

IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN TERHADAP ELEKTABILITAS CALON PRESIDEN 2024

SKRIPSI

**OLEH
ALFI RAHMA DANI
NIM. 200601110055**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**IMPLEMENTASI ALGORITMA *SUPPORT*
VECTOR MACHINE UNTUK ANALISIS SENTIMEN
TERHADAP ELEKTABILITAS CALON PRESIDEN 2024**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
ALFI RAHMA DANI
NIM. 200601110055**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

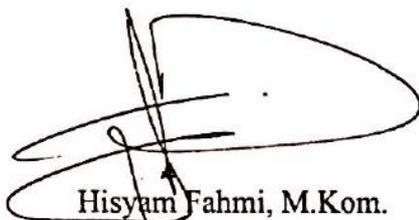
IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN TERHADAP ELEKTABILITAS CALON PRESIDEN 2024

SKRIPSI

Oleh
Alfi Rahma Dani
NIM.200601110055

Telah Disetujui Untuk Diuji
Malang, 24 Juni 2024

Dosen Pembimbing I



Hisyam Fahmi, M.Kom.
NIP. 19890727 201903 1 018

Dosen Pembimbing II



Ach. Nashiehuddin, M.A.
NIP. 19730705 200003 1 002

Mengetahui,

Konita Program Studi Matematika



Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

IMPLEMENTASI ALGORITMA *SUPPORT VECTOR MACHINE* UNTUK ANALISIS SENTIMEN TERHADAP ELEKTABILITAS CALON PRESIDEN 2024

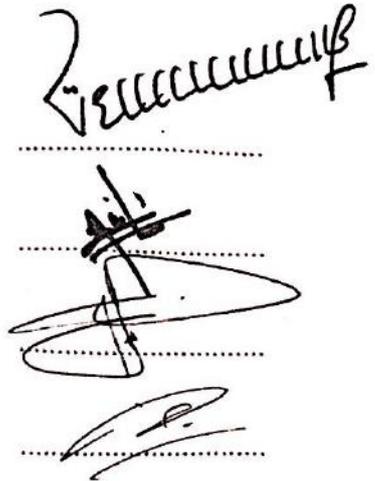
SKRIPSI

Oleh
Alfi Rahma Dani
NIM.200601110055

Telah Dipertahankan di Depan Dewan Penguji Skripsi dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan untuk Memperoleh Gelar Sarjana Matematika(S.Mat.)

Tanggal, 26 Juni 2024

Ketua Penguji : Evawati Alisah, M.Pd.
Anggota Penguji 1 : M. Nafie Jauhari, M.Si.
Anggota Penguji 2 : Hisyam Fahmi, M.Kom.
Anggota Penguji 3 : Ach. Naschichuddin, M.A.



Mengetahui,
Ketua Program Studi Matematika


Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Alfi Rahma Dani

NIM : 200601110055

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Implementasi Algoritma Support Vector Machine Untuk
Analisis Sentimen Terhadap Elektabilitas Calon Presiden 2024

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 26 Juni 2024

Yang membuat pernyataan,



Alfi Rahma Dani

NIM. 200601110055

MOTO

“God has perfect timing, never early, never late, It takes a little patience and it takes a lot faith, but it’s a worth the wait.”

“Trust to Allah for everything no matter what. You lose trust to Allah, you win you trust to Allah, you gain you trust to Allah, you have a problem you trust to Allah, things are not going your way, you thank him even more and you talk to him, thats a very good habit to talk to Allah.”

"Letakan aku dalam hatimu, maka aku akan meletakanmu dalam hatiku"
(Q.S Al-Baqarah:152)

PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

Ayah dan Ibu tercinta Syamsul Huda dan Nur Siti yang senantiasa memberikan doa, dukungan, nasihat dan motivasi terbaik untuk kesuksesan penulis. Kedua adik penulis Aqila Syahida Anaila dan Ayunda Zalfa Az-zahra yang dengan kasih sayangnya senantiasa mendukung penuh penulis dalam segala hal. Serta sahabat-sahabat penulis yang selalu memberikan bantuan dan semangat dalam menyelesaikan skripsi ini.

KATA PENGANTAR

Alhamdulillah robbil ‘alamin, puji syukur senantiasa peneliti ucapkan kepada Allah Swt. atas limpahan rahmat, taufik, serta karunia-Nya sehingga skripsi ini dapat diselesaikan dengan lancar. Shalawat dan salam tidak lupa senantiasa tercurahkan kepada baginda Nabi Agung Muhammad saw. yang telah membawa kita dari zaman jahiliyyah ke zaman yang terang benderang yakni agama Islam. Dalam pengerjaan skripsi ini tidak lepas dari doa, bantuan, bimbingan, serta arahan dari berbagai pihak, karena itu peneliti mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Zainuddin, M.A., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Ibu Dr. Sri Harini M.Si., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Ibu Dr. Elly Susanti, M.Sc., selaku ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak Hisyam Fahmi, M.Kom., selaku dosen pembimbing I yang senantiasa memberikan bimbingan, arahan, dan ilmu kepada peneliti dalam menyelesaikan skripsi ini.
5. Bapak Ach. Nashichuddin, M.A., selaku dosen pembimbing II yang senantiasa memberikan bimbingan, arahan, dan ilmu kepada peneliti dalam menyelesaikan skripsi ini.
6. Seluruh civitas academica Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
7. Teristimewa dari peneliti untuk Ayah dan Ibu, Syamsul Huda dan Nur Siti selaku kedua orang tua peneliti dan kedua adik peneliti Aqila dan Ayunda, juga keluarga besar, yang senantiasa memberikan dukungan kepada peneliti baik moril dan materiil sehingga peneliti dapat menyusun dan menyelesaikan laporan skripsi ini dengan baik.
8. Teman teman terbaik peneliti terutama Konsorsium Komputasi, serta Eka, dan Fathun yang selalu bersedia menemani, membantu dan mendukung peneliti secara langsung dan tidak langsung.

9. Penghuni “Rumah Kami” yakni Lia, Alisa, Shoha yang mendukung dan memberi saran yang membangun.
10. Teman teman kepanitiaan dan keorganisasian baik di lingkungan kampus maupun luar kampus yang telah memberikan banyak pembelajaran yang luar biasa sehingga peneliti mendapatkan motivasi untuk terus berkembang dan belajar.
11. Seluruh teman teman seperjuangan MAHATMA’20
12. Kepada semua pihak yang tidak dapat disebutkan satu persatu, yang telah membantu menyelesaikan skripsi baik moril maupun materiil.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun dengan ikhlas, penulis mempersembahkan sebagai bentuk apresiasi dan penghormatan kepada semua pihak yang telah memberikan kontribusi, dukungan, dan inspirasi kepadapenulis. Semoga skripsi ini bermanfaat bagi pembaca .

Malang, 26 Juni 2024

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	iv
MOTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xiv
DAFTAR LAMPIRAN	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث.....	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	7
1.5 Batasan Masalah	8
BAB II KAJIAN TEORI	9
2.1 Teori Pendukung	9
2.1.1 Machine Learning	9
2.1.2 Text Mining	9
2.1.3 Term Frequency-Inverse Document Frequency (TF-IDF).....	12
2.1.4 Support Vector Machine (SVM)	13
2.1.5 Confusion Matrix	18
2.1.6 Cross Validation	20
2.1.7 Natural Language Processing	21
2.1.8 Analisis Sentimen.....	21
2.1.9 Twitter (X).....	22
2.1.10 Pemilihan Presiden	23
2.1.11 Elektabilitas Calon Presiden.....	24
2.2 Pemimpin Yang Baik Dalam Pandangan Islam	24
2.3 Kajian Topik Dengan Teori Pendukung	26
BAB III METODE PENELITIAN	28
3.1 Jenis Penelitian	28
3.2 Data dan Sumber Data	28
3.3 Pelabelan Data	29
3.4 Teknik Analisis Data	30
BAB IV PEMBAHASAN	35
4.1 Pengumpulan Data.....	35
4.2 Preprocessing	36

4.3 Spliting Data	41
4.4 Klasifikasi Dengan Algoritma SVM	41
4.4.1 Perhitungan Manual TF-IDF	41
4.4.2 Perhitungan Manual Menggunakan Metode SVM	44
4.4.3 Hasil Klasifikasi menggunakan Metode SVM	48
4.5 Evaluasi	49
4.5.1 Confusion Matrix	49
4.5.2 Cross Validation	51
4.6 Elektabilitas Calon Presiden	52
4.7 Analisis Sentimen Terhadap Elektabilitas Calon Pemimpin Dalam Pandangan Islam	54
BAB V KESIMPULAN	56
5.1 Kesimpulan	56
5.2 Saran	56
DAFTAR PUSTAKA	58
LAMPIRAN.....	61
RIWAYAT HIDUP	70

DAFTAR TABEL

Tabel 2.1	Confusion Matrix	18
Tabel 2.2	Rumus perhitungan Confusion Matrix.....	19
Tabel 3.1	Kriteria Pelabelan Data	29
Tabel 4.1	Sampel Data Komentar Twitter	36
Tabel 4.2	Hasil Cleaning Data Komentar Twitter	37
Tabel 4.3	Hasil Case Folding Data Komentar Twitter.....	38
Tabel 4.4	Hasil Tokenizing Data Komentar Twitter.....	38
Tabel 4.5	Hasil Normalisasi Data Komentar Twitter.....	39
Tabel 4.6	Hasil Stopword Removal Pada Data Komentar Twitter	40
Tabel 4.7	Hasil Stemming Data Komentar Twitter	41
Tabel 4.8	Pembagian Data Training dan Testing	41
Tabel 4.9	Kalimat Sample TF-IDF	42
Tabel 4.10	Menghitung Frekuensi Kemunculan Kata Pada Dokumen.....	42
Tabel 4.11	Hasil perhitungan TF-IDF dan Normalisasi.....	43
Tabel 4.12	Sampel Data Perhitungan Training	44
Tabel 4.13	Hasil Perhitungan Persamaan Dengan Kernel Linear.....	45
Tabel 4.14	Nilai Weight	46
Tabel 4.15	Sampel Data Perhitungan Testing.....	47
Tabel 4.16	Hasil Perhitungan Kernel Linear Dengan Data Test.....	47
Tabel 4.17	Confusion Matrix Dari Keseluruhan Data	49
Tabel 4.18	Confusion Matrix Anies.....	50
Tabel 4.19	Confusion Matrix Prabowo.....	50
Tabel 4.20	Confusion Matrix Ganjar	50
Tabel 4.21	Hasil Cross Validation	52

DAFTAR GAMBAR

Gambar 2.1 Pemetaan Data Input ke Fitur Space	17
Gambar 3.1 Flowchart Analisis Data	30
Gambar 3.2 Flowchart Preprocessing	30
Gambar 3.3 Flowchart Spliting Data	32
Gambar 3.4 Flowcart Pemodelan Dengan Support Vector Machine	33
Gambar 3.5 Flowchart Evaluasi Model	34
Gambar 4.1 Grafik Jumlah Data	35
Gambar 4.2 Hasil Klasifikasi	49

DAFTAR SIMBOL

Simbol-simbol yang digunakan pada penelitian ini adalah:

w	:	Parameter <i>hyperplane</i>
x	:	Titik data masukan
a_i	:	Nilai bobot pada setiap data
$K(x, x_i)$:	Fungsi Kernel
b	:	Parameter <i>hyperplane</i> (nilai bias)

DAFTAR LAMPIRAN

Lampiran 1 Kode Program Crawling Data Twitter.....	61
Lampiran 2 Kode Program Analisis Sentimen Metode SVM.....	61
Lampiran 3 Dataset Calon Presiden.....	68

ABSTRAK

Dani, Alfi Rahma, 2024. **Implementasi Implementasi Algoritma *Support Vector Machine* Untuk Analisis Sentimen Terhadap Elektabilitas Calon Presiden 2024**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing (I) Hisyam Fahmi, M.Kom. (II) Ach. Nashichuddin, M.A.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, SVM, Twitter, Pemilihan Presiden.

Penelitian ini membahas analisis sentimen untuk mengukur elektabilitas calon Presiden dalam Pemilihan Umum 2024 dengan menggunakan algoritma *Support Vector Machine* (SVM). Data dikumpulkan melalui proses *crawling* dari Twitter (X) dengan kata kunci terkait calon presiden Anies Baswedan, Prabowo Subianto, dan Ganjar Pranowo. Setiap kata kunci menghasilkan 2.000 tweet, dengan total 6.000 data. Proses preprocessing mencakup penghapusan karakter tidak penting, transformasi huruf, *tokenizing*, normalisasi, *stopword removal*, dan *stemming*. Model dilatih dengan SVM dan dievaluasi menggunakan *K-Fold Cross Validation* untuk mengukur performa model. Hasil evaluasi menunjukkan bahwa model memiliki rata-rata presisi sebesar 71,59%, recall sebesar 71,39%, *F1-Score* sebesar 71,33%, dan akurasi sebesar 71,39%. Dari hasil analisis sentimen, elektabilitas Anies Baswedan mencapai 38,44%, Prabowo Subianto 35,38%, dan Ganjar Pranowo 28,18%. Hasil penelitian ini menunjukkan bahwa algoritma SVM efektif dalam mengklasifikasikan data teks untuk analisis sentimen dengan akurasi yang cukup baik. Elektabilitas calon presiden dapat diukur dan dianalisis menggunakan data dari media sosial, memberikan gambaran mengenai opini publik terhadap masing-masing calon presiden.

ABSTRACT

Dani, Alfi Rahma, 2024. **Implementation of Support Vector Machine Algorithm for Sentiment Analysis of 2024 Presidential Candidate Electability**. Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisor (I) Hisyam Fahmi, M.Kom. (II) Ach. Nashichuddin, M.A.

Keywords: Sentiment Analysis, Support Vector Machine, SVM, Twitter, Presidential Election.

This research discusses sentiment analysis to measure the electability of presidential candidates in the 2024 General Election using the Support Vector Machine (SVM) algorithm. Data was collected through a crawling process from Twitter (X) with keywords related to presidential candidates Anies Baswedan, Prabowo Subianto, and Ganjar Pranowo. Each keyword generated 2.000 tweets, totaling 6.000 data. The preprocessing process includes removal of unimportant characters, letter transformation, tokenizing, normalization, stopword removal, and stemming. The model will be trained with SVM and evaluated using K-Fold Cross Validation to measure model performance. The evaluation results show that the model has an average precision of 71,59%, recall of 71,39%, F1-Score of 71,33%, and accuracy of 71,39%. From the sentiment analysis results, Anies Baswedan's electability reached 36,44%, Prabowo Subianto 35,38%, and Ganjar Pranowo 28,18%. The results of this study show that the SVM algorithm is effective in classifying text data for sentiment analysis with fairly good accuracy. The electability of presidential candidates can be measured and analyzed using data from social media, providing an overview of public opinion towards each presidential candidate.

مستخلص البحث

دايني، ألفي رحمة، ٢٠٢٤. تطبيق تطبيق خوارزمية آلة دعم المتجهات الداعمة لتحليل المشاعر الانتخابية للمرشح الرئاسي لعام ٢٠٢٤. الأطروحة بحت العلمي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا، مالك إبراهيم الإسلامية الحكومية مالانج. المشرف (١) هشام فهمي، الماجستير (٢) أحمد ناصح الدين الماجستير

الكلمات المفتاحية تويتير، الانتخابات الرئاسية، *SVM*، تحليل المشاعر، آلة دعم المتجهات المساندة

يناقش هذا البحث تحليل المشاعر لقياس مدى قابلية المرشحين للرئاسة في الانتخابات العامة لعام ٢٠٢٤ باستخدام بكلمات (X) مُجمعت البيانات من خلال عملية زحف من تويتير. (*SVM*) خوارزمية آلة دعم المتجهات مفتاحية تتعلق بالمرشحين الرئاسيين أنيس باسويدان وبرابوو سوبيانتو وجانجار برانوو. أنتجت كل كلمة مفتاحية ألفي تغريدة، أي ما مجموعه ستة آلاف بيانه، وتشمل عملية المعالجة المسبقة إزالة الأحرف غير المهمة، وتحويل وتقييمه *SVM* الحروف، والتمييز، والتطبيع، وإزالة كلمات التوقف، والوقف، وسيتم تدريب النموذج باستخدام لقياس أداء النموذج. تُظهر نتائج التقييم أن *K-Fold Cross Validation* باستخدام التحقق المتقاطع ودقة، $F1-Score$ 71.33% النموذج يتمتع بمتوسط دقة 71.59%، واستدعاء 71.39%، ودرجة ومن نتائج تحليل المشاعر، بلغت نسبة انتخاب أنيس باسويدان 36.44%، وبرابوو سوبيانتو، 71.39% فعالة في تصنيف البيانات *SVM* وجانجار برانوو 28.18%. تُظهر نتائج هذه الدراسة أن خوارزمية، 35.38% النصية لتحليل المشاعر بدقة جيدة إلى حد ما. يمكن قياس قابلية المرشحين للرئاسة للانتخاب وتحليلها باستخدام بيانات من وسائل التواصل الاجتماعي، مما يوفر نظرة عامة على الرأي العام تجاه كل مرشح رئاسي.

