

ANALISIS SENTIMEN BAKAL CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE PADA TWITTER

Oleh:

Wilson Winarto¹, Izmy Alwiah Musdar^{2*}, Hasniati³

^{1,2,3}Program Studi Teknik Informatika, STMIK Kharisma Makassar

²Program Studi Sistem Informasi, UIN Alauddin Makassar

e-mail: wilsonwinarto_20@kharisma.ac.id, ²izmyalwiah@gmail.com, hasniati@kharisma.ac.id

Abstrak: Indonesia merupakan negara demokrasi dengan penduduk yang sangat banyak dan pengguna Twitter terbesar di dunia. Pemilihan presiden 2024 di Indonesia menjadi suatu topik yang menarik bagi para pengguna Twitter. Tweet masyarakat yang berkaitan dengan para calon presiden dapat digunakan untuk melihat gambaran opini masyarakat terhadap para calon presiden. Banyaknya jumlah tweet yang masuk mengenai para calon presiden mendorong perlunya metode yang membantu untuk melihat opini masyarakat secara efektif. Salah satu metode yang dapat digunakan untuk mengklasifikasikan opini masyarakat secara efektif adalah Support Vector Machine (SVM). Metode ini digunakan untuk mengklasifikasikan suatu opini masyarakat termasuk dalam sentimen positif atau negatif dengan mencari hyperlane terbaik dari kedua kelas klasifikasi. Penambahan fungsi Kernel pada Support Vector Machine berguna untuk mengatasi data yang tidak terpisah secara linear. Penggunaan Metode K-Fold Cross Validation diperuntukan agar data dapat secara bergantian menjadi data uji sehingga meningkatkan akurasi. Pembobotan dilakukan menggunakan Term Frequency Document Inverse Frequency (TF-IDF). Evaluasi sistem dilakukan menggunakan confusion matrix untuk mengukur akurasi sistem dalam melakukan klasifikasi rata-rata akurasi menggunakan kernel linear. Hasil dari klasifikasi didapatkan Anies Baswedan mendapatkan akurasi sebesar 78,28% dan presisi sebesar 81,298%. Ganjar Pranowo mendapatkan akurasi sebesar 82,494% dan presisi sebesar 85,642%. Prabowo Subianto mendapatkan akurasi sebesar 83,904% dan presisi sebesar 86,22%.

Kata kunci: Sentimen Analisis, Pemilihan Presiden, Tweet, Klasifikasi, SVM.

Abstract: Indonesian is a democratic country with a huge population and the largest Twitter users in the world. The 2024 presidential election in Indonesia is an interesting topic for Twitter users. Public tweets related to presidential candidates can be used to see a picture of public opinion on presidential candidates. The large number of incoming tweets about presidential candidates encourages the need for methods that help to see public opinion effectively. One method that can be used to classify public opinion effectively is Support Vector Machine (SVM). This method will classify whether a public opinion will belong to a positive or negative sentiment by finding the best hyperlane from both classification classes. The addition of the Kernel function to the Support Vector Machine is useful for dealing with data that is not linearly separated. The use of the K-Fold Cross Validation Method is intended so that data can alternately become test data so as to increase accuracy. Weighting is done using Term Frequency Document Inverse Frequency (TF-IDF). System evaluation was carried out using confusion matrix to measure the accuracy of the system in classifying the average accuracy using a linear kernel. The results of the classification obtained by Anies Baswedan get an accuracy of 78.28% and a precision of 81.298%. Ganjar Pranowo received an accuracy of 82.494% and a precision of 85.642%. Prabowo Subianto received an accuracy of 83.904% and a precision of 86.22%.

Keywords: Sentiment Analysis, Presidential Election, Tweets, Classification, SVM.

* Corresponding author : Izmy Alwiah Musdar (izmyalwiah@gmail.com)

1. PENDAHULUAN

Perkembangan teknologi saat ini memudahkan pengguna media sosial untuk dapat mencurahkan isi hatinya atau pendapat mengenai suatu hal melalui platform dan twitter merupakan salah satu platform media sosial yang banyak digunakan oleh masyarakat. Dengan adanya banyaknya data yang tersebar dan tersedia di Twitter, analisis sentimen menjadi lebih mudah dilakukan. Data tweet yang dikumpulkan dapat digunakan untuk mengetahui opini masyarakat tentang calon presiden. Analisis sentimen digunakan untuk mengevaluasi perasaan atau opini yang terkandung dalam tweet tentang calon presiden.

