

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PENERAPAN METODE SVM PADA KLASIFIKASI SENTIMEN TERHADAP ANIES BASWEDAN SEBAGAI BAKAL CALON PRESIDEN 2024

TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik
Pada Jurusan Teknik Informatika



Oleh

RAMADANU PUTRA

NIM. 11950115180



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU**

2023

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSETUJUAN

**PENERAPAN METODE SVM PADA KLASIFIKASI
SENTIMEN TERHADAP ANIES BASWEDAN SEBAGAI
BAKAL CALON PRESIDEN 2024**

TUGAS AKHIR

Oleh

RAMADANU PUTRA
NIM. 11950115180

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 13 Juli 2023

Pembimbing I,



Yusra, S.T., M.T.
NIP. 19840123 201503 2 001

Pembimbing II,



Muhammad Fikry, S.T., M.Sc.
NIP. 19801018 20070 1 002



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

**PENERAPAN METODE SVM PADA KLASIFIKASI
SENTIMEN TERHADAP ANIES BASWEDAN SEBAGAI
BAKAL CALON PRESIDEN 2024**

Oleh

RAMADANU PUTRA
NIM. 11950115180

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 13 Juli 2023
Mengesahkan,
Ketua Jurusan,

Dekan,

Dr. Hartono, M.Pd.
NIP. 19640301 199203 1 003

Iwan Iskandar, M.T.
NIP. 19821216 201503 1 003

DEWAN PENGUJI

Ketua : Surya Agustian, S.T., M.Kom.
Pembimbing I : Yusra, S.T., M.T.
Pembimbing II : Muhammad Fikry, S.T., M.Sc.
Penguji I : Dr. Lestari Handayani, M.Kom.
Penguji II : Reski Mai Candra, S.T., M.Sc.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Ramadanu Putra
NIM : 11950115180
Tempat/Tgl. Lahir : Duri, 26 Desember 2000
Fakultas : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi : Penerapan Metode SVM pada Klasifikasi Sentimen terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024

Menyatakan dengan sebenar-benarnya bahwa:

1. Penulisan Skripsi dengan judul sebagaimana tersebut diatas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu, Skripsi saya ini, saya nyatakan bebas dari plagiat
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 17 Juli 2023
Yang membuat pernyataan



Ramadanu Putra
NIM. 11950115180

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

Assalammu'alaikum wa rohmatullohi wa barokatuh.

Alhamdulillah robbil'alamin, Puji syukur penulis ucapkan kepada Allah SWT yang telah memberikan nikmat dan kesempatan sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul **“Penerapan Metode SVM pada Klasifikasi Sentimen terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024”** sebagai salah satu syarat untuk menyelesaikan pendidikan Strata 1 (S1) Teknik Informatika pada Fakultas Sains dan Teknologi di Universitas Islam Negeri Sultan Syarif Kasim Riau. Dalam penyusunan tugas akhir ini penulis mendapat banyak doa dan dukungan, bimbingan, arahan, serta masukan dari berbagai pihak. Oleh karena itu penulis mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag. selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd. selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M.T. selaku Kepala Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Fadhilah Syafria, S.T., M.Kom selaku Koordinator Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Ibu Yusra, S.T., M.T. selaku Dosen Pembimbing I Tugas Akhir, yang telah memberikan ilmu, arahan serta bimbingan dalam proses penyelesaian Tugas Akhir.
6. Bapak Muhammad Fikry S.T., M.Sc selaku Dosen Pembimbing II Tugas Akhir, yang telah memberikan ilmu, arahan serta bimbingan dalam penyelesaian Tugas Akhir.
7. Ibu Dr. Lestari Handayani, S.T., M.Kom selaku Dosen penguji I Tugas Akhir, yang telah banyak memberikan saran demi kemajuan dan penyempurnaan dalam pengerjaan Tugas Akhir.
8. Bapak Reski Mai Candra, S.T., M.Sc selaku Dosen penguji II Tugas Akhir,

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang telah banyak memberikan saran demi kemajuan dan penyempurnaan dalam pengerjaan Tugas Akhir.

9. Ayahanda dan Ibunda yang tiada hentinya memberikan do'a, dukungan, motivasi, biaya, cinta dan kasih sayang yang tulus untuk kesuksesan anak-anaknya.
10. Teman-teman seperjuangan Jurusan Teknik Informatika terkhusus kelas TIF G angkatan 2019 yang selalu membantu, dan selalu ada buat teman-temannya semoga langkah kita selalu dimudahkan Allah SWT.
11. Untuk semua teman-teman yang tidak bisa disebutkan satu persatu. Mereka selalu memberikan bantuan dengan perannya masing-masing, sehingga penulis dapat menyelesaikan laporan ini.

Kami menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat kami harapkan untuk kesempurnaan laporan ini. Akhirnya kami berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya.