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen atau yang biasa dikenal dengan istilah *opinion mining* merupakan salah satu cabang penelitian dari *text mining* yang bertujuan untuk menentukan persepsi atau subjektivitas publik terhadap suatu topik pembahasan. Analisis sentimen biasa digunakan untuk pendekatan analitis yang digunakan untuk menganalisis teks. Tujuan dari analisis sentimen ini adalah untuk menentukan subjektivitas opini, hasil ulasan dari suatu data yang berisi kumpulan kalimat. Berdasarkan analisis sentimen, opini dari seseorang dapat diklasifikasikan ke dalam berbagai kategori berdasarkan ukuran data dan jenis dokumen (Wongkar & Angdresey., 2019). Hasil dari analisis sentiment akan mengelompokkan polaritas dari teks yang ada dalam dokumen untuk mengetahui pendapat yang dikemukakan dalam dokumen atau kalimat bersifat positif atau negatif (Pang et al., 2004)

Klasifikasi merupakan salahsatu bagian dari *machine learning* yang dapat digunakan untuk berbagai macam tujuan seperti analisis data, pengambilan keputusan, dan prediksi. Klasifikasi sendiri memiliki berbagai macam metode yakni *logistic regression*, *naïve bayes*, dan *support vector machine*. Dalam pemilihan algoritma klasifikasi yang tepat, perlu memperhatikan berbagai faktor, seperti karakteristik data dan tujuan dari klasifikasi. Dengan memilih algoritma yang tepat nantinya akan memberikan hasil klasifikasi dengan tingkat keakuratan yang tinggi (Alpaydn., 2014).

Support Vector Machine (SVM) adalah metode non parametrik yang digunakan untuk tujuan regresi dan klasifikasi. *Support Vector Machine* (SVM)

pertama kali dikembangkan oleh Vapnik dan juga rekannya Bernhard Boser dan Isabelle Guyon pada tahun 1992 (Ritonga & Purwaningsih, 2018). SVM digunakan untuk mencari nilai *hyperplane* menggunakan support vector dan nilai margin dengan konsep klasifikasi yang menggunakan pembelajaran mesin (*supervised learning*), dimana *hyperplane* tersebut berperan dalam membedakan sentimen yang memprediksi kelas berdasarkan pola atau model dari hasil proses pelatihan. (Mulyawan & Slamet., 2021).

Algoritma SVM memiliki beberapa kelebihan, diantaranya yaitu efektif digunakan untuk mengklasifikasikan data dengan dimensi yang tinggi, bahkan jika jumlah sampel relative sedikit. Selain itu, SVM juga memaksimalkan margin antar kelas kelas data sehingga menghasilkan model yang memiliki performa yang baik pada data baru. SVM juga memiliki beberapa kernel yang bisa disesuaikan dengan proses klasifikasi yang diinginkan. Dari beberapa kelebihan di atas maka peneliti mengambil metode SVM untuk dijadikan metode karena sesuai dengan data yang akan digunakan dan juga dirasa efektif untuk melakukan klasifikasi.

Beberapa penelitian sebelumnya yang pernah dilakukan terkait klasifikasi yaitu penerapan metode *Support Vector Machine* untuk analisis sentimen pengguna Twitter terhadap Vaksin Covid-19, menghasilkan nilai akurasi tertinggi sebesar 93% jika diuji menggunakan *confusion matrix*. Selain itu hasil evaluasi terhadap algoritma *Support Vector Machine* dengan *k-fold cross validation* didapatkan akurasi tertinggi sebesar 93% dan rata-rata sebesar 83.5% (Dani Mulyawan & Slamet., 2021). Penelitian lainnya yang terkait yakni membandingkan algoritma *Naïve Bayes*, SVM, dan k-NN untuk analisis sentiment gadget, dari ketiga algoritma tersebut memperlihatkan bahwa algoritma SVM memiliki performa

terbaik dengan hasil rata-rata akurasi model klasifikasi yaitu sebesar 96,43%, dibandingkan dengan algoritma *Naïve Bayes* yang hanya sebesar 83,54%, dan k-NN sebesar 59,68% (Iskandar & Nataliani., 2021). Dari beberapa literatur tersebut menunjukkan bahwa SVM merupakan algoritma yang memiliki performa yang baik untuk klasifikasi sentimen. Kemampuan SVM dalam memisahkan data dengan baik dan memaksimalkan margin antara kelas positif dan negatif menjadikan algoritma ini dianggap sangat baik untuk memahami opini dan persepsi publik terhadap berbagai isu. Diharapkan penggunaan algoritma SVM dalam penelitian skripsi ini nantinya dapat membantu menghasilkan analisis yang lebih akurat.

Selama dua dekade terakhir, media sosial telah menjadi bagian besar dari kehidupan sehari-hari masyarakat. Media sosial adalah saluran komunikasi tempat ide dan pemikiran dibagikan, sehingga media sosial membentuk jejaring sosial yang membuka kemungkinan bagi orang-orang di mana saja untuk terhubung (Emelie et al., 2023). Salah satu bagian dari media sosial adalah Twitter yang pada bulan Juli 2023 telah berubah nama menjadi (X). Twitter (X) adalah layanan jejaring sosial yang membantu penggunanya mengirim dan membaca pesan berbasis teks hingga 140 karakter. Pada awal tahun 2013, pengguna Twitter (X) mengirimkan lebih dari 500 juta kicauan per hari. Twitter (X) menjadi salah satu media sosial yang sering digunakan masyarakat sebagai wadah dalam berekspresi dan mengeluarkan pendapat. Pengguna Twitter (X) di Indonesia mencapai 25,30 juta pada Juli 2023, sehingga tak heran jika Indonesia menduduki peringkat 4 pengguna Twitter(X) terbanyak di dunia (Databooks., 2023). Twitter (X) dapat menjadi salah satu sumber data tentang pendapat masyarakat atau opini. Data yang ada tersebut dapat diolah, sehingga dapat digunakan sebagai salah satu studi sosial.

Pada tanggal 14 Februari 2024 telah terselanggara proses pemilihan Presiden. Hal ini sudah diumumkan oleh Komisi Pemilihan Umum (KPU) melalui portal resminya. Namun, dalam perjalanan menuju pemilihan Presiden ini nama-nama calon Presiden sudah mulai banyak diperbincangkan, hal tersebut tidak lepas juga dari peran media sosial. Dunia maya yang begitu bebas dan sulit dikontrol, membuat semua orang bebas berpendapat atau beropini tentang calon Presiden jagoannya masing-masing. Beberapa nama yang sering kali masuk dalam survei memunculkan banyak opini-opini masyarakat, tidak hanya opini yang positif atau netral tapi juga yang negatif. Beberapa nama yang sering disebut-sebut dalam Twitter, yakni Anies Baswedan, Ganjar Pranowo, dan Prabowo Subianto. Nama-nama tersebut akan menjadi keyword atau kata kunci dalam pencarian *tweet* yang terkait dengan Calon Presiden Indonesia 2024.

Pemilihan Presiden juga erat kaitannya dengan cara pandang kita terhadap memilih pemimpin dalam Islam. Dalam masyarakat, Presiden memiliki peran yang sangat penting, seorang Presiden harus menjadi pemimpin dalam suatu negara, menjadi teladan dan mampu menjadi panutan bagi seluruh rakyat yang dipimpinnya. Dalam Islam sendiri pemimpin yang baik harus memiliki sifat jujur, adil, bijaksana, dan taat kepada Allah dan Rasul. Seorang pemimpin yang taat akan mengarahkan masyarakatnya ke jalan yang benar. Selain itu ia akan membuat keputusan yang terbaik bagi rakyatnya, agar rakyatnya hidup bahagia dan makmur. Oleh karena itu seorang pemimpin harus dipilih dengan penuh kehati-hatian dalam Al Qur'an surat An Nisa Ayat 58 yang berbunyi (Kemenag RI, 2022):

إِنَّ اللَّهَ يَأْمُرُكُمْ أَنْ تُؤَدُّوا الْأَمَانَاتِ إِلَىٰ أَهْلِهَا وَإِذَا حَكَمْتُمْ بَيْنَ النَّاسِ أَنْ تَحْكُمُوا بِالْعَدْلِ ۗ إِنَّ اللَّهَ نِعِمَّا يَعِظُكُمْ بِهِ ۗ إِنَّ اللَّهَ كَانَ سَمِيعًا بَصِيرًا

“Sungguh, Allah menyuruhmu menyampaikan amanat kepada yang berhak menerimanya, dan apabila kamu menetapkan hukum di antara manusia hendaknya kamu menetapkannya dengan adil. Sungguh, Allah sebaik-baik yang memberi pengajaran kepadamu. Sungguh, Allah Maha Mendengar, Maha Melihat”

Dalam surat ini, Allah Subhaanahu wa Ta'aala memerintahkan hamba-hamba-Nya menunaikan amanat secara sempurna, tidak dikurangi dan tidak ditunda-tunda. Yang digolongkan amanat dalam hal ini adalah amanat untuk beribadah (seperti shalat, zakat, puasa dsb), amanat jabatan, harta dan rahasia serta perkara-perkara yang hanya diketahui oleh Allah (Tafsirweb, 2020). Contoh menunaikan amanat dalam jabatan yang dalam konteks presiden adalah dengan memenuhi kewajiban, memenuhi amanat dalam menjadi pemimpin dengan menjaga apa yang menjadi hak dari rakyatnya.

Pada tahap pemilihan Presiden ini sering kali pihak-pihak tim sukses mengalami kesulitan terkait sasaran kampanye, dan jarang melihat polarisasi yang terjadi dalam dunia maya. Maka dari hal tersebut analisis sentimen dapat digunakan untuk melihat popularitas dan elektabilitas dari calon pemimpin di mata masyarakat. Analisis sentimen juga dapat digunakan untuk menyerap berbagai aspirasi dan masukan dari masyarakat terkait pemerintah. Nantinya hasil dari analisis sentimen dapat menunjukkan bagaimana citra dari calon presiden, apakah cenderung positif atau negatif dan juga hasil dari analisis ini dapat digunakan untuk tim sukses dari setiap calon Presiden untuk menargetkan kampanye pada masyarakat agar tepat sasaran.

Berdasarkan pemaparan di atas, maka penulis tertarik untuk melakukan sebuah penelitian dengan judul “Implementasi Algoritma *Support Vector Machine* Untuk Analisis Sentimen Terhadap Elektabilitas Calon Presiden 2024”. Pada penelitian ini analisis sentimen dilakukan untuk melihat dan mengambil informasi

masyarakat terhadap calon Presiden tahun 2024. Dengan adanya tweet yang berhubungan dengan calon Presiden Indonesia 2024 yang telah ditetapkan oleh KPU, maka dapat menjadi salah satu cara untuk melakukan sosial media analitik yang memuat informasi mengenai suatu topik apakah cenderung positif atau negatif. Diharapkan nantinya, hasil analisis sentimen terhadap calon Presiden 2024 ini dapat digunakan untuk mengetahui isu-isu yang paling penting bagi masyarakat, sehingga nantinya dapat memberikan pemahaman yang lebih baik tentang dinamika politik yang terjadi saat ini. Selain itu, analisis opini masyarakat di media sosial Twitter diharapkan dapat membantu untuk mengetahui respon masyarakat Indonesia terhadap Pemilu 2024, serta menghindari berita hoax, opini negatif, dan ujaran kebencian.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat permasalahan yang akan diteliti sebagai berikut.

1. Bagaimana tingkat akurasi metode *Support Vector Machine* dalam melakukan analisis sentimen pada data calon Presiden 2024 di Twitter (X)?
2. Bagaimana estimasi elektabilitas calon Presiden dalam Pemilihan Umum 2024 menurut analisis sentimen data Twitter (X)?

1.3 Tujuan Penelitian

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat permasalahan yang akan diteliti sebagai berikut.

1. Untuk mengetahui tingkat akurasi metode *Support Vector Machine* dalam

melakukan analisis sentimen pada data calon Presiden 2024 di Twitter (X).

2. Untuk mengetahui estimasi elektabilitas calon Presiden 2024 menurut analisis sentimen dari data Twitter (X).

1.4 Manfaat Penelitian

Sesuai dengan tujuan penelitian, maka manfaat penelitian ini dibedakan berdasarkan beberapa kepentingan pihak yaitu:

1. Manfaat bagi peneliti

Penelitian ini merupakan salah satu kegiatan untuk menambah wawasan peneliti tentang bagaimana hasil analisis sentimen terhadap elektabilitas calon Presiden 2024 menggunakan algoritma *Support Vector Machine*.

2. Manfaat bagi pembaca

- a. Bagi mahasiswa, penelitian ini dapat digunakan sebagai bahan rujukan untuk mata kuliah pada konsorsium komputasi.
- b. Bagi peneliti akademis, penelitian ini dapat digunakan untuk pengembangan penelitian-penelitian lainnya.
- c. Bagi praktisi, penelitian ini dapat digunakan untuk membantu mengetahui polaritas sentiment pengguna aplikasi Twitter, sehingga nantinya dapat menjadi rujukan untuk penelitian selanjutnya.

3. Manfaat bagi instansi

Penelitian ini dapat meningkatkan pengembangan kurikulum keilmuan matematika khususnya pada konsorsium komputasi.

1.5 Batasan Masalah

Agar tidak terjadi perluasan masalah, maka penelitian ini memfokuskan pada penggunaan data yang meliputi:

1. Objek yang digunakan pada penelitian ini adalah data tweet dari pengguna Twitter terkait calon Presiden yakni Anies Baswedan, Prabowo Subianto, dan Ganjar Pranowo, berdasarkan penetapan calon Presiden oleh KPU, dengan jumlah tweet yang sama untuk semua calon Presiden.
2. Data *crawling* yang digunakan hanya pada periode tertentu, dimulai dari penetapan calon presiden pada 14 November 2023 sampai dengan 14 Desember 2023.
3. Data yang digunakan adalah tweet dari pengguna Twitter yang mengandung salah satu keyword berikut : #prabowo, #prabowosubianto, #Ganjarpranowo, #Ganjar, #Ganjar_pranowo, #aniesbaswedan, #anies, @prabowo, @anies, @Ganjar.
4. Pelabelan Twitter bersifat netral, tanpa adanya unsur majas.
5. Pengklasifikasian Tweet berdasarkan dua kategori yakni negatif dan positif.
6. Library yang digunakan dalam proses *Preprocessing* adalah Sastrawi.
7. Hasil mungkin tidak sesuai dengan KPU, dikarenakan data yang digunakan terbatas pada rentang waktu tertentu, dan hanya berdasar pada satu platform Twitter saja.

BAB II

KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 *Machine Learning*

Machine learning adalah cabang dari *artificial intelligence* atau kecerdasan buatan yang bertujuan untuk memungkinkan mesin melakukan pekerjaannya secara terampil dengan menggunakan perangkat *intelligence software* (Giuseppe Bonaccorso., 2017).

Machine learning adalah cabang dari teknologi berupa sistem berbasis kecerdasan buatan, yang memiliki kemampuan untuk mempelajari pola dari data dan mengembangkan secara otomatis pengalaman dalam pelatihan tanpa ada program secara khusus. Pola dari data yang didapatkan digunakan untuk membuat keputusan. *Machine learning* sering dikategorikan dalam beberapa algoritma seperti *supervised*, *semi-supervised* atau *unsupervised algorithms* (Dinakar et al., 2021)

2.1.2 *Text Mining*

Text mining adalah proses menemukan informasi penting dari teks. Informasi ini dapat berupa pola, tren, atau hubungan yang tidak secara eksplisit dinyatakan dalam teks. *Text mining* sangat berguna untuk menganalisis data tekstual dalam jumlah besar, seperti media sosial, ulasan produk, dan dokumen penelitian. Dengan *text mining*, kita dapat menemukan pola dan hubungan yang sulit ditemukan secara manual (Weiss, 2005).

Text mining adalah proses analisis dari suatu text yang dilakukan secara otomatis menggunakan komputer untuk mendapatkan informasi penting. Informasi ini dapat berupa ringkasan dari sejumlah data teks, atau informasi yang tidak secara eksplisit dinyatakan dalam teks. *Text mining* dapat membantu untuk memahami isi teks secara lebih cepat dan akurat daripada membaca secara manual. *Text mining* memiliki kesamaan dengan *data mining* dalam hal tujuannya, yaitu untuk menemukan informasi berharga dari data. Namun, *text mining* memiliki perbedaan dalam penanganan data teks, yaitu data teks yang tidak terstruktur. Untuk mengubah data teks yang tidak terstruktur menjadi data yang terstruktur, terdapat beberapa tahapan perlu dilakukan (Muttaqin & Bachtiar, 2016).

1. *Crawling*

Crawling dapat diartikan sebagai proses pengambilan sejumlah besar data dari web. Data tersebut dengan cepat disimpan dalam suatu tempat penyimpanan lokal dan mengindeksnya berdasar sejumlah kata kunci (F. A. Pozzi, 2012). Mesin pencari web beroperasi dengan menyimpan informasi tentang banyak halaman web, yang diambil langsung dari situs web dan untuk penelitian ini akan mengambil opini dari akun X.