Dari beberapa teknik analisis data dalam data *science*, analisis sentimen digunakan karena dapat menghemat waktu analisis data untuk mengetahui tingkat sentimen suatu topik. Hal ini dikarenakan analisis sentimen memanfaatkan perkembangan teknologi melalui berbagai *tools* khusus [1]. Analisis sentimen terhadap bakal calon presiden adalah proses menganalisis pandangan, opini, dan perasaan masyarakat terhadap bakal calon presiden yang bersangkutan, berdasarkan data teks dari berbagai sumber seperti media sosial, ulasan berita, dan platform online lainnya. Melalui analisis sentimen ini, dapat diidentifikasi bagaimana publik merespon dan merasa terhadap bakal calon presiden, apakah sentimen tersebut cenderung positif, atau negatif. Hal ini memberikan wawasan yang berharga bagi bakal calon presiden dan tim kampanye mereka dalam memahami popularitas, reputasi, dan isu-isu yang diangkat oleh masyarakat, serta dapat membantu mengarahkan strategi kampanye dan komunikasi untuk mengatasi permasalahan yang muncul dan memperkuat dukungan dari pemilih. Meskipun analisis sentimen dapat memberikan wawasan tambahan, keputusan pemilihan tetap bergantung pada berbagai faktor lain, termasuk visi, program, dan rekam jejak calon presiden, serta partisipasi dan penilaian akhir dari pemilih dalam proses pemilihan yang demokratis.

Analisis sentimen merupakan proses yang digunakan untuk mengevaluasi perasaan atau opini yang terkandung dalam teks [2]. Pada penelitian ini analisis sentimen di terapkan untuk mengetahui sentimen publik terkait calon presiden 2024. Data opini diperoleh dari aplikasi twitter dengan mengumpulkan tweet selama rentang waktu tertentu

Analisis sentimen dapat menggunakan algoritma *Naïve Bayes* [3][4][5][6], atau *Support Vector Machine* [7][4][5][8][9][10]. Pada penelitian ini peneliti menggunakan algoritma SVM karena memiliki tingkat akurasi cukup baik dalam mengkategorikan tweet [11][4][5]. Dikarenakan data tweet merupakan jenis teks yang informal dengan banyak sekali *noise* dan tidak mengindahkan aturan tata bahasa. Oleh karena itu, dalam penelitian ini digunakan Algoritma *Support Vector Machine* (SVM) untuk Sentimen Analisis Bakal Calon Presiden 2024.

2. PREPROCESSING

Preprocessing adalah Teknik yang digunakan untuk mengubah data mentah dalam format yang berguna dan efisien. Preprocessing dilakukan untuk menghilangkan kecacatan atau kekurangan pada data dan mempersiapkan data untuk tahap analisis selanjutnya.

Cleansing

Cleansing atau pembersihan data adalah tahapan preprocessing pada pengolahan data yang bertujuan untuk membersihkan data dari kesalahan, duplikasi, atau ketidakkonsistenan yang mungkin terjadi pada data mentah.

Case Folding

Case folding adalah salah satu Teknik dalam tahap preprocessing pada pengolahan teks yang bertujuan untuk mengubah semua karakter dalam teks menjadi huruf kecil (*lowercase*). Teknik ini biasa digunakan untuk mengurangi jumlah fitur dalam teks dan mencegah duplikasi fitur karena perbedaan karakter besar kecil (*Case Sensitivity*)

Stopword Removal

Stopwords adalah kata-kata yang umumnya tidak memiliki nilai makna yang signifikan pada konteks tertentu dan sering muncul dalam teks. Kata-kata seperti “dan”, “atau”, “yang”, “dari”, dan sebagainya.

Tokenizing

Tokenization adalah proses pemisahan teks atau dokumen menjadi unit-unit yang lebih kecil yang disebut token. Tujuan dari tokenization adalah untuk memproses dan menganalisis teks dengan lebih efisien, karena setiap *token* dianggap sebagai unit terkecil dalam teks yang dapat diolah oleh komputer.

Stemming

Stemming adalah proses pemangkasan akhiran kata dalam Bahasa yang digunakan agar hanya tersisa kata dasarnya. Stemming dilakukan untuk mengurangi varian kata dalam dokumen teks, sehingga memudahkan proses analisis dan perhitungan frekuensi kata.

