram.com/septiarboengoe	2/24/2024, 5:58:59 PM
3	5870918800	citra_meia2	1791932208879831	@septiarboengoe yuth dan aku jajan krip aku	https://www.instagram.com/citra_meia2	2/24/2024, 6:01:29 PM
4	2158929361	septiarboengoe	1805286285560443	@citra_meia2 mana ada aku gik nemu curang	https://www.instagram.com/septiarboengoe	2/24/2024, 6:02:47 PM

Gambar 4. 3 Hasil Crawling Data

Pada gambar 4.3 adalah hasil crawling data dalam bentuk soft file csv dari jumlah keseluruhan data yang tercatat

berjumlah 3573 data yang berjumlah 7 atribut diantaranya User Id, Username, Comment Id, Comment Text, Profil URL, Avatar URL, dan Date dari 7 atribut yang ditampilkan. setelah itu proses pembersihan data dan menyeleksi atribut teks.

4.2.2 Pre-Processing Data

Sebelum melakukan tahapan *preprocessing* terlebih dahulu import library yang dibutuhkan untuk memanggil dataset yang akan di proses di *google collab*. Setelah library berhasil di import, selanjutnya memanggil dataset yang telah didapat dari proses crawling data. Pada gambar 4.3 merupakan hasil pemanggilan dataset yang berformat CSV dengan nama *comments2.csv*.

1. Cleansing

	Comment Text	cleansing
0	@septianboengsoe nahhh itu tuhnh	nahhh itu tuhnh
1	@septianboengsoe yok cari fakta knp gibran bisa...	yok cari fakta knp gibran bisa lolos
2	@citra_melat2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...
3	@septianboengsoe yudh deh aku jelsain kma aku...	yudh deh aku jelsain kma aku juga tau anda ml...
4	@citra_melat2 mana ada aku gak nerima curang	mana ada aku gak nerima curang
...
3568	Ntar dulu... Jgn gegabah	Ntar dulu jgn gegabah
3569	Tenang pak... Tenang pak... jgn bunuz2 KPU lagi...	Tenang pak Tenang pak jgn bunuz2 KPU lagi ngltu...
3570	@nurisepitka terus presiden harus miskin gtu?	terus presiden harus miskin gtu?
3571	ditunggu 300 FK nya	ditunggu 300 FK nya
3572	Gasss	Gasss

Gambar 4.4 Pembersihann Data

Gambar 4.4 menunjukkan hasil pembersihan data pada atribut teks untuk menghilangkan atau menghapus bagian yang tidak diperlukan seperti URL, hashtag, mention, emoticon, dan karakter spesial lainnya. Setelah data dibersihkan, langkah selanjutnya adalah melakukan *pre-processing* yang lebih mendetail, seperti *case folding*, *tokenisasi*, *stemming*, dan *penghapusan stop words*.

2. Case Folding

	Comment Text	cleansing	case_folding
0	@septianboengsoe nahhh itu tuhnh	nahhh itu tuhnh	nahhh itu tuhnh
1	@septianboengsoe yok cari fakta knp gibran bisa...	yok cari fakta knp gibran bisa lolos	yok cari fakta knp gibran bisa lolos
2	@citra_melat2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...
3	@septianboengsoe yudh deh aku jelsain kma aku...	yudh deh aku jelsain kma aku juga tau anda ml...	yudh deh aku jelsain kma aku juga tau anda ml...
4	@citra_melat2 mana ada aku gak nerima curang	mana ada aku gak nerima curang	mana ada aku gak nerima curang
...
3568	Ntar dulu... Jgn gegabah	Ntar dulu jgn gegabah	ntar dulu jgn gegabah
3569	Tenang pak... Tenang pak... jgn bunuz2 KPU lagi...	Tenang pak Tenang pak jgn bunuz2 KPU lagi ngltu...	tenang pak tenang pak jgn bunuz2 kpu lagi ngltu...
3570	@nurisepitka terus presiden harus miskin gtu?	terus presiden harus miskin gtu?	terus presiden harus miskin gtu?
3571	ditunggu 300 FK nya	ditunggu 300 FK nya	ditunggu 300 FK nya
3572	Gasss	Gasss	gasss