2. Praproses Teks

Tahap awal dalam pengolahan teks, yaitu praproses teks, dilakukan untuk mengubah dokumen menjadi data terstruktur sesuai kebutuhan agar dapat lebih lanjut diolah dalam proses *text mining*, terutama dalam konteks klasifikasi data. Tujuan dari tahap praproses teks dalam klasifikasi adalah

untuk meningkatkan akurasi klasifikasi data. Berikut beberapa tahapan dalam praproses teks (Kurniawan, 2017) :

a. *Cleansing*

Cleansing merupakan tahap penting dalam penelitian yang melibatkan analisis data teks. Proses ini menghilangkan karakter yang tidak penting seperti emoticon, angka, tanda baca, simbol, spasi berlebih, enter, tautan, dan tagar, untuk meningkatkan kualitas data, mempermudah analisis data, dan meningkatkan akurasi dan efisiensi analisis.

b. *Case folding*

Proses menyamakan huruf dalam dokumen. Ini dilakukan untuk membuat pencarian lebih mudah. Tidak semua dokumen teks menggunakan huruf kapital secara konsisten. Maka, *case folding* diperlukan untuk mengubah teks secara keseluruhan dalam dokumen menjadi bentuk standar, biasanya dalam bentuk huruf kecil.

c. *Tokenizing*

Proses membagi teks yang berasal dari kalimat atau paragraf menjadi bagian tertentu dikenal sebagai tokenisasi (Manning et al., 2008). Sebagai contoh, jika kalimat "Azizah senang sekali keliling Papua" ditokenisasi, akan dihasilkan lima token yaitu "Azizah", "senang", "sekali", "keliling", dan "papua". Biasanya, spasi dan tanda baca digunakan untuk membedakan token satu sama lain. Hasil dari *tokenizing* bermanfaat untuk analisis teks lebih lanjut.

d. Normalisasi

Tahap Normalisasi merupakan langkah penting dalam pengolahan data teks. Pada tahap ini, seluruh kata yang memiliki kesalahan ejaan, noise, atau kata tidak baku diubah menjadi bentuk baku. Kata-kata yang mengandung noise sendiri mengacu pada unsur-unsur bahasa daerah juga bahasa gaul (*slangword*) yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) (Septian et al., 2019).

e. *Stopword Removal*

Stopword adalah kata umum yang sering muncul tetapi tidak memiliki pengaruh yang signifikan atau tidak memiliki makna, seperti kata "tidak", "misal", "atau" (Iskandar & Nataliani, 2021). Maka dari itu penghapusan *stopword* ini diperlukan. Namun sebelum dilakukannya proses ini peneliti harus membuat daftar kata *stopword* berdasarkan pada dataset yang ada. *Stopwords removal* bertujuan untuk mengeleminasi kata umum tersebut untuk meningkatkan hasil analisis dan mengurangi beban komputasi.

f. *Stemming*

Proses untuk mendapatkan kata dasar dengan menghilangkan awalan, akhiran, sisipan, dan *confixes*. Istilah kombinasi awalan dan akhiran juga digunakan disebut dengan *stemming*.

2.1.3 *Term Frequency-Inverse Document Frequency (TF-IDF)*

Data bersih yang telah melalui praproses text selanjutnya akan dihitung bobot per-kata dengan menggunakan metode TF-IDF. Pembobotan kata adalah sebuah cara yang untuk memberikan skor pada kemunculan suatu kata di dalam

sebuah dokumen. *Term Frequency* (TF) , menunjukkan frekuensi kemunculan kata, sedangkan *Inverse Document Frequency* (IDF) menunjukkan bobot dari per-kata. Semakin jarang kata muncul di dokumen, semakin besar bobotnya (Madasu & Elango, 2020). Perhitungan TF ditunjukkan pada Persamaan berikut:

$$TF = \frac{\text{frekuensi kemunculan sebuah kata}(t)\text{dalam dokumen } (d)}{\text{Jumlah kata dalam dokumen}(d)} \quad (2.1)$$

Perhitungan IDF dapat ditunjukkan dalam Persamaan berikut ini:

$$IDF_{(w)} = \ln \left(\frac{\text{Jumlah total dokumen}(d) + 1}{\text{Jumlah total dokumen yang mengandung kata} + 1} \right) + 1 \quad (2.1)$$

Selanjutnya, Perhitungan TF-IDF atau bobot dokumen untuk setiap kata, dapat ditunjukkan pada Persamaan berikut ini:

$$V = TF \times IDF_{(w)} \quad (2.3)$$

Kemudian hasil dari V akan dinormalisasikan dengan Euclidian norm. Persamaan ini berfungsi sebagai normalisasi nilai bobot yang diperoleh, supaya nilai bobot berada pada range yang sama.

$$V_{norm} = \frac{V_n}{\sqrt{V_1^2 + V_2^2 + \dots + V_n^2}}, \text{ dimana } n \text{ merupakan indeks dari vektor} \quad (2.4)$$

2.1.4 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah sebuah algoritma tergolong baru, namun algoritma ini menunjukkan performa yang baik dalam berbagai aplikasi, seperti klasifikasi teks, pengenalan tulisan tangan, dan bioinformatika. Algoritma

ini bekerja dengan menganalisis data dan mengidentifikasi pola klasifikasi melalui proses pelabelan. (Suhardjono & Abdul, 2019).

Algoritma SVM disebut sangat cepat dan begitu efektif untuk proses pengolahan klasifikasi teks, dalam isitilah geometris disebut sebuah klasifikasi biner, dimana terdapat *hyperplane* yang menentukan ruang fitur antar titik- titik yang mewakili situasi positif maupun negatif (Rahman Isnain et al., 2021). Algoritma SVM memiliki Persamaan umum *hyperplane* sebagai berikut (Liang, 2021) :

$$f(x) = w \cdot x + b \quad (2.5)$$

Keterangan

- w : Parameter *hyperplane* yang dicari (gerak yang tegak lurus antara garis *hyperplane* & titik *support vector machine*)
- x : Titik data masukan
- b : Parameter *hyperplane* (nilai bias)

Hyperplane terbaik adalah hyperplane yang terletak di tengah di antara dua set objek dari dua kelas. Untuk itu perlu dicari hyperplane terbaik dengan mendapatkan nilai margin terbesar. Margin terbesar atau Max margin dapat di temukan dengan memaksimalkan jarak antaraa hyperplane dan titik dekatnya. Selain itu, batasan harus ditambahkan ke setiap kelas untuk setiap data kelas agar mencegah memasuki batas. Data akan masuk kedalam kelas negatif jika memenuhi

$$w \cdot x_i + b = -1, \quad \text{jika } y = -1$$

Dan data akan masuk kedalam kelas positif jika

$$w \cdot x_i + b = 1, \quad \text{jika } y = +1$$

Dalam kasus non-linier support vector seringkali kelas positif dan kelas negatif yang berdimensi dua tidak dapat dipisahkan secara sempurna oleh hyperplane secara linier. Namun, setelah dilakukan transformasi data ke dalam ruang berdimensi tiga, data kelas positif dan kelas negatif dapat dipisahkan oleh hyperplane secara linier. Sehingga, cara untuk mendapatkan hyperplane dengan batas terbesar dan dengan kasus non-linier support vector dapat menggunakan pemrograman kuadratik yang perhitungannya dapat dilihat pada Persamaan berikut:

$$\text{Max margin} = \min \frac{1}{2} \|w\|^2$$

Dalam *Quadratic Programming* terdapat beberapa metode, salah satunya ialah *Lagrange Multiplier* yang perhitungannya dapat dilihat pada Persamaan

$$\text{Min } L_p(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i [y_i(w \cdot x_i + b) - 1] \quad (2.6)$$

Koefisien dalam *Lagrange Multipliers* bernilai lebih dari sama dengan nol. Nilai dari *Lagrange Multipliers* dapat dimaksimalkan dengan meminimalkan nilai dari variabel \vec{w} dan b serta memaksimalkan nilai variabel a_i melalui Persamaan:

$$L_{(a)} = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K(x_i x_j) \quad (2.7)$$

$$\text{Dengan syarat } a_i \geq 0, (i = 1, 2, \dots, n), \sum_{i=1}^n a_i y_i = 0$$

Selanjutnya, untuk mendapatkan hyperplane yang optimal tidak mudah karena nilai vektor pembobot w belum diketahui. Untuk menemukan nilai vektor pembobot w dalam Hilbert Space dapat ditulis dalam Persamaan:

$$w_i = \sum_{i=1}^N a_i y_i \phi(x_i) \quad (2.8)$$

Kemudian nilai b didapatkan melalui Persamaan:

$$b = \frac{1}{NSV} \sum_{x_j \in SV} \left(\frac{1}{y_1} - \sum_{x_j \in SV} (a_j y_j K(x_j, x_i)) \right) \quad (2.9)$$

Selanjutnya akan di substitusikan nilai a , $bias$ dan w pada Persamaan fungsi pemisah optimal atau hyperplane, sehingga didapatkan fungsi keputusan untuk menentukan kelas dari data yang diformulasikan dengan Persamaan:

$$f(x) = \text{sign}(w \cdot x + b)$$

$$f(x) = \text{sign}(\sum_{i=1}^{NSV} a_i y_i K(x_i, x_d) + b) \quad (2.10)$$

Keterangan

a : Nilai Lagrange Multiplier data support vector

NSV : Jumlah *support vector*

a : Data *support vector*

b : nilai bias

y : Kelas data

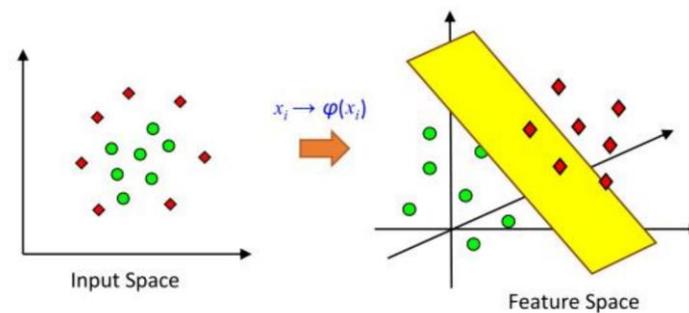
x_d : Data Uji

$K(x_j, x_i)$: Fungsi Kernel

sign : Fungsi Signum atau fungsi tanda

Banyak penelitian yang menerapkan konsep SVM dikarenakan Tingkat Akurasi pada model yang dihasilkan oleh proses peralihan SVM sangat berhubungan dengan fungsi kernel dan parameter yang digunakan. Tujuan dari pembelajaran SVM ialah menemukan nilai *margin* maksimum untuk memisahkan dua kelas data (Rahma Yustihan & Pandu Adikara, 2021). Pada awalnya teknik *machine learning* terbatas pada asumsi kelinearan, padahal umumnya

permasalahan yang terjadi di dalam dunia nyata bersifat *non-linier*, dimana data tidak dapat dipisahkan secara sempurna dengan garis lurus (*hyperplane*). Akibatnya, hasil algoritma-algoritma tersebut terbatas pada kasus-kasus linear. Untuk mengatasi keterbatasan ini, *Support Vector Machine (SVM)* menggunakan fungsi kernel untuk memetakan data ke ruang fitur (*fitur space*) berdimensi tinggi, sehingga memungkinkan *hyperplane* memisahkan data secara lebih efektif.



Gambar 2.1 Pemetaan Data Input ke *Fitur Space*

Beberapa fungsi kernel yang umumnya digunakan dalam SVM adalah sebagai berikut (Han et al., 2012) :

1. Kernel Linear

$$K(x_i, x_j) = x_i^T x_j \quad (2.11)$$

2. Kernel Polynomial

$$K(x_i, x_j) = (x_i^T x_j + 1)^p \quad (2.12)$$

3. Kernel *Radial Basic Function* (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.13)$$

dengan $\gamma = \frac{1}{2\sigma^2}$

4. Kernel Sigmoid

$$K(x_i, x_j) = \tanh(\alpha(x_i, x_j) + \vartheta) \quad (2.14)$$

Maka, SVM dapat digunakan secara efisien dalam menyelesaikan kasus non linear dengan memilih fungsi kernel yang sesuai.

2.1.5 Confusion Matrix

Confusion Matrix adalah matrix yang digunakan untuk mengevaluasi model. Hasil dari model di klasifikasikan berdasarkan jumlah data uji yang benar dan salah. Dengan ini, kualitas kinerja model klasifikasi dapat dinilai (Normawati & Prayogi, 2021). *Confusion Matrix* merupakan tabel matriks yang menggambarkan kinerja dari model klasifikasi yaitu pada contoh sebagai berikut:

Tabel 2.1 Confusion Matrix

Data		Aktual	
		<i>True</i>	<i>False</i>
Prediksi	<i>True</i>	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
	<i>False</i>	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

Keterangan:

True Positive (TP) : Jumlah data yang bernilai positif diklasifikasikan benar sebagai positif.

True Negative (TN) : Jumlah data yang bernilai negatif tetapi diklasifikasikan sebagai negatif.

False Positive (FP) : Jumlah data yang bernilai negatif tetapi diklasifikasikan sebagai positif.

False Negative (FN) : Jumlah data yang bernilai positif tetapi diklasifikasikan sebagai negatif.

Confusion matrix dapat menghitung beberapa metrik untuk mengevaluasi kinerja model, seperti akurasi, presisi, *recall*, dan *F1-Score*. Selanjutnya, pada tabel di bawah terdapat konsep rumus untuk menghitung performa menggunakan *confusion matrix* yaitu sebagai berikut (Azhari et al., 2021):

Tabel 2.2 Rumus perhitungan *Confusion Matrix*

Metrik	Rumus
<i>Accuracy</i>	$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$
<i>Precision</i>	$\frac{TP}{TP + FP} \times 100\%$
<i>Recall</i>	$\frac{TP}{TP + FN} \times 100\%$
<i>F1-Score</i>	$\frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\%$

Keterangan

Accuracy : Nilai akurasi dari rasio data tweet terdeteksi melalui pengujian.

Nilai akurasi ini dapat menunjukkan seberapa jauh perbedaan antara nilai prediksi sistem dan prediksi manusia.

Precision : Nilai yang didasarkan pada ketepatan dari system mengenai informasi dari data untuk membandingkan prediksi *true* positif dan total nilai positif. Pada *precission* nilai prediksi positif berbanding dengan jumlah nilai yang psoitif.

Recall : Nilai tingkat keberhasilan yang menunjukkan data positif dan negative yang benar. Nilai *Recall* ini dihasilkan dari jumlah nilai *true* positif dibanding nilai aktual positif t (Irwansyah Saputra, 2022).

F1-Score : Nilai perbandingan yang didapatkan dari rata-rata precision dan recall yang dibobotkan t (Irwansyah Saputra, 2022).

2.1.6 Cross Validation

Cross Validation merupakan teknik yang digunakan untuk mengevaluasi kinerja model pada dataset. Teknik ini membagi data menjadi beberapa bagian, melatih model pada beberapa bagian, dan mengujinya pada bagian yang tidak digunakan untuk pelatihan. Proses ini diulang beberapa kali, dan hasil rata-ratanya digunakan sebagai perkiraan kinerja model yang lebih akurat dibandingkan metode evaluasi lainnya. *Cross Validation* merupakan metode evaluasi yang penting dalam *machine learning* dan membantu para peneliti dan praktisi untuk mengembangkan model yang lebih akurat (Veronica, E., et al. 2020).

Salah satu jenis *Cross Validation* yang populer adalah *K-Fold Cross Validation*. Dalam metode ini, data dibagi menjadi k bagian yang sama besar. Kemudian, model dilatih pada $k-1$ bagian dan diuji pada bagian yang tersisa. Proses ini diulang k kali, dengan setiap bagian data mendapatkan kesempatan untuk menjadi data uji. Peneliti dapat memilih nilai k yang berbeda, seperti 5, 10, 15, atau 20, tergantung pada dataset dan kebutuhan penelitian.

2.1.7 Natural Language Processing

Natural Language Processing (NLP) adalah suatu bidang ilmu komputer yang mempelajari tentang bagaimana cara komputer memahami, memproses, dan berpikir seperti manusia. Analisis sentimen adalah salah satu teknik NLP yang paling sering digunakan.

Bidang ilmu komputer atau yang dikenal sebagai *Natural Language Processing* (NLP) lebih khusus lagi berfokus pada proses perintah dan kemampuan komputer untuk memahami bahasa tertulis dan lisan *Natural Language Processing* menggabungkan model statistik, *machine learning*, *deep learning*, dan pemodelan bahasa manusia berbasis aturan linguistik atau *neural network*. Teknologi ini memungkinkan komputer untuk sepenuhnya "memahami" bahasa manusia dari data ucapan atau teks serta maksud dan perasaan penulis (IBM-Amerika Serikat, 2020).

2.1.8 Analisis Sentimen

Analisis sentiment pada dasarnya adalah proses untuk memahami opini atau perasaan yang diungkapkan dalam teks. Analisis sentimen ini dapat mengekspresikan perasaan emosional dari manusia melalui tulisan berbentuk teks, apakah perasaan emosional yang tergambar cenderung sedih, senang, atau marah. Analisis sentiment juga dapat digunakan untuk menentukan pendapat terkait produk, merek ataupun perseorangan, apakah memiliki kecenderungan opini yang didapat positif, negatif, atau netral (Kurniawan, 2017).

Pada realitanya, analisis sentimen Twitter digunakan untuk menganalisis pendapat orang tentang faktor-faktor yang penting untuk sukses di pasar. Program

analisis sentimen Twitter menggunakan pendekatan pembelajaran mesin yang dianggap lebih akurat dalam menganalisis sentimen, bersama dengan teknik pemrosesan *Natural Language Processing* (NLP). Analisis sentimen atau penggalan opini didefinisikan sebagai studi komputasi tentang pendapat, perilaku, dan emosi orang terhadap entitas.(Wijaya et al., n.d, 2019).

2.1.9 Twitter (X)

Twitter (X) adalah situs web yang dibuat pada tahun 2006, Twitter(X) juga merupakan platform media sosial yang paling banyak digunakan, yang memungkinkan pengguna untuk membuat postingan dengan batas 280 karakter beserta gambar, video, tautan web, dan tautan ke platform media sosial lain yang opsional(Ferguson et al., 2014). Pengguna dapat menggunakan retweet dan suka untuk mendukung postingan. Hashtag adalah alat promosi yang digunakan di Twitter(X) untuk menyoroti dan memfokuskan pada inti dari diskusi, serta membantu mengidentifikasi tweet yang ditandai dengan hashtag yang sama (Pershad et al., 2018).

Twitter (X) dapat digunakan untuk berbagi foto dan video, hyperlink, ke sumber online dan membuat poling singkat. Setelah diluncurkan Twitter(X) terus berkembang dari layanan yang kurang dikenal menjadi fenomena di seluruh dunia dengan dampak besar pada berita terkait berbagai bidang diantaranya bisnis, politik, hiburan, olahraga, dan Pendidikan dan juga bidang lainnya (Rosell-Aguilar, 2018).