3. ALGORITMA SVM

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vladimir N. Vapnik dan Alexey Ya. Chervonenkis pada tahun 1992 sebagai rangkaian harmonis konsep-konsep unggulan dalam bidang *pattern recognition*. SVM adalah algoritma machine learning yang bekerja atas prinsip *Structural Risk Minimization (SRM)* dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah *class* pada *input space* [12]. Bisa dilihat dari Gambar 1 menunjukkan model dari Algoritma SVM [7].

Adapun kelebihan dan kelemahan dari algoritma SVM adalah sebagai berikut [7]:

Kelebihan:

- Cocok untuk ruang dimensi tinggi
- Efektif untuk kasus dimana jumlah dimensi lebih besar dari jumlah sampel
- Hemat memori, karena menggunakan training point dari fungsi keputusan (support vector)
- Bekerja relatif baik ketika ada margin pemisahan yang jelas antar kelas.

Kelemahan:

- Algoritma SVM tidak cocok untuk dataset dalam jumlah yang besar karena membutuhkan waktu training yang lama.
- SVM tidak bekerja dengan baik ketika dataset memiliki lebih banyak noise misalnya kelas target terjadi tumpang tindih.
- Jika jumlah fitur untuk setiap titik data melebihi jumlah sampel data training, SVM memiliki performa yang kurang baik.
- Karena support vector classifier bekerja dengan meletakkan titik data di atas dan di bawah hyperplane, tidak ada kejelasan probabilistik untuk klasifikasi tersebut. Hal ini dapat menyebabkan beban komputasi yang tinggi.

Fungsi matematis yang digunakan pada proses transformasi algoritma SVM dikenal dengan istilah fungsi kernel (kernel function). Fungsi kernel digunakan untuk mengklasifikasikan data non-linier. Caranya adalah dengan mengubah data non-linear menjadi data linier dan kemudian membentuk hyperplane [7].

Ada tiga jenis fungsi kernel yaitu [13]:

1. Kernel Linear

Kernel linear adalah fungsi kernel yang paling sederhana. Kernel digunakan ketika data yang dianalisis sudah terpisah secara linier. Kernel linier cocok ketika terdapat banyak fitur dikarenakan pemetaan ke ruang dimensi yang lebih tinggi tidak benar-benar meningkatkan kinerja. Persamaan untuk fungsi kernel adalah:

$$K(x, xi) = \text{sum}(x * xi)$$

2. Kernel Polynomial

Kernel Polynomial adalah fungsi kernel yang cocok untuk digunakan dalam SVM dan kernelisasi lainnya, di mana kernel mewakili kesamaan vektor sampel pelatihan dalam ruang fitur. Kernel Polynomial juga cocok untuk memecahkan masalah klasifikasi pada dataset pelatihan yang dinormalisasi. Persamaan untuk fungsi kernel polynomial adalah:

$$K(x, xi) = 1 + \text{sum}(x * xi)^d$$

3. Kernel RBF (Radial Basic Function)

Kernel RBF atau juga disebut kernel Gaussian adalah konsep kernel yang paling banyak digunakan untuk memecahkan masalah klasifikasi data yang tidak dapat dipisahkan secara linear. Kernel ini dikenal memiliki performa yang baik dengan parameter tertentu, dan hasil dari peltihan memiliki nilai error yang kecil dibandingkan dengan kernel lainnya. Rumus persamaan untuk fungsi kernel RBF adalah:

$$K(x, xi) = 1 + sum(x * xi)^d$$

4. K-FOLD CROSS VALIDATION

K-fold cross validation merupakan salah satu dari Teknik yang difungsikan untuk memilah data menjadi train data serta test data. Teknik ini banyak diterapkan peneliti karena didapati mengurangi bias yang didapatkan didalam pengambilan sebuah sampel. K-fold cross validation berlaku continue membagi data-data menjadi train data dan test data, sehingga setiap data mendapat kesempatan menjadi test data [6]. Contohnya pada gambar 1 [14]:



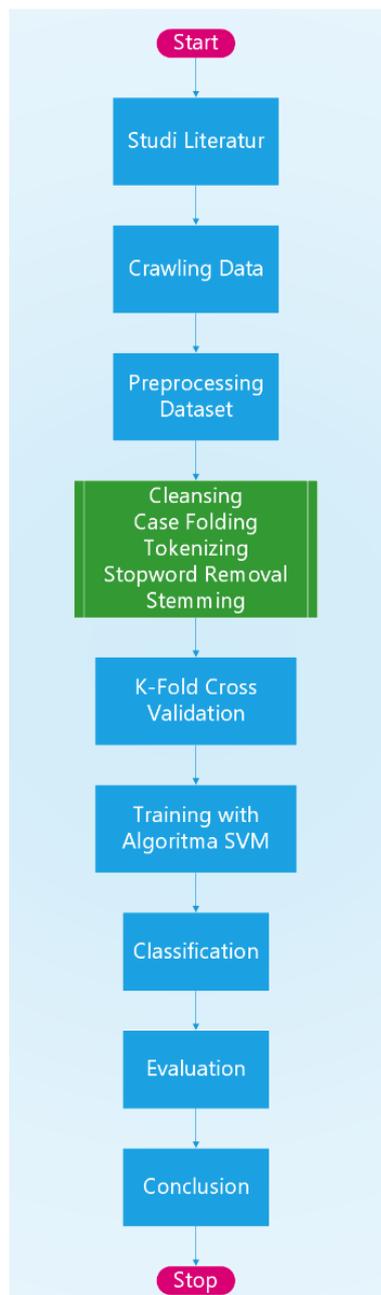
Gambar 1. K-Fold Cross Validation
(sumber: mltut.com, 2022)

5. METODE PENELITIAN

Penelitian ini dimulai dari tanggal 24 Mei 2023 sampai dengan 31 Juli 2023. Pada penelitian ini Jenis data yang saya gunakan ialah Data *Primer* yang bersumber dari tweet pengguna aplikasi twitter. Data tersebut berupa id, tweet, lokasi, dan *username* dan disimpan kedalam file berformat csv.

Pengumpulan data dilakukan dengan cara studi literatur, yaitu dengan mengumpulkan dan mempelajari referensi dari berbagai sumber yang berkaitan dengan Sentimen Analisis, Twitter API, Teknik Crawling, Pre-Processing, Support Vector Machine (SVM), Kernel Linear, dan K-Fold Cross Validation. Selain itu, untuk dataset yang digunakan, diperoleh melalui proses crawling menggunakan Twitter API dan library tweepy version 3.10. Dataset yang berhasil dikumpulkan disimpan dalam file csv.

Penelitian ini dilakukan dalam beberapa tahap. Tahapan penelitian dimulai dari studi literatur, identifikasi dan perumusan masalah, pengambilan dataset, pra pemrosesan dataset, pelatihan dataset, pengujian dan penyusunan kesimpulan diakhir penelitian. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Alur Penelitian

- 1) Studi literatur dilakukan agar mendapatkan informasi awal terkait topik penelitian yang dipilih.
- 2) Pengambilan data dari Twitter menggunakan API Twitter.
- 3) Data tersebut dilakukan *preprocessing* dengan beberapa tahapan yaitu *cleansing, case folding, tokenizing, stopwords removal, stemming*.
- 4) Teknik pemilahan data digunakan sebanyak 5 *fold* yaitu terdapat 5 variasi data yang berbeda untuk proses klasifikasi. Setiap proses *training* dan *testing* menggunakan jumlah data *training* sebanyak 80% dan jumlah data *testing* sebanyak 20% yang berbeda setiap *fold*-nya.
- 5) Data tweet diubah menjadi bentuk numerik dengan metode pembobotan TF-IDF. Hasil TF-IDF digunakan untuk proses *training* dan klasifikasi.
- 6) Pembuatan model dengan menggunakan algoritma *Support Vector Machine (SVM)* dengan penggunaan fungsi kernel Linear. Implementasi SVM menggunakan modul *svm* dari *library scikit-learn*.
- 7) Pada tahapan klasifikasi, data uji diberikan label secara otomatis sesuai dengan model yang dihasil pada proses *training*.
- 8) Pada tahapan evaluasi, dilakukan pengukuran peformasi algoritma menggunakan akurasi dan presisi.
- 9) Pada tahapan kesimpulan, implementasi *wordcloud*, dan ekstraksi topik untuk mengetahui topik yang sedang hangat dibicarakan oleh para pengguna twitter.