Gambar 4.5 Hasil Case Folding

Pada gambar 4.5 menunjukkan hasil yang setelah dilakukan Cleansing dilanjut dengan *Case Folding* pada atribut teks untuk merubah semua huruf besar menjadi huruf kecil. Contoh kalimat yang ada pada gambar "Waktu kampanye di GBK gua kn blm dpt uang padahal gua pendukung oke gasss, Timses nya janji bkln dpt setelah nyoblos, t t t t tapi smpe sekarang hilal nya belum keliatan tuh uang.", menjadi "waktu kampanye di gbk gua kn blm dpt uang padahal gua pendukung oke gasss, timses nya janji bkln dpt setelah nyoblos, t t t t tapi smpe sekarang hilal nya belum keliatan tuh uang."

3. Tokenizing

	Comment Text	cleansing	case_folding	tokenize
0	@septianboengsoe nahhh itu tuhnh	nahhh itu tuhnh	nahhh itu tuhnh	[nahhh, itu, tuhnh]
1	@septianboengsoe yok cari fakta knp gibran bisa...	yok cari fakta knp gibran bisa lolos	yok cari fakta knp gibran bisa lolos	[yok, cari, fakta, knp, gibran, bisa, lolos]
2	@citra_melat2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	[tp, mungkin, mba, nya, bim, paham, politik, gpp, sih, k...]
3	@septianboengsoe yudh deh aku jelsain kma aku...	yudh deh aku jelsain kma aku juga tau anda ml...	yudh deh aku jelsain kma aku juga tau anda ml...	[yudh, deh, aku, jelsain, kma, aku, juga, tau, anda, ml...]
4	@citra_melat2 mana ada aku gak nerima curang	mana ada aku gak nerima curang	mana ada aku gak nerima curang	[mana, ada, aku, gak, nerima, curang]
...
3568	Ntar dulu... Jgn gegabah	Ntar dulu jgn gegabah	ntar dulu jgn gegabah	[ntar, dulu, jgn, gegabah]
3569	Tenang pak... Tenang pak... jgn bunuz2 KPU lagi...	Tenang pak Tenang pak jgn bunuz2 KPU lagi ngltu...	tenang pak tenang pak jgn bunuz2 kpu lagi ngltu...	[tenang, pak, tenang, pak, jgn, bunuz2, kpu, lagi, ngltu, de...]

Gambar 4.6 Hasil Tokenizing

Pada gambar 4.6 merupakan tahapan tokenisasi setelah dilakukan case folding kalimat "yok cari fakta knp gibran bisa lolos" yang jika ditokenisasikan menjadi "[yok', 'cari', 'fakta', 'knp', 'gibran', 'bisa', 'lolos']" dari kalimat tersebut telah terbagi menjadi kata token yang terpisah.

4. Stopword removal

	Comment Text	cleansing	case_folding	tokenize	filtering/stopword removal
0	@septianboengsoe nahhh itu tuhnh	nahhh itu tuhnh	nahhh itu tuhnh	[nahhh, itu, tuhnh]	[nahhh, tuhnh]
1	@septianboengsoe yok cari fakta knp gibran bisa...	yok cari fakta knp gibran bisa lolos	yok cari fakta knp gibran bisa lolos	[yok, cari, fakta, knp, gibran, bisa, lolos]	[yok, cari, fakta, knp, gibran, lolos]
2	@citra_melat2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	[tp, mungkin, mba, nya, bim, paham, politik, gpp, sih, k...]	[tp, mba, nya, bim, paham, politik, gpp, sih, k...]
3	@septianboengsoe yudh deh aku jelsain kma aku...	yudh deh aku jelsain kma aku juga tau anda ml...	yudh deh aku jelsain kma aku juga tau anda ml...	[yudh, deh, aku, jelsain, kma, aku, juga, tau, anda, ml...]	[yudh, deh, jelsain, kma, tau, ml, cari, fa...]
4	@citra_melat2 mana ada aku gak nerima curang	mana ada aku gak nerima curang	mana ada aku gak nerima curang	[mana, ada, aku, gak, nerima, curang]	[gak, nerima, curang]
...
3568	Ntar dulu... Jgn gegabah	Ntar dulu jgn gegabah	ntar dulu jgn gegabah	[ntar, dulu, jgn, gegabah]	[ntar, jgn, gegabah]
3569	Tenang pak... Tenang pak... jgn bunuz2 KPU lagi...	Tenang pak Tenang pak jgn bunuz2 KPU lagi ngltu...	tenang pak tenang pak jgn bunuz2 kpu lagi ngltu...	[tenang, pak, tenang, pak, jgn, bunuz2, kpu, lagi, ngltu, de...]	[tenang, tenang, jgn, bunuz2, kpu, ngltung, de...]

