

PENERAPAN MACHINE LEARNING PADA ANALISIS SENTIMEN TWITTER SEBELUM DAN SESUDAH DEBAT CALON PRESIDEN DAN WAKIL PRESIDEN TAHUN 2024

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi

Oleh:



ZAIRY CINDY DWINNIE

12050324081



FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2024

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



LEMBAR PERSETUJUAN

**PENERAPAN MACHINE LEARNING PADA ANALISIS
SENTIMEN TWITTER SEBELUM DAN SESUDAH DEBAT
CALON PRESIDEN DAN WAKIL PRESIDEN TAHUN 2024**

TUGAS AKHIR

Oleh:

ZAIRY CINDY DWINNIE

12050324081

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 3 Juni 2024

Ketua Program Studi

Eki Saputra, S.Kom., M.Kom.
NIP. 198307162011011008

Pembimbing

Dr. Rice Novita, S.Kom., M.Kom.
NIP. 198511272023212032

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

PENERAPAN MACHINE LEARNING PADA ANALISIS SENTIMEN TWITTER SEBELUM DAN SESUDAH DEBAT CALON PRESIDEN DAN WAKIL PRESIDEN TAHUN 2024

TUGAS AKHIR

Oleh:

ZAIRY CINDY DWINNIE
12050324081

Telah dipertahankan di depan sidang dewan penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau di Pekanbaru, pada tanggal 7 Mei 2024

Pekanbaru, 7 Mei 2024
Mengesahkan,

Ketua Program Studi



Dr. Harsono, M.Pd.
NIP. 1964030119992031003



Eki Saputra, S.Kom., M.Kom.
NIP. 198307162011011008

DEWAN PENGUJI:

Ketua : Eki Saputra, S.Kom., M.Kom.

Sekretaris : Dr. Rice Novita, S.Kom., M.Kom.

Anggota 1 : Mustakim, ST., M.Kom.

Anggota 2 : M. Afdal, ST., M.Kom.

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

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

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

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

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

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

Lampiran Surat :
Nomor : Nomor 25/2021
Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Zainy Cindy Dwinnie
NIM : 12050324081
Tempat/Tgl. Lahir : Dumai, 8 Mei 2002
Fakultas/Paseasarjana : Sains dan Teknologi
Prodi : Sistem Informasi

Judul Disertasi/Thesis/Skripsi/Karya Ilmiah lainnya*:

Penerapan Machine Learning Pada Analisis Sentimen Twitter terhadap
Sebelum dan sesudah Debat Calon Presiden dan Wakil Presiden 2024.

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan ~~Disertasi/Thesis/Skripsi/Karya Ilmiah~~ lainnya* dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu ~~Disertasi/Thesis/Skripsi/Karya Ilmiah~~ lainnya* saya ini, saya nyatakan bebas dari plagiat.
4. Apa bila dikemudian hari terbukti terdapat plagiat dalam penulisan ~~Disertasi/Thesis/Skripsi/(Karya Ilmiah lainnya)*~~ saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

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

Pekanbaru, 10 Juni 2024
Yang membuat pernyataan



NIM : 12050324081

*pilih salah satu sesuai jenis karya tulis

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan menyebut nama Allah yang maha pengasih lagi maha penyayang

Assalamu 'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah Rabbil 'Alamin, segala puji bagi Allah Subhanahu Wa Ta'ala sebagai bentuk rasa syukur atas segala nikmat yang telah diberikan tanpa ada kekurangan sedikitpun. Shalawat beserta salam tak lupa pula kita ucapkan kepada Nabi Muhammad Shallallahu 'Alaihi Wa Sallam dengan mengucapkan *Allahumma Sholli 'ala Sayyidina Muhammad Wa 'ala Ali Sayyidina Muhammad*. Semoga kita semua selalu senantiasa mendapat syafa'at-Nya di dunia maupun di akhirat, *aamiin ya rab-bala 'alaamiin*. Kupersembahkan hadiah istimewa karya kecil ini sebagai salah satu bentuk bakti, rasa terima kasih, dan hormatku kepadamu orang tuaku tercinta, Ayah dan Ibu.

Ayah dan Ibu tersayang, terima kasih untuk setiap perjuangan yang di usahakan, doa yang selalu di berikan disetiap sujud panjang, membimbing, dan mendorong saya dalam kebaikan dan selalu ada saat keadaan tersulit sekalipun. Terima kasih untuk segala pengorbanan yang dilakukan. Sampai kapanpun tiada rasa, tiada cara yang dapat membalas semua yang telah kalian lakukan. Untuk itu saya anakmu ini selalu mendoakan yang terbaik agar Ayah dan Ibu bahagia dunia dan akhirat, serta diberikan tempat istimewa di sisi-Nya kelak sehingga kita bisa berkumpul kembali bersama-sama di Jannah-Nya.

Terima kasih saya ucapkan kepada Ibu Rice Novita selaku pembimbing Tugas Akhir saya yang telah membimbing dari awal perkuliahan hingga saat ini dengan setulus hati. Terima kasih saya ucapkan kepada Bapak dan Ibu Dosen Sistem Informasi yang selama ini sudah mewariskan ilmu, motivasi, dan arahan untuk menyelesaikan studi di Program Studi Sistem Informasi ini. Terima kasih juga saya ucapkan kepada kakak-kakak tingkat dan teman-teman saya yang telah membantu dan memberikan semangat dalam menyelesaikan Tugas Akhir ini.

Wassalamu 'alaikum Warahmatullahi Wabarakaatuh.