Wassalamu'alaikum wa rohmatullohi wa barokatuh.

Pekanbaru, Juli 2023

Penulis

Penerapan Metode SVM pada Klasifikasi Sentimen terhadap Anies Baswedan sebagai Bakal Calon Presiden 2024

Ramadanu Putra^{1*}, Yusra², Muhammad Fikry³

¹Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl. HR. Soebrantas No.Km. 15, Kec. Simpang Baru, Kota Pekanbaru, Riau 28293

e-mail: ¹11950115180@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id

Submitted Date: 2023-06-01

Reviewed Date: 2023-06-01

Revised Date: 2023-06-08

Accepted Date: 2023-06-16

Abstract

Twitter is one of the increasingly popular and rapidly growing platforms for sharing opinions and activities. In the context of the 2024 presidential election in Indonesia, Anies Baswedan's name has been proposed as one of the presidential candidates by the National Democratic Party (NasDem) on October 3, 2022. Therefore, this study aims to perform sentiment classification on Twitter social media related to user opinions on Anies Baswedan as a 2024 presidential candidate. This research uses the Support Vector Machine (SVM) method to perform sentiment classification. The dataset used consists of 3400 Twitter data that has been labeled positive (2130 tweets) and negative (1270 tweets). The labeling process is done manually using crowdsourcing techniques, and the level of agreement obtained with the kappa metric is 0.68, indicating a strong level of agreement. Before classification, text preprocessing was performed on the data. The dataset was divided into training data and test data with a ratio of 90:10. To overcome data imbalance, data balancing is carried out using the SMOTE (Synthetic Minority Oversampling Technique) method and a threshold of 5 is selected. The SVM model with RBF kernel using $C = 9$ and $\gamma = 2$ parameter pairs has produced good performance in the validation and evaluation stages. The test results show accuracy of 90.61%, precision of 90.67%, recall of 90.61% and f1-score of 90.61%.

Keywords: Classification; Presidential Candidates; Sentiment; Support Vector Machine; RBF Kernel

Abstrak

Twitter adalah salah satu platform yang semakin populer dan berkembang pesat untuk berbagi pendapat dan aktivitas. Dalam konteks pemilihan presiden 2024 di Indonesia, nama Anies Baswedan telah diusung sebagai salah satu bakal calon presiden oleh Partai Nasional Demokrat (NasDem) pada tanggal 3 Oktober 2022. Oleh karena itu, penelitian ini bertujuan untuk melakukan klasifikasi sentimen pada media sosial Twitter terkait pendapat pengguna terhadap Anies Baswedan sebagai bakal calon presiden 2024. Penelitian ini menggunakan metode *Support Vector Machine* (SVM) untuk melakukan klasifikasi sentimen. *Dataset* yang digunakan terdiri dari 3400 data Twitter yang telah diberi label positif (2130 *tweet*) dan negatif (1270 *tweet*). Proses pelabelan dilakukan secara manual dengan menggunakan teknik *crowdsourcing*, dan tingkat kesepakatan yang diperoleh dengan metrik Kappa adalah sebesar 0,68, menunjukkan tingkat kesepakatan tergolong kuat. Sebelum melakukan klasifikasi, dilakukan proses *text preprocessing* pada data. *Dataset* dibagi menjadi data latih dan data uji dengan perbandingan 90:10. Untuk mengatasi ketidakseimbangan data, dilakukan penyeimbangan data dengan Metode SMOTE (*Synthetic Minority Oversampling Technique*) dan dilakukan pemilihan *threshold* sebesar 5. Model SVM dengan *kernel* RBF yang menggunakan pasangan parameter $C=9$ dan $\gamma=2$ telah menghasilkan performa yang baik dalam tahap validasi dan evaluasi. Hasil pengujian menunjukkan akurasi sebesar 90,61%, presisi sebesar 90,67%, recall sebesar 90,61% dan f1-score sebesar 90,61%.



Keywords: Klasifikasi; Bakal Calon Presiden; Sentimen; Support Vector Machine; Kernel RBF

1. Pendahuluan

Teknologi Internet berkembang dengan pesat setiap tahunnya. Dengan perkembangan teknologi Internet, sangat membantu pemenuhan kebutuhan akan informasi. Perkembangan teknologi turut berdampak pada peningkatan penggunaan Internet yang sangat cepat. Berdasarkan APJII, Pada tahun 2023 terdapat 215,63 juta pengguna Internet di Indonesia (APJII, 2023).