Gambar 4.7 Hasil Stopword removal

Pada gambar 4.7 merupakan tahapan *stopword* menghilangkan kata yang tidak berguna seperti dan, pada, sebagai dll. Seperti hasil dari kalimat berikut "mana ada aku gak nerima curang" yang jika ditokenisasikan menjadi "[gak', 'nerima', 'curang']" kata yang dihilangkan yaitu 'mana', 'ada', 'aku', karena dianggap tidak terlalu memiliki arti penting dalam proses text mining dan kalimat tersebut telah terbagi menjadi kata token yang terpisah.

5. Stemming

	Comment Text	cleansing	case_folding	tokenize	filtering/stopword removal	stemming_data
0	@septianboengsoe nahhh itu tuhnh	nahhh itu tuhnh	nahhh itu tuhnh	[nahhh, itu, tuhnh]	[nahhh, tuhnh]	nahhh tuhnh
1	@septianboengsoe yok cari fakta knp gibran bisa...	yok cari fakta knp gibran bisa lolos	yok cari fakta knp gibran bisa lolos	[yok, cari, fakta, knp, gibran, bisa, lolos]	[yok, cari, fakta, knp, gibran, lolos]	yok cari fakta knp gibran lolos
2	@citra_melat2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	[tp, mungkin, mba, nya, bim, paham, politik, gpp, sih, k...]	[tp, mba, nya, bim, paham, politik, gpp, sih, k...]	tp mba nya bim paham politik gpp sih iar aja ...
3	@septianboengsoe yudh deh aku jelsain kma aku...	yudh deh aku jelsain kma aku juga tau anda ml...	yudh deh aku jelsain kma aku juga tau anda ml...	[yudh, deh, aku, jelsain, kma, aku, juga, tau, anda, ml...]	[yudh, deh, jelsain, kma, tau, ml, cari, fa...]	yudh deh jelsain kma tau ml cari fakta bryk...
4	@citra_melat2 mana ada aku gak nerima curang	mana ada aku gak nerima curang	mana ada aku gak nerima curang	[mana, ada, aku, gak, nerima, curang]	[gak, nerima, curang]	gak nerima curang
5	@septianboengsoe lahnh coba lu scroll keatas bel	lahnh coba lu scroll keatas bel	lahnh coba lu scroll keatas bel	[lahnh, coba, lu, scroll, keatas, bel]	[lahnh, coba, lu, scroll, keatas, bel]	lahnh coba lu scroll atas bel

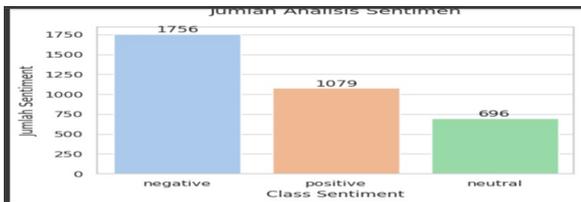
Gambar 4.8 Hasil Tahapan Stemming

Pada gambar 4.8 di atas, ditunjukkan proses hasil *stemming* di mana kata-kata dengan imbuhan diubah menjadi kata dasar pada dataset. Sebagai contoh, kata 'keatas' setelah proses stemming menjadi 'atas'. Proses ini bertujuan untuk menyederhanakan kata-kata dan mengurangi variasi dalam dataset agar analisis sentimen atau pemrosesan teks menjadi lebih efektif dan akurat.