KATA PENGANTAR

Assalamu 'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillah Rabbil 'Alamin, bersyukur kehadiran Allah Subhanahu Wa Ta'ala atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Shalawat serta salam kita ucapkan kepada Nabi Muhammad *Shalallahu 'Alaihi Wa Sallam* dengan mengucapkan *Allahumma Sholli'Ala Sayyidina Muhammad Wa'Ala Ali Sayyidina Muhammad*. Tugas Akhir ini dibuat sebagai salah satu syarat untuk mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi Universitas Islam Negeri Sultan Syarif Kasim Riau. Pada penulisan Tugas Akhir ini, terdapat beberapa pihak yang sudah berkontribusi dan mendukung penulis baik berupa materi, moril, dan motivasi. Penulis ingin mengucapkan banyak terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd sebagai Dekan Fakultas Sains dan Teknologi.
3. Bapak Eki Saputra, S.Kom., M.Kom sebagai Ketua Program Studi Sistem Informasi, Fakultas Sains dan teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Siti Monalisa, ST., M.Kom sebagai Sekretaris Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
5. Ibu Dr. Rice Novita, S.Kom., M.Kom sebagai Dosen pembimbing Proposal dan Tugas Akhir yang telah berkenan membimbing, mengarahkan dan meluangkan waktu, tenaga dan pikiran guna membantu penulisan dalam menyelesaikan Tugas Akhir ini.
6. Para Dosen Program Studi Sistem Informasi Fakultas Sains dan Teknologi Sultan Syarif Kasim Riau yang memberikan ilmu yang bermanfaat serta memberikan semangat untuk meraih cita-cita dan masa depan yang cerah.
7. Seluruh Pegawai dan Staf Fakultas Sains dan Teknologi yang telah membantu dan mempermudah proses administrasi selama perkuliahan ini.
8. Keluarga hebat dan membanggakan, keluarga *Puzzle Research Data Technology* (PREDATECH), yang terdiri dari dosen pembimbing terhebat (Pak Mustakim, Pak Inggih, Pak Afdal, Buk Rice dan dosen lainnya) dan seluruh alumni serta anggota aktif PREDATECH yang tidak dapat penulis sebutkan satu persatu.
9. Teristimewa untuk kedua orang tua penulis yang selalu menjadi motivasi

Hak Cipta Dilindungi Undang-Undang

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

untuk penulis dalam menyelesaikan segala urusan dalam perjalanan karir penulis.

10. Kepada saudara kembar penulis yang selalu membantu dan memberikan motivasi selama perkuliahan hingga penyelesaian Tugas Akhir.
11. Kepada keluarga serta saudara-saudara penulis yang terus memberikan semangat dan motivasi kepada penulis sehingga penulis dapat menyelesaikan Tugas Akhir.
12. Para Dosen Program Studi Sistem Informasi Fakultas Sains dan Teknologi Sultan Syarif Kasim Riau yang memberikan ilmu yang bermanfaat serta memberikan semangat untuk meraih cita-cita dan masa depan yang cerah.
13. Kepada Kakak-kakak senior. Terima kasih atas motivasi dan arahnya.
14. Serta semua pihak yang namanya tidak dapat disebutkan satu persatu, yang telah banyak membantu dalam pelaksanaan penelitian Tugas Akhir dan menyelesaikan Tugas Akhir ini.

Semoga segala doa dan dorongan yang telah diberikan selama ini menjadi amal kebajikan dan mendapat balasan setimpal dari Allah *Subhanahu Wa Ta'ala*. Penulis menyadari bahwa penulisan Tugas Akhir ini masih banyak terdapat kekurangan dan jauh dari kata sempurna. Untuk itu kritik dan saran yang membangun sangat diharapkan untuk kesempurnaan Tugas Akhir ini dan semoga laporan ini bermanfaat bagi kita semua. Akhir kata penulis ucapkan terima kasih.

Wassalamu'alaikum Warahmatullahi Wabarakaatuh.

Pekanbaru, 3 Juni 2024

Penulis,

ZAIRY CINDY DWINNIE

NIM. 12050324081

UIN SUSKA RIAU

Jurnal
Media Informatika Budidarma



Diterbitkan Oleh :



STMik Budi Darma Medan

Jl. Sisingamangaraja No.338 Simpang Limun Medan

Telp. 061-7875998

<http://www.stmik-budidarma.ac.id>

Jurnal Media Informatika Budidarma	Volume : No.	Halaman:	Medan	ISSN 2548-8368 (media online)
------------------------------------	--------------	----------	-------	-------------------------------------

Hak Cipta Dilindungi Undang-Undang

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

Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum dan Sesudah Debat Calon Presiden dan Wakil Presiden Tahun 2024

Zairy Cindy Dwinnie*, Rice Novita

Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ^{1,*}12050324081@students.uin-suska.ac.id, ²rice.novita@uin-suska.ac.id

Email Penulis Korespondensi: 12050324081@students.uin-suska.ac.id

Abstrak—Pemilihan Presiden 2024 menjadi topik terhangat pada dua tahun belakang. KPU telah mengkonfirmasi bahwasannya ada 3 kandidat Calon Presiden dan Wakil Presiden. Untuk itu, sebagai momentum bagi pemilih untuk menilai kandidat Presiden dan Wakil Presiden 2024, KPU melaksanakan Debat Pilihan Presiden 2024 yang didasarkan pada UU Nomor 7 Tahun 2017 tentang Pemilihan Umum. Berdasarkan informasi yang dipaparkan dilaman kpu.go.id, debat akan diselenggarakan sebanyak 5 kali dengan 3 debat calon presiden dan 2 debat calon wakil presiden. Untuk itu perlu dilakukan analisis untuk mengetahui bagaimana sentimen publik dalam bentuk positif, negatif dan netral di Twitter terhadap ketiga calon pasangan Presiden dan Wakil Presiden 2024 pada saat sebelum dan sesudah debat diselenggarakan. Tujuannya untuk memperkirakan dukungan atau ketidaksetujuan publik terhadap ketiga pasangan calon. Pada penelitian ini menggunakan tiga algoritma sebagai perbandingan akurasi klasifikasi yaitu algoritma Support Vector Machine, Random Forest dan Logistic Regression. Dimana data yang digunakan adalah data tweet pada twitter terkait sebelum dan sesudah debat sebanyak 30 dataset dengan jumlah keseluruhan 9000 data. Dari hasil klasifikasi, didapat rata-rata akurasi pada ketiga algoritma yaitu SVM dan Random Forest sebesar 78%, dan Logistic Regression sebesar 79%. Didapat polaritas tertinggi dari klasifikasi ketiga algoritma terdapat pada kelas positif. Hal ini menandakan bahwa algoritma Logistic Regression memberikan kinerja lebih baik dalam mengklasifikasikan sentimen Twitter terkait pasangan calon presiden dan wakil presiden 2024.

