

Performa Machine Learning Terhadap Analisis Sentimen Pemilu 2024 Melalui Media Digital

Diterima: 10 Juni 2024
Revisi: 10 Juli 2024
Terbit: 1 Agustus 2024

¹Miranda Putri Arifin, ²Umi Mahdiyah, ³Ahmad Bagus Setiawan
¹⁻³ Universitas Nusantara PGRI Kediri
*¹mirandaputriarifin@gmail.com, ²umimahdiyah@unpkediri.ac.id,
³ahmadbagus@unpkediri.ac.id*

Abstrak—Pemanfaatan media sosial untuk kepentingan publik sudah banyak digunakan, salah satunya untuk kampanye pemilu presiden 2024. Analisis sentimen bertujuan untuk komputasi kalimat opini berupa teks dari media aplikasi *twitter*. Dengan menggunakan proses bantuan *machine learning* didapatkan data *crawling* yang kemudian diolah dengan *teks preprocessing* yaitu *cleaning, case folding, normalization, stopword remover, tokenization, dan stemming* yang kemudian dilanjutkan ke tahap training dan testing kemudian evaluasi dengan Support Vector Machine (SVM). Hasil akhir evaluasi yang didapatkan dengan pemrosesan 50 data dan 100 data dimana perolehan 80% untuk hasil akurasi 50 data sedangkan 70% untuk hasil akurasi 100 data. Hasil yang berbeda ini membuktikan bahwa jumlah banyaknya data dan pemrosesan yang maksimal atau belum maksimal mempengaruhi perolehan hasil yang didapat.

Kata Kunci—Analisis Sentimen, SVM, *Twitter*

Abstract—*Abstract The use of social media for public interests has been widely used, one of them for the 2024 presidential campaign. Sentiment analysis is aimed at computing opinions sentences in the form of text from the media of the twitter application. Using a machine learning aid process, crawling data is obtained which is then processed with preprocessing text cleaning, case folding, normalization, stopword remover, tokenization, and stemming that is then proceeded to the training and testing phase and then evaluated with Support Vector Machine. (SVM). The final evaluation results obtained with processing 50 data and 100 data whoever the result is 80% for 50 data accuracy results while 70% for 100 data accurate results. These different results prove that the maximum amount of data and processing or not maximum influence the obtaining results.*

Keywords—*Sentiment Analyst, SVM, Twitter*

This is an open access article under the CC BY-SA License.



Penulis Korespondensi:

Miranda Putri Arifin,
Teknik Informatika,
Universitas Nusantara PGRI Kediri,
Email: mirandaputriarifin@gmail.com
ID Orcid: [<https://orcid.org/register>]
Handphone: 082133055947

I. PENDAHULUAN

Pemanfaatan media sosial untuk kepentingan publik sudah banyak digunakan. Salah satunya yaitu pada saat kampanye pemilu presiden 2024 ini. Pemilihan umum presiden yang

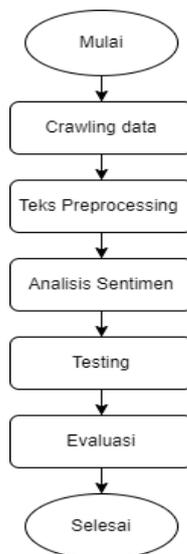
dilaksanakan pada bulan Februari 2024 ini merupakan suatu hal yang penting untuk negara demokrasi, dikarenakan proses pemilihan Presiden NKRI yang dilaksanakan dalam setiap 5 tahun sekali. Kandidat Capres dan Cawapres meliputi Anies Baswedan – Muhaimin Iskandar, Prabowo Subianto – Gibran Rakabuming Raka, dan Ganjar Pranowo – Mahfud MD. Permasalahan yang timbul dengan adanya kampanye debat setiap paslon ini mendapatkan perhatian publik sehingga pemanfaatan teknologi analisis sentimen ini digunakan [1].