Media sosial menjadi sarana komunikasi di masyarakat. Pada saat ini, hampir seluruh masyarakat Indonesia menggunakan media sosial sebagai sarana untuk mengakses dan membagikan informasi kepada khalayak umum (Reyvaldi et al., 2022). Salah satu media sosial yang paling sering digunakan oleh masyarakat Indonesia sampai saat ini, yaitu Twitter. Twitter adalah sebuah platform media sosial microblogging yang memungkinkan pengguna untuk menulis dan mempublikasikan aktivitas serta pendapat. Berdasarkan laporan Datareportal, Twitter sebagai platform media sosial berada di urutan ke-enam yang paling sering digunakan di Indonesia pada tahun 2023 dengan jumlah pengguna sebanyak 24 juta (Kem, 2023).

Pemilu Presiden dijadwalkan pada Rabu, 14 Februari 2024 untuk masa bakti 2024-2029 yang bertujuan memilih Presiden dan Wakil Presiden Republik Indonesia (Purnawan, 2022). Beberapa kandidat yang tepat menggantikan presiden Indonesia pada tahun 2024, menjadi pembicaraan media berita. Anies Baswedan menjadi salah satu nama bakal calon presiden yang telah diusung oleh Partai Nasional Demokrat (NasDem) pada tanggal 3 Oktober 2022 (Dirgantara, 2022).

Sentimen atau pendapat masyarakat pengguna Twitter dapat dilihat melalui *tweet* atau cuitan terhadap Anies Baswedan sebagai bakal calon presiden 2024. *Tweet* tersebut dapat dianalisis untuk memperoleh informasi sentimen masyarakat terhadap Anies Baswedan sebagai bakal calon presiden 2024. Klasifikasi sentimen merupakan sebuah metode untuk menentukan nilai sentimen atas sebuah objek dengan tujuan mengetahui pendapat terhadap suatu masalah apakah pendapat tersebut positif atau negatif (Pohan et al., 2022).

Machine Learning merupakan cabang ilmu komputer yang memungkinkan komputer untuk mempelajari pola dan membuat keputusan berdasarkan data, tanpa perlu diprogram secara

eksplisit. Salah satu algoritma yang digunakan dalam *machine learning* adalah *Support Vector Machine* (SVM), yang termasuk ke dalam kategori *supervised learning*. Prinsip kerja SVM didasarkan pada konsep *Structural Risk Minimization*, Metode ini bertujuan untuk menemukan *hyperplane* terbaik yang dapat secara optimal memisahkan dua kelas atau lebih dalam ruang input (Pohan et al., 2022)

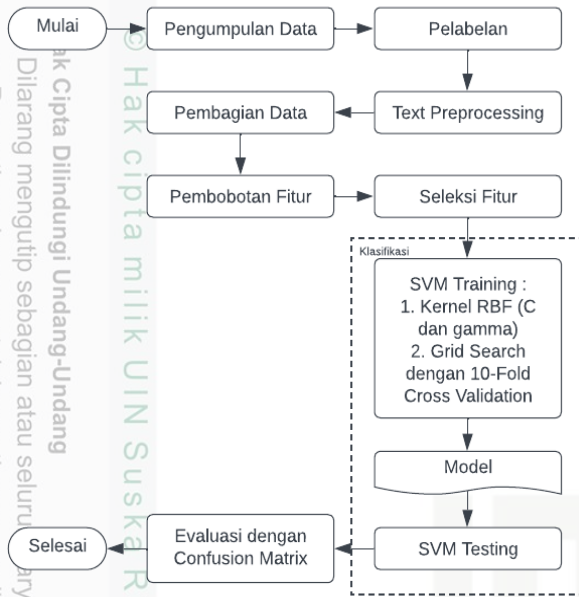
Penelitian mengenai klasifikasi sentimen terhadap tokoh politik telah dilakukan sebelumnya yang dapat dijadikan referensi dalam penelitian ini. Pada penelitian (Fitriyani & Hartanto, 2020), metode SVM digunakan untuk analisis sentimen terhadap Anies Baswedan didapatkan hasil akurasi 95,9%, presisi 94,49%, dan recall 96,4%. Pada penelitian (Herlinawati et al., 2020), metode *Naïve Bayes* Dan SVM diterapkan untuk analisis sentimen terhadap Joko Widodo diperoleh hasil algoritma SVM dengan akurasi 95.42%, dan *Naïve Bayes* dengan akurasi 94.07%. Kemudian penelitian (Nugroho et al., 2021), diketahui bahwa metode SVM dengan *kernel* linear memiliki kinerja yang lebih baik dibandingkan metode *Naïve Bayes*, Akurasi SVM mencapai 82%, sedangkan *Naïve Bayes* hanya mencapai 69%.

Terdapat penelitian lain yang juga melakukan klasifikasi sentimen dengan metode SVM. Pada penelitian (Yunas et al., 2018), metode SVM digunakan untuk klasifikasi *tweet e-commerce* didapatkan hasil akurasi sebesar 94%. Pada penelitian (Chairunnisa et al., 2022) metode SVM diterapkan untuk klasifikasi sentimen ulasan pengguna aplikasi PeduliLindungi. Hasilnya didapatkan akurasi sebesar 93%.