Kata Kunci: Text Mining; Debat Pilpres 2024; Support Vector Machine; Random Forest; Logistic Regression

Abstract—The 2024 Presidential Election has become the hottest topic in the past two years. The KPU has confirmed that there are 3 candidates for President and Vice President. For this reason, as a momentum for voters to assess the 2024 Presidential and Vice Presidential candidates, the KPU is holding the 2024 Presidential Choice Debate which is based on Law Number 7 of 2017 concerning General Elections. Based on the information presented on the kpu.go.id page, the debate will be held 5 times with 3 presidential candidate debates and 2 vice presidential candidate debates. For this reason, it is necessary to carry out an analysis to find out how public sentiment is positive, negative, and neutral on Twitter towards the three candidates for President and Vice President in 2024 before and after the debate was held. The aim is to estimate public support or disapproval of the three candidate pairs. This research uses three algorithms as a comparison of classification accuracy, namely the Support Vector Machine algorithm, Random Forest, and Logistic Regression. Where the data used is tweet data on Twitter related to before and after the debate as many as 30 datasets with a total of 9000 data. From the classification results, the average accuracy obtained for the three algorithms, namely SVM and Random Forest, was 78%, and Logistic Regression was 79%. The highest polarity obtained from the classification of the three algorithms is in the positive class. This indicates that the Logistic Regression algorithm provides better performance in classifying Twitter sentiment regarding the 2024 presidential and vice presidential candidate pairs.

Keywords: Text Mining; 2024 Presidential Election Debate; Support Vector Machine; Random Forest; Logistic Regression

1. PENDAHULUAN

Negara-negara demokratis menganggap pemilu sebagai simbol dan ukuran demokrasi itu sendiri. Hasil pemilihan parlemen berlangsung dalam suasana terbuka dan bebas, kebebasan berserikat dianggap sebagai partisipasi masyarakat[1]. Pemilu di Indonesia telah berlangsung sejak tahun 1955 untuk memilih lembaga legislatif. Di tingkat nasional tingkat, masyarakat Indonesia tidak memilih presiden sampai tahun 2004. Hingga saat itu, merupakan kali pertama pemilihan presiden dan anggota Majelis Permusyawaratan Rakyat dilakukan pada hari yang sama[2]. Peristiwa Politik terpenting akan terjadi dalam di tahun 2024. Dimana habisnya periode Joko Widodo sebagai presiden sekaligus Pemilihan Presiden Indonesia tahun 2024. Pemilihan presiden yang diadakan setiap 5 tahun sekali adalah bagian dari proses demokrasi khususnya di Indonesia[3][4].

Dengan kemajuan teknologi informasi dan komunikasi saat ini, penyebaran berita online di kalangan masyarakat tersebar begitu cepat dan mudah[5][6]. Media sosial adalah alat komunikasi yang hampir digunakan seluruh kalangan Masyarakat Indonesia sebagai cara untuk menggunakan dan berbagi menginformasikan kepada publik[7]-[8]. Dunia maya adalah dunia yang bebas dan sulit dikendalikan di mana siapa pun dapat dengan bebas berbicara atau mengekspresikan pendapatnya tentang kandidatnya. Pendapat yang diungkapkan masyarakat bisa positif, netral atau bahkan negatif[9]. Berita mengenai Pemilihan Presiden Tahun 2024 menjadi isu terhangat di media sosial dalam 2 tahun belakangan. Melalui media sosial, partai politik dapat saling berkomunikasi dan membentuk citra sehingga terciptanya kedekatan dengan pemilih[10]. Twitter merupakan salah satu media sosial terbesar kelima di Indonesia diantara pengguna media sosial[11]. Publik dan maupun para pendukung dari ketiga Pasangan Calon dan Wakil Presiden menjadikan Twitter sebagai media yang banyak membahas mengenai Pemilihan Presiden 2024 serta menjadi hot topic pada dua tahun belakangan ini.



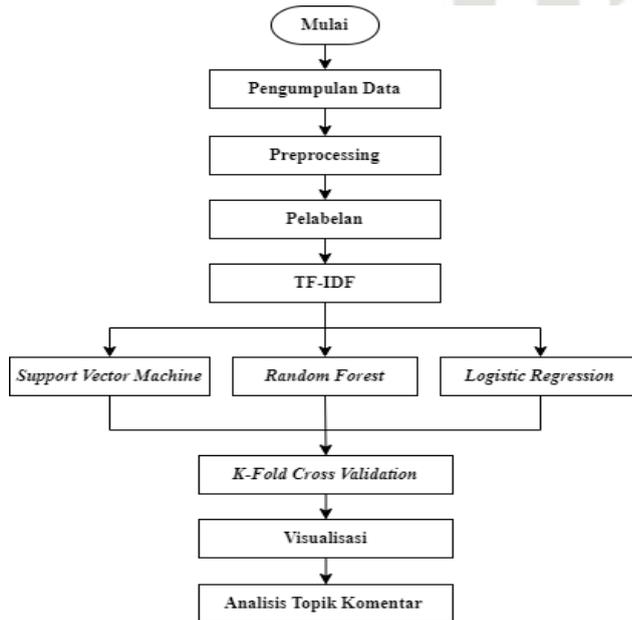
Berdasarkan UU Nomor 7 Tahun 2017 tentang Pemilihan Umum, Komisi Pemilihan Umum (KPU) menyelenggarakan debat calon presiden dan wakil presiden pemilu 2024 sebanyak 5 kali selama masa kampanye. Calon presiden berdebat sebanyak 3 kali dan calon wakil presiden 2 kali[12]. Lantas, banyak netizen dan pendukung menggunakan Twitter untuk mengaungkan dukungan, memberikan kritik ataupun berbagai bentuk komentar positif, negatif, dan netral lainnya. Bentuk-bentuk dari komentar tersebut membahas berbagai bidang seperti keadaan ekonomi, kesehatan, kinerja saat ini dan lain sebagainya. Sehingga perlu dilakukan penelitian mengenai analisis sentimen publik kepada ketiga calon Presiden dan Wakil Presiden 2024 berdasarkan pada komentar atau tweet publik pada twitter sebelum dan sesudah debat dilaksanakan, untuk mengetahui dukungan maupun ketidaksetujuan dari ketiga calon tersebut.

Pada penelitian sebelumnya, dilakukan analisis sentimen terhadap calon presiden 2019, dimana Wiliam dan [13] menggunakan algoritma SVM untuk mengklasifikasikan opini masyarakat di Twitter. Hasil dari klasifikasi ini didapat akurasi sebesar 86% [13]. Penelitian lainnya membahas mengenai partai politik peserta pemilu 2019 dengan menggunakan metode Lexicon inSet. Dimana dari hasil penelitian ini data berhasil memiliki prediksi yang banyak 1.493 atau 86% telah berhasil diproses oleh sistem dan memiliki akurasi sebesar 61% [14]. Menurut [15], algoritma SVM sebesar 88%, dan algoritma logistic sebesar 86%. Algoritma logistic regression memiliki akurasi yang lebih tinggi dari kedua algoritma lainnya. Dari penelitian ini ditemukan bahwa persentase sentimen negatif lebih tinggi dari pada sentimen positif terhadap Bank BSI [15]. Penelitian oleh Ramadanu pada tahun 2023, melakukan analisis sentimen terhadap salah satu calon presiden 2024. Penelitian dilakukan dengan menerapkan algoritma Support Vector Machine dengan kernel RBF dan mendapatkan akurasi terbaik sebesar 90,98% menggunakan kombinasi nilai $C=9$ dan $\gamma=2$ [7].