2.1.10 Pemilihan Presiden

Pemilu adalah Pemilihan Umum untuk memilih anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, Dewan Perwakilan Rakyat Daerah, Presiden dan Wakil Presiden, Gubernur dan Wakil Gubernur, Bupati dan Wakil Bupati, Walikota dan Wakil Walikota. Pemilihan Presiden (PILPRES) merupakan pemilihan umum untuk memilih Presiden dan Wakil Presiden, hal tersebut didasarkan pada UU No. 7 tahun 2017 tentang Pemilihan Umum. (Pemerintah Pusat, 2017). Sementara itu PILPRES diselenggarakan setiap lima tahun sekali sesuai masa jabatan Presiden bersamaan dengan pemilihan anggota dewan lainnya sesuai dengan undang-undang.

Pemilihan Umum di Indonesia telah memiliki sejarah panjang, namun pemilihan umum langsung baru dimulai pada tahun 2004, setelah berakhirnya era Orde Baru. Pemilihan umum Presiden tahun 2024 merupakan kesempatan penting untuk memperkuat demokrasi di Indonesia. Dalam momen pemilihan Presiden juga terdapat DPT (Daftar Pemilih Tetap) Dimana di tahun 2024 pemilih yang termasuk DPT sebanyak 204,8 juta orang. Berdasarkan data KPU tersebut, jumlah pemilih muda mencapai 106.358.447 jiwa. Rinciannya, pemilih berusia 17 tahun sebanyak 0,003 persen atau sekitar 6 ribu jiwa. Kemudian pemilih dengan rentang usia 17 tahun hingga 30 tahun mencapai 31,23 persen atau sekitar 63,90 juta jiwa. Lalu disusul dengan Pemilih dengan 31 tahun hingga 40 tahun sebanyak 20,70 persen atau sekitar 42,39 juta jiwa.

2.1.11 Elektabilitas Calon Presiden

Elektabilitas merupakan sebuah indikator penting dalam pemilihan Presiden. Istilah ini merujuk pada tingkat popularitas dan dukungan yang diperoleh seorang calon Presiden dari pemilih. Sederhananya, elektabilitas mencerminkan kemungkinan atau potensi seorang calon Presiden untuk memenangkan pemilihan berdasarkan tingkat dukungan yang mereka peroleh. Dalam hal analisis sentimen elektabilitas akan diukur dari *cross validation*, seperti yang dilakukan pada penelitian sebelumnya terkait prediksi pemilihan presiden pada tahun 2019-2024, menggunakan analisis pada Twitter. Pada penelitian tersebut elektabilitas seorang calon presiden dikatakan baik jika nilai prediksi dan nilai aktual pada True positif lebih banyak dari pada pasangan calon presiden lainnya. Persamaan untuk menghitung presentase nilai elektabilitas adalah sebagai berikut:

$$Elektabilitas_{Calon\ Presiden} = \frac{True\ Positif_{Calon\ Presiden}}{Jumlah\ total\ True\ Positif} \quad (2.14)$$

2.2 Pemimpin Yang Baik Dalam Pandangan Islam

Pemilihan Presiden adalah salah satu momen penting dalam kehidupan berbangsa dan bernegara. Dalam momen ini, rakyat memiliki kesempatan untuk menentukan pemimpin yang akan memimpin negara selama lima tahun ke depan. Oleh karena itu, pemilihan Presiden harus dilakukan dengan kehati-hatian dan penuh pertimbangan. Dalam Islam sendiri memilih pemimpin haruslah berdasarkan prinsip prinsip yang melandasi seseorang untuk memilih pemimpin diantaranya adalah hadits tentang etika dalam memilih seorang pemimpin (Lidwa Pustaka, 2020).

حَدَّثَنَا إِسْحَاقُ بْنُ إِبْرَاهِيمَ الْحَنْظَلِيُّ أَخْبَرَنَا عَيْسَى بْنُ يُونُسَ حَدَّثَنَا الْأَوْزَاعِيُّ عَنْ يَزِيدَ بْنِ يَزِيدَ بْنِ جَابِرٍ عَنْ زُرَيْقِ بْنِ حَيَّانَ عَنْ مُسْلِمِ بْنِ قَرْظَةَ عَنْ عَوْفِ بْنِ مَالِكٍ عَنْ رَسُولِ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ خَيْرُ أُمَّتِكُمُ الَّذِينَ تُحِبُّوهُمْ وَيُحِبُّونَكُمْ وَيُصَلُّونَ عَلَيْكُمْ وَتُصَلُّونَ عَلَيْهِمْ وَشَرَّارُ أُمَّتِكُمُ الَّذِينَ تُبْغِضُوهُمْ وَيُبْغِضُونَكُمْ وَتَلْعَنُونَهُمْ وَيَلْعَنُونَكُمْ قِيلَ يَا رَسُولَ اللَّهِ أَفَلَا تُنَادِيهِمْ بِالسَّيْفِ فَقَالَ لَا مَا أَقَامُوا فِيكُمْ الصَّلَاةَ وَإِذَا رَأَيْتُمْ مِنْ وَلَا تَكُمْ شَيْئًا تَكْرَهُونَهُ فَارْكَرُوهَا عَمَلَهُ وَلَا تَنْزِعُوا يَدًا مِنْ طَاعَةٍ

"Telah menceritakan kepada kami Ishaq bin Ibrahim Al Handlali telah mengabarkan kepada kami Isa bin Yunus telah menceritakan kepada kami Al Auza'i dari Yazid bin Yazid bin Jabir dari Ruzaiq bin Hayyan dari Muslim bin Qaradlah dari 'Auf bin Malik dari Rasulullah shallallahu 'alaihi wasallam, beliau bersabda: "Sebaik-baik pemimpin kalian adalah mereka mencintai kalian dan kalian mencintai mereka, mereka mendo'akan kalian dan kalian mendo'akan mereka. Dan sejelek-jelek pemimpin kalian adalah mereka yang membenci kalian dan kalian membenci mereka, mereka mengutuk kalian dan kalian mengutuk mereka." Beliau ditanya, "Wahai Rasulullah, tidakkah kita memerangi mereka?" maka beliau bersabda: "Tidak, selagi mereka mendirikan shalat bersama kalian. Jika kalian melihat dari pemimpin kalian sesuatu yang tidak baik maka bencilah tindakannya, dan janganlah kalian melepas dari ketaatan kepada mereka."(H.R MUSLIM – 3447).

Hadits ini secara umum berbicara tentang kriteria pemimpin yang baik dan pemimpin yang tidak baik. Pemimpin yang baik adalah pemimpin yang mencintai rakyatnya dan rakyat juga mencintai pemimpinnya, kedua belah pihak saling mendoakan dalam kebaikan. Sementara pemimpin yang tidak baik adalah pemimpin yang sangat dibenci oleh rakyatnya dan juga membenci rakyatnya serta saling mendoakan celaka (Nuril, 2020). Dalam Al-Qur'an juga terdapat berbagai petunjuk mengenai karakteristik pemimpin yang baik, salah satunya disebutkan dalam surat An-Nisa, ayat 58 yang artinya (Kemenag RI, 2022):

"Sesungguhnya Allah menyuruh kamu menyampaikan amanat kepada yang berhak menerimanya, dan (menyuruh kamu) apabila menetapkan hukum di antara manusia supaya kamu menetapkan dengan adil. Sesungguhnya Allah memberi pengajaran yang sebaik-baiknya kepadamu. Sesungguhnya Allah adalah Maha Mendengar lagi Maha Melihat."

Dalam surah ini dijelaskan bahwa sebagai pemimpin yang baik maka akan amanah dalam menyampaikan kepada yang berhak dan menetapkan hukum dengan adil (IbnuKatsir, 1992). Seorang pemimpin harus memegang amanah, menjaga kepercayaan rakyat, bertanggung jawab, dan mengelola sumber daya dengan bijak serta transparan. Hal ini mencakup integritas dalam menjalankan tugas tanpa menyalahgunakan kekuasaan untuk kepentingan pribadi. Selain itu seorang pemimpin harus memiliki sikap adil dalam membuat keputusan berdasarkan kebenaran tanpa memihak, menegakkan hukum secara konsisten, dan melindungi hak-hak rakyat, terutama yang lemah dan tertindas.

2.3 Kajian Topik Dengan Teori Pendukung

Analisis sentimen, sebuah disiplin ilmu yang berkembang pesat, membuka gerbang untuk memahami dan mengekstrak opini, emosi, dan perasaan dari teks. Teknik ini menjanjikan manfaat yang signifikan dalam berbagai bidang, seperti analisis media sosial, analisis ulasan pelanggan, dan analisis umpan balik. Didalam analisis sentimen, klasifikasi sentiment merupakan suatu hal yang sangat penting. Proses ini melibatkan kategorisasi teks ke dalam kategori tertentu, seperti positif, negatif, atau netral. Hal ini membantu kita untuk mengidentifikasi dan memahami sentimen yang terkandung dalam teks. Diantara berbagai metode klasifikasi yang tersedia, *Support Vector Machine* (SVM) telah menjadi salah satu algoritma terbaik dalam *machine learning* karena efektivitasnya terutama dalam klasifikasi data. Algoritma ini memanfaatkan *hyperplane* untuk memisahkan data points dengan kelas berbeda. Tetapi disisi lain algoritma ini memiliki beberapa kekurangan, seperti sensitivitas terhadap outlier, pemilihan kernel yang rumit, dan kompleksitas

komputasi untuk dataset besar, namun SVM tetap populer karena mudah di implementasikan, dan efisien.

Penelitian terkait metode SVM, salah satunya adalah penelitian terkait analisis sentimen terhadap penggunaan aplikasi Shopee, diperoleh nilai akurasi sebesar 98%, nilai presisi sebesar 98%, nilai *recall* sebesar 98%, dan nilai *f-measure* sebesar 98%. Hasil penelitian ini membuktikan hasil klasifikais menggunakan algoritma *Support Vector Machine* terbukti mampu menghasilkan akurasi yang cukup baik (Idris et al., 2023).

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan penelitian kuantitatif dengan pendekatan deskriptif dan studi literatur. Penelitian kuantitatif merupakan penelitian yang menggunakan populasi atau sampel tertentu dengan mengumpulkan data dan mengolah data yang bersifat statistik. Sedangkan pendekatan deskriptif merupakan metode yang digunakan untuk menganalisis hasil dari penelitian dan studi literatur yang digunakan dimulai dengan pencarian dan pengumpulan referensi pada buku, jurnal ilmiah, tesis, dan disertasi yang memiliki sumber kredible dan sesuai tujuan penelitian.

3.2 Data dan Sumber Data

Jenis data yang digunakan pada penelitian ini merupakan data sekunder karena diperoleh secara tidak langsung, tetapi melalui pengambilan data dari sebuah alat. Data yang digunakan dalam penelitian berupa data unggahan atau Twitter (*X*). Pengambilan data dilakukan dengan metode *crawling* pada media sosial Twitter(*X*) dalam bentuk tweets. Pengambilan data dilakukan dengan menggunakan code pada Python.

Data yang di *crawling* terkait isu ‘Pemilihan Presiden 2024’ dan ‘Calon Presiden 2024’, kata kunci yang digunakan yaitu “Probowo”, ”Anies”, ”Ganjar”. Kata kunci yang digunakan dinilai dapat menjangkau opini masyarakat Indonesia di Twitter(*X*). Data tweets yang dikumpulkan sebanyak 2000 data pada setiap *keyword*

calon Presiden. Pengambilan data dilakukan pada rentang bulan November sampai dengan bulan Desember 2023.

3.3 Pelabelan Data

Pelabelan data merupakan kegiatan pemrosesan data dengan cara penandaan data. Pada tahap ini data Twitter (*X*) akan di labeli menjadi dua kategori, yaitu positif dan negatif. Pelabelan data dilakukan secara manual oleh peneliti kemudian hasil pelabelan data tersebut akan diperiksa oleh validator data. Dalam hal ini validasi data akan dibantu oleh dua ahli bahasa yang saat ini berprofesi sebagai guru Bahasa Indonesia, dan juga seorang guru Bahasa Indonesia untuk penutur asing. Selanjutnya seluruh komentar Twitter (*X*) akan dilakukan pelabelan data secara manual dengan berdasarkan kriteria pada tabel 3.1 (Bahri et al., 2018).

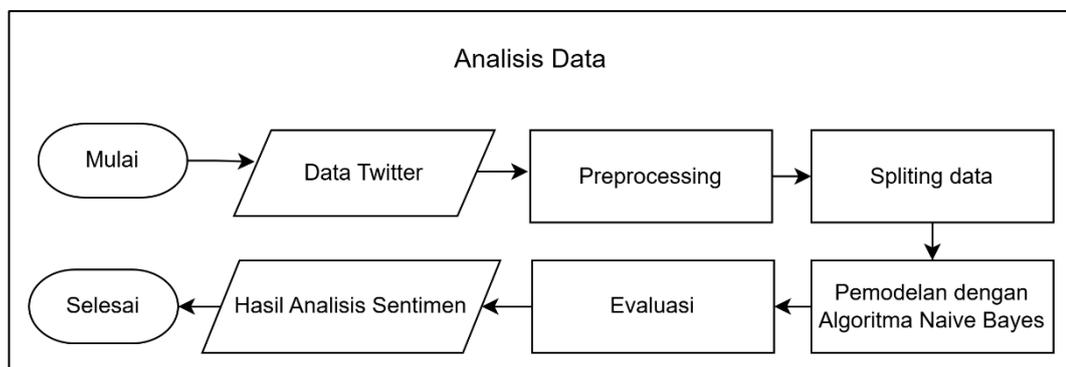
Tabel 3.1 Kriteria Pelabelan Data

Label Positif	Label Negatif
Suatu kalimat di kategorikan positif jika memiliki kalimat yang mengandung optimise, prasangka baik, unsur do'a, motivasi, ajakan menuju hal yang baik, mengekspresikan rasa senang dan mendukung calon Presiden.	Suatu kalimat di kategorikan negatif jika mengandung prasangka buruk, mengolok, menghasut, memiliki kalimat pesimis, mengandung ketidakpercayaan, menjelekkkan, mengkritik, mengekspresikan kekecewaan, tidak senang, mengandung keluhan, dan mengandung kalimat yang menjatuhkan calon Presiden dan mendukung calon Presiden lainnya.

Jika data sudah dilakukan pelabelan dengan memenuhi kriteria yang ada maka selanjutnya data tersebut akan masuk ke dalam tahap *preprocessing*.

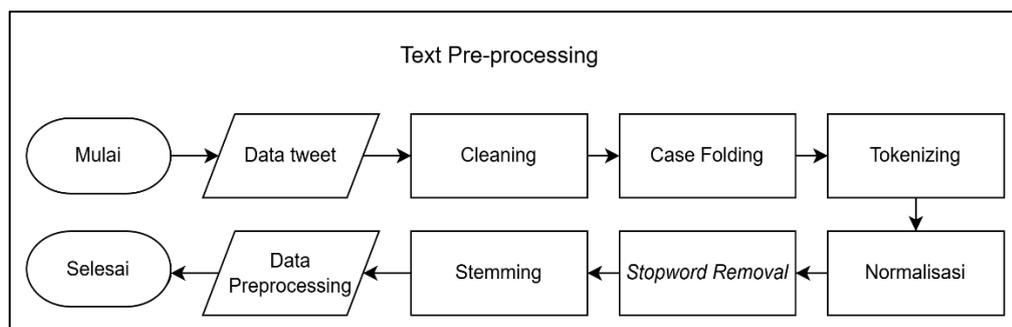
3.4 Teknik Analisis Data

Tahapan yang dilakukan untuk mendapatkan hasil dari analisis sentiment pemilihan presiden dengan menggunakan metode *Support Vector Machine* dapat dilihat pada gambar 3.1.



Gambar 3.1 *Flowchart Analisis Data*

1. Preprocessing



Gambar 3.2 *Flowchart Preprocessing*

a. Pengumpulan data

Dalam pengumpulan data diambil dengan proses *crawling*, berkaitan dengan isu 'Pemilihan Presiden 2024' dan 'Calon Presiden 2024', dengan keyword 'Prabowo', 'Anies', dan 'Ganjar'. Setiap keyword akan diambil 2.000 data

tweet sehingga nantinya akan menghasilkan 6.000 data Twitter(X).

b. *Cleaning*

Proses ini akan menghilangkan karakter yang tidak penting dalam suatu kalimat, seperti karakter *emoticon*, angka, tanda baca, simbol, spasi berlebih, tautan juga pagar. Hal ini ditujukan agar teks lebih bersih dan mempermudah mesin untuk melakukan klasifikasi.

c. *Case folding*

Proses ini mentransformasikan dari semua dataset dari pemilihan Presiden menjadi huruf kecil. Dari *casefolding* tujuan utamanya yakni untuk menyamakan konteks dan istilah didalam dataset.

d. *Tokenizing*

Setelah melalui tahap *case folding*, selanjutnya kalimat dalam dataset akan melalui proses *tokenizing*, di mana kalimat akan dijadikan token atau dipecah per-kata sehingga nantinya akan menjadi bagian bagian yang lebih kecil agar lebih mudah dipahami oleh mesin. Jika suatu kalimat tidak dilakukan *tokenizing* nantinya akan sulit bagi mesin untuk melakukan klasifikasi terhadap kata.

e. Normalisasi

Setelah proses *Tokenization*, data Twitter akan memasuki proses normalisasi, di mana kata yang memiliki kesalahan ejaan, *noise*, menggunakan bahasa daerah dan bahasa gaul (*slangword*) nantinya akan disesuaikan dengan kata yang sesuai dengan kamus Bahasa Indonesia dan ejaan yang benar.

f. *Stopword Removal*

Setelah proses normalisasi, data tweet akan dihilangkan kata-kata yang tidak

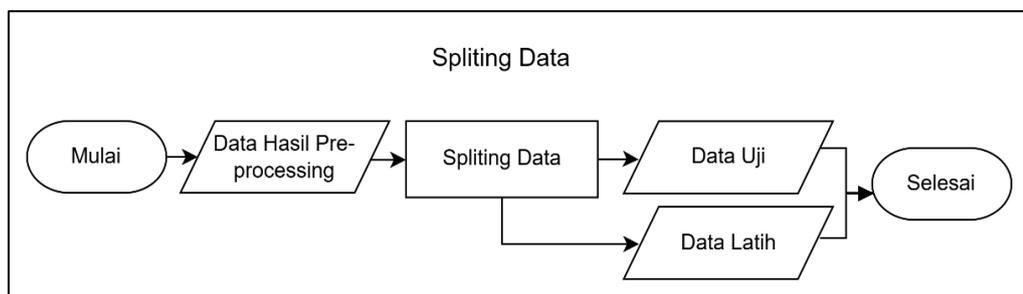
mengandung makna atau arti. Kata-kata dan juga kata konjungsi akan di hapuskan untuk mereduksi data.

g. *Stemming*

Setelah dilakukan proses *stopword removal*, data akan di *stemming*. Dalam proses ini data akan dicari kata dasarnya dan dihilangkan *prefix-suffix*. Hal ini dilakukan untuk meningkatkan akurasi temu kembali informasi (*retrieval information*).