6. HASIL DAN PEMBAHASAN

6.1 Crawling

Modul operasi *regular expression (RegEx)* dari *Python* digunakan untuk mengerjakan proses prapengolahan atau normalisasi teks. Hasil *crawling* didokumentasikan dan disimpan di repositori Github. Pengumpulan data yang dilakukan menggunakan rentang waktu *tweet* dari tanggal 24 Mei 2023 sampai dengan 31 Juli 2023. Jumlah data yang didapatkan yakni 16.936 tweet dengan 5814 tweet dari Anies Baswedan, 5875 tweet dari Ganjar Pranowo, dan 5247 tweet dari Prabowo Subianto sebelum Prapengolahan, setelah indentifikasi data duplikasi maka jumlah data yang didapatkan yakni 10448 tweet dengan 2565 tweet dari Anies Baswedan, 3913 tweet dari Ganjar Pranowo, dan 3970 tweet dari Prabowo Subianto. Hasil dari pengambilan data ini ditampilkan pada Tabel 1.

Tabel 1: Hasil Penarikan Data

Dataset	Jumlah Data	
	Sebelum Prapengolahan	Setelah Prapengolahan
Anies Baswedan	5814	2565
Ganjar Pranowo	5875	3913
Prabowo Subianto	5247	3970
TOTAL	16936	10448

Pada proses penarikan data peneliti juga mengambil beberapa atribut data lainnya yang dianggap penting untuk diperlukannya peninjauan ulang terhadap data tweet yang telah dikumpulkan sebelumnya, tetapi untuk penggunaan pada tahapan selanjutnya atribut yang digunakan hanya teks. Jenis atribut data tweet yang berhasil dicrawl disajikan pada Tabel 2.

Tabel 2: Atribut Dataset Tweet Twitter

Nama Atribut	Penjelasan Atribut	Tipe Data Atribut
id	Menentukan ID pengguna.	Numeric
created_at	Waktu dan tanggal dari pembuatan tweet tersebut	Date and time
text	Tweet dari pengguna Twitter	String
screen_name	Nama dari akun yang membuat tweet tersebut	String
location	Lokasi User	String

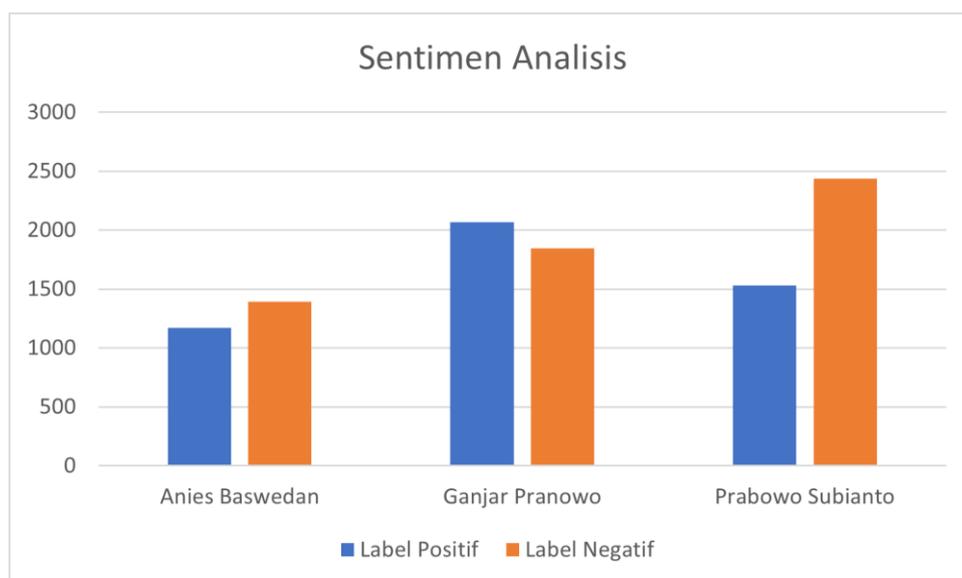
6.2 Preprocessing

Pada tahap preprocessing dilakukan dengan beberapa langkah dimulai dari cleansing, case fold, tokenizing, stopwords removal, dan stemming. Data hasil preprocessing ditunjukkan pada Tabel 3.