6. Pelabelan

Setelah Pelaksanaan Tahapan Preprocessing dilaksanakan dan menghasilkan preprocessing dengan variable *Stemming_data* yang merupakan data tidak terstruktur dan tidak memiliki kelas/label, agar dapat digunakan untuk proses mining dengan algoritma SVM, maka perlu dilakukan proses pelabelan data ntuk menentukan kata-kata yang merepresentasikan kata positif dan negatif dan Menghitung jumlah kata positif dan negatif pada setiap satu kalimat dengan bantuan dari

kamus inset lexicon positif dan negatif yang bersumber dari penelitian dan repository. Pelabelan diklasifikasikan kedalam 3 kelas yaitu positif, negatif dan netral.



Gambar 4. 9 Hasil Jumlah Sentimen

Pada gambar 4.9 di atas, ditunjukkan bahwa jumlah sentimen negatif lebih banyak, dengan total 1756 dokumen. Sentimen positif berjumlah 1079 dokumen, sementara sentimen netral berjumlah 696 dokumen. Hal ini menunjukkan bahwa sentimen negatif lebih dominan dibandingkan dengan sentimen lainnya. Jika dipresentasikan, hasilnya adalah 49.7% negatif, 30.6% positif, dan 19.7% netral, seperti yang ditampilkan pada gambar 4.10. Dominasi sentimen negatif ini mengindikasikan bahwa sebagian besar teks yang dianalisis cenderung memiliki pandangan atau opini yang negatif



Gambar 4. 10 Diagram Hasil Klasifikasi Sentimen Dalam Bentuk Persen

4.2.3 Pelatihan Model Menggunakan SVM

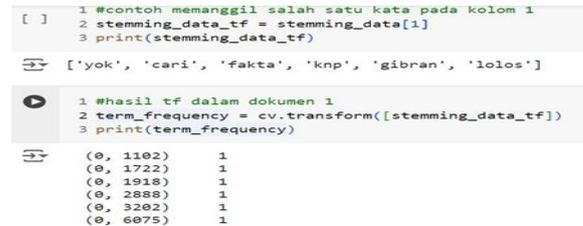
1. Pembobotan dengan TF-IDF

Setelah pelaksanaan proses pelabelan dilanjutkan dengan Tahapan Pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) yang menggunakan perhitungan modul *scikit-learn* yaitu *TfidfVectorizer*. Dimana hasil dari tahap TF-IDF yaitu dari 3531 dataset setelah dilakukan pembobotan kata atau kemunculan kata berjumlah 6109 kata



Gambar 4. 11 kamus (dictionary) yang menghubungkan kata-kata (terms) dengan indeks uniknya

Pada gambar 4.11 menjelaskan kamus ini digunakan dalam *CountVectorizer* atau *TfidfVectorizer* dari library *scikit-learn*, yang mengubah koleksi dokumen teks menjadi matriks istilah-dokumen. Setiap kata atau term dalam teks diberikan indeks unik untuk mengidentifikasi kata tersebut di dalam vektor atau matriks. Misalnya, kata "nahhh" memiliki indeks 3718, "tuhhh" memiliki indeks 5729, dan seterusnya



Gambar 4. 12 Hasil TF

Pada gambar 4.12 merupakan hasil TF dari kalimat "yok cari fakta knp gibran lolos", terdapat ada 6 kata dengan masing-masing nomor indeksnya. Contoh jika dalam hasil yang diberikan (0, 1102) 1, ini berarti dalam dokumen pertama (indeks 0), kata "cari" muncul sekali dan begitupun seterusnya

word	frequency
nahhh	1
tuhhh	2
yok	4
cari	57
fakta	61
knp	20
gibran	92
lolos	15

Gambar 4. 13 Hasil Frequency Kata

Gambar 4.13 di atas merupakan gambar "Term Frequency" yang menunjukkan frekuensi kemunculan kata-kata tertentu dalam sebuah korpus atau dataset teks