2. *Splitting Data*

Data akan dibagi menjadi dua yakni data uji dan data latih. Data latih digunakan untuk membuat model dan data uji digunakan untuk pelatihan akurasi model. Tahapan *splitting* data ditunjukkan pada Gambar 3.3.

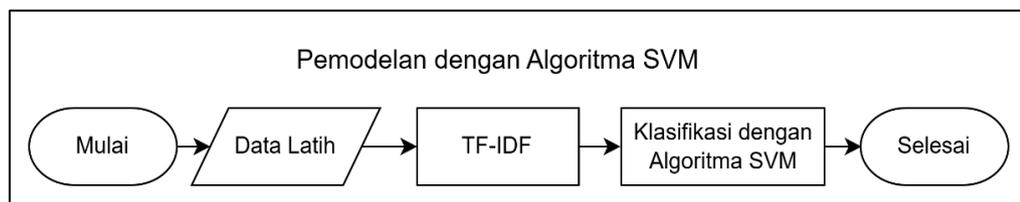


Gambar 3.3 *Flowchart Splitting Data*

- a. Data hasil *preprocessing* digunakan sebagai tahap awal pembagian data.
- b. Data *splitting* digunakan untuk mengambil secara acak data yang telah di *preprocessing*
- c. Data latih (*training*) dan data uji (*testing*).

3. Pemodelan dengan algoritma *Support Vector Machine*

Pada metode *support vector machine* dilakukan proses yaitu *TF-IDF* pada setiap dokumen dengan mengasumsikan setiap dokumen tidak memiliki rangkaian kata yang unik yang hanya terjadi pada kelas tertentu. Proses yang dilakukan dalam implementasi algoritma ini adalah dengan memberikan model dari data latih yang sudah diberi label sentimennya kemudian model akan belajar dari data tersebut untuk memprediksi sentimen pada data uji. Tahapan ditunjukkan pada Gambar 3.4.

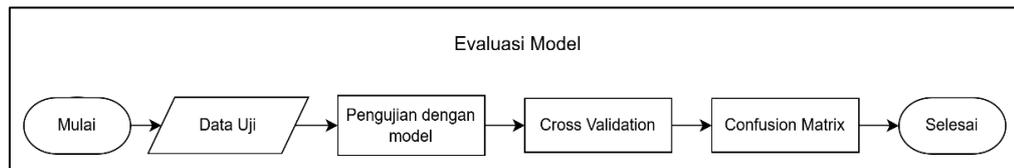


Gambar 3.4 *Flowchart* Pemodelan Dengan *Support Vector Machine*

- a. Data latih digunakan untuk proses klasifikasi dengan algoritma *support vector machine*.
 - b. Dilakukan proses *TF-IDF* untuk pembobotan kata pada data latih.
 - c. Klasifikasi dengan algoritma *Support Vector Machine* menggunakan data latih.
 - d. Hasil klasifikasi.
- ### 4. Evaluasi

Pada penelitian ini evaluasi yang dilakukan ditujukan untuk mengetahui tingkat akurasi dan seleksi fitur terhadap algoritma *Support Vector Machine*. Validasi bertujuan untuk melihat akurasi model yang digunakan dengan hasil yang sebelumnya. Teknik akurasi yang digunakan dalam penelitian akan di ukur menggunakan *confussion matrix* dan *Cross Validation*. Tahapan evaluasi model

ditunjukkan pada Gambar 3.5.



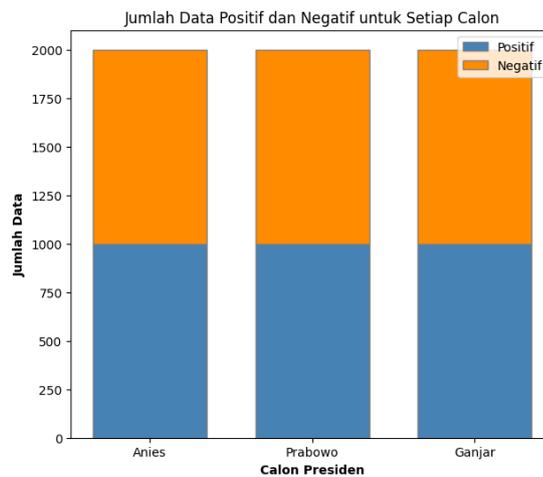
Gambar 3.5 *Flowchart* Evaluasi Model

- a. Data uji digunakan untuk proses evaluasi.
- b. Hasil model yang digunakan akan divalidasi dengan data uji.
- c. *Cross Validation* juga akan mencari model paling optimal dengan melakukan proses latih dan uji berkali-kali berdasarkan data yang telah dipartisi dengan metode ini.
- d. *Confusion matrix* akan menentukan seberapa akurat model dalam mengklasifikasikan data uji.
- e. Hasil evaluasi dengan *confusion matrix* dan *Cross Validation*.

BAB IV PEMBAHASAN

4.1 Pengumpulan Data

Dataset yang dikumpulkan merupakan hasil dari hasil *scraping* sebanyak 6000 komentar dari *platform* Twitter terkait dengan calon Presiden Indonesia tahun 2024. Proses pengambilan data ini menggunakan tools *Google Collaboratory* dengan bahasa pemrograman Python, dalam rentang waktu November- Desember 2023. Setelah pengambilan data selanjutnya dilakukan pelabelan data dan validasi oleh ahli bahasa, didapatkan komentar positif sebanyak 1000 data dan komentar negatif sebanyak 1000 pada setiap dataset komentar calon Presiden. Visualisasi perbandingan antara sentimen positif dan negatif dapat dilihat pada Gambar 4.1.



Gambar 4.1 Grafik Jumlah Data

Pada Gambar 4.1, terlihat bahwa data dengan sentimen positif sebesar 50% dan data dengan sentimen negatif sebesar 50% dari jumlah keseluruhan data komentar yang digunakan. Tabel 4.1 menyajikan sampel data komentar.

Tabel 4.1 Sampel Data Komentar Twitter

No.	Komentar	Label
1.	@RizkiR4madani @Ganjarpranowo Pinter ya pak Ganjar ngomong nya.... 😊😊 Tp biasanya yg seperti ini hasilnya alasan doang..	NEGATIF
2.	Seluruh masyarakat Indonesia siap mendukung Ganjar Pranowo #sahabatGanjar #relawanGanjar #capres2024 #Ganjarpranowofor2024 #GanjarPilihanRakyat #GanjarCapres https://t.co/5bDosiS8Qr	POSITIF
3.	@Melihat_Indo Pertunjukan memukau dari @Ganjarpranowo dalam debat pertama kali ini memang patut diapresiasi. Paparan visi misi yang gamblang, sikap tenang dalam tanya jawab, dan apresiasi positif tertinggi dari masyarakat menunjukkan kesiapan yang luar biasa.	POSITIF
4	@kikysaputrii Pak Ganjar mendengar suara dari Sabang sampai Merauke, itu Wadas ga didengar?	NEGATIF
5	@Gus_Raharjo Gagasannya gak pernah salah dan gak pernah gagal kalau pak Ganjar mah, pasti sukses terus ya pak., #GanjarSikatKorupsi	POSITIF

4.2 Preprocessing

Preprocessing adalah sebuah tahapan untuk mengubah sebuah dokumen yang tidak terstruktur menjadi lebih terstruktur dengan cara menghilangkan atribut yang tidak dibutuhkan, sehingga data menjadi sistematis dan meminimalisi *noise*. Penggunaan dari *preprocessing* diperlukan karena data yang diambil berupa ulasan pengguna Twitter yang tidak ada batasan dalam penggunaan kata dalam kalimat sehingga mempengaruhi hasil dari proses klasifikasi. Data yang tidak terstruktur akan diolah menjadi data terstruktur dengan proses *preprocessing*. Pada tahap ini

akan dibantu dengan menggunakan library *sastrawi* dimana library ini biasa digunakan dalam *Natural Language Processing* (NLP) dalam bahasa Indonesia.

1. *Cleaning*

Pada tahap ini, ini akan menghilangkan karakter yang tidak penting dalam suatu kalimat, seperti karakter *emoticon*, angka, tanda baca, simbol, spasi berlebih, tautan juga pagar. Hal ini ditujukan agar teks lebih bersih dan mempermudah mesin untuk melakukan klasifikasi. dapat dilihat hasil dari sebelum dan sesudah proses *Cleaning/Cleansing* pada data teks komentar Twitter pada tabel 4.2.

Tabel 4.2 Hasil *Cleaning* Data Komentar Twitter

Sebelum	Sesudah
@RizkiR4madani @ganjarpranowo Pinter ya pak Ganjar ngomong nya...sama kayak Anies... 😊😊 Tp biasanya yg seperti ini hasilnya alasan doang..	Pinter pak Ganjar ngomong nya sama kayak Anies biasanya seperti ini hasilnya alasan doang
Seluruh masyarakat Indonesia siap mendukung Ganjar Pranowo #sahabatganjar #relawanganjar #capres2024 #ganjarpranowofor2024 #GanjarPilihanRakyat #GanjarCapres https://t.co/5bDosiS8Qr	Seluruh masyarakat Indonesia siap mendukung Ganjar Pranowo sahabatganjar relawanganjar capres ganjarpranowofor GanjarPilihanRakyat GanjarCapres

2. *Case Folding*

Sebelum semua ulasan dipecah menjadi kata-kata, ulasan akan diubah terlebih dahulu menjadi huruf kecil (*LowerCase*) untuk menghilangkan

perbedaan seperti kata “Ganjar” (terdapat huruf kapital) dan “ganjar” yang dapat mempengaruhi hasil akurasi di akhir. Pada Tabel 4.3 dapat dilihat hasil dari sebelum dan sesudah proses *Case Folding* pada data teks komentar Twitter.

Tabel 4.3 Hasil *Case Folding* Data Komentar Twitter

Sebelum	Sesudah
Pinter pak Ganjar ngomong nya sama kayak Anies biasanya seperti ini hasilnya alasan doang	pinter pak ganjar ngomong nya sama kayak anies biasanya seperti ini hasilnya alasan doang
Seluruh masyarakat Indonesia siap mendukung Ganjar Pranowo sahabatganjar relawanganjar capres ganjarpranowofor GanjarPilihanRakyat GanjarCapres	seluruh masyarakat indonesia siap mendukung ganjar pranowo sahabatganjar relawanganjar capres ganjarpranowofor ganjarpilihanrakyat ganjarcapres

3. *Tokenizing*

Pada tahap ini digunakan untuk membagi kalimat, paragraf atau dokumen menjadi bagian-bagian yang lebih kecil, yaitu token atau kata terpisah yang berdiri sendiri. Pada Tabel 4.4 dapat dilihat hasil dari sebelum dan sesudah proses *Tokenizing* pada data teks komentar Twitter.

Tabel 4.4 Hasil *Tokenizing* Data Komentar Twitter

Sebelum	Sesudah
pinter pak ganjar ngomong nya sama kayak anies biasanya seperti ini hasilnya alasan doang	['', 'pinter', 'pak', 'ganjar', 'ngomong', 'nya', 'sama', 'kayak', 'anies', 'biasanya', 'seperti', 'ini', 'hasilnya', 'alasan', 'doang', '']

seluruh masyarakat indonesia siap mendukung ganjar pranowo sahabatganjar relawanganjar capres ganjarpranowofor ganjarpilihanrakyat ganjarcapres	['seluruh', 'masyarakat', 'indonesia', 'siap', 'mendukung', 'ganjar', 'pranowo', 'sahabatganjar', 'relawanganjar', 'capres', 'ganjarpranowofor', 'ganjarpilihanrakyat', 'ganjarcapres', '']
---	---

4. Normalisasi

Pada tahap ini merupakan tahap untuk mengubah kalimat yang tidak baku atau *slangword* menjadi kalimat baku yang sesuai dengan KBBI (Kamus Besar Bahasa Indonesia). Kamus *slangword* yang digunakan diambil dari <https://github.com/luthfilkhairi/kamus/blob/Slangword-indonesian.xlsm>.

Pada Tabel 4.5 dapat dilihat hasil dari sebelum dan sesudah proses normalisasi pada data teks komentar Twitter.

Tabel 4.5 Hasil Normalisasi Data Komentar Twitter

Sebelum	Sesudah
['', 'pinter', 'pak', 'ganjar', 'ngomong', 'nya', 'sama', 'kayak', 'anies', 'biasanya', 'seperti', 'ini', 'hasilnya', 'alasan', 'doang', '']	['', 'pintar', 'bapak', 'ganjar', 'mengomong', 'nya', 'sama', 'seperti', 'anies', 'biasanya', 'seperti', 'ini', 'hasilnya', 'alasan', 'saja', '']
['seluruh', 'masyarakat', 'indonesia', 'siap', 'mendukung', 'ganjar', 'pranowo', 'sahabatganjar', 'relawanganjar', 'capres', 'ganjarpranowofor', 'ganjarpilihanrakyat', 'ganjarcapres', '']	['seluruh', 'masyarakat', 'indonesia', 'siap', 'mendukung', 'ganjar', 'pranowo', 'sahabatganjar', 'relawanganjar', 'capres', 'ganjarpranowofor', 'ganjarpilihanrakyat', 'ganjarcapres', '']

5. *Stopword Removal*

Setelah proses normalisasi, data Twitter akan dihilangkan kata-kata yang tidak mengandung makna atau arti. Kata-kata dan juga kata konjungsi akan di hapuskan untuk mereduksi data. Proses *Stopword Removal* akan dibantu library *sastrawi* yang juga menyediakan daftar kata *stopword* dalam bahasa Indonesia. Pada Tabel 4.6 dapat dilihat hasil dari sebelum dan sesudah proses *Stopword Removal* pada tabel berikut.

Tabel 4.6 Hasil *Stopword Removal* Pada Data Komentar Twitter

Sebelum	Sesudah
['', 'pintar', 'bapak', 'ganjar', 'mengomong', 'nya', 'sama', 'seperti', 'anies', 'biasanya', 'seperti', 'ini', 'hasilnya', 'alasan', 'saja', '']	['', 'pintar', 'ganjar', 'mengomong', 'nya', 'hasilnya', 'alasan', '']
['seluruh', 'masyarakat', 'indonesia', 'siap', 'mendukung', 'ganjar', 'pranowo', 'sahabatganjar', 'relawanganjar', 'capres', 'ganjarpranowofor', 'ganjarpilihanrakyat', 'ganjarcapres', '']	['masyarakat', 'indonesia', 'mendukung', 'ganjar', 'pranowo', 'capres', 'ganjarpilihanrakyat', 'ganjarcapres', '']

6. *Stemming*

Proses *Stemming* teks yang digunakan untuk menemukan kata dasar. Pada Tabel 4.7 dapat dilihat hasil dari sebelum dan sesudah *stemming* pada data teks komentar Twitter.

Tabel 4.7 Hasil *Stemming* Data Komentar Twitter

Sebelum	Sesudah
['', 'pintar', 'ganjar', 'mengomong', 'nya', 'hasilnya', 'alasan', '']	pintar ganjar omong hasil alasan
['masyarakat', 'indonesia', 'mendukung', 'ganjar', 'pranowo', 'capres', 'ganjarpilihanrakyat', 'ganjarcapres', '']	masyarakat indonesia dukung ganjar pranowo capres ganjarpilihanrakyat ganjarcapres

4.3 *Splitting Data*

Pada tahap ini data akan dibagi menjadi data latih (*training*) dan data uji (*testing*), data dibagi dengan rasio 80% untuk *training* dan 20% untuk *testing*. Pada Tabel 4.8 ditunjukkan pembagian dari dataset.

Tabel 4.8 Pembagian Data *Training* dan *Testing*

Calon Presiden	<i>Training</i>	<i>Testing</i>
Anies	1600	400
Prabowo	1600	400
Ganjar	1600	400
Total Data	4800	1200

Data *training* akan digunakan melatih model sedangkan untuk data *testing* akan digunakan untuk melatih akurasi dari model.

