

Sentimen Analisis Presiden Terpilih Menggunakan Algoritma Support Vector Machine

Ricky Arie Kurniawan^{1a)}, Fernandi Indi Nizar G^{2a)}, Muhammad Rizky Pribadi^{3a)}

^{a)}Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang
Jl. Rajawali No.14, 9 Ilir, Kec. Ilir Tim. II, Kota Palembang, Sumatera Selatan

Corresponding author: rickyariekurniawan@mhs.mdp.ac.id

Abstract. Elections are essentially an arena that accommodates candidates in a political contest to gain power through the participation of the people to make choices and as a channel for the social and political rights of the society itself. In the 2024 elections, specifically for the presidential and vice-presidential elections in Indonesia, the results showed that the presidential candidate of the Republic of Indonesia with ballot number 2, Prabowo Subianto and Gibran Rakabuming Raka, were elected as president and vice president of the Republic of Indonesia for the 2024 – 2029 period, with 58% of the vote. This has sparked various responses from all Indonesians, especially on the social media platform Twitter (X). This study aims to conduct a sentiment analysis of the reactions of Indonesians regarding the election of Prabowo Subianto and Gibran Rakabuming Raka as president and vice president of the Republic of Indonesia for the 2024 – 2029 period. This research uses the Support Vector Machine (SVM) algorithm to classify Twitter data and understand public sentiment towards the elected president in the 2024 elections. The collected data was randomly selected to yield 427 data points, consisting of 26 positive tweets, 3 negative tweets, and 398 neutral tweets. The results of the study show that tweets about the elected president of the Republic of Indonesia received a neutral sentiment of 96% from the public on Twitter. The experiments conducted produced the highest accuracy of 96% using the Support Vector Machine (SVM) model with SMOTE implementation to handle imbalanced data. These results can be used to gauge public opinion and expectations of the elected president for the progress of the Republic of Indonesia over the next 5 years.

Keywords: President, Sentiment analysis, Support Vector Machine (SVM), Prabowo Subianto, Twitter.

Abstrak. Pemilihan umum sejatinya merupakan arena yang mawadahi para calon kandidat dalam kontestasi politik yang meraih kekuasaan dengan partisipasi rakyat untuk menentukan pilihan dan sebagai penyalur hak sosial dan politik masyarakat itu sendiri. Pada pemilu tahun 2024, khususnya untuk pemilihan presiden dan wakil presiden Indonesia, hasil menunjukkan bahwa kandidat presiden Republik Indonesia nomor urut 2, yaitu Prabowo Subianto dan Gibran Rakabuming Raka, terpilih sebagai presiden dan wakil presiden Republik Indonesia untuk periode 2024 – 2029 dengan perolehan suara sebesar 58%. Hal ini memicu beragam tanggapan dari seluruh masyarakat Indonesia, khususnya di media sosial Twitter (X). Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap tanggapan masyarakat Indonesia terkait terpilihnya Prabowo Subianto dan Gibran Rakabuming Raka sebagai presiden dan wakil presiden Republik Indonesia periode 2024 – 2029. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan data Twitter dan mengetahui sentimen masyarakat terhadap presiden terpilih pada pemilu 2024. Data yang terkumpul diseleksi secara acak untuk menghasilkan 427 data dari total data, yang terdiri dari 26 tweet positif, 3 tweet negatif, dan 398 tweet netral. Hasil penelitian menunjukkan bahwa tweet mengenai presiden terpilih Republik Indonesia mendapat sentimen netral sebesar 96% dari publik di Twitter. Percobaan yang dilakukan menghasilkan akurasi tertinggi sebesar 96% pada model Support Vector Machine (SVM) dengan implementasi SMOTE untuk menangani data yang tidak seimbang. Hasil ini dapat digunakan untuk melihat opini publik dan harapan terhadap presiden terpilih untuk kemajuan Negara Republik Indonesia selama 5 tahun ke depan.

Kata kunci: Presiden, Sentimen analisis, Support Vector Machine (SVM), Prabowo Subianto, Twitter.