Rumusan permasalahan pada penelitian ini adalah bagaimana menerapkan metode *Support Vector Machine* (SVM) dalam melakukan klasifikasi sentimen terhadap Anies Baswedan sebagai bakal calon Presiden 2024. Penelitian ini menggunakan *dataset* berjumlah 3400 *tweet* yang telah dikelompokkan menjadi label positif dan label negatif. Penelitian ini bertujuan untuk memperoleh akurasi, presisi, recall, dan f1-score yang optimal dalam melakukan klasifikasi sentimen terhadap Anies Baswedan sebagai bakal calon Presiden 2024.

2. Metodologi Penelitian

Gambar 1 memperlihatkan Tahapan Penelitian.



Gambar 1 Tahapan Penelitian

Berikut ini adalah penjelasan Gambar 1 terkait tahapan penelitian:

a. Pengumpulan Data

Pada penelitian ini, pengumpulan data dilakukan menggunakan teknik *crawling* dengan menggunakan bahasa pemrograman Python dan Twitter API. *Dataset* yang dikumpulkan terdiri dari *tweet* berbahasa Indonesia yang mengandung kata kunci terkait Anies Baswedan sebagai bakal calon presiden 2024.

b. Pelabelan

Proses pelabelan data *tweet* dilakukan secara manual dengan melibatkan 2 orang atau lebih sebagai annotator. Untuk mencapai kesepakatan dalam pelabelan, digunakan metode Fleiss Kappa yang mengukur tingkat kesepakatan antara dua atau lebih annotator (Indraini & Ernawati, 2022). Berikut merupakan tabel tingkat kesepakatan berdasarkan standar interpretasi nilai Kappa dapat dilihat pada Tabel 1

Tabel 1 Tingkat Kesepakatan

Fleiss Kappa	Agreement Level
< 0.00	Poor
0.00 to 0.20	Slight
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 1.00	Perfect

Tabel 1 memberikan panduan umum untuk menginterpretasikan Nilai Fleiss Kappa dalam konteks kesepakatan antara dua atau lebih

annotator yang melibatkan data kategorikal. Semakin tinggi nilai Fleiss Kappa, semakin tinggi tingkat kesepakatan antara annotator tersebut. Interpretasi nilai Kappa dapat membantu dalam mengevaluasi tingkat kesepakatan dan keandalan hasil penilaian atau pengukuran yang dilakukan oleh annotator yang berbeda.

c. Text Preprocessing

Tweet yang telah dikumpulkan akan melalui beberapa tahapan. Tujuannya adalah untuk mengubah data menjadi lebih terstruktur, sehingga memudahkan pengolahan data. Berikut ini adalah penjelasan tahapan *text preprocessing*:

Cleaning, yaitu upaya untuk menghilangkan *noise* dari teks. Entitas *tweet* yang dibersihkan yaitu mention, hashtag dan URL.

Case Folding, yaitu mengubah semua huruf menjadi huruf kecil. Tujuannya adalah agar kata-kata dengan penggunaan huruf kecil yang sama dapat diperlakukan secara seragam dalam analisis teks, sehingga meningkatkan akurasi dan konsistensi hasil analisis.

Tokenizing, yaitu memisahkan teks menjadi token-token yang bermakna. Tujuannya memungkinkan pemahaman lebih lanjut tentang struktur dan komposisi teks.

Normalisasi, yaitu mengubah kata tidak baku menjadi kata baku. Tujuannya memastikan konsistensi dan keseragaman.

Negation Handling, yaitu untuk menangani negasi terhadap suatu kata. Proses *negation handling* penting dalam analisis sentimen, karena dapat mempengaruhi interpretasi dan evaluasi keseluruhan teks ketika ada negasi yang terlibat.

Stopword Removal, yaitu untuk menghapus kata hubung. Dengan menghapus kata-kata hubung yang umum, kita dapat fokus pada kata-kata kunci yang memberikan makna dan informasi yang lebih penting dalam teks yang dianalisis.

Stemming, yaitu untuk mengubah kata ke dalam bentuk dasarnya. Hal ini bertujuan untuk mengurangi variasi kata yang sebenarnya memiliki makna yang sama secara dasar.

d. Pembagian Data

Tujuan pembagian data latih dan uji adalah untuk mengevaluasi kinerja model yang telah dilatih dan menguji kemampuannya dalam menggeneralisasi pola pada data yang belum pernah dilihat sebelumnya. Data dibagi menjadi 2 subset, yaitu 90% data latih dan 10% data uji.