4.4 Klasifikasi Dengan Algoritma SVM

4.4.1 Perhitungan Manual TF-IDF

Dalam penelitian ini, Algoritma dari *machine learning* memerlukan masukan data dalam bentuk numerik agar pelatihan dan pembuatan model dapat

dilakukan. Untuk itu akan dilakukan *TF-IDF* untuk mendapatkan data numerik. Perhitungan data pada setiap kata akan dilakukan untuk mendapatkan bobot dari jumlah kemunculan *term* dalam satu dokumen (*TF*) dan jumlah kemunculan term dalam kumpulan dokumen (*IDF*), yang dihitung dengan Persamaan (2.1) dan (2.2). Setelah itu akan dilakukan perhitungan pada Persamaan (2.3) untuk mendapatkan nilai *TF* dan *IDF*, pembobotan pada setiap term dihitung menggunakan Persamaan (2.4) untuk mendapatkan bobot (V_{norm}) masing-masing term dalam kumpulan dokumen.

Sampel data yang akan digunakan untuk proses *TF-IDF* ditunjukkan pada tabel 4.9.

Tabel 4.9 Kalimat Sample *TF-IDF*

D	Kalimat
D1	Janji palsu tambah beban rakyat
D2	Hebat bisa jadi presiden
D3	Calon pemimpin hebat

Setelah dokumen sampel didapatkan kemudian setiap dokumen akan dipecah menjadi term atau kata. Dari kata tersebut akan dihitung frekuensi kemunculannya. x_1, x_2 dan x_3 merupakan frekuensi kemunculan kata dari setiap dokumen, dimana x_1 merupakan kata yang didapatkan dari D1, x_2 merupakan kata yang didapatkan dari D2, dan x_3 merupakan kata yang didapatkan dari D3. Hasil perhitungan frekuensi dapat dilihat pada tabel 4.10

Tabel 4.10 Menghitung Frekuensi Kemunculan Kata Pada Dokumen

Kata (Term)	freq			TF			df	$\frac{D + 1}{df + 1}$	IDF
	x_1	x_2	x_3	x_1	x_2	x_3			
Janji	1	0	0	0,2	0	0	1	2,00	1,693
Palsu	1	0	0	0,2	0	0	1	2,00	1,693
Tambah	1	0	0	0,2	0	0	1	2,00	1,693

Beban	1	0	0	0,2	0	0	1	2,00	1,693
Rakyat	1	0	0	0,2	0	0	1	2,00	1,693
Hebat	0	1	1	0	0,25	0,333	2	1,33	1,288
Bisa	0	1	0	0	0,25	0	1	2,00	1,693
Jadi	0	1	0	0	0,25	0	1	2,00	1,693
Presiden	0	1	0	0	0,25	0	1	2,00	1,693
Calon	0	0	1	0	0	0,333	1	2,00	1,693
Pemimpin	0	0	1	0	0	0,333	1	2,00	1,693
Total freq	5	4	3						

Setelah dilihat frekuensi kemunculan kata dari setiap dokumen selanjutnya akan dilakukan perhitungan dengan mengalikan nilai TF dan IDF dari setiap term dengan Persamaan (2.3). Kemudian akan dilakukan normalisasi dengan *Euclidian norm* (2.4). Persamaan ini berfungsi sebagai normalisasi nilai bobot yang diperoleh, supaya nilai bobot berada pada range yang sama. Hasil dari TF-IDF dan normalisasi V dapat dilihat pada Tabel 4.11.

Tabel 4.11 Hasil perhitungan TF-IDF dan Normalisasi

TF x IDF			Normalized		
x_1	x_2	x_3	x_1	x_2	x_3
0,339	0	0	0,447	0	0
0,339	0	0	0,447	0	0
0,339	0	0	0,447	0	0
0,339	0	0	0,447	0	0
0,339	0	0	0,447	0	0
0	0,322	0,429	0	0,402	0,474
0	0,423	0	0	0,529	0
0	0,423	0	0	0,529	0
0	0,423	0	0	0,529	0

0	0	0,564	0	0	0,623
0	0	0,564	0	0	0,623

4.4.2 Perhitungan Manual Menggunakan Metode SVM

Pada perhitungan manual, dilakukan klasifikasi SVM menggunakan kernel linear yang didefinisikan pada Persamaan (2.5). Pada proses pelatihan ini, hal pertama yang dilakukan adalah mencari nilai hyperlane terbaik dengan menghitung Persamaan menggunakan kernel linear. Pada contoh akan digunakan tiga data yakni dua data untuk *training* dan satu data untuk *testing*. Data x_1 dan x_2 yang telah dilakukan normalisasi akan digunakan sebagai data *training*. Ulasan positif dari data *training* akan diberi label 1 dan ulasan negatif akan diberi label -1. Tahapan *training* akan dilakukan seperti contoh pada tabel 4.12.

Tabel 4.12 Sampel Data Perhitungan *Training*

<i>term</i>	x_1	x_2
<i>term</i> ₁	0,447	0
<i>term</i> ₂	0,447	0
<i>term</i> ₃	0,447	0
<i>term</i> ₄	0,447	0
<i>term</i> ₅	0,447	0
<i>term</i> ₆	0	0,402
<i>term</i> ₇	0	0,529
<i>term</i> ₈	0	0,529
<i>term</i> ₉	0	0,529
<i>term</i> ₁₀	0	0
<i>term</i> ₁₁	0	0
Y	-1	1

Dari sampel data perhitungan *training* pada Tabel (2.10), selanjutnya akan dihitung nilai dari matriks kernel K dengan Persamaan linier $K(x_i, x_j) = x_i x_j^T$.

Hasil dari perhitungan kernel dapat dilihat pada tabel 4.13.

Tabel 4.13 Hasil Perhitungan Persamaan Dengan Kernel Linear

$x_1 x_1^T$	$x_1 x_2^T$	$x_2 x_1^T$	$x_2 x_2^T$
0,999045	0	0	1,001127

Selanjutnya akan dihitung nilai hyperplane terbaik (a) dengan menggunakan Persamaan dualitas langrange pada Persamaan (2.7).

$$L_{(a)} = \sum_{i=1}^l a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K(x_i x_j)$$

$$0 = (a_1 + a_2) - \frac{1}{2} \{ (a_1 a_1 y_1 y_1 x_1 x_1^T) + (a_1 a_2 y_1 y_2 x_2 x_1^T) + (a_2 a_1 y_2 y_1 x_1 x_2^T) + (a_2 a_2 y_2 y_2 x_2 x_2^T) \}$$

$$0 = (a_1 + a_2) - \frac{1}{2} \{ (a_1 a_1 (1.0,999045)) + (a_1 a_2 (-1.0)) + (a_2 a_1 (-1.0)) + (a_2 a_2 (1.1,001127)) \}$$

$$\text{Dengan } \sum_{i=1}^n a_i y_i = 0$$

$$a_1(-1) + a_2(1) = 0$$

$$-a_1 + a_2 = 0$$

$$a_1 = a_2$$

$$0 = (a_1 + a_1) - \frac{1}{2} \{ (a_1 a_1 (1.0,999045)) + (a_1 a_1 (-1.0)) + (a_1 a_1 (-1.0)) + (a_1 a_1 (1.1,001127)) \}$$

$$0 = (2a_1) - 2,000172 a_1 a_1$$

$$2a_1 = 2,000172 a_1 a_1$$

$$2 = 2,000172 a_1$$

$$a_1 = 0,9999140074$$

Dan nilai $a_1 = a_2 = 0,9999140074$, Kemudian akan dilakukan pencarian nilai w (*weight*) menggunakan Persamaan (2.8). hasil yang didapatkan ditunjukkan pada Tabel 4.14.

Tabel 4.14 Nilai Weight

w_1	w_2
-0,9998959	1,0000267

Selanjutnya akan dilakukan perhitungan nilai bias menggunakan Persamaan (2.10). dan didapatkan nilai sebagai berikut:

$$\begin{aligned}
 b &= \frac{1}{NSV} \sum_{x_j \in SV} \left(\frac{1}{y_1} - \sum_{x_j \in SV} (a_j y_j K(x_j, x_i)) \right) \\
 b &= \frac{1}{NSV} \left\{ \left(\frac{1}{y_1} - (a_1 y_1 x_1 x_1^T) + (a_2 y_2 x_1 x_2^T) \right) \right. \\
 &\quad \left. + \left(\frac{1}{y_2} - (a_1 y_1 x_2 x_1^T) + (a_2 y_2 x_2 x_2^T) \right) \right\} \\
 b &= \frac{1}{2} \left(\frac{1}{-1} - ((0,9999140074 \cdot -1,0,9999140074) \right. \\
 &\quad \left. + (0,9999140074 \cdot 1,0) \right. \\
 &\quad \left. + \left(\frac{1}{1} - ((0,9999140074 \cdot -1,0) \right. \right. \\
 &\quad \left. \left. + (0,9999140074 \cdot 1,1,001127) \right) \right) \\
 b &= -0,001040
 \end{aligned}$$

Setelah mendapatkan hasil nilai a , *weight* dan bias dari proses pelatihan, selanjutnya akan dilakukan klasifikasi dengan menggunakan data *testing*. Hal ini dilakukan agar dapat mengetahui apakah data komentar Twitter terkait

calon Presiden Indonesia 2024 akan masuk ke dalam kelas positif atau negatif. Sampel yang akan digunakan untuk data *testing* adalah data x_3 yang telah dilakukan normalisasi. Berikut tahapan tahapan untuk mendapatkan hasil *testing* dapat dilihat pada tabel 4.15.

Tabel 4.15 Sampel Data Perhitungan *Testing*

<i>term</i>	x_{test}
$term_1$	0
$term_2$	0
$term_3$	0
$term_4$	0
$term_5$	0
$term_6$	0,474
$term_7$	0
$term_8$	0
$term_9$	0
$term_{10}$	0,623
$term_{11}$	0,623
Y	0

Pada matrix x_{Test} belum memiliki label, maka akan diberikan label untuk dilakukan *testing*. Selanjutnya akan dihitung dot product, dan akan dihitung nilai dari matriks kernel K dengan Persamaan linier $K(x_i, x_j) = x_i x_j^T$ pada data *testing*. Hasil dari perhitungan kernel dapat dilihat pada Tabel 4.16.

Tabel 4.16 Hasil Perhitungan Kernel Linear Dengan Data Test

$x_{test}x_1^T$	$x_{test}x_2^T$
0	0,190548

Selanjutnya akan dilakukan perhitungan untuk mendapatkan fungsi pemisah optimal dengan menggunakan Persamaan (2.10).

$$f(x_{Test}) = \text{sign}\left(\sum_{i=1}^{nsv} a_i y_i K(x_i, x_{test}) + b\right)$$

$$f(x_{Test}) = \text{sign}\left((a_1 y_1 K(x_1, x_{test}) + a_1 y_2 K(x_2, x_{test}) + b)\right)$$

$$f(x_{Test}) = \text{sign}\left((0,9999140074 \cdot -1 \cdot 0) + (0,9999140074 \cdot 1 \cdot (0,190548) + -0,001040)\right)$$

$$f(x_{Test}) = \text{sign}(0,189491)$$

$$f(x_{Test}) = +1$$

Berdasarkan perhitungan *testing* dari sample x_{Test} maka data tersebut di klasifikasikan sebagai kelas positif karena memiliki nilai pemisah optimal lebih dari nol, maka akan di klasifikasikan sebagai kelas positif. Selanjutnya proses perhitungan diatas akan diterapkan pada seluruh data *training* untuk mendapatkan model SVM yang optimal untuk mendapatkan klasifikasi dari analisis sentimen.

4.4.3 Hasil Klasifikasi menggunakan Metode SVM

Klasifikasi untuk keseluruhan data calon presiden menggunakan bahasa pemrograman Python pada *notebook google collab*. Untuk mencari akurasi terbaik, parameter yang akan digunakan pada metode SVM dengan adalah $C = 1$, $\gamma = 1$, dan kernel yang akan digunakan yakni kernel RBF. Pada penelitian ini rasio pembagian dataset dibagi menjadi dua bagian 80:20. 80% adalah data *training* dan 20% adalah data *testing*.

Selanjutnya data akan dilakukan *training* untuk mendapatkan model. Setelah model klasifikasi SVM didapatkan, dilakukan proses uji performa model

yang dilakukan dengan menggunakan code Python pada *google colab*. Akurasi yang didapatkan ditunjukkan pada Gambar 4.2.

	precision	recall	f1-score	support
0	0.70	0.79	0.74	600
1	0.76	0.66	0.71	600
accuracy			0.73	1200
macro avg	0.73	0.72	0.72	1200
weighted avg	0.73	0.72	0.72	1200

Gambar 4.2 Hasil Klasifikasi

4.5 Evaluasi

4.5.1 *Confusion Matrix*

Berikut merupakan hasil *confusion matrix* yang didapatkan dari model klasifikasi menggunakan SVM. Ditunjukkan pada Tabel 4.17.

Tabel 4.17 *Confusion Matrix* Dari Keseluruhan Data

Label		Prediksi	
		Positif	Negatif
Aktual	Positif	472	128
	Negatif	202	398

Berdasarkan Tabel 4.17 dari *confusion matrix* menunjukkan bahwa 472 benar diprediksi sebagai sentimen positif, 128 data sentimen positif diprediksi sebagai sentimen negatif, 398 data sentimen negatif diprediksi benar dan 202 sentimen negatif diprediksi salah sebagai sentimen positif. Tabel *confusion matrix* dari setiap calon Presiden akan ditunjukkan pada tabel 4.18, 4.19, dan 4.20.

Tabel 4.18 *Confusion Matrix* Anies

Label		Prediksi	
		Positif	Negatif
Aktual	Positif	173	28
	Negatif	98	102

Tabel 4.19 *Confusion Matrix* Prabowo

Label		Prediksi	
		Positif	Negatif
Aktual	Positif	167	33
	Negatif	27	173

Tabel 4.20 *Confusion Matrix* Ganjar

Label		Prediksi	
		Positif	Negatif
Aktual	Positif	133	67
	Negatif	77	123

Berdasarkan Tabel 4.18, 4.19 dan 4.20 hasil *confusion matrix* diatas terlihat bahwa model berhasil memprediksi 173 data sentimen positif benar dan 102 sentimen negatif untuk dataset Anies Baswedan, kemudian data sisanya diprediksi salah. Untuk dataset Prabowo model berhasil memprediksi 167 data sentimen positif benar dan 173 sentimen negatif benar, sisanya diprediksi salah. Model juga berhasil memprediksi 133 Sentimen positif untuk dataset Ganjar, 123 data sentimen negatif diprediksi benar, dan sisanya salah . Kemudian, dari hasil prediksi diatas dapat digunakan untuk mendapatkan nilai akurasi, presisi *recall*, dan *F1-score* dengan Persamaan yang terdapat pada Tabel (2.2) sehingga didapatkan hasil sebagai berikut:

1. Nilai Akurasi

$$\text{Akurasi} = \frac{472 + 398}{472 + 398 + 202 + 128} \times 100\% = 72,5\%$$

2. Nilai Presisi

$$\text{Presisi} = \frac{472}{472 + 202} \times 100\% = 70\%$$

3. Nilai *Recall*

$$\text{Recall} = \frac{472}{472 + 128} \times 100\% = 78,67\%$$

4. Nilai *F1-score*

$$\text{F1 - Score} = \frac{2 \times 70 \times 78,67}{70 + 78,67} \times 100\% = 74\%$$

Berdasarkan hasil perhitungan diatas, dapat disimpulkan bahwa hasil analisis sentimen terhadap data calon Presiden Indonesia 2024, menggunakan SVM memiliki akurasi sebesar 73%. Hal ini membuktikan bahwa metode SVM mampu mengklasifikasikan diagnosis hipertensi dengan cukup baik. Kemudian untuk hasil nilai presisi sebesar 70%, yang artinya dari 674 data komentar yang diprediksi oleh model SVM, sebanyak 472 komentar diprediksi benar memiliki sentimen positif. Kemudian untuk nilai recall didapatkan sebesar 78,67%, yang artinya dari 600 komentar Twitter , sebanyak 472 komentar diprediksi benar oleh model SVM. Terakhir nilai F1-Score diperoleh sebesar 74% yang artinya model SVM mampu menyeimbangkan antara presisi dan *recall* yang didapatkan dari data Twitter calon Presiden Indonesia 2024.

4.5.2 *Cross Validation*

Model klasifikasi analisis sentimen menggunakan algoritma SVM yang telah dibuat memiliki hasil dan performa yang cukup baik. Namun, sebelum

digunakan pada data yang lebih besar, model klasifikasi perlu dievaluasi lebih lanjut sehingga performa dapat diukur dengan lebih akurat. Metode evaluasi yang akan digunakan adalah *5-Fold Cross Validation* yang akan membagi data 5 bagian dengan setiap bagian akan diuji terhadap model klasifikasi yang akan dibuat dari 4 bagian sisanya. Jadi, model klasifikasi akan melalui tahapan pelatihan dan pengujian sebanyak 5 kali sehingga performa yang didapat dapat diukur dengan lebih akurat terhadap seluruh data. Berikut merupakan hasil evaluasi yang didapatkan dari *5-Fold Cross Validation*:

Tabel 4.21 Hasil Cross Validation

K	Presisi (%)	Recall (%)	F1-Score (%)	Akurasi (%)
1	69,86	69,58	69,47	69,58
2	71,3	71,16	71,1	71,16
3	73,15	73,08	73	73
4	70,73	70,58	70,5	70,5
5	72,92	72,57	72,47	72,5
Rata-rata	71,59	71,39	71,33	71,39

Berdasarkan tabel diperoleh rata-rata presisi sebesar 71,59% *recall* sebesar 71,39 *fl score* sebesar 71,33 dan rata-rata akurasi sebesar 71,39. Hasil ini menunjukkan bahwa model memiliki performa yang cukup dan akurat dalam mengklasifikasikan data calon Presiden Indonesia 2024.