Analisis sentimen adalah cabang dari data mining yang bertujuan untuk komputasi kalimat opini yang berupa teks dari *platform* media sosial seperti aplikasi *twitter* untuk berbagi opini yang sangat melekat pada kehidupan masyarakat sekarang ini. Pengolahan analisis sentimen menggunakan *machine learning* untuk mengolah semua proses dari data mentah yang diambil dari aplikasi *twitter* menggunakan kode token dari *inspect* website *twitter* [2].

Machine Learning sangat dibutuhkan dalam pengolahan data yang sangat banyak seperti ini. *Python* merupakan bahasa pemrograman tingkat tinggi yang sangat menjadi perhatian dan mudah digunakan untuk sekarang ini. Dengan *library package* bersifat *open source* memudahkan akses yang luas di berbagai bidang dan fungsi [3].

II. METODE

Alur metodologi penelitian meliputi *crawling data*, teks *preprocessing*, analisis sentimen, testing dan evaluasi. Metode yang digunakan untuk penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Alur Metodologi Penelitian

2.1 Tahap Crawling data

Pengambilan data dari media sosial *Twitter* menggunakan kode rahasia unik aplikasi yang tersedia pada situs website aplikasi tersebut untuk mencari data *tweet* untuk

kebutuhan pengembangan studi [4]. Hasil dari crawling data ini peneliti menggunakan *Jupyter notebook* untuk proses pengambilan data dengan kata kunci “#debatcapres #capres2024 #pemiluserentak #pemilu2024, #prabowogibran, #ganjarmahfud, #aniesmuhaimin” [5].

Berikut merupakan hasil *crawling* data dari *twitter*, selanjutnya hasil *crawling* data tersebut akan dilakukan proses *text preprocessing* : “Masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada Pemilu 2024. Ikuti pembahasannya di #BedahEditorial di kanal youtube metrotvnews <https://t.co/26sQpy7em3> #Melawan #Kecurangan #Hakangket #Pemilu2024 #Pilpres2024 #EditorialMalam <https://t.co/NraPobWP44>”.

2.2 Tahap Teks Preprocessing

Tahap Teks Preprocessing ini yaitu *text mining* dimana mengolah data berupa kalimat menggunakan NLTK dengan proses didalamnya yaitu *Cleaning*, *Case folding*, *Tokenization*, *Normalization*, *Stopword* dan *Stemming* [6].

2.2.1 Cleaning

Proses *cleaning* dilakukan dengan membaca dataframe kemudian memfilter data guna menghilangkan spasi yang berlebih, dari simbol-simbol yang tidak perlu (username, tanda baca, simbol dan link) [7].

Berikut merupakan contoh kalimat setelah dibersihkan (*cleaning*) : “Masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada Pemilu Ikuti pembahasannya di BedahEditorial di kanal youtube metrotvnews Melawan Kecurangan Hakangket Pemilu Pilpres Editorial Malam”.

2.2.2 Case folding

Proses mengubah bentuk kata-kata tweet agar menjadi seragam atau menjadi huruf kecil semua (*lowercase*) [8].

Berikut merupakan contoh kalimat setelah di ubah : “masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada pemilu ikuti pembahasannya di bedaheditorial di kanal youtube metrotvnews melawan kecurangan hakangket pemilu pilpres editoriamalam”.

2.2.3 Normalization

Proses memperbaiki kata yang salah dalam sebuah teks atau kalimat dengan memperhatikan ejaan kalimat yang benar dan kalimat yang harusnya terdiri dari beberapa kata atau berdiri sendiri.

Berikut merupakan contoh kalimat setelah di ubah : “masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada pemilu ikuti pembahasannya youtube melawan kecurangan hak angket pemilu pilpres”.

2.2.4 Stopword Remover

Proses menghilangkan kata yang tidak sesuai dengan topik yang dimana akan mempengaruhi proses akurasi, dengan memperhatikan kata yang banyak munculnya.