e. Pembobotan Fitur

Pembobotan fitur dengan TF-IDF (Term Frequency-Inverse Document Frequency) dilakukan untuk mengubah kata menjadi representasi numerik yang dapat digunakan oleh model SVM. Tujuannya adalah untuk memungkinkan SVM mempelajari pola dan hubungan antara fitur-fitur numerik tersebut. SVM membutuhkan input berupa data numerik untuk melatih model. Dalam perhitungan nilai TF-IDF pada data latih dan data uji, jumlah total dokumen yang digunakan merujuk pada jumlah data latih. Dalam proses ini, hasil nilai TF-IDF akan menentukan relevansi atau signifikansi fitur atau kata tersebut pada korpus (Pravina et al., 2019). Kata-kata dengan bobot yang lebih tinggi akan dianggap lebih relevan atau signifikan dan mampu memberikan kontribusi yang lebih besar dalam proses analisis atau klasifikasi data.

f. Seleksi Fitur

Pada penelitian ini, dilakukan seleksi fitur pada model yang dihasilkan dengan metode frequency-based dan menentukan nilai ambang batas (*threshold*) pada Document Frequency (DF).

g. Klasifikasi Support Vector Machine

SVM (*Support Vector Machine*) adalah sebuah metode yang banyak digunakan untuk klasifikasi dan regresi, baik dalam masalah linear maupun non-linear. Salah satu keunggulan utama SVM adalah kemampuannya dalam melakukan pemisahan linear pada data input yang memiliki dimensi tinggi, bahkan ketika data tersebut bersifat non-linear. Keunggulan ini dicapai dengan menggunakan fungsi kernel yang diperlukan dan memilih kernel yang sesuai dengan baik dalam proses klasifikasi (Herlinawati et al., 2020).

h. Evaluasi dengan Confusion Matrix

Pada penelitian ini *Confusion matrix* digunakan untuk mengevaluasi kinerja model. *Confusion matrix* merupakan tabel yang digunakan untuk menganalisis hasil prediksi suatu model dengan membandingkan prediksi terhadap data yang sebenarnya.

3. Hasil dan Pembahasan

Hasil penelitian yang telah dilakukan ialah bagaimana menerapkan metode SVM pada klasifikasi sentimen terhadap Anies Baswedan sebagai bakal calon presiden 2024.

3.1 Pengumpulan Data

Dataset yang digunakan berisikan informasi terkait sentimen masyarakat terhadap Anies Baswedan sebagai bakal calon presiden 2024. Data *tweet* diambil dari rentang waktu 21 Oktober – 23 November 2022 sebanyak 5355 *tweet*. Setelah itu, dilakukan seleksi data dengan menghapus *tweet* duplikat dan *tweet* yang bukan opini, sehingga data akhirnya terdiri dari 3400 *tweet*.

3.2 Pelabelan

Data *tweet* dalam penelitian ini diberi label ke dalam dua kelas sentimen, yakni sentimen positif dan sentimen negatif. Pelabelan dilakukan secara manual dengan teknik *crowdsourcing*. Hasil pelabelan pada data *tweet* dapat disebut dengan kesepakatan 5 annotator. Kemudian, Fleiss Kappa digunakan untuk mengukur tingkat kesepakatan yang telah dicapai. Dengan menggunakan teknik pemberian label ini, terdapat potensi untuk mendapatkan label dengan kualitas yang lebih baik. Tabel 2 memperlihatkan Pelabelan Data 5 annotator.

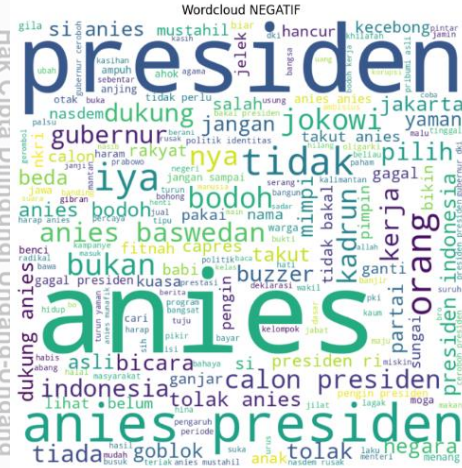
Tabel 2 Pelabelan data 5 annotator

Tweet	1	2	3	4	5	Label
Pembohong seperti Anies tak layak jadi Presiden #AniesNasdemOut	N	N	N	P	N	Negatif
@marlina_idha @aniesbaswedan Anies Baswedan The next leader of Indonesia Presiden RI 2024	P	P	P	P	P	Positif

Berdasarkan Tabel 2, hasil pelabelan data dilakukan oleh 5 annotator, dan kemudian label mayoritas yang diberikan akan diambil. Setelah menghitung kesepakatan dengan menggunakan 5 annotator, diperoleh nilai Kappa sebesar 0,68. Maka tingkat kesepakatannya dikategorikan sebagai *Substantial*, yang menunjukkan tingkat kesepakatan tergolong kuat.