PENDAHULUAN

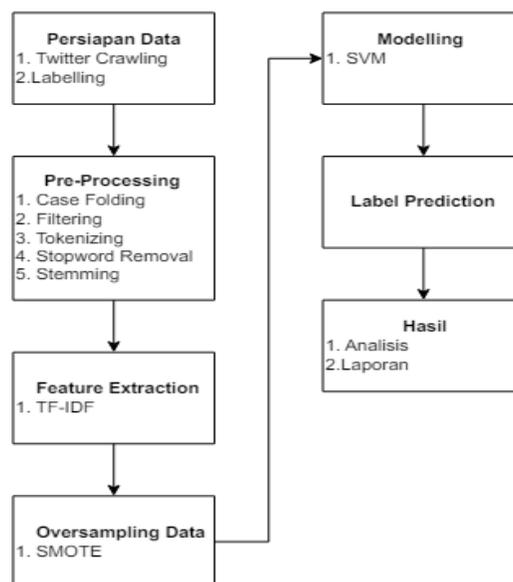
Pemilihan umum sebagai sarana pelaksanaan kedaulatan rakyat yang dilaksanakan secara langsung, umum, bebas, jujur, dan adil dengan menjamin prinsip perwakilan, akuntabilitas, dan legitimasi dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia tahun 1945 [1]. Pemilihan umum sejatinya merupakan sebuah arena yang mewadahi para calon kandidat dalam kontestasi politik yang meraih kekuasaan partisipasi rakyat untuk menentukan pilihan dan sebagai penyalur hak sosial dan politik masyarakat itu sendiri [2].

Di dalam pemilu tahun 2024 khususnya untuk pemilihan presiden dan wakil presiden Indonesia mendapati hasil bahwa kandidat presiden Republik Indonesia yang bernomor urut 2 yaitu Prabowo Subianto dan Gibran Rakabuming Raka terpilih sebagai presiden dan wakil presiden Republik Indonesia untuk periode 2024 – 2029. Dengan memperoleh suara sebesar 58%, mendapati beragam tanggapan dari seluruh masyarakat Indonesia khususnya di media sosial twitter (X).

Berdasarkan masalah tersebut, menjadi dasar penulis untuk melakukan penelitian yang berjudul “SENTIMEN ANALISIS PRESIDEN TERPILIH MENGGUNAKAN ALGORITMA SVM”. Dengan menganalisis beragam tanggapan masyarakat Indonesia terkait terpilihnya Prabowo Subianto dan Gibran Rakabuming Raka sebagai presiden dan wakil presiden Republik Indonesia periode 2024 – 2025. Dalam penelitian ini bertujuan untuk mengetahui kinerja algoritma Support Vector Machine (SVM) dalam mengklasifikasi data Twitter dan kemudian mengetahui tanggapan masyarakat untuk presiden terpilih Republik Indonesia pada pemilu 2024.

METODOLOGI PENELITIAN

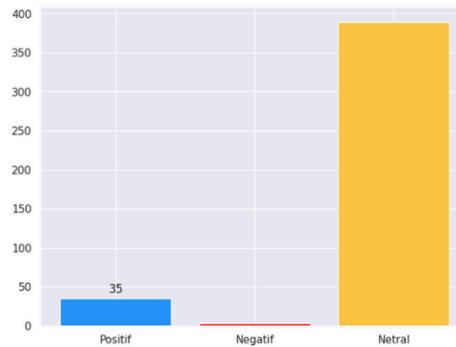
Metodologi penelitian ini menggunakan Python sebagai bahasa pemrograman dan beberapa library untuk penambangan teks. Ada enam langkah dalam penelitian: 1) Data Preparation yang mencakup crawl dan labeling data, 2) Pre-processing, 3) Fitur Extraction, 4) Over Sampling, 5) Modelling, 6) Label Prediction, 7) Hasil, serta analisis dan pelaporannya dapat dilihat pada Gambar 1.



GAMBAR 1. Metodologi Penelitian

Data Preparation

Data pengujian dan pelatihan dikumpulkan dengan mengambil data dari Twitter API. Pengumpulan data dilakukan mulai dari 14 Februari 2024 ketika pemilu dilaksanakan di Indonesia hingga penghitungan suara real-count pada tanggal 20 Maret 2024. Data yang sudah terkumpul akan diseleksi secara acak untuk menghasilkan 427 data dari total data. Dari pelabelan manual, ada 26 tweet positif, 3 tweet negatif, dan 398 tweet netral. Gambar 2 adalah ilustrasi distribusi dan informasi tweet.