4.6 Elektabilitas Calon Presiden

Nilai elektabilitas didapatkan dari hasil *True Positive confusion matrix* yang didapatkan setelah melakukan klasifikasi menggunakan metode SVM. Pada setiap calon presiden akan dihitung nilai elektabilitasnya untuk mengetahui tingkat

keterpilihan dari dataset Twitter. Perhitungan presentase elektabilitas didapatkan dari Persamaan (2.14). Berikut hasil elektabilitas dari calon presiden.

1. Elektabilitas Anies Baswedan

$$\frac{TP}{Jumlah\ Total\ TP} \times 100\% = \frac{172}{472} \times 100\% = 36,44\%$$

2. Elektabilitas Prabowo Subianto

$$\frac{TP}{Jumlah\ Total\ TP} \times 100\% = \frac{167}{472} \times 100\% = 35,38\%$$

3. Elektabilitas Ganjar Pranwo

$$\frac{TP}{Jumlah\ Total\ TP} \times 100\% = \frac{133}{472} \times 100\% = 28,18\%$$

Dari hasil perhitungan di atas dapat dilihat bahwa elektabilitas Anies Baswedan adalah yang paling tinggi diantara ketiga calon Presiden yang lainnya yakni sebesar 36,44%. Disusul dengan elektabilitas Prabowo Subianto sebesar 35,38% dan Ganjar sebesar 28,18%. Menurut hasil analisis sentimen menggunakan SVM menggunakan data twitter ini Anies Baswedan akan diprediksi memiliki tingkat elektabilitas yang lebih tinggi dibanding calon presiden yang lainnya.

Hasil dari Analisis sentimen ini berbanding terbalik dengan hasil keputusan KPU yang di keluarkan pada tanggal 21 Maret 2024 dikarenakan bebarapa faktor. Dimana terdapat faktor pendukung calon Presiden jarang yang menggunakan media sosial Twitter, faktor yang lainnya yaitu dalam rentang waktu bulan November sampai dengan Desember pendukung salah satu calon sedang gencar dalam melakukan kampanye sehingga hasil komentar positif yang dihasilkan dalam periode tersebut lebih banyak dan pola komentar yang didapatkan hampir sama.

4.7 Analisis Sentimen Terhadap Elektabilitas Calon Pemimpin Dalam Pandangan Islam

Integrasi hasil analisis sentimen dengan kajian keislaman memberikan panduan yang jelas dan menyeluruh dalam proses pemilihan presiden. Prinsip-prinsip Islam yang berkaitan dengan keadilan, verifikasi informasi, serta mengedepankan kepentingan umum dapat diterapkan dalam analisis sentimen untuk memastikan bahwa proses pemilihan presiden dilakukan dengan cara yang adil, objektif, dan membawa manfaat bagi seluruh masyarakat. Dalam hadits yang disebutkan, Rasulullah SAW menekankan pentingnya memiliki pemimpin yang mencintai rakyatnya dan dicintai oleh rakyatnya, serta bagaimana reaksi rakyat terhadap pemimpin yang tidak baik. Prinsip ini dapat diterjemahkan dalam analisis sentimen dengan mengukur tingkat sentimen positif dan negatif yang diterima oleh calon presiden dari masyarakat, sehingga dapat menggambarkan kecintaan dan kepercayaan timbal balik antara calon pemimpin dan rakyat.

Selain itu, dalam Surat Al-Hujurat ayat 6 menekankan pentingnya memeriksa kebenaran informasi sebelum menyebarkannya, yang relevan dalam konteks analisis sentimen dengan memastikan data yang dianalisis valid dan berasal dari sumber yang dapat dipercaya. Hal ini penting untuk menghindari kesalahan informasi yang dapat menimbulkan ketidakadilan dan konflik. Dalam proses analisis sentimen, teknik pemrosesan bahasa alami (NLP) dan pembelajaran mesin digunakan untuk menyaring informasi palsu atau tidak akurat, sehingga analisis sentimen memberikan gambaran yang tepat tentang opini publik. Dengan demikian, analisis sentimen tidak hanya menjadi alat teknis untuk mengukur opini publik tetapi juga mencerminkan nilai-nilai Islam dalam mencari kebenaran dan keadilan.

Secara keseluruhan, integrasi analisis sentimen dengan kajian keislaman memberikan pendekatan yang menyeluruh dan bermakna dalam pemilihan presiden. Pendekatan ini memastikan bahwa proses pemilihan tidak hanya adil dan objektif tetapi juga selaras dengan prinsip-prinsip Islam yang mengutamakan keadilan, kebenaran, dan kesejahteraan umum. Hasil analisis sentimen yang dilakukan dengan prinsip-prinsip ini dapat digunakan untuk mendukung keputusan yang membawa manfaat bagi kepentingan umum, mencerminkan aspirasi rakyat, dan mempromosikan pemimpin yang benar-benar dicintai dan dipercaya oleh masyarakat. Dengan demikian, pemilihan presiden tidak hanya menjadi proses politik tetapi juga merupakan cerminan dari nilai-nilai keislaman yang mulia, menggabungkan teknologi modern dengan kebijaksanaan spiritual untuk mencapai hasil yang terbaik bagi bangsa dan negara

BAB V

KESIMPULAN

5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan pada bab sebelumnya maka diperoleh kesimpulan sebagai berikut:

1. Tingkat akurasi yang diperoleh dari analisis sentimen pada data calon presiden menggunakan metode SVM sebesar 73%. Hal ini menunjukkan bahwa metode SVM mampu mengklasifikasikan data text dengan cukup baik.
2. Hasil estimasi analisis sentimen elektabilitas calon Presiden dalam Pemilihan Umum menurut data komentar Twitter(X) menunjukkan model terbaik menggunakan metode SVM, didapatkan hasil elektabilitas dari Anies Baswedan sebesar 36,44%, Prabowo Subianto sebesar 35,38% dan Ganjar Pranowo sebesar 28,18%. Dari analisis sentimen menggunakan SVM ini diprediksi Anies Baswedan memiliki elektabilitas yang paling tinggi di antara dua calon Presiden lainnya.

5.2 Saran

Adapun beberapa hal yang dapat dijadikan acuan untuk penyempurnaan penelitian selanjutnya maka penulis memberikan saran berikut:

1. Pada metode yang digunakan, penelitian selanjutnya diharapkan dapat menggunakan metode *deep learning* untuk menghasilkan keakuratan yang lebih baik.
2. Melakukan percobaan dengan menggunakan kamus lexicon.

3. Melakukan percobaan dengan membandingkan pelabelan dataset secara manual dan dataset dengan pelabelan menggunakan vander.
4. Dalam pengembangan selanjutnya, perlu dilakukan pengujian dengan menggunakan metode atau algoritma lainnya sebagai bahan perbandingan.

DAFTAR PUSTAKA

- Alpaydin, E. (2014). *Ethem Alpaydin-Introduction to Machine Learning-The MIT Press (2014)*. 3.
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2), 640.
- Bahri, S., Bahri, P., & Lal, S. (2018). A Novel *approach of Sentiment Classification using Emoticons*. *Procedia Computer Science*, 132, 669–678.
- Dani Mulyawan, M., & Slamet, I. (2021). *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika 2021*.
- Databoks Katadata. (2023), Januari 31 Terus bertambah, pengguna Twitter capai 421 juta akun pada akhir 2023.
- Dinakar, K., Reichart, R., & Lieberman, H. (2021). *Modeling the Detection of Textual Cyberbullying. Proceedings of the International AAAI Conference on Web and Social Media*, 5(3), 11–17. \
- Emelie, 2023: "*Dampak Media Sosial Pada Kehidupan Sehari-hari Masyarakat.*" *Jurnal Ilmiah Komunikasi*, Vol. 10, No. 1 (2023): 1-15
- F. A. Pozzi, E. Fersini, E. Messina and B. Liu, in *Sentiment Analysis In Social Network, United States*, Todd Green, 2017,
- Ferguson, C., Inglis, S. C., Newton, P. J., Cripps, P. J. S., Macdonald, P. S., & Davidson, P. M. (2014). *Social media: A tool to spread information: A case study analysis of Twitter conversation at the Cardiac Society of Australia & New Zealand 61st Annual Scientific Meeting 2013*. *Collegian*, 21(2),
- Giuseppe Bonaccorso. (2017). *Machine Learning Algorithms*. *Packt Publishing Ltd*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques*. *Journal of Chemical Information and Modeling* (Vol. 3). Elsevier. <https://doi.org/10.1016/C2009-0-61819-5>.
- IBM-United States. (2020). What is natural language processing? *Ibm.Com*
- IbnuKatsir, (1992). *Tafsir al-Qur'an al-'Adzim*, Mesir: Daar al-Fikr., Jilid I
- Idris, Irma. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5.
- Irwansyah Saputra, D. A. K. (2022). *Machine Learning untuk Pemula*. *Informatika Bandung*.

- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*.
- Kemenag RI. (2022). *Qur'an Kemenag. Lajnah Pentashihan Mushaf Al-Qur'an*. Diambil dari <https://quran.kemenag.go.id/>.
- Kurniawan, T. (2017). Implementasi Text mining pada Analisis Sentimen Pengguna Twitter Menggunakan Naïve Bayes Classifier dan Support Vector Machine. Institut Teknologi Sepuluh Nopember.
- Liang, S. (2021). *Comparative Analysis of SVM, XGBoost and Neural Network on Hate Speech Classification*. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*.
- Lidwa Pustaka. (2020). "Kitab Shahih Muslim", (Kitab 9 Imam, ver. 1.2).
- Madasu, A., & Elango, S. (2020). *Efficient feature selection techniques for sentiment analysis. Multimedia Tools and Applications*.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Mulyawan, M. D., & Slamet, I. (2021). Analisis Sentimen Terkait Vaksin Covid-19 Pada Data Twitter Menggunakan Support Vector Machine.
- Muttaqin, F. A., & Bachtiar, A. M. (2016). Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak 'Dodo Kids Browser'. *J. Ilm. Komput. Dan Inform.*
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI (Vol. 5, Issue 2))*.
- Nuril, F. (2020). Etika Memilih Pemimpin Studi Pemahaman Masyarakat Desa Karas Sedan Rembang (Studi Living Hadis). Undergraduate thesis, IAIN KUDUS.
- Pang and Lillian Lee.(2004). A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pages 271–278, Barcelona, Spain.
- Pemerintah Pusat. (2017). UU Nomor 7 Tahun 2017 Tentang Pemilihan Umum. <https://Peraturan.Go.Id/Id/Uu-No-7-Tahun-2017>.
<https://peraturan.go.id/id/uu-no-7-tahun-2017>
- Pershad, Y., Hange, P., Albadawi, H., & Oklu, R. (2018). Social Medicine: Twitter in Healthcare. *Journal of Clinical Medicine*.
- Pusat Data Ekonomi dan Bisnis Indonesia Databoks. <https://databoks.katadata.co.id/>

- Rahma Yustihan, S., & Pandu Adikara, P. (2021). Analisis Sentimen berbasis Aspek terhadap Data Ulasan Rumah Makan menggunakan Metode Support Vector Machine (SVM) . [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id), 5(3).
- Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM. *JDMSI*, 2(1), 31–37. <https://t.co/NfhmfMjtXw>
- Ritonga, A. S., & Purwaningsih, E. S. (2018). PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI KUALITAS PENGELASAN SMAW (SHIELD METAL ARC WELDING). In *Jurnal Ilmiah Edutic* (Vol. 5, Issue 1).
- Rosell-Aguilar, F. (2018). *Twitter as a formal and informal language learning tool: from potential to evidence. In Innovative language teaching and learning at university: integrating informal learning into formal language education* .
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49.
- Suhardjono, G. W. , & Abdul, H. (2019). Prediksi Waktu Kelulusan Mahasiswa Menggunakan Svm Berbasis Pso. *Bianglala Informatika*.
- Tafsirweb. (2020). Tafsir as-Sa'di / Syaikh Abdurrahman bin Nashir as-Sa'di, Surat Al-Hujurat 6. <https://tafsirweb.com/9776-surat-al-hujurat-ayat-6/>. Diakses pada tanggal 20 Desember 2023
- Tafsirweb. (2020). Hidayatul Insan bi Tafsiril Qur'an/ Ustadz Marwan Hadidi bin Musa, M.Pd.I, Surat An-Nisa 58. <https://Tafsirweb.Com/1590-Surat-an-Nisa-Ayat-58/>.
- Wijaya, Hadi Sasmita (2019). Analisis Sentimen Pendapat Pengalaman Kuliah Hybrid di Universitas Multimedia Nusantara dengan Multinomial Naive Bayes. Universitas Multimedia Nusantara, *Proceedings of 2019 5th International Conference on New Media Studies (CONMEDIA 2019)*: October 09-11, 2019, Kuta, Bali, Indonesia.
- Veronica, E., et al. (2020). *Cross-validation in machine learning: A practical guide*.
- Weiss, S. M. (2005). *Text mining : predictive methods for analyzing unstructured information*. Springer.
- Wongkar, M., & Angdresey, A. (2019). Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter. 2019 Fourth International Conference on Informatics and Computing (ICIC), 1–5. <https://doi.org/10.1109/ICIC47613.2019.8985884>

LAMPIRAN

Lampiran 1 Kode Program *Crawling* Data Twitter

```
#@title Twitter Auth Token
twitter_auth_token = ''

# Import required Python package
!pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
!sudo apt-get update
!sudo apt-get install -y ca-certificates curl gnupg
!sudo mkdir -p /etc/apt/keyrings
!curl -fsSL https://deb.nodesource.com/gpgkey/nodesource-
repo.gpg.key | sudo gpg --dearmor -o
/etc/apt/keyrings/nodesource.gpg

!NODE_MAJOR=20 && echo "deb [signed-
by=/etc/apt/keyrings/nodesource.gpg]
https://deb.nodesource.com/node_${NODE_MAJOR}.x nodistro main" |
sudo tee /etc/apt/sources.list.d/nodesource.list

!sudo apt-get update
!sudo apt-get install nodejs -y
!node -v

# Crawl Data
filename = ''
search_keyword = ''
limit = 10000
!npx -y tweet-harvest@2.6.0 -o "{filename}" -s
"{search_keyword}" --tab "LATEST" -l {limit} --token
{twitter_auth_token}
```

Lampiran 2 Kode Program Analisis Sentimen Metode SVM

```
#Install package
import pandas as pd
import numpy as np
import requests
import seaborn as sns
import matplotlib.pyplot as plt
import re
import string
import nltk
from sklearn.feature_extraction.text import TfidfVectorizer,
TfidfTransformer, CountVectorizer
nltk.download('stopwords')
```

```

from sklearn.feature_extraction.text import CountVectorizer
from nltk.tokenize import word_tokenize
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from nltk.corpus import stopwords
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory

import matplotlib.pyplot as plt
from wordcloud import WordCloud
from sklearn.metrics import (confusion_matrix,
ConfusionMatrixDisplay)

# SEE THE FILES INSIDE DIR
def load_files_in_directory(repo_url, directory_path,
token=None):
    api_url = f"{repo_url}/contents/{directory_path}"
    headers = {"Authorization": f"Bearer {token}"} if token
else {}
    response = requests.get(api_url, headers=headers)
    files_inside = []
    if response.status_code == 200:
        files = [file_info['name'] for file_info in
response.json()]
        file_contents = {}
        for file_name in files:
            file_content =
requests.get(f"{api_url}/{file_name}", headers=headers).text
            file_contents[file_name] = file_content
            for file, content in file_contents.items():
                files_inside.append(file)

        # delete unused files
        files_inside.pop(0)
        files_inside.pop(0)
        return files_inside
    else:
        return f"Failed to fetch directory {directory_path}
from {repo_url}. Status code: {response.status_code}"

# Example usage:
repo_url = "name repository in Github"

ganjar_path = "Dataset/GANJARPRE"
anies_path = "Dataset/ANIESPRE"
prabowo_path = "Dataset/PRABOWOPRE"

github_token = "....." # Replace with your actual GitHub
personal access token

ganjar_files = load_files_in_directory(repo_url, ganjar_path ,
github_token)

```

```

anies_files = load_files_in_directory(repo_url, anies_path ,
github_token)
prabowo_files = load_files_in_directory(repo_url, prabowo_path
, github_token)

all_files = ganjar_files + anies_files + prabowo_files
all_files

url_data = 'https://github.com/Sentiment-Analysis-on-Twitter-
RI-President-Electability/raw/main/Dataset'
def load_csv_files(list_file, file_url, delimiter="," ,
header='infer'):
    dataframes = []
    for i in range(len(list_file)):
        if list_file[i].endswith('csv'):
            file_path = f"{file_url}/{list_file[i]}?raw=true"
            df = pd.read_csv(file_path, delimiter=delimiter)
            dataframes.append(df)
        else:
            return "no files"
    return dataframes

ganjar = load_csv_files(ganjar_files, url_ganjar,
delimiter=",")
anies = load_csv_files(anies_files, url_anies)
prabowo = load_csv_files(prabowo_files, url_prabowo)
df_ganjar = ganjar[0]
df_anies = anies[0]
df_prabowo = prabowo[0]
print("Dataset ganjar") # ganti dengan dataset yang ingin di
tampilkan
print(ganjar[0].head(2))

# check duplicates and NaN
print('GANJAR') #ganti dengan dataset yang ingin di tampilkan
print(f'duplicated data: {df_ganjar.duplicated().sum()}')
print(f'missing data: {df_ganjar.isna().sum().sum()}')
print()

# CLEAN DATASET
def clean_dataframe(dataset):
    dataset = dataset.dropna()
    dataset = dataset.drop_duplicates()
    dataset = dataset.reset_index(drop=True)
    return dataset
df_ganjar = clean_dataframe(df_ganjar) #ganti dengan dataset
yang ingin dilakukan cleaning

# Tahap cleaning text
def remove_punct(Text):
    # menghilangkan alfabet "n" yang tidak sinkron diawal kata.
    Text = re.sub(r'\b(n)([a-zA-Z])', r'\2', Text)
    # Menghapus karakter baris baru