D1	tf	idf	tf-idf
yok	0,166666667	2,945837726	0,490972954
cari	0,166666667	1,792022862	0,298670477
fakta	0,166666667	1,762567883	0,293761314
knp	0,166666667	2,246867722	0,374477954
gibran	0,166666667	1,58410989	0,264018315
lolos	0,166666667	2,371806459	0,395301076

Gambar 4. 14 Hasil TF-IDF

Pada gambar 4.13 menunjukkan hasil nilai yang diperoleh dari perhitungan tf-idf yang menggunakan tools *google collab*. Sedangkan kata "yok" memiliki nilai *tfidf* 0.490972954 dan seterusnya menunjukkan pasangan indeks term dan nilai TF-IDF-nya yang hasilnya dikalikan. Nilai TF-IDF memberikan bobot pada setiap term berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang term tersebut muncul di dokumen lain. Rumus perhitungan TF, IDF dan TFIDF kata "yok", "cari", "fakta", "knp", "gibran", "lolos", dengan nomor indeks 6075, 1102, 2888, 1722, 1918, 3202:

4.2.4 Pemodelan Dengan Metode Support Vector Machine

Setelah melewati pembobotan TF-IDF, selanjutnya tahapan pembagian data dilakukan untuk mengevaluasi performa model dari SVM, yang dibagi menjadi data training dan data testing dimana didalam penelitian ini menggunakan pembagian data 90:10. Hasil dari Nilai X train : (3177, 6051), Nilai X Test : (354,6051) Nilai Y train (3177) dan nilai dari Y Test (354).

	precision	recall	f1-score	support
negative	0.88	0.85	0.86	176
neutral	0.65	0.76	0.70	70
positive	0.83	0.80	0.82	108
accuracy			0.81	354
macro avg	0.79	0.80	0.79	354
weighted avg	0.82	0.81	0.82	354

Gambar 4. 15 Hasil dari pemodelan algoritma svm

Pada gambar 4.14 diatas mengindikasikan bahwa nilai akurasi dengan pembagian 90% data training dan 10% data testing adalah 81%. Pada kelas "negative" memiliki precision 0.88, recall 0.85, dan F1-score 0.86. Kelas "neutral" memiliki precision 0.65, recall 0.76, dan F1-score 0.70, sedangkan Kelas "positive" memiliki precision 0.83, recall 0.80, dan F1-score 0.82.

4.2.5 Confusion Matrix

Setelah didapatkannya hasil klasifikasi sentimen menggunakan SVM perlu dilakukan pengujian menggunakan *confusion matrix* untuk mengetahui informasi yang diprediksi dan hasil yang sebenarnya dan yang digunakan adalah data testing yaitu 354 data testing. Perhitungan dari *confusion matrix* adalah accuracy, recall, dan precision. Berikut merupakan hasil pengujian menggunakan *confusion matrix*.



Gambar 4. 16 Pengujian Confusion Matrix

Dari Gambar 4.15 diatas dapat dijelaskan bahwa memiliki 149 data sebagai true negatif, 53 data sebagai data true netral dan 86 data sebagai true positif. Pada perhitungan manual *akurasi*, *presisi* dari matriks diatas sebagai berikut:

$$Accuracy = \frac{\text{Jumlah prediksi benar}(TN + TP + TN)}{\text{Jumlah Prediksi}(TN + FN + FP + TN)} \times 100\%$$

$$Accuracy = \frac{86 + 53 + 149}{149+16+11+11+53+6+10+12+86} \times 100\%$$

$$Accuracy = \frac{288}{354} \times 100\%$$

$$Accuracy = 0.813 \times 100\% = 81\%$$

4.2.6 Visualisasi



Gambar 4. 17 Wordcloud ulasan negative

Pada gambar 4.16 merupakan hasil visualisasi dari ulasan negative Terlihat jelas bahwa frekuensi kemunculan sebuah kata meningkat seiring dengan ukurannya di dalam wordcloud. Beberapa istilah yang sering muncul dalam ulasan positif terhadap aplikasi yaitu "bukti", "curang", "salah", "kalah", dan lainnya.