Berikut merupakan contoh kalimat setelah di ubah : “masyarakat gantungkan harapan hak angket lawan kecurangan pemilu ikuti pembahasannya kanal youtube melawan kecurangan hak angket pemilu pilpres”.

2.2.5 Tokenization

Proses pemotongan kata menjadi kata tunggal, dengan memisahkan semua kalimat dalam dataset yang sangat banyak dan menjadi kata yang berdiri sendiri.

Berikut merupakan contoh kalimat setelah di ubah : “('masyarakat', 'gantungkan', 'harapan', 'hak', 'angket', 'lawan', 'kecurangan', 'pemilu', 'ikuti', 'pembahasannya', 'kanal', 'youtube', 'melawan', 'kecurangan', 'hak', 'pilpres')”.

2.2.6 Stemming

Proses mencari kata dasar yang banyak muncul dari proses stopwords, maka disini akan menyisakan kata yang sudah dibersihkan yang nantinya digunakan untuk proses analisis.

Berikut merupakan contoh kalimat setelah di ubah : “masyarakat gantung harap hak angket lawan curang milu ikut bahas bedaheditorial kanal youtube lawan curang hak angketpemilu pilpres”.

2.3 Tahap Analisis Sentimen

Pada tahap analisis sentimen ini, hal yang dilakukan yaitu untuk mengetahui polaritas kalimat untuk memudahkan pelabelan ke kelas positif atau kelas negatif. Berbagai cara bisa dilakukan pada tahap analisis ini [9]. Bisa dilakukan dengan cara manual atau menggunakan bantuan *machine learning*. Pada pelabelan pada analisis ini dilakukan dengan cara manual yaitu dengan mengamati berapa banyak kata yang muncul untuk mengidentifikasi kalimat positif dan negatif, maka dikelompokkan dan dijadikan sebagai acuan bahwa jika satu kalimat tersebut terdapat kata-kata yang sudah disusun, maka dikelompokkan ke label positif dan label negatif [10].

2.4 Tahap Training dan Testing

Pada tahap ini, nantinya semua data yang telah melalui proses preprocessing dan analisis sentimen, akan dilakukan pembagian data *training* dan *testing* dari data

keseluruhan untuk proses evaluasi permodelan dengan *machine learning* [11]. Pengujian dari 50 data dan 100 data akan dibagi menjadi 2 proses, yaitu proses data latih dan proses data uji, dimana untuk 50 data membagi 20% data uji dan 80% data training, untuk 100 data membagi 10% data uji dan 90% data training [12].

2.5 Tahap Evaluasi Model Klasifikasi Performa Machine Learning (SVM)

Model klasifikasi ini memiliki 2 keluaran kelas yaitu kelas positif dan kelas negatif untuk model klasifikasi yang akan diproses dengan menggunakan model *Support Vector Machine (SVM)* yang dimana ini merupakan metode pembelajaran yang cocok untuk klasifikasi opini publik karena menghasilkan akurasi tinggi [13]. Proses evaluasi menggunakan *confussion matrix* yang berfungsi untuk melihat performa dari model klasifikasi yang sudah diproses dan menentukan akurasinya [14].

III. HASIL DAN PEMBAHASAN

3.1 Hasil dari Semua Proses Processing

Pada bagian ini akan ditampilkan hasil dan pembahasan mengenai *crawling data, teks preprocessing*, analisis sentimen, training dan testing, kemudian yang terakhir yaitu evaluasi. Berikut merupakan tampilan semua perubahan data pada tabel 1 :