Penelitian ini menerapkan algoritma Supervised Learning dalam melakukan analisis sentimen publik terhadap debat calon Presiden dan wakil presiden 2024. Dengan menerapkan beberapa algoritma seperti Support Vector Machine, Random Forest dan Logistic Regression. Beberapa penelitian sebelumnya memaparkan ketiga algoritma ini menghasilkan akurasi yang cukup baik [16]. Ketiga algoritma ini diketahui sangat baik dan diusulkan untuk diimplementasikan pada analisis sentimen, dikarenakan ketiga algoritma ini mampu menangani data dengan jumlah fitur yang besar seperti teks [17][18][19]. Hasil dari analisis sentimen opini pada debat calon presiden dan wakil presiden Indonesia tahun 2024, dapat dijadikan sebagai perkiraan tingkat dukungan dan ketidaksetujuan masyarakat terhadap ketiga calon tersebut [20]. Komisi Pemilihan Umum dapat menjadikan hasil analisis tersebut sebagai bahan pertimbangan strategi kampanye dan tindakan lain yang diperlukan untuk memenangkan pemilu presiden dimasa yang akan datang [21]. Selain itu, analisis sentimen ini dapat memberikan informasi berharga kepada masyarakat umum, media, dan pengamat politik untuk memahami dinamika politik yang akan terjadi pada pemilihan Presiden Indonesia pada tahun 2024 dan pemilihan selanjutnya.

2. METODOLOGI PENELITIAN

Pada penelitian ini, ada beberapa tahapan yang dilakukan pada penelitian ini, yaitu pengumpulan data, preprocessing, pelabelan, penerapan algoritma untuk melakukan klasifikasi sentimen dan analisis topik pada tweet. Adapun tahapan dalam penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metodologi Penelitian

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

Berdasarkan Gambar 1 diatas dapat dilihat beberapa tahapan yang dilakukan. Dimulai dengan mengumpulkan data komentar/tweet pada twitter dengan teknik crawling data menggunakan Twitter API. Data yang diambil terkait opini publik terhadap ketiga pasangan calon presiden dan wakil presiden 2024 sebelum dan sesudah dilaksanakannya debat. Setelah itu, memasuki tahap preprocessing data dengan melakukan pembersihan dan seleksi data. Selanjutnya data dilakukan labeling sentimen positif, negatif dan netral secara manual oleh pakar bahasa Indonesia. Setelah proses labeling data, dilakukan pembobotan kata menggunakan TF-IDF. Setelah tahap persiapan data selesai, barulah data dapat dilakukan klasifikasi menggunakan tiga algoritma machine learning yaitu Support Vector Machine, Random Forest dan Logistic Regression. Klasifikasi dilakukan menggunakan bahasa pemrograman python. Setelah klasifikasi sentimen, dapat dihasil visualisasi dan kategori topik pada komentar.

2.1 Text Mining

Text mining merupakan teknik pengambilan informasi dari teks dengan mengidentifikasi pola dan tren disebut sebagai penambangan teks. Proses ini melibatkan pemanfaatan teknik seperti penandaan, anotasi, dan asosiasi untuk mendapatkan data dari sumber teks, dengan maksud untuk analisis atau prediksi[22].

2.2 Analisis Sentimen

Analisis sentimen adalah proses mengidentifikasi polaritas teks dalam suatu dokumen atau jenis dokumen lain untuk memberi label kategori sebagai positif, negatif, atau netral. Analisis sentimen mengumpulkandata untuk mengatur dan mengorganisasikan teks ke dalam entitas tertentu seperti layanan, produk, orang, fenomena, dan objek lainnya[23].

2.3 Term Frequency-Inverse Document Frequency (TF-IDF)

Pembobotan kata TF-IDF adalah merupakan salah satu cara pemberian nilai bobot terhadap data yang telah dilakukan pengekstrakan atau dapat disebut pemberian nilai angka pada data yang tidak berbentuk angka[23]. Dalam proses ini, nilai hasil TF-IDF menentukan kepentingan atau makna suatu fitur atau kata dalam korpus[7]. Pada algoritma TF-IDF digunakanlah rumus untuk menghitung bobot (W) masing-masing dokumen pada kata kunci menggunakan rumus berikut[23]:

$$W_{dt} = tf_{dt} * Id_{ft} \quad (1)$$

2.4 Support Vector Machine

Support Vector Machine atau yang biasa disebut dengan SVM merupakan salah satu algoritma Machine Learning yang tergolong Supervised Learning[24]. SVM menggunakan kernel pada ruang pencarian berdimensi tinggi. Tujuan SVM adalah mencari hyperplane optimal yang dapat memisahkan dua kelas dengan sebaik mungkin[25]. SVM mengelompokkan instance pelatihan ke dalam salah satu dari dua kelas dengan menetapkan perbedaan (hyperplanes) antar kelas sehingga margin antara batas dan satu kelas dimaksimalkan. Kelas dari sebuah instance baru ditentukan oleh bagian dari hyperplane tersebut[15].

$$w \cdot x + b = 0 \quad (2)$$

2.5 Random Forest

Random forest adalah sekumpulan metode pembelajaran yang menggunakan pohon keputusan sebagai pengklasifikasi dasar yang dibangun dan digabungkan. Beberapa aspek penting dari metode random forest antara lain melakukan pengambilan sampel terkontrol untuk membangun pohon prediksi[26]. Berikut persamaan algoritma random forest.

$$F_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} n_{ij}}{\sum_{k \in \text{all nodes}} n_{ik}} \quad (3)$$

2.5 Logistic Regression

Logistic regression digunakan untuk menemukan hubungan antara input dengan probabilitas hasil output[27]. Logistic Regression mampu untuk mendapatkan model terbaik dan sederhana untuk menjelaskan hubungan antara keluaran dari variabel respons (y) dengan variabel-variabel prediktornya (x) pada analisis data kategori[19].

$$P(Y = 1 | X) = \frac{1}{1 + e^{-x}} \quad (4)$$

2.6 K-Fold Cross Validation

K-Fold Cross Validation adalah metode yang digunakan untuk mengurutkan data menjadi data latih dan data uji[23]. Cross Validation adalah salah satu opsi pengambilan sampel ulang data paling umum digunakan prediksi kesalahan model dan parameterisasi dari modelnya[28]. Selain itu, K-Fold Cross Validation menerapkan partisi data dengan membagi jumlah yang sama menjadi k kumpulan data. Proses ini dilakukan sebanyak k kali selama pelatihan dan pengujian berulang[23].