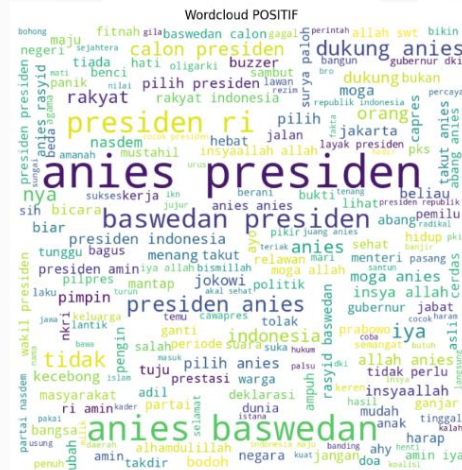
Pelabelan yang dilakukan secara manual menghasilkan label positif sebanyak 2130 *tweet* dan label negatif sebanyak 1270 *tweet*. Gambar 2 dan 3 memperlihatkan kata-kata yang sering muncul pada *tweet* berlabel negatif dan *tweet* berlabel positif.

1. Diarangi mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Diarangi mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2 Wordcloud Negatif

Pada Gambar 2, terdapat beberapa kata yang sering muncul seperti "tolak", "buzzer", "kadrun", "bodoh", "gagal", dan beberapa kata lainnya.



Gambar 3 Wordcloud Positif

Pada Gambar 3, terdapat beberapa kata yang sering muncul seperti "pimpin", "menang", "pilih", "dukung", "prestasi", dan beberapa kata lainnya.

3.3 Text Preprocessing

Tabel 3 berisikan hasil dari tahapan *text preprocessing* yang telah dilakukan.

Tabel 3 Tahapan Text Preprocessing

Tahapan Preprocessing	Sebelum Perubahan	Setelah Perubahan
Cleaning	Pembohong seperti Anies tak layak jadi Presiden #AniesNasdemOut #TenggelamkanNasdem	Pembohong seperti Anies tak layak jadi Presiden

	#TenggelamkanAnies Nasdem #GoodByeNasdemAnies #AniesGabenerPembohong https://t.co/FmiMf23o2j	
Case Folding	Pembohong seperti Anies tak layak jadi Presiden	pembohong seperti anies tak layak jadi presiden
Tokenizing		['pembohong', 'seperti', 'anies', 'tak', 'jadi', 'presiden']
Normalisasi	['pembohong', 'seperti', 'anies', 'tak', 'layak', 'jadi', 'presiden']	['pembohong', 'seperti', 'anies', 'tidak', 'layak', 'jadi', 'presiden']
Negation Handling	['pembohong', 'seperti', 'anies', 'tidak', 'layak', 'jadi', 'presiden']	['pembohong', 'seperti', 'anies', 'jelek', 'jadi', 'presiden']
Stopword Removal	['pembohong', 'seperti', 'anies', 'jelek', 'jadi', 'presiden']	['pembohong', 'anies', 'jelek', 'presiden']
Stemming	['pembohong', 'anies', 'jelek', 'presiden']	['bohong', 'anies', 'jelek', 'presiden']

3.4 Pembagian Data

Pada tabel 4 memperlihatkan perbandingan jumlah data latih dan data uji setelah data dibagi.

Tabel 4 Pembagian Dataset

Pembagian Tweet	Kelas Tweet	Jumlah	Total Data
Data latih (90% dataset)	Positif	1917	3060
	Negatif	1143	
Data uji (10% dataset)	Positif	213	340
	Negatif	127	

3.5 Penyeimbangan Data

Setelah dilakukan pembagian *dataset*, terjadi ketidakseimbangan kelas atau label pada data training. Hal ini dapat berdampak pada performa model yang dihasilkan, karena model cenderung memprediksi kelas mayoritas dan mengabaikan kelas minoritas. Oleh karena itu, diperlukan penyeimbangan *dataset* menggunakan Metode SMOTE (*Synthetic Minority Oversampling Technique*). Teknik ini mensintesis sampel baru agar jumlah sampel kelas minoritas menjadi setara dengan kelas mayoritas, sehingga dapat menangani ketidakseimbangan kelas tersebut (Putri & Kharisudin, 2022). Pada tabel 5 memperlihatkan jumlah data setelah *oversampling*.

Tabel 5 Data setelah *oversampling*

Pembagian Tweet	Kelas Tweet	Jumlah	Total Data
Data latih (90% dataset)	Positif	1917	3834
	Negatif	1917	
Data uji (10% dataset)	Positif	213	426
	Negatif	213	

3.6 Pembobotan Fitur

Dalam konteks ini, hanya menampilkan 5 fitur dan 10 baris dari hasil pembobotan fitur TF-IDF. Tabel 6 dan 7 memperlihatkan hasil pembobotan fitur TF-IDF training dan TF-IDF testing.