GAMBAR 2. Tweet Distibution

Pre-processing

Pre-processing mencakup lima langkah: 1) case folding, 2) document filtering, 3) tokenizing, 4) stopword removal, 5) stemming. NLTK dan sastrawi digunakan dalam metodologi penelitian ini untuk pre-processing. Case Folding adalah proses konversi huruf yang sebelumnya terdapat huruf kapital menjadi huruf kecil semua [3]. Huruf ‘a’ sampai ‘z’ saja yang diterima untuk proses case folding. Filtering adalah tahap menghilangkan atau menghapus kata-kata yang tidak diperlukan atau dibutuhkan Tokenization adalah proses pemisahan teks dari kalimat atau paragraf menjadi bagian-bagian tertentu[4]. Pada tahap ini membagi teks dari kalimat berdasarkan spasi dan tanda baca untuk tahap analisis teks pada tahap selanjutnya. Stopword Removal adalah tahap membuang kata-kata yang tidak relevan didalam suatu kalimat yang sering muncul dan tidak ada artinya [5]. Untuk bahasa Indonesia kata-kata Stopword meliputi “yang”, “dengan”, “ke”, “dari”. Stemming adalah proses mengubah seluruh kata yang ada di dalam token menjadi kata dasar yang selanjutnya akan di proses menggunakan Sastrawi literature [6]. Table 1 menampilkan proses konversi kalimat yang terjadi di Tweet dengan sentimen netral setelah tahap pra-pemrosesan.

TABEL 1. Data Pre-processing

Original Tweet	@stillwithgemoy Presiden terpilih Prabowo Subianto nyatakan sikap tegas lanjutkan bangun Ibukota Nusantara
Case Folding	@stillwithgemoy presiden terpilih prabowo subianto nyatakan sikap tegas lanjutkan bangun ibukota nusantara
Filtering	stillwithgemoy presiden terpilih prabowo subianto nyatakan sikap tegas lanjutkan bangun ibukota nusantara
Tokenization	“stillwithgemoy”, “presiden”, “terpilih”, “prabowo”, “subianto”, “nyatakan”, “sikap”, “tegas”, “lanjutkan”, “bangun”, “ibukota”, “nusantara”
Stopword Removal	“stillwithgemoy”, “presiden”, “pilih”, “prabowo”, “subianto”, “nyata”, “sikap”, “lanjut”, “ibukota”, “nusantara”
Stemming	“stillwithgemoy”, “presiden”, “prabowo”, “subianto”, “bangun”, “ibukota”, “nusantara”

Fitur Extraction

Fitur Extraction adalah tahap yang digunakan sebagai dasar klasifikasi, sistem menghitung jumlah kemunculan term dan frekuensi invers dokumen menggunakan metode TF (Term Frequency) (1) dan TF-IDF (Term Frequency-Inverse Document Frequency) (2). Dalam Metode TF-IDF, term frekuensi berfokus pada pembobotan kata dengan menentukan frekuensi kemunculan kata dalam sebuah dokumen [7]. Rumus TF-IDF ditampilkan berikut ini.

$$w_{tf\ t, d} = \begin{cases} 1 + \log_{10} tf\ t, d & , \text{if } tf\ t, d > 0 \\ 0 & , \text{if } tf\ t, d = 0 \end{cases} \quad (1)$$

$$idf = \log\left(\frac{N}{df}\right) \quad (2)$$

Note:

$tf(t, d)$ = frequency of term t that appears in document d

N = The total of documents in the collection of selection document

df = The number of documents that contain term t

Oversampling Data

Synthetic Minority Over-sampling Technique (SMOTE) adalah metode pengambilan sampel data yang digunakan. SMOTE adalah pendekatan over-sampling yang digunakan untuk pemecahan masalah dan ketidakseimbangan kumpulan data [8]. Pengambilan sampel yang berlebihan akan meningkatkan ukuran data yang memungkinkan untuk menghasilkan lebih banyak fitur untuk pelatihan model dan dapat membantu meningkatkan akurasi model.

Modelling

Untuk membuat model pengklasifikasi, K-Fold Cross Validation digunakan dalam kumpulan data. Model Support Vector Machine diterapkan ke dalam training dataset dan diimplementasikan ke dalam test dataset untuk menghitung akurasi. Keakuratan model akan ditunjukkan pada bagian berikutnya.

Label Prediction

Hanya 427 data tweet dari presiden terpilih dataset yang diberi label secara manual. Model dipilih dari tahap sebelumnya untuk diterapkan dalam memprediksi label. Prediksi label ditampilkan pada bagian berikutnya

HASIL DAN PEMBAHASAN

Pada bagian ini menampilkan hasil evaluasi model Support Vector Machine (SVM) dalam dua skenario, yaitu tanpa SMOTE dan dengan SMOTE. Dari manual pelabelan, terdapat 26 tweet positif, 3 tweet negatif, dan 398 tweet netral. Tabel 2 menunjukkan keakuratannya, tabel 3 menunjukkan presisi, tabel 4 menunjukkan perolehan, tabel 5 menunjukkan F1 skor algoritma yang digunakan untuk klasifikasi. Secara keseluruhan model klasifikasi menggunakan model Support Vector Machine (SVM), SMOTE meningkatkan akurasi masing-masing model klasifikasi.