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



3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data yang diambil dari platform twitter terkait ketiga pasangan calon presiden dan wakil presiden 2024. Data diambil dengan kata kunci nama dan nomor urut dari ketiga pasangan. Data yang dikumpulkan diambil pada rentang waktu kelima debat dilaksanakan yaitu pada bulan Desember 2023 hingga Januari 2024.

3.1 Data Awal

Data yang digunakan pada penelitian ini sebanyak 9000 data dengan 600 data mengenai masing-masing calon presiden dan wakil presiden di ke lima debat. Dan 600 data dibagi menjadi 300 data sebelum dan 300 data setelah debat atau 30 dataset.

Tabel 1. Data Awal

No	Tweet
1	@mazzini_gsp @prabowo Mereka semua saling nyerang pak, cuman counter attack pak anies yg paling kuat
2	@NgalimanMS @aniesbaswedan @cakimiNOW @DPP_PKB @EmHasanuddin @Jazilul_Fawaid @zainul_munas @HasanRofiqi4 @Mursyidan10 Anies memang nomor 1 dalam debat capres pertama
3	@jr205newlivery Nope. Di survey 02 unggul di Jabar. Kayaknya isu Anies ga terlalu ngaruh di akar rumput.
4	@PartaiSocmed Mas Anies @aniesbaswedan tolong admin dibalik layar Socmed ini diberikan data nya mas. Sepertinya orang ini masih mengigau seperti Prabowo.
...	...
300	@sniperruben45 @maymarmas @tvOneNews Bilang aja lu benci, ya kan. Nyatanya semua dirangkul oleh Anies, basis Islam, basis Kristen, basis Hindu, basis Budha

Dapat dilihat berdasarkan Tabel 1, bentuk data awal yang diambil pada twitter masing memiliki simbol dan angka serta penulisan kalimat masih memiliki keberagaman bentuk penulisan (huruf besar dan huruf kecil).

3.2 Preprocessing

Beberapa tahapan preprocessing yang dilakukan adalah pembersihan data berupa penghapusan angka, simbol, emoticon dan url. Lalu dilakukan seleksi data terkait topik penelitian. Selanjutnya tokenize untuk pemecahan kalimat menjadi kata dan stopword untuk penghapusan kata yang tidak memiliki nilai positif dan negatif. Dan stemming untuk mengubah kata berimbuhan menjadi kata dasarnya. Hasil preprocessing data dapat dilihat pada Tabel 2.

Tabel 2. Hasil Preprocessing

No	Tweet
1	nyerang cuman counter attack anies yg kuat
2	anies nomor debat capres
3	nope survey unggul jabar kayak isu anies ga ngaruh akar rumput
4	mas anies tolong admin balik layar socmed data nya mas orang igau prabowo.
...	...
300	bilang aja lu benci ya rangkul anies basis islam basis kristen basis hindu basis budha

Dapat dilihat berdasarkan tabel 2 diatas, bentuk data setelah melewati tahap preprocessing menjadi lebih ringkas dan penulisan telah seragam atau lowercase. Data hanya terdiri dari kata-kata yang memiliki makna positif, negatif dan netral.

3.3 Pelabelan

Data tweet yang diambil terkait ketiga pasangan calon presiden dan wakil presiden 2024 adalah data tanpa sentimen. Pelabelan dilakukan secara manual dibantu oleh pakar bahasa. Dalam proses pelabelan tweet, maka dapat dihasilkan sentimen publik terkait topik penelitian ini yang bernilai positif, negatif dan netral. Berikut data hasil pelabelan pada Tabel 3 yang terdiri oleh kolom Tweet/komentar dan Sentimen.

Tabel 3. Hasil Pelabelan

No	Tweet	Sentimen
1	nyerang cuman counter attack anies yg kuat	positif
2	anies nomor debat capres	positif
3	nope survey unggul jabar kayak isu anies ga ngaruh akar rumput	positif
4	mas anies tolong admin balik layar socmed data nya mas orang igau prabowo.	positif
300	bilang aja lu benci ya rangkul anies basis islam basis kristen basis hindu basis budha	positif

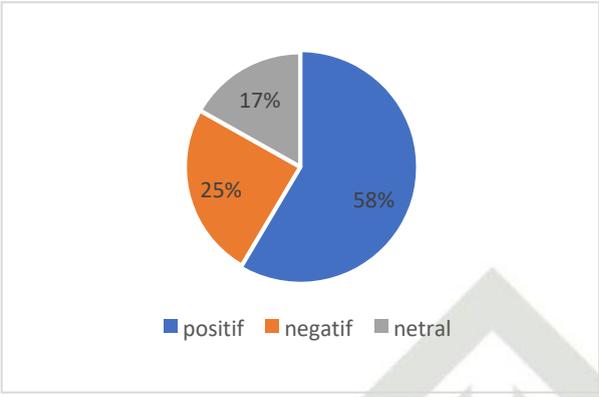
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar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



Hak Cipta Diindungi Undang-Undang

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

Pelabelan dilakukan sebanyak lima kali setara dengan banyaknya debat yang diselenggarakan. Dari hasil pelabelan manual 9000 data oleh pakar seperti salah satu data diatas, didapatkan pada tweet twitter memiliki 5265 tweet sentimen positif, 2224 tweet sentimen negatif, dan 1511 tweet sentimen netral.

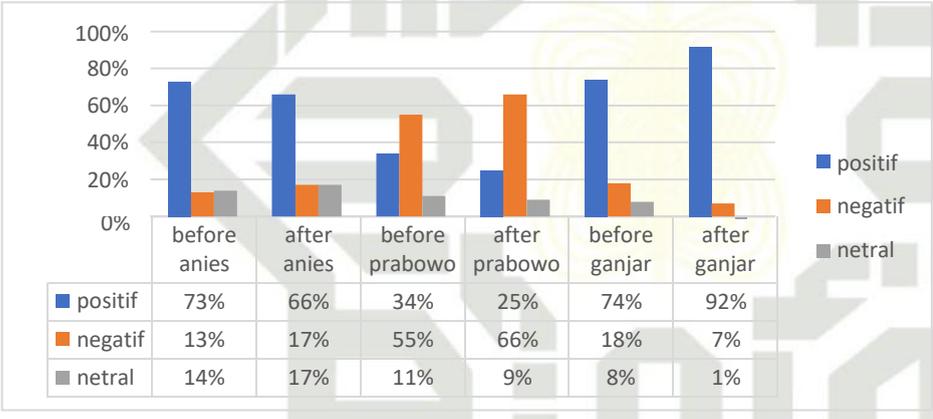


Gambar 2. Hasil Seluruh Pelabelan

Berdasarkan Gambar 2, dapat dilihat sentimen tertinggi ada pada label positif sebesar 58%, sentimen label negatif sebesar 25%, dan sentimen terendah pada label netral sebesar 17%.