Tabel 6 Hasil TF-IDF Training

index	anies	nasionalis	ri	bakal	hancur
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.1602
10	0.0	0.0	0.0	0.0	0.0

Tabel 7 Hasil TF-IDF Testing

index	anies	nasionalis	ri	bakal	hancur
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0
9	0.0	0.3902	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.2722

3.7 Seleksi Fitur

Pada penelitian ini, *threshold* yang digunakan yaitu 5. Hanya fitur-fitur yang muncul pada 5 dokumen atau lebih yang dipertahankan sebagai fitur pada model yang dihasilkan. Penentuan nilai *threshold* ini didasarkan pada evaluasi performa model dan eksperimen yang telah dilakukan.

3.8 SVM Training

Pada penelitian ini, digunakan *kernel* RBF (*Radial Basis Function*) untuk memetakan data ke dalam ruang fitur yang lebih tinggi. *Kernel* RBF memungkinkan SVM untuk mengatasi masalah pemisahan linier yang kompleks dengan memproyeksikan data ke dalam ruang fitur yang tidak linier (Widayani & Harliana, 2021).

Pada tahap SVM training, dilakukan *grid search* dengan *10-fold cross validation* pada data latih menggunakan *kernel* RBF serta berbagai kombinasi nilai parameter C dan γ . Rentang nilai yang digunakan untuk C dan γ yaitu $1 \leq C \leq 10$ dan $1 \leq \gamma \leq 10$. Selain itu, juga digunakan rentang *threshold* antara 2 hingga 20. Tabel 6 memperlihatkan rataan akurasi dari kombinasi nilai parameter dan *threshold*.

Tabel 8 Rataan akurasi dari kombinasi nilai parameter dan threshold

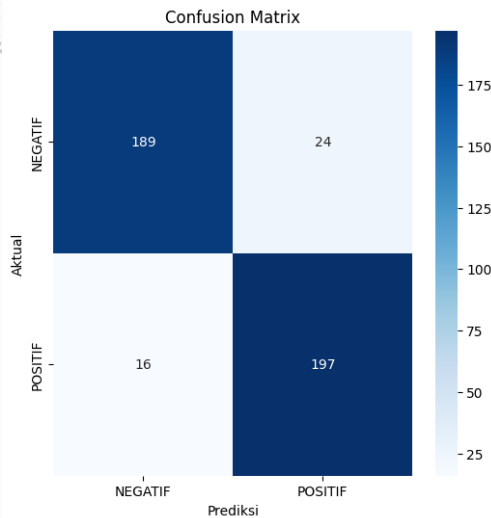
Threshold	C	γ	Akurasi
2	9	2	90,82%
3	9	2	90,90%
4	7	2	90,95%
5	9	2	90,98%
6	7	2	90,77%
7	9	3	90,53%
8	8	3	90,74%
9	8	3	90,66%
10	8	3	90,58%
11	8	3	90,51%
12	8	3	90,40%
13	5	3	90,35%
14	6	3	90,38%
15	6	3	90,51%
16	7	3	90,32%
17	2	3	90,32%
18	5	3	90,48%
19	5	3	90,45%
20	5	3	90,35%

Berdasarkan Tabel 6 diperoleh hasil validasi terbaik menggunakan *grid search* dengan *10-fold cross validation*. Hasil tersebut menunjukkan

bahwa parameter terbaik adalah $C=9$ dan $\gamma=2$ pada *threshold* 5 dengan rata-rata akurasi sebesar 90,98%. Model dengan akurasi terbaik dari kombinasi nilai C dan γ akan diaplikasikan pada tahap pengujian SVM.

3.9 SVM Testing

Pada tahap pengujian SVM, model terbaik yang telah diperoleh sebelumnya diterapkan dengan menerapkan *threshold*. Evaluasi dilakukan dengan *Confusion matrix* terhadap 426 data uji. Gambar 4. Memperlihatkan hasil confusion matrix.



Gambar 4 Hasil Confusion Matrix

Berdasarkan Gambar 4, Hasil *Confusion matrix* menunjukkan bahwa model berhasil memprediksi 221 positif dan 205 negatif. Hasil matrix menunjukkan True Positif (TP) sebanyak 197, False Positif (FP) sebanyak 24, True Negatif (TN) sebanyak 189 dan False Negatif (FN) sebanyak 16. Hasil pengujian menunjukkan akurasi sebesar 90,61%, presisi sebesar 90,67%, recall sebesar 90,61% dan f1-score sebesar 90,61%.

4. Kesimpulan

Pada penelitian ini, metode SVM diterapkan untuk melakukan klasifikasi sentimen terhadap Anies Baswedan sebagai bakal calon presiden 2024. Dataset yang digunakan berjumlah 3400 dengan 2130 *tweet* untuk sentimen positif dan 1270 *tweet* untuk sentimen negatif. Metode SVM dengan *kernel* RBF berhasil menghasilkan model terbaik dengan akurasi 90,98% menggunakan kombinasi nilai $C=9$ dan $\gamma=2$. Setelah dilakukan pengujian, diperoleh hasil akurasi sebesar 90,61%, presisi sebesar 90,67%, recall sebesar 90,61% dan f1-score sebesar 90,61%.