TABEL 2. Skore of accuracy evaluation

Classifier	Tanpa SMOTE	Dengan SMOTE	Increase
SVM	95%	96%	1%

TABEL 3. Skore of precision evaluation

Classifier	Tanpa SMOTE	Dengan SMOTE	Increase
SVM	65%	95%	30%

TABEL 4. Skore of recall evaluation

Classifier	Tanpa SMOTE	Dengan SMOTE	Increase
SVM	44%	96%	52%

TABEL 5. Skore of recall evaluation

Classifier	Tanpa SMOTE	Dengan SMOTE	Increase
SVM	0.57	0.97	0.40

Hasil sentimen analisis menggunakan model *Support Vector Machine* (SVM) dengan SMOTE dari dataset menunjukkan bahwa tweet Presiden terpilih menadapat sentimen netral sebesar 96%. Visualiasi dari kata-kata dalam tweet dilakukan menggunakan Word Cloud. Word Cloud untuk presiden terpilih dapat dilihat pada gambar 4.



GAMBAR 4. Word Cloud untuk sentimen analisis Presiden terpilih

KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa tweet presiden terpilih Republik Indonesia mendapat sentimen netral dari publik di Twitter, dengan sentimen netral sebesar 96%. Berdasarkan percobaan yang telah dilakukan, kami menemukan hasil akurasi tertinggi sebesar 96% pada model Support Vector Machine (SVM) dengan implementasi SMOTE untuk menangani data yang tidak seimbang. Hasil ini dapat digunakan untuk melihat opini publik untuk harapan presiden terpilih pada periode selanjutnya untuk kemajuan Negara Republik Indonesia 5 tahun kedepan.

DAFTAR PUSTAKA

- [1] A. B. Susilo and A. Sa'bani, "Mahar Politik Sebagai Dari Unsur Tindak Pidana," *Seminar Nasional Hukum Universitas Negeri Semarang*, vol. 4, p. 2, 2018.
- [2] J. Simamora, "Menyongsong Rezim Pemilihan Umum Serentak," *Rechtsvinding*, vol. ., p. 1, 2014.
- [3] G. Darmawan, S. Alam and M. I. Sulisty, "Jurnal Ilmiah Teknik dan Ilmu Komputer," *ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTORE MENGGUNAKAN METODE NAÏVE BAYES*, vol. 2, p. 3, 2023.
- [4] C. D. Manning, P. Raghavan and H. Schütze, "Cambridge University Press Cambridge," *Introduction To Information Retrieval*, 2009.
- [5] R. Rinandyaswara, Y. A. Sari and M. T. Furqon, "Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes," *Pembentukan Daftar Stopword*

- Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes*, p. 1, 2022.
- [6] B. Siswanto and Y. Dani, "in 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)," *Sentiment Analysis about Oximeter as Covid-19 Detection Tools on Twitter Using Sastrawi Library*, no. 10.1109/ICITACEE53184.2021.9617216. , pp. 161-164, 2021.
- [7] N. Hikmah and N. Ariyanti, "Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF," *Jurnal Teknologi Informasi dan Multimedia*, vol. 4, pp. 136 - 148, 2022.
- [8] W. Priatna, "Dampak Pengambilan Sampel Data untuk Optimalisasi Data Tidak Seimbang pada Klasifikasi Penipuan Transaksi E-Commerce," *Indonesian Journal of Computer Science*, vol. 13, p. 3072, 2024.
- [9] D. W. Seno and A. Wibowo, "Analisis Sentimen Data Twitter Tentang Pasangan Capres-Cawapres Pemilu 2019 Berbasis Metode Lexicon Dan Support Vector Machine," *Jurnal Ilmiah FIFO*, no. <http://dx.doi.org/10.22441/fifo.2019.v11i2.004> , pp. 144 - 155, 2019.
- [10] A. Handayani and I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM* , 2023.
- [11] M. Raihan, F. Sya, U. Enri and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes," *Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes*, vol. 9, no. 10.30865/jurikom.v9i2.3989., pp. 265 - 273, 2022.