3.3.1 Pelabelan Debat 1

Berikut dapat dilihat pada Gambar 3 hasil pelabelan data debat pertama Pilihan Presiden oleh Calon Presiden 2024 pada dataset sebelum dan sesudah debat calon presiden 2024.

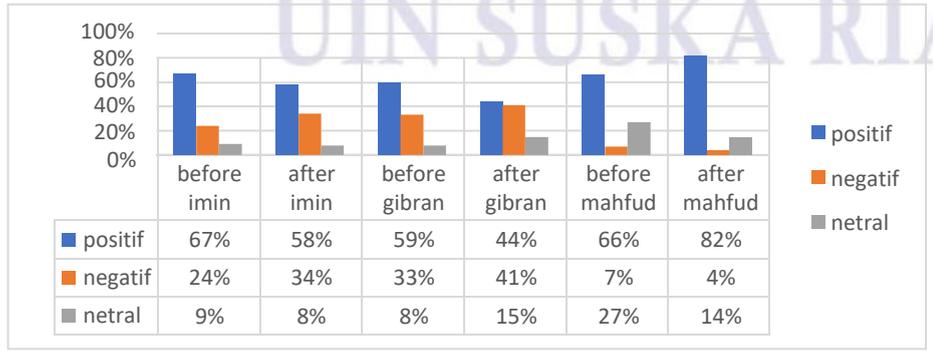


Gambar 3. Hasil Pelabelan Debat 1

Dari hasil pelabelan pada gambar diatas, dapat dilihat kelas positif tertinggi terdapat pada data setelah debat calon presiden Ganjar dengan nilai 92%. Terjadi peningkatan sentimen positif dari data sebelum debat Ganjar. Kelas negatif tertinggi terdapat pada calon presiden Prabowo 66%. Terjadi peningkatan sentimen negatif setelah debat Prabowo. Dan kelas netral tertinggi terdapat pada data sesudah debat Anies dengan nilai sebesar 17%.

3.3.2 Pelabelan Debat 2

Berikut dapat dilihat pada Gambar 4 hasil pelabelan data debat kedua Pilihan Presiden oleh Calon Wakil Presiden 2024 pada dataset sebelum dan sesudah debat calon presiden 2024.



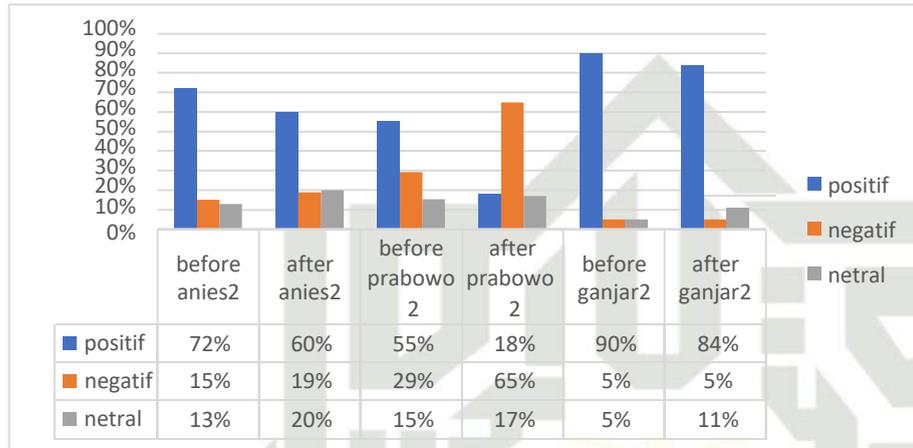
Gambar 4. Hasil Pelabelan Debat 2



Dari hasil pelabelan pada gambar 4 diatas, dapat dilihat kelas positif tertinggi terdapat pada data setelah debat calon wakil presiden Mahfud dengan nilai 82%. Terjadi peningkatan sentimen positif dari data sebelum debat mahfud. Kelas negatif tertinggi terdapat pada calon wakil presiden gibran 41%. Terjadi peningkatan sentimen negatif setelah debat Gibran. Kelas netral tertinggi terdapat pada data sebelum debat mahfud. Sentimen positif pada data setelah debat cak imin menurun dari data sentimen sebelum debat dan terjadi peningkatan pada sentimen negatif menjadi 34%.

3.3 Pelabelan Debat 3

Berikut dapat dilihat pada Gambar 5 hasil pelabelan data debat ketiga Pilihan Presiden oleh Calon Presiden 2024 pada dataset sebelum dan sesudah debat calon presiden 2024.

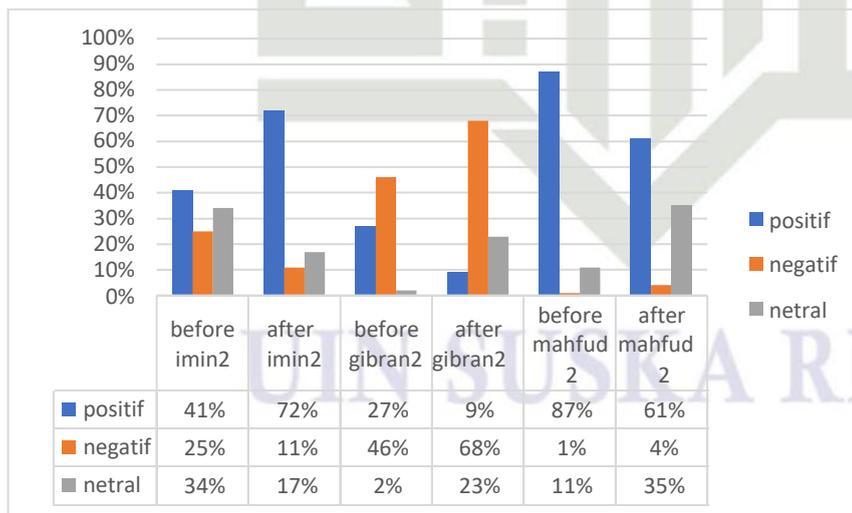


Gambar 5. Hasil Pelabelan Debat 3

Dari hasil pelabelan pada gambar 5 diatas, dapat dilihat kelas positif tertinggi terdapat pada data setelah debat calon presiden Ganjar dengan nilai 90% dan mengalami penurunan sentimen positif setelah debat Ganjar. Kelas negatif tertinggi terdapat pada calon presiden Prabowo sebesar 65%. Terjadi peningkatan sentimen negatif setelah debat Prabowo. Dan kelas netral tertinggi terdapat pada data sesudah debat Anies dengan nilai sebesar 19%. Dan sentimen positif pada data setelah debat Anies dan Prabowo kembali menurun seperti pada debat pertama.