Tabel 3: Sample Preprocessing

Tahap	Sampel
Crawling	RT @Yurissa_Samosir: Kader Partai PAN di Riau Tetap Dukung ANIES BASWEDAN https://t.co/GqMb91ra9l
	Haus akan pujian jelang Pilpres! Kelompok Anies Baswedan buat berita palsu dengan foto editan, mengaku diundang ker... https://t.co/hMGpF7TWub
Cleansing	Kader Partai Riau Tetap Dukung ANIES BASWEDAN
	Haus akan pujian jelang Pilpres Kelompok Anies Baswedan buat berita palsu dengan foto editan mengaku diundang
Case Fold	kader partai riau tetap dukung anies baswedan
	haus akan pujian jelang pilpres kelompok anies baswedan buat berita palsu dengan foto editan mengaku diundang
Tokenizing	['kader', 'partai', 'riau', 'tetap', 'dukung', 'anies', 'baswedan']
	['haus', 'akan', 'pujian', 'jelang', 'pilpres', 'kelompok', 'anies', 'baswedan', 'buat', 'berita', 'palsu', 'dengan', 'foto', 'editan', 'mengaku', 'diundang']
Stopwords Removal	['kader', 'partai', 'riau', 'dukung']
	['haus', 'pujian', 'jelang', 'pilpres', 'kelompok', 'berita', 'palsu', 'foto', 'editan', 'mengaku', 'diundang']
Stemming	['kader', 'partai', 'riau', 'dukung']
	['haus', 'puji', 'jelang', 'pilpres', 'kelompok', 'berita', 'palsu', 'foto', 'edit', 'aku', 'undang']

Setelah dataset melalui *preprocessing*, maka dilakukan pelabelan sentimen untuk setiap tweet secara manual. Label yang diberikan berupa sentimen positif atau negatif. Rangkuman hasil pelabelan ditunjukkan pada Gambar 3.



Gambar 3. *Sentiment Positive & Negative*

6.3 K-fold Cross Validation

Untuk mengetahui performa model dengan lebih baik dan spesifik, model klasifikasi kemudian dievaluasi menggunakan *k-fold cross validation*. Data yang digunakan untuk evaluasi adalah data latih (teks dan label) dari pengujian awal. Algoritma yang digunakan memanfaatkan model *K-Fold* dari *scikit-learn*. Dalam penelitian ini, jumlah iterasi *k-fold* yang digunakan adalah 5. Artinya untuk setiap calon, data latih yang digunakan dalam pengujian awal dibagi menjadi 80% untuk data latih dan 20% data uji evaluasi. Untuk setiap iterasi, *k-fold cross validation* akan mengacak kombinasi data latih dan data uji evaluasi.

6.4 Training SVM

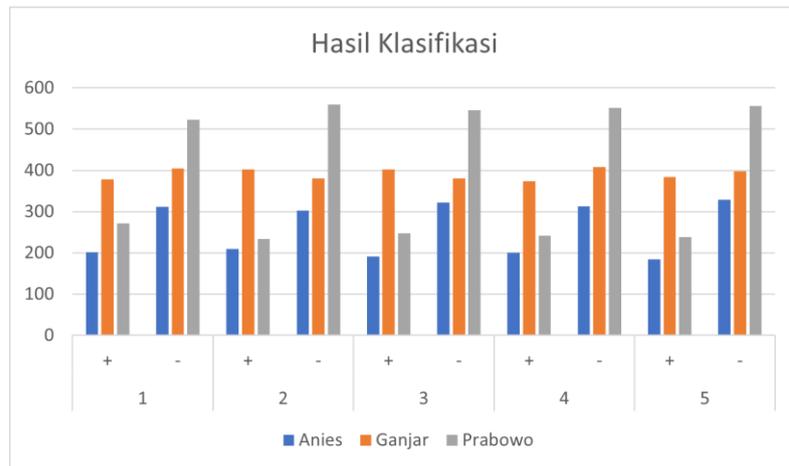
Perubahan data atribut yang telah melewati metode pembobotan TF-IDF dapat berubah menjadi matrik numerik dengan dimensi jumlah tweet x jumlah korpus kata. Setelah atribut diubah menjadi bentuk numerik dilanjutkan dengan perubahan data dari label dengan menggunakan metode *encoder*. Setiap label diubah menjadi numerik dimana pemetaan perubahan data label tersebut dapat ditampilkan pada Tabel 4.

Tabel 4: *Hasil Pemetaan Konversi Label*

Sebelum	Sesudah
Negatif	0
Positif	1

6.5 Classification

Proses klasifikasi adalah proses penentuan kelas dari setiap tweet yang belum diketahui kelasnya. Penentuan kelas dilakukan secara otomatis oleh model yang telah ditraining pada tahap Training SVM. Hasil proses klasifikasi ditunjukkan pada Gambar 4.