5. Saran

Setelah mengevaluasi hasil penelitian yang telah dilakukan, terdapat berbagai potensi pengembangan yang dapat dieksplorasi. Berikut adalah beberapa saran yang dapat diberikan untuk pengembangan penelitian ini:

- Mempertimbangkan *kernel* SVM yang berbeda untuk mengetahui mana yang paling cocok untuk kasus klasifikasi sentimen terhadap Anies Baswedan.
- Melakukan optimasi parameter untuk meningkatkan kinerja model.
- Mempertimbangkan jumlah data yang lebih banyak, guna meningkatkan akurasi dari hasil analisis.

References

- APJII, T. (2023). *Survei APJII Pengguna Internet di Indonesia Tembus 215 Juta Orang*. <https://apjii.or.id/berita/d/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang>
- Chairunnisa, C., Ernawati, I., & Santoni, M. M. (2022). Klasifikasi Sentimen Ulasan Pengguna Aplikasi PeduliLindungi di Google Play Menggunakan Algoritma Support Vector Machine dengan Seleksi Fitur Chi-Square. *JURNAL INFORMATIK Edisi Ke-18*, 1, 69–79. <https://doi.org/10.52958/iftk.v17i4.4594>
- Dirgantara, A. (2022). *Nasdem Resmi Deklarasikan Anies Baswedan Jadi Capres 2024*. <https://nasional.kompas.com/read/2022/10/03/10440681/nasdem-resmi-deklarasikan-anies-baswedan-jadi-capres-2024>
- Fitriyani, N., & Hartanto, A. (2020). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Support Vector Machine. *MEANS (Media Informasi Analisa Dan Sistem)*, 5(1), 8–12. <https://doi.org/10.54367/means.v5i1.615>
- Herlinawati, N., Yuri, Y., Siti, F., Gata, W., & Samudi. (2020). Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine. *CESS (Journal of Computer Engineering System and Science)*, 5(2), 293–298. <https://doi.org/10.24114/cess.v5i2.18186>
- Indraini, A. N., & Ernawati, I. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, 14(1),

68–80.

<https://doi.org/10.22441/fifo.2022.v14i1.007>
 Kem, S. (2023). *DIGITAL 2023: INDONESIA*.
<https://datareportal.com/reports/digital-2023-indonesia>

Nugroho, G., Murdiansyah, D., & Lhaksmana, K. (2021). Analisis Sentimen Pemilihan Presiden Amerika 2020 di Twitter Menggunakan Naïve Bayes dan Support Vector Machine. *E-Proceeding of Engineering*, 8(5), 10106–10115. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15727/15440>

Pohan, R., Ratnawati, D., & Arwani, I. (2022). Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4924–4931. <http://j-ptiik.ub.ac.id>

Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(3), 2789–2797. <http://j-ptiik.ub.ac.id>

Purnawan, H. (2022). *Sah, Pemilu 14 Februari 2024 Disepakati DPR, Pemerintah, dan Penyelenggara Pemilu*.
<https://bawaslu.go.id/id/berita/sah-pemilu->

14-februari-2024-disepakati-dpr-pemerintah-dan-penyelenggara-pemilu-1

Putri, M. I., & Kharisudin, I. (2022). Analisis Sentimen Pengguna Aplikasi Marketplace Tokopedia Pada Situs Google Play Menggunakan Metode Support Vector Machine (SVM), Naïve Bayes, dan Logistic Regression. *PRISMA, Prosiding Seminar Nasional Matematika*, 5, 759–766. <https://journal.unnes.ac.id/sju/index.php/prisma/>

Reyvaldi, P., Septian, R., Alkholis, Z., Widyanto, M. R., Prayoga, M. A., Zebua, Y., Daeli, E. K., Ajisaputro, H. E., Saputri, R. R., Hakim, A. F., & Prasetyo, S. M. (2022). Pengenalan Internet Sehat Untuk Membangun Generasi Yang Cerdas, Modern Dan Religius Di Asrama Yatim & Dhu'afa Yayasan Sahabat Yatim Mandiri. *JATIMIKA*, 3(2), 337–340. <http://openjournal.unpam.ac.id/index.php/JATIMIKA/article/view/20837>

Widayani, W., & Harliana. (2021). Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains Dan Informatika*, 7(1), 20–27. <https://doi.org/10.34128/jsi.v7i1.268>

Yunas, A. H., Yusra, & Fikry, M. (2018). Klasifikasi Tweet E-Commerce dengan Menggunakan Metode Support Vector Machine. *Jurnal CoreIT*, 4(2). <https://doi.org/10.24014/coreit.v4i2.5205>