```

```

Text = re.sub(r'\n', ' ', Text)
# Menghapus URL atau tautan web
Text = re.sub('((www\.[^\s]+)|(https?:\/\/[^\s]+))', ' ',
Text)
# Menghapus karakter tanda kutip ganda
Text = re.sub('&quot;', " ", Text)
# Menghapus angka-angka
Text = re.sub(r"\d+", " ", str(Text))
# Menghapus kata-kata yang hanya terdiri dari satu huruf
Text = re.sub(r"\b[a-zA-Z]\b", "", str(Text))
# Menghapus karakter non-alphanumeric dan non-spasi
Text = re.sub(r"[^\w\s]", " ", str(Text))
# Menggabungkan dua atau lebih kemunculan karakter
berurutan yang sama menjadi dua kemunculan
Text = re.sub(r'(\.)\1+', r'\1\1', Text)
# Mengganti dua atau lebih spasi berturut-turut dengan satu
spasi
Text = re.sub(r"\s+", " ", str(Text))
# Menghapus karakter hastag
Text = re.sub(r'#', ' ', Text)
# Menghapus karakter selain huruf-huruf alfanumerik
Text = re.sub(r'[^a-zA-z0-9]', ' ', str(Text))
# Menghapus kata-kata yang terdiri dari satu atau dua
karakter
Text = re.sub(r'\b\w{1,2}\b', ' ', Text)
# Mengganti dua atau lebih spasi berturut-turut dengan satu
spasi
Text = re.sub(r'\s\s+', ' ', Text)
return Text
df_ganjar['cleaning'] = df_ganjar['full_text'].apply(lambda x:
remove_punct(x))

#Case Folding
df_ganjar['case_folding'] = df_ganjar['cleaning'].str.lower()

# Tokenization untuk membagi kata
def tokenization(text):
    text = re.split(r'\W+', text.lower())
    return text
df_ganjar['tokenisasi'] =
df_ganjar['case_folding'].apply(lambda x: tokenization(x))

# Snormalisasi untuk mengubah kata tidak baku
kamus_slangword =
pd.read_csv('https://raw.githubusercontent.com/sentiment_analys
is/main/data/slangwords.csv')

# kata normalisasi manual (TAMBAHIN KATA SENDIRI SESUAI DOC)
kata_normalisasi_dict = { }
for index, row in kamus_slangword.iterrows():
    if row[0] not in kata_normalisasi_dict:
        kata_normalisasi_dict[row[0]] = row[1]

```

```

def normalisasi_kata(document):
    return [kata_normalisasi_dict[term] if term in
kata_normalisasi_dict else term for term in document]
df_ganjar['konversi_slang'] =
df_ganjar['tokenisasi'].apply(lambda x: normalisasi_kata(x))

#stopwordremoval untuk mengurangi jumlah kata yang dianggap ga
penting
# Daftar stopwords bahasa Indonesia
stopword = nltk.corpus.stopwords.words('indonesian')

# Tambahkan kata-kata ke daftar stopwords
additional_stopwords = [] # tambahin nama buat semua capres
stopword.extend(additional_stopwords)

# Fungsi untuk menghapus stopwords, termasuk yang telah
ditambahkan
def remove_stopwords(Text):
    Text = [word for word in Text if word not in stopword]
    return Text

# Contoh penggunaan
df_ganjar['stopword_removal'] =
df_ganjar['konversi_slang'].apply(lambda x:
remove_stopwords(x))

#mengubah dari bentuk token menjadi bentuk kalimat kembali
def fit_stopwords(Text):
    Text = np.array(Text)
    Text = ' '.join(Text)
    return Text
df_ganjar['cleaned_text'] =
df_ganjar['stopword_removal'].apply(lambda x: fit_stopwords(x))

# Stemming
factory = StemmerFactory()
stemmer = factory.create_stemmer()
def stemming(text):
    return stemmer.stem(text)
df_ganjar['stemming'] = df_ganjar['cleaned_text'].apply(lambda
x: stemming(x))

# split into train test data
from sklearn.model_selection import train_test_split
# GANJAR
X_ganjar = df_ganjar['stemming']
y_ganjar = df_ganjar['Label']

#dibagi menjadi 80:20 untuk train dan test
X_train_ganjar, X_test_ganjar, y_train_ganjar, y_test_ganjar =
train_test_split(X_ganjar, y_ganjar, test_size=0.2,
random_state=42, stratify=y_ganjar)

```

```

#Labelencoder
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
# Step 1: Split training and testing data
X_all = df_ganjar['stemming']
y_all = df_ganjar['Label']
# Step 2: Transform the dataset using TF-IDF
tfidf_vectorizer = TfidfVectorizer()
X_tfidf = tfidf_vectorizer.fit_transform(X_all)
# Step 3: Apply LabelEncoder to the target variable
label_encoder = LabelEncoder()
y = label_encoder.fit_transform(y_all)

#Klasifikasi dengan Algoritma SVM
from sklearn.model_selection import GridSearchCV
from sklearn.svm import SVC
# Define the hyperparameters to tune
param_grid = {'C': [0.1, 1, 10, 100],
              'kernel': ['linear', 'rbf', 'sigmoid'],
              'gamma': [0.1,0.01,0.001,1,'scale', 'auto']}

# Create an SVC classifier
svc_classifier = SVC()
# Perform Grid Search with cross-validation
grid_search = GridSearchCV(svc_classifier, param_grid, cv=3,
scoring='accuracy')
grid_search.fit(X_train, y_train)
# Get the best hyperparameters
best_params = grid_search.best_params_
best_params

# TRAINING PROCESS
# Create a SV classifier
from sklearn.svm import SVC
cf = SVC(C=1, gamma=1, kernel='rbf')

# Train the model
cf.fit(X_train, y_train)

# Make predictions on the test set
y_pred = cf.predict(X_test)

# CLASSIFICATION REPORT
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))

#confusion matrix
pred_GANJAR = cf.predict(X_GANJAR)
cm = confusion_matrix(y_GANJAR, pred_GANJAR)
class_labels = list(all_data['Label'].unique())

```

```

# Create a ConfusionMatrixDisplay object and customize the plot
disp = ConfusionMatrixDisplay(confusion_matrix=cm,
display_labels=class_labels)

# Plot the confusion matrix with labels and title
fig, ax = plt.subplots(figsize=(8, 6))
disp.plot(ax=ax)
plt.xlabel('label prediksi')
plt.ylabel('label sebenarnya')
plt.title('GANJAR Confusion Matrix ')
plt.show()

# STRATIFIED KFOLD CROSS VALIDATION
from sklearn.model_selection import RepeatedStratifiedKFold
from sklearn.metrics import accuracy_score, f1_score,
precision_score, recall_score

# Assuming you have your dataset X and labels y

# Specify the number of splits and repeats
n_splits = 5 # Number of folds
n_repeats = 1 # Number of repeats

# Specify NuSVC parameters
C = 1

# Create a RepeatedStratifiedKFold object
rskf = RepeatedStratifiedKFold(n_splits=n_splits,
n_repeats=n_repeats, random_state=42)

# Initialize SVC with linear kernel and specified parameters
model = SVC(kernel=' ', C= , gamma= )

acc = []
prec = []
rec = []
f1score = []

# Perform repeated stratified cross-validation
for train_index, test_index in rskf.split(X_tfidf, y):
    X_train, X_test = X_tfidf[train_index], X_tfidf[test_index]
    y_train, y_test = y[train_index], y[test_index]

    # Train the SVC model
    model.fit(X_train, y_train)

    # Make predictions on the test set
    predictions = model.predict(X_test)

    # Evaluate accuracy
    accuracy = accuracy_score(y_test, predictions)
    precision = precision_score(y_test, predictions,
average="macro")

```

```

recall = recall_score(y_test, predictions, average="macro")
f1 = f1_score(y_test, predictions, average="macro")

acc.append(accuracy)
prec.append(precision)
rec.append(recall)
f1score.append(f1)

print(f"Accuracy:{accuracy}, Precision:{precision},
Recall:{recall}, F1:{f1}")

print(f"Average accuracy : {np.mean(acc)}")
print(f"Average precision : {np.mean(prec)}")
print(f"Average recall : {np.mean(rec)}")
print(f"Average f1 score : {np.mean(f1score)}")

```

Lampiran 3 Dataset Calon Presiden yang sudah dilakukan pelabelan manual berikut isi dari dataset calon presiden yang telah dilakukan pelabelan manual, untuk data lengkap dilampirkan pada link <https://tinyurl.com/Dataset-Calon-Presiden>

full_text	Label
Anies Pasti Menang Satu Putaran, yakin banget	POSITIF
@moncheriyoo sumoah gue gahabis pikir ya tuh pas anies udah kasih data panjang lebar yang ditangkep cuman kok salahin angin ANJIR KAYAK NGOMONG SM ORANG TUA YG KERAS KEPALA	NEGATIF
anies dihari Rakyat Indonesia, pasti akan menang satu PUTARAN	POSITIF
@PartaiSocmed nah mending gini gpp memang begitu tum @PartaiSocmed ga usa tantrum baru di kritik dikit aja kok uda ngos-ngosan. pak anies di kritik sebanyak itu aja masi santai tapi beliau masi dengan nilainya beda sama anda Qobul aja dulu	POSITIF
@Muhammad_Saewad Fahri Hamzah semakin hari semakin bodoh tidak punya idealisme partai malah ngomong Anies	NEGATIF
@OposisiCerdas Dia ini ratunya partai pnyebar hoax ya psi punya dendam apa si sm pak anies smpe segitunya pngen ngejatuhin pak anies	POSITIF
@detikcom Ini tuh sinyal (kode keras) untuk capres yang menolak IKN alamat nyungsep. Anies menang ibu kata DKi Jakarta !!! Iya gpp. Kalau ibu kota tetep di IKN dong..!!! *timhoreeee..	NEGATIF
@aintyoursm @tubirfess Banyak orang (yg sering kemakan framing politik identitas) belum tau bahwa Anies sebenarnya gak se-kanan itu	NEGATIF
@NasDem Omong kosong semua ucapan mu Anies	NEGATIF
@MichelAdam7__ Dukungan kami terhadap Anies Baswedan dan Gus Imin karena dedikasi mereka dalam membangun infrastruktur kota yang lebih baik.	POSITIF

@sobat_anies Orang bodoh disuruh bicara ya gini jadinya	NEGATIF
@r_yulia @sobat_anies @GibraltarNc Anis Rasyid Baswedan PRESIDEN RI 2024	POSITIF
@DrEvaChaniago Ga bs komen yg enak...jadi ga mau komen. Ntar klo 2024 Anies udah jadi Presiden bisa bebas bicara. Ga takut dijemput subuh2	POSITIF
YAKIN PASTI MENANG	POSITIF
@atok_senger Anies Gagal Memimpin Jakarta	NEGATIF
Anies pembohong kelas kakap tukang ngibulin rakyat pembawa kehancuran Anies sonde layak jadi Presiden https://t.co/NQEvmuYyUW	NEGATIF
@Muhammad_Saewad Anies Rasyid Baswedan presiden Indonesia 2024 aminnn	POSITIF
@Metro_TV @aniesbaswedan Tidak perlu umbar janji yang penting tindakan terpenuhi itu adalah pak anies	POSITIF
@rustamaji @kafiradikalis Dulu saya jg pendukung anis bgt. Setelah anis jd pengkhianat g punya integritas dan cuma jadi bonekanya brewok. Malah saya pengen anies kalah.	NEGATIF
@PartaiSocmed jelas2 Anies bilang dan di tegesin ama mba nana waktu itu (tahun 2019) Anies tanggung jawab ama amanat rakyat ampe 5 tahun.. tuntas.. ga kaya sebelumnya yg baru 2 tahun dah ninggalin buat jd presiden..	POSITIF
Anies - Cak Imin pemimpin yang mengakar pada nilai-nilai kebersamaan. #sayaAMIN2024	POSITIF
Pendukung Anies tua hitam jelek bahasa islami bgt	NEGATIF
@tempodotco bisalah ANIES TUKANG NGIBUL DAN CUCI TANGAN. NATURALISASI NGIBUL NIPU REKLAMASI NIPU DAN MENJIJIKKAN CUMA GANTI.NAMA KORUPSI MJD KELEBIHAN BAYAR.	NEGATIF
ORANG BODOH PILIH ANIES PEMBOHONG	NEGATIF
@txtdariyeji Gue ga bilang anies jelek ya. Eksekusi dia di program JAKLINGKO dan JAKI itu bagus banget. Patut apresiasi. Cuma klo adu program dan eksekusi sama ahok ya beda kelas lah. Seandainya ahok muslim lu pada rela kali dia jadi presiden 3 periode.	NEGATIF
@dantsus @171cakmus @gm_gm Yg bilang usulan dr tim AMIN itu kan dr kubu 02 sedangkan Anies sendiri sudah membantah kalau tim nya tidak pernah memberi usulan diajak diskusi saja tidak pernah. Kalau sudah tolol karena dukung prabodoh dan gibranjing minimal jgn bego lah	POSITIF
@ekowboy2 Anies Jangan coba2 ganggu gagasan Jokowi...ambblas nanti surveynya....buat dan perkuat gagasan orisinilmu sendiri aja (klo ada) sibuuuk mengkoreksi Jokowi tapi gagasan orisinil kagak ada ini kesannya Anies sok paling pinter tapi minterin	NEGATIF
@tempodotco @msaid_didu Setuju pak Anies Baswedan	POSITIF
@ishqmee Anies terlalu banyak teori makanya gagal bikin JIS #MulutManisAnies	NEGATIF

RIWAYAT HIDUP



Penulis, Alfi Rahma Dani, lahir di Malang pada tanggal 25 Juli 2001, Penulis lahir dari pasangan Bapak Syamsul Huda dan Ibu Nur Siti sebagai anak pertama dari tiga bersaudara. Penulis menempuh pendidikan dasar di SDN Jambearjo 01, kemudian melanjutkan pendidikan menengah pertama di SMPN 1 Bululawang dan lulus pada tahun 2017. Setelah itu, penulis menyelesaikan pendidikan menengah atas di SMAN 1 Bululawang dan lulus pada tahun 2020. Pada tahun yang sama penulis melanjutkan studi di Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang mengambil program studi Matematika di Fakultas Sains dan Teknologi. Selama masa kuliah, selain menyelesaikan tugasnya sebagai mahasiswi, penulis juga merupakan penerima Beasiswa dari Bank Indonesia yang tergabung pada Generasi Baru Indonesia (GenBI Malang). Penulis juga aktif mengikuti berbagai kegiatan organisasi di dalam kampus, termasuk kepanitiaan dalam KOMET 2021 dan 2022, dilain waktu juga penulis menjabat sebagai ketua ARTMOSTFAIR tahun 2022. Selain itu penulis juga merupakan pengurus HMPS “Integral” Matematika pada tahun 2021 dan menjadi pengurus DEMA Fakultas Sains dan Teknologi pada tahun 2022. Penulis juga pernah menjadi asisten praktikum Pengenalan Pemrograman Komputer selama satu semester dan mengikuti kegiatan di luar kampus seperti pelatihan dan seminar.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Alfi Rahma Dani
NIM : 200601110055
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Implementasi Algoritma Support Vector Machine Untuk Analisis Sentimen Terhadap Elektabilitas Calon Presiden 2024
Pembimbing I : Hisyam Fahmi, M.Kom.
Pembimbing II : Ach. Nashichuddin, M.A.

No	Tanggal	Hal	Tanda Tangan
1.	25 Januari 2024	Konsultasi Topik dan Data	1.
2.	01 Februari 2024	Konsultasi Bab I, II, dan III	2.
3.	06 Februari 2024	Konsultasi Bab I, II, dan III	3.
4.	13 Februari 2024	Konsultasi Bab I, II, dan III	4.
5.	17 Februari 2024	Konsultasi Bab I, II, dan III	5.
6.	25 Februari 2024	ACC Bab I, II, dan III	6.
7.	26 Februari 2024	Konsultasi Kajian Agama Bab I dan II	7.
8.	28 Februari 2024	ACC Kajian Agama Bab I dan II	8.
9.	04 Maret 2024	ACC Seminar Proposal	9.
10.	08 Maret 2024	Konsultasi Revisi Seminar Proposal	10.
11.	30 April 2024	Konsultasi Bab IV dan V	11.
12.	23 Mei 2024	Konsultasi Bab IV dan V	12.
13.	31 Mei 2024	Konsultasi Bab IV dan V	13.
14.	08 Juni 2024	Konsultasi Bab IV dan V	14.



**KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

15.	12 Juni 2024	ACC Bab IV dan V	15.
16.	13 Juni 2024	Konsultasi Kajian Agama Bab IV	16.
17.	15 Juni 2024	ACC Kajian Agama Bab IV	17.
18.	19 Juni 2024	ACC Seminar Hasil	18.
19.	19 Juni 2024	ACC Seminar Hasil lanjutan	19.
20.	19 Juni 2024	Konsultasi Revisi Seminar Hasil	20.
21.	22 Juni 2024	ACC Sidang Skripsi	21.
22.	26 Juni 2024	ACC Keseluruhan	22.

Malang, 26 Juni 2024

Mengetahui,

Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005