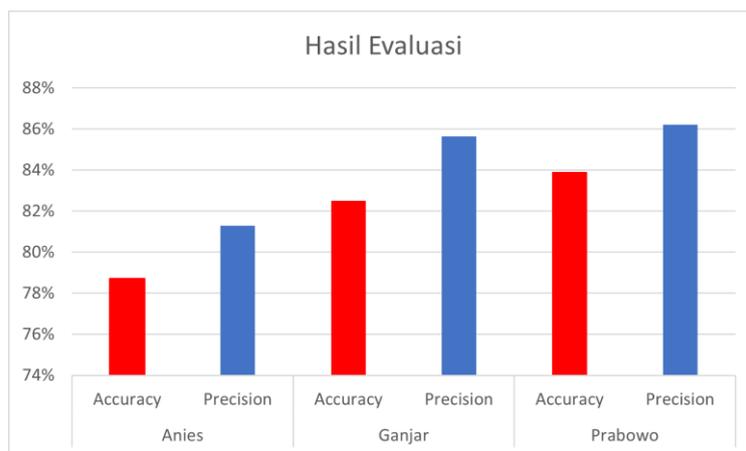


Gambar 4: Classification

Hasil klasifikasi pada Diagram Gambar 4 menunjukkan bahwa untuk 5 fold jumlah tweet dengan label negative terbanua adalah tweet tentang Prabowo. Jumlah label positif dan negatif untuk tweet terkait Ganjar cenderung seimbang, sedangkan tweet terkait Anies untuk semua fold memiliki jumlah label negatif yang lebih besar dibandingkan label positif.

6.6 Evaluation

Setelah proses klasifikasi yang menggunakan model klasifikasi, maka selanjutnya penulis menghasilkan confusion matrix agar mempermudah visualisasi data dan mengukur tingkat akurasi dan presisi data. Berdasarkan hasil visualisasi dengan confusion matrix, maka penulis mengukur akurasi dan presisi hasil klasifikasi untuk setiap *fold*. Akurasi dan presisi hasil klasifikasi disajikan pada diagram Gambar 5.



Gambar 5: Evaluation

Hasil Evaluasi pada Diagram Gambar 5 menunjukkan bahwa untuk tingkat akurasi dan presisi tertinggi terdapat pada tweet tentang Prabowo sebesar 83,904% dan 86,22%, berikutnya pada tweet tentang Ganjar sebesar 82,494% dan 85,642% lalu tingkat akurasi dan presisi terendah terdapat pada tweet tentang Anies sebesar 78,28% dan 81,298%.

6.7 Conclusion

Pada tahapan Conclusion penulis melakukan evaluasi dimulai dari calon Anies Baswedan, Ganjar Pranowo, dan Prabowo Subianto menggunakan teknik *Wordcloud* dan *Topic Extraction* yang dibicarakan pengguna melalui tweet terkait dengan calon presiden. Hal ini dilakukan untuk mengetahui topik yang dibicarakan oleh pengguna twitter tentang masing-masing calon presiden, dimulai dari Gambar 6 sampai dengan Gambar 11.

6.7.1 Anies Baswedan



Gambar 6: *Wordcloud Anies Baswedan Positive & Negative*

- Topik 0 : lebih elektabilitas naik smrc agusyudhoyono pdemokrat demokrat bersama pasang milik
- Topik 1 : indonesia rakyat baik aniespendukunghilafah kelam aniesnasdemout butuh pimpin ubah bangsa
- Topik 2 : dukung deklarasi jawa pilpres purnawirawan polri rawan partai warga ratus
- Topik 3 : persen gubernur mantan jakarta jawa neruskepdkhe ketua prabowountukindonesia simulasi semuadukungprabowo
- Topik 4 : presiden calon rasyid alumni wakil jokowi menang deklarasi satu prediksi
- Topik 5 : ubah koalisi satu cawapres partai usung nasdem bacapres capres umum
- Topik 6 : demokrat pdemokrat agusyudhoyono menang kriteria peluang bacawapres aku bicara partai
- Topik 7 : pilih dajjal orang selamat cerdas cawapres agama responden warga ogah
- Topik 8 : survei capres kalah hasil unggul airlangga duet puan suara simulasi
- Topik 9 : jejak rekam jakarta jokowi pimpin orang guys aniespresiden bangun bukti

Gambar 7: *Extraction Topic Anies Baswedan*

Gambar 6 menampilkan visualisasi *output* sentimen positif dan negatif Anies Baswedan, pada gambar tersebut terdapat kata-kata yang paling sering muncul yaitu kata “dukung”, “ubah”, “presiden”, “indonesia”, “partai”, “jakarta”, dan lain-lain.