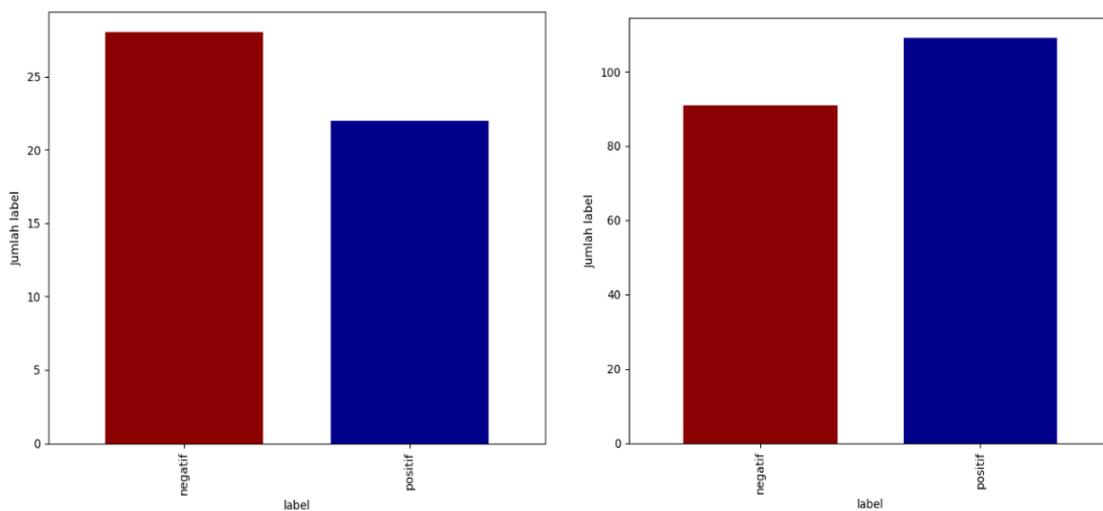
Tabel 1. Tampilan Data Hasil Semua *Preprocessing*

No.	Data Hasil Semua <i>Preprocessing</i>	
	Tahap <i>Preprocessing</i>	Hasil <i>Preprocessing</i>
1	<i>Crawling</i>	Masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada Pemilu 2024. Ikuti pembahasannya di #BedahEditorial di kanal youtube metrotvnews https://t.co/26sQpy7em3 #Melawan #Kecurangan #Hakangket #Pemilu2024 #Pilpres2024 #EditorialMalam https://t.co/NraPobWP44
2	<i>Cleaning</i>	Masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada Pemilu Ikuti pembahasannya di BedahEditorial di kanal youtube metrotvnews Melawan Kecurangan Hakangket Pemilu Pilpres EditorialMalam
3	<i>Case Folding</i>	masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada pemilu ikuti pembahasannya di bedaheditorial di kanal youtube metrotvnews melawan kecurangan hakangket pemilu pilpres editoriamalam
4	<i>Normalization</i>	masyarakat gantungkan harapan pada hak angket guna lawan kecurangan pada pemilu ikuti pembahasannya youtube melawan kecurangan hak angket pemilu pilpres
5	<i>Stopword Remover</i>	masyarakat gantungkan harapan hak angket lawan kecurangan pemilu ikuti pembahasannya kanal youtube melawan kecurangan hak angket pemilu pilpres
6	<i>Tokenization</i>	['masyarakat', 'gantungkan', 'harapan', 'hak', 'angket', 'lawan', 'kecurangan', 'pemilu', 'ikuti', 'pembahasannya', 'kanal', 'youtube', 'melawan', 'kecurangan', 'hak', 'pilpres']

Data Hasil Semua <i>Preprocessing</i>		
No.	Tahap <i>Preprocessing</i>	Hasil <i>Preprocessing</i>
7	<i>Stemming</i>	masyarakat gantung harap hak angket lawan curang milu ikut bahas bedaheditorial kanal youtube lawan curang hak angketpemilu pilpres

3.2 Analisis Sentimen

Pada proses ini menentukan sentimen positif dan negatif pada data yang telah dilakukan pemrosesan, untuk 50 data dan 100 data pelabelan diperoleh kelas positif atau kelas negatif yang di dapat pada gambar 2 kelas positif dan negatif :



Gambar 2. Kelas Perolehan Banyaknya Kelas Positif dan Negatif

3.3 Evaluasi Menggunakan Machine Learning Nilai Confussion Matrix-Cross Validation

Pada proses ini, banyak penelitian yang sudah dilakukan sebelumnya. Hasil dari perhitungan penelitian lainnya juga berbeda beda. Penggunaan metode yang berbeda beda mulai dari Naïve Bayes, Linear Regression, SVM, dan lainnya, didapatkan nilai akurasi dan hasil berbeda-beda, tergantung dengan analisis atau pelabelan yang dilakukan [15].