3.3.4 Pelabelan Debat 4

Berikut dapat dilihat pada Gambar 6 hasil pelabelan data debat keempat Pilihan Presiden oleh Calon Wakil Presiden 2024 pada dataset sebelum dan sesudah debat calon presiden 2024.



Gambar 6. Hasil Pelabelan Debat 4

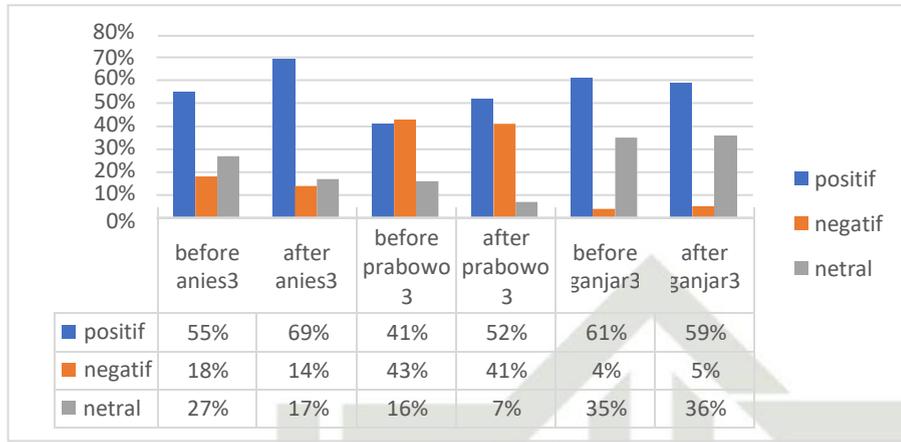
Dari hasil pelabelan pada gambar 6 diatas, dapat dilihat kelas positif tertinggi terdapat pada data sebelum debat calon wakil presiden Mahfud dengan nilai 87%. Terjadi penurunan sentimen positif dari data setelah debat mahfud. Kelas negatif tertinggi terdapat pada calon wakil presiden gibran sebesar 68%. Terjadi peningkatan sentimen negatif setelah debat Gibran. Kelas netral tertinggi terdapat pada data setelah debat mahfud. Sentimen positif pada data setelah debat cak imin meningkat menjadi sebesar 72%.

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



3.5 Pelabelan Debat 5

Berikut dapat dilihat pada Gambar 7 hasil pelabelan data debat kelima Pilihan Presiden oleh Calon Presiden 2024 pada dataset sebelum dan sesudah debat calon presiden 2024.



Gambar 7. Hasil Pelabelan Debat 5

Dari hasil pelabelan pada gambar 7, dapat dilihat kelas positif tertinggi terdapat pada data setelah debat calon presiden Anies dengan nilai 69% dimana mengalami peningkatan sentimen positif dari data sebelum debat Anies. Kelas negatif tertinggi terdapat pada calon presiden Prabowo sebesar 43%. Terjadi peningkatan sentimen positif dan penurunan sentimen negatif setelah debat Prabowo. Dan kelas netral tertinggi terdapat pada data setelah debat Ganjar dengan nilai sebesar 59% dan terjadi penurunan sentimen positif setelah debat.

3.4 Analisa dan Hasil Klasifikasi Ketiga Algoritma Machine Learning

Pada penelitian ini, klasifikasi data dilakukan dengan menggunakan algoritma machine learning yaitu Support Vector Machine, Random Forest dan Logistic Regression. Pada penelitian ini dilakukan analisis pada ke 5 debat Pilihan Presiden 2024. Dimana pada masing-masing debat menggunakan 6 dataset tweet setelah (before) dan sesudah (after) debat dari masing-masing pasangan calon Presiden dan Wakil Presiden 2024. Pengujian setiap 30 dataset dilakukan dengan pembagian data uji dan latih menggunakan K-Fold Cross Validation dengan nilai K=10. Akurasi yang didapat merupakan nilai k tertinggi pada rentang K1 sampai dengan K10 di setiap dataset. Berikut dapat dilihat pada Tabel 5, hasil akurasi dari ketiga algoritma pada 30 dataset yang digunakan.

Tabel 5. Akurasi Ketiga Algoritma

Debat	Dataset	Algoritma		
		SVM	RN	LR
Debat 1	before anies	75%	79%	79%
	after anies	79%	75%	71%
	before prabowo	71%	75%	75%
	after prabowo	79%	75%	79%
	before ganjar	87%	79%	87%
	after ganjar	92%	83%	92%
Debat 2	before imin	76%	83%	75%
	after imin	76%	79%	75%
	before gibran	76%	75%	79%
	after gibran	67%	62%	67%
	before mahfud	76%	79%	79%
	after mahfud	85%	92%	87%
Debat 3	before anies 2	83%	75%	79%
	after anies 2	71%	67%	71%
	before prabowo 2	71%	79%	67%
	after prabowo 2	75%	75%	75%
	before ganjar 2	92%	92%	92%
	after ganjar 2	87%	87%	87%
Debat 4	before imin 2	71%	62%	67%
	after imin 2	79%	75%	79%
	before gibran 2	71%	67%	71%
	after gibran 2	71%	71%	71%
	before mahfud 2	75%	87%	92%

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



Debat	Dataset	Algoritma		
		SVM	RN	LR
	after mahfud 2	92%	71%	83%
	Beforeanies 3	67%	67%	62%
	afteranies 3	75%	75%	75%
Debat 5	before prabowo 3	62%	71%	67%
	after prabowo 3	92%	96%	92%
	before ganjar 3	92%	92%	87%
	after ganjar 3	87%	87%	96%
Akurasi rata-rata ketiga algoritma		78%	78%	79%