6.7.2 Ganjar Pranowo



Gambar 8: *Wordcloud Ganjar Pranowo Positive & Negative*

- Topik 0 : indonesia maju bangun depan bangsa butuh wujud hasil modal ganjarmentangtotal
- Topik 1 : jawa gubernur jateng persen bangun warga jalan desa sekolah provinsi
- Topik 2 : dukung deklarasi masyarakat rawan partai pilpres perindo capresganjar ganjarpranowo milenial
- Topik 3 : presiden calon bacapres juang pdip pilih capresganjar partai terus usung
- Topik 4 : jokowi lanjut program terus perintah komitmen bangun presidut ganti kerja
- Topik 5 : semarang festival hadir giat daring warga acara gpfestivalsemarang kota sambut
- Topik 6 : menang pilpres optimis juang persentase onic Pertamina onad drastis satu
- Topik 7 : ganjarcapres ganjarpranowo warga ganjarpresiden kalbar pilih baik sambut milu bersamaganjar
- Topik 8 : capres pilih survei erick thohir cawapres elektabilitas damping hasil pdip
- Topik 9 : pimpin sosok rakyat prestasi butuh anak peduli nilai muda toleransi

Gambar 9: *Extraction Topic Ganjar Pranowo*

DAFTAR PUSTAKA

- [1] S. Bhatia, M. Sharma, and K. K. Bhatia, "Sentiment Analysis and Mining of Opinions," *Stud. Big Data*, vol. 30, no. May, pp. 503–523, 2018, doi: 10.1007/978-3-319-60435-0_20.
- [2] "Analisis Sentimen (Sentiment Analysis): Definisi, Tipe dan Cara Kerjanya." <https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/> (accessed Aug. 04, 2023).
- [3] M. Raihan, F. Sya, U. Enri, and T. N. Padilah, "Analisis Sentimen Terhadap Bakal Calon Presiden 2024 dengan Algoritma Naïve Bayes," vol. 9, no. 2, pp. 265–273, 2022, doi: 10.30865/jurikom.v9i2.3989.
- [4] P. Studi and M. Informatika, "Perbandingan Naïve Bayes Classifier dan Support Vector Machine Untuk Klasifikasi Judul Artikel," vol. 1, no. 2, pp. 90–93, 2016.
- [5] N. Wayan *et al.*, "Naïve Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis," pp. 2–4, 2013.
- [6] T. Ridwansyah, "KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *Media Online*, vol. 2, no. 5, pp. 178–185, 2022, [Online]. Available: <https://djournals.com/klik>.
- [7] "Penjelasan Lengkap Algoritma Support Vector Machine (SVM) - Trivusi," 2022. <https://www.trivusi.web.id/2022/04/algoritma-svm.html> (accessed May 13, 2023).
- [8] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, pp. 19–25, 2017, doi: 10.29207/resti.v1i1.11.
- [9] D. T. Lukmana, S. Subanti, and Y. Susanti, "Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine," no. 2002, pp. 154–160, 2019.
- [10] H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Menggunakan Algoritma Support Vector Machine," vol. 4, no. 1, 2020.
- [11] F. Rahutomo, P. Y. Saputra, and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film MenggunaRahutomo, F., Saputra, P. Y. and Fidyawan, M. A. (2018) 'Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine', *Jurnal Informatika Polinema*," *J. Inform. Polinema*, vol. 4, no. 2, p. 93, 2018.
- [12] D. A. Pisner and D. M. Schnyer, "Support vector machine," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 101–121, 2019, doi: 10.1016/B978-0-12-815739-8.00006-7.
- [13] "Apa itu Kernel Trick? Pengertian dan Jenis-jenis Fungsi Kernel SVM - Trivusi." <https://www.trivusi.web.id/2022/04/fungsi-kernel-svm.html> (accessed May 14, 2023).
- [14] "K Fold Cross-Validation in Machine Learning? How does K Fold Work?" <https://www.mltut.com/k-fold-cross-validation-in-machine-learning-how-does-k-fold-work/> (accessed May 13, 2023).

LAMPIRAN DATASET GITHUB CODE

<https://github.com/wilsonkharisma/Skripsi.git>