Pada penelitian ini, pada percobaan pengukuran evaluasi menggunakan *confussion matrix*, dimana pada penentuan performa ini nantinya menghasilkan *recall*, *precision*, dan *accuracy*. Untuk percobaan menggunakan 50 data dan 100 data ini menghasilkan *f-1 score* 70% dan 80% yang dimana ini merupakan hasil performa yang cukup.

Rumus untuk menentukan nilai *Confussion Matrix* dan Hasil *Cross Validation* dari 50 data dan 100 data di dapat :

Tabel 2. Hasil *Confussion Matrix* 50 data

<i>Actual</i>	<i>Prediction</i>	
	0	1
0	6	1
1	1	2

Keterangan hasil perhitungan :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{6+2}{6+2+1+1} = \frac{8}{10} = 0,8 \times 100\% = 80\% \dots\dots\dots (1)$$

$$Presisi = \frac{TP}{FP+TP} = \frac{2}{2+1} = \frac{2}{3} = 0,67 \times 100\% = 67\% \dots\dots\dots (2)$$

$$Recall = \frac{TP}{Fn+TP} = \frac{2}{2+1} = \frac{2}{3} = 0,67 \times 100\% = 67\% \dots\dots\dots (3)$$

$$F - 1 \text{ Score} = \frac{(2 \times recall \times precision)}{(recall + precision)} = \frac{(2 \times 0,67 \times 0,67)}{(0,67 + 0,67)} = 0,67 \times 100\% = 67\% \dots (4)$$

Tabel 3. Hasil *Confussion Matrix* 100 data

<i>Actual</i>	<i>Prediction</i>	
	0	1
0	3	0
1	3	4

Keterangan hasil perhitungan :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+4}{3+4+0+3} = \frac{7}{10} = 0,7 \times 100\% = 70\% \dots\dots\dots(1)$$

$$Presisi = \frac{TP}{FP+TP} = \frac{4}{4+0} = \frac{4}{4} = 1,0 \times 100\% = 100\% \dots\dots\dots(2)$$

$$Recall = \frac{TP}{Fn+TP} = \frac{4}{4+3} = \frac{4}{7} = 0,57 \times 100\% = 57\% \dots\dots\dots(3)$$

$$F - 1 \text{ Score} = \frac{(2 \times recall \times precision)}{(recall + precision)} = \frac{(2 \times 0,57 \times 1,0)}{(0,57 + 1,0)} = 0,72 \times 100\% = 72\% \dots(4)$$

IV. KESIMPULAN

Pelatihan Model dengan Evaluasi menggunakan model algoritma *machine learning Support Vector Machine* (SVM) pada data berbeda yang digunakan menghasilkan nilai dari *confussion matrix* dan mendapat akurasi dari hasil proses *cross-validation* yang dilakukan untuk tahap uji model dan evaluasi model dari jumlah data yang berbeda, 50 data dan 100 data. Hasil pengujian *cross-validation* pada 50 data adalah didapatkan akurasi 80%, presisi 67%, *recall* 67% dan *f1-score* 67%. Untuk hasil pengujian *cross-validation* pada 100 data didapatkan akurasi 70%, presisi 100%, *recall* 57%, dan *f1-score* 72% yang dimana membuktikan bahwa banyaknya data dan juga pemrosesan yang maksimal maupun belum maksimal, akan mempengaruhi hasil evaluasi yang dilakukan. Penelitian selanjutnya dapat menerapkan analisis sentimen dalam bidang sosial media lainnya dengan pemanfaatan berbagai teknologi berupa video atau gambar yang berkembang pesat dan meluas di era sekarang ini.