Gambar 4. 18 Wordcloud ulasan positive

Gambar 4.17 menampilkan hasil visualisasi ulasan positive. Dimana kata "terima", "presiden "percaya", "menang", "dukung", sering digunakan dalam ulasan positive.



Gambar 4. 19 Wordcloud ulasan netral

Sementara itu, Gambar 4.18 menampilkan hasil visualisasi ulasan neutral. Kata "unggul", "rakyat", "suara", "bang", dan "Indonesia" sering digunakan dalam ulasan neutral.

Hasil dari penelitian ini menggunakan data sebanyak 3572 yang sudah diolah dari mengecek data duplikat sebanyak 41, menghapus data komentar yang kosong, dilakukan pre-processing sehingga menjadi 3531. Kemudian dilakukan pelabelan dengan hasil negative 1756 data, positive 1079 data, dan neutral 696. Kemudian dilakukan TF-IDF yang memiliki 6109 kata unik, dilanjutkan melakukan pemodelan SVM yang memperoleh hasil accuracy 81% dengan presisi dan recall untuk sentimen negatif masing-masing 88% dan 85%, untuk sentimen netral 65% dan 76%, serta untuk

sentimen positif 83% dan 80%. Ini menunjukkan bahwa model SVM yang dikembangkan mampu mengklasifikasikan sentimen dengan akurasi yang baik.

5. KESIMPULAN

Dalam penelitian ini, menentukan sentimen masyarakat dengan cara komentar-komentar pengguna Instagram yang berkaitan dengan Calon Presiden 2024 seperti Ganjar Pranowo, Prabowo Subianto, dan Anis Baswedan dianalisis menggunakan algoritma Support Vector Machine (SVM) dan metodologi Knowledge Discovery in Database.

1. Sebanyak 3572 data komentar diperoleh melalui web crawling dari situs exportcomments.com. Setelah melalui proses preprocessing, didapatkan 3351 data yang digunakan dan dibagi menjadi tiga kelas: 1756 data kelas negatif, 1079 data kelas positif, dan 696 data kelas netral.

2. Selanjutnya setelah dilakukan tahapan pre-processing akan dilanjutkan tahapan pelatihan model SVM, dimana pelatihannya menggunakan perhitungan TF-IDF untuk menghitung bobot kata.

3. Penggunaan algoritma SVM dengan kernel linear pada tahap data mining mampu mengklasifikasikan komentar dengan baik. Pengujian dilakukan dengan empat skenario pembagian data 90:10 menggunakan confusion matrix untuk menilai skor. Hasil terbaik diperoleh pada skenario 90:10 dengan akurasi tertinggi sebesar 81%, serta nilai presisi sentimen negatif 88%, recall 85%, untuk nilai presisi sentimen positif 83%, Reccal 80% dan F1-score dari sentimen negatif 86%, positif 82% yang nilainya signifikan.

6. SARAN

Saran untuk pengembangan lebih lanjut dari penelitian ini sebagai berikut :

1. Mencoba menggunakan metode klasifikasi lainnya yang belum digunakan pada penelitian ini
2. Menggunakan ekstraksi fitur lainnya pada proses klasifikasi sentimen.
3. Dapat dipertimbangkan untuk menggunakan sumber data yang lebih beragam, termasuk komentar dari media sosial lainnya.

7. DAFTAR PUSTAKA

Rahmawati, S. (2023). Implementasi Algoritma BERT untuk Analisis Sentimen Ulasan Pengguna Aplikasi PeduliLindungi (Doctoral dissertation, UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA).

Samsudiney, (2019). "Penjelasan Sederhana tentang apaituSVM?https://medium.com/@samsudiney/penjelasan-sederhana-tentangapa-itu-svm-149fec72bd02

Saraswati, R., & Prasakti, D. A. (2020). Kepercayaan Masyarakat Terhadap Hasil Quick Count Pada Pemilihan Presiden Dan Legislatif Tahun 2019. *Jurnal Kewarganegaraan*, 4(1), 25-30.