Dari 30 dataset, masing-masing dataset dilakukan pengujian pada ketiga algoritma. Dari hasil uji klasifikasi ini didapat polaritas dari masing-masing algoritma yaitu pada model Support Vector Machine, polaritas terbanyak ada pada kelas positif dengan jumlah prediksi kelas negatif pada 5 dataset. Model Random Forest dengan polaritas terbanyak ada pada kelas positif dengan jumlah prediksi kelas negatif pada 7 dataset. Dan pada model Logistic Regression didapat polaritas terbanyak pada kelas positif dengan jumlah prediksi kelas negatif pada 8 dataset. Dapat disimpulkan polaritas hasil prediksi tertinggi terdapat pada kelas positif. Pada Tabel 5 akurasi diatas, dapat dilihat bahwa dari ketiga algoritma telah mengklasifikasikan sentimen dari seluruh dataset dengan baik. Akurasi tertinggi didapat pada algoritma Logistic Regression dengan akurasi rata-rata sebesar 79%. Dan dua algoritma lainnya memiliki akurasi yang sama dengan rata-rata akurasi sebesar 78%. Perbedaan akurasi dari ketiga algoritma cukup kecil sehingga dapat disimpulkan ketiga algoritma bekerja dengan cukup baik pada dataset yang digunakan.

3.5 Visualisasi Data

Dalam visualisasi data tweet perlu dilakukan pemecahan data menjadi empat kategori: keseluruhan, positif, negatif, dan netral. Visualisasi dilakukan dengan pemrograman python dengan menginput library wordCloud. Hasil visualisasi data dapat dilihat pada Gambar 8.



Gambar 8. Visualisasi Data Keseluruhan

Berdasarkan visualisasi pada gambar 8 dapat dilihat, kata yang sering muncul pada tweet adalah nama dari ketiga kandidat pasangan calon presiden dan wakil presiden. Kata anies, prabowo, cak imin merupakan tiga kata teratas yang paling banyak muncul pada tweet.

3.6 Analisa Kategori Topik pada Data

Untuk mengetahui topik pembahasan pada masing-masing data sentimen twitter. Dari pengelompokkan kategori, terdapat 11 topik yang dibahas publik melalui tweet mengenai calon presiden dan wakil presiden 2024. Kategori topik dapat dilihat pada Tabel 6.

Tabel 6. Kategori Topik

No.	Topik
1.	Politik
2.	Sosial
3.	Pendidikan
4.	Pembangunan
5.	Kesehatan
6.	Keagamaan
7.	Teknologi dan Informasi
8.	Seni dan Budaya
9.	Lingkungan
10.	Pertahanan
11.	Ekonomi

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

Dari hasil analisis kategori yang dilakukan, tweet terkait ketiga pasangan calon presiden dan wakil presiden 2024 sebelum dan sesudah debat, lebih dari 90% didominasi oleh tweet kategori politik. Dan 10% dari data tweet membahas 10 topik lainnya. Berikut kategori topik pada tweet dapat dilihat Tabel 7, yang terdiri dari kolom topik pada data tweet, pembahasan mengenai topik, jumlah tweet yang membahas topik, serta persentase dari masing-masing kategori topik pada data tweet.

Tabel 7. Kategori Topik pada Data Tweet

No.	Topik	Pembahasan	Jumlah Tweet	Persentase
1.	Politik	dukungan publik kepada paslon, pemilu 2024, debat pilpres 2024, pelanggaran hukum HAM, mahkamah konstitusi, politik dinasti, kebijakan pemerintahan, keadilan, oligarki	8172	90,7%
2.	Sosial	toleransi agama, etnis tionghoa, bantuan sosial, HAM, ramah, sopan, lgbt	302	3,35%
3.	Ekonomi	sgie, food estate, hilirisasi, lapangan pekerjaan, kenaikan pajak, kesejahteraan rakyat, nelayan dan petani, pertambangan nikel, kredit bank	269	2,98%
4.	Pembangunan	IKN, transportasi, rusun, fasilitas kesehatan, gereja	73	0,8%
5.	Pendidikan	beasiswa pendidikan sarjana, SDM unggul, guru	61	0,7%
6.	Kesehatan	olahraga, stunting, susu, rumah sakit, dokter	43	0,5%
7.	Keagamaan	agama, tadarus, sholat, natal	27	0,3%
8.	Teknologi dan Informasi	Internet, media sosial	23	0,25%
9.	Seni dan Budaya	pelestarian budaya dan adat, karya anak bangsa	17	0,2%
10.	Lingkungan	daur ulang, inflasi hijau	10	0,1%
11.	Pertahanan	militer dan alusista	3	0,03%

Berdasarkan Tabel 7, dapat kita ketahui melalui data tweet sebelum dan sesudah debat, terdapat 3 topik teratas yang paling sering dibahas publik. Pada urutan pertama topik politik paling mendominasi diantara topik lainnya dengan jumlah 8172 tweet atau 90,7%. Di urutan kedua topik sosial dengan jumlah 302 tweet atau sebesar 3,35%. Dan di urutan ketiga, topik ekonomi dengan jumlah 269 tweet atau sebesar 2,98%.

4. KESIMPULAN

Dari hasil pengolahan dan analisis data yang telah dilakukan dengan bahasa pemrograman Python, dapat disimpulkan bahwa ke 30 dataset dengan jumlah keseluruhan data sebanyak 9000 data, dihasilkan rata-rata akurasi tertinggi pada algoritma Logistic Regression dengan akurasi sebesar 79%. Kedua algoritma lainnya juga memiliki akurasi yang cukup tinggi dengan rata-rata akurasi sebesar 78%. Akurasi didapat dari K-Fold Cross Validation dengan K=10 pada setiap data. Dari hasil klasifikasi dengan ketiga algoritma, didapatkan polaritas tertinggi terdapat pada kelas positif. Dari hasil analisis diatas dapat disimpulkan ketiga algoritma telah bekerja dengan cukup baik. Dan berdasarkan hasil analisis kategori topik pembahasan pada tweet terkait sebelum dan sesudah debat, terdapat tiga topik teratas yang cukup memiliki nilai signifikan tertinggi dari kategori topik lainnya. Topik tertinggi membahas seputar isu-isu politik pada pemilu sebesar 90%. Kedua topik sosial sebesar 3,35% dan ketiga topik ekonomi 2,98%.

UCAPAN TERIMAKASIH

Terima kasih saya ucapkan kepada teman-teman, kakak-kakak tingkat, pihak KPU Provinsi Riau, serta Ibu