DAFTAR PUSTAKA

- [1] A. P. Nardilasari, A. L. Hananto, S. S. Hilabi, T. Tukino, dan B. Priyatna, “Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter,” *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 8, no. 1, hal. 11–18, 2023, doi: 10.31328/jointecs.v8i1.4265.
- [2] S. A. Anwar Hridoy, M. T. Ekram, M. S. Islam, F. Ahmed, dan R. M. Rahman, “Localized twitter opinion mining using sentiment analysis,” *Decis. Anal.*, vol. 2, no. 1, 2015, doi: 10.1186/s40165-015-0016-4.
- [3] R. Vindua dan A. U. Zailani, “Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, hal. 479, 2023, doi: 10.30865/jurikom.v10i2.5945.
- [4] N. Chadha, R. C. Gangwar, dan R. Bedi, “Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey,” *Int. J. Comput. Appl.*, vol. 131, no. 11, hal. 28–31, 2015.
- [5] V. Karyukin *dkk.*, “On the development of an information system for monitoring user opinion and its role for the public,” *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00660-w.
- [6] I. Kurniawan dan A. Susanto, “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019,” *J. Eksplora Inform.*, vol. 9, no. 1, hal. 1–10, 2019.
- [7] C. Prianto, N. H. Harani, dan I. Firmansyah, “Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter,” *J. Media Inform. Budidarma*, vol. 3, no. 4, hal. 405, 2019, doi: 10.30865/mib.v3i4.1549.
- [8] M. R. F. Sya’bani, U. Enri, dan T. N. Padilah, “Analisis Sentimen Terhadap Bakal Calon Presiden 2024 Dengan Algoritme Naïve Bayes,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, hal. 265–273, 2022.
- [9] A. Averina, H. Hadi, dan J. Siswantoro, “Analisis Sentimen Multi-Kelas Untuk Film Berbasis Teks Ulasan Menggunakan Model Regresi Logistik,” *Teknika*, vol. 11, no. 2, hal. 123–128, 2022, doi: 10.34148/teknika.v11i2.461.
- [10] S. Bhatia, M. Sharma, dan K. K. Bhatia, “Sentiment analysis and mining of opinions,” *Internet things big data Anal. Towar. next-generation Intell.*, hal. 503–523, 2018.
- [11] F. N. Arintowati dan A. B. Wahyudi, “PENANDA TANGGAPAN POSITIF DAN NEGATIF DALAM AKUN INSTAGRAM @nadiemmakarim,” *Prasi*, vol. 17, no. 1, hal. 55–67, 2022, doi: 10.23887/prasi.v17i1.37524.
- [12] H. Utama, E. Daniati, dan A. Masruro, “WEAK SUPERVISION DENGAN PENDEKATAN LABELING FUNCTION UNTUK ANALISIS SENTIMEN,” vol. 3, no. 1, hal. 49–57, 2024.
- [13] G. A. BUNTORO, R. ARIFIN, G. N. SYAIFUDDIIN, A. SELAMAT, O. KREJCAR, dan H. FUJITA, “Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia’s 2019 Presidential Election,” *IIUM Eng. J.*, vol. 22, no. 1, hal. 78–92, 2021, doi: 10.31436/IIUMEJ.V22I1.1532.
- [14] M. R. A. Nasution dan M. Hayaty, “Perbandingan akurasi dan waktu proses algoritma K-NN dan SVM dalam analisis sentimen twitter,” *J. Inf.*, vol. 6, no. 2, hal. 226–235, 2019.
- [15] B. P. Zen, D. Wicaksana, dan H. Alfidzar, “Analisis Sentimen Tweet Vaksin Covid 19 Sinovac Menggunakan Metode Support Vector Machine,” *J. Data Min. dan Sist. Inf.*, vol. 3, no. 2, hal. 21–27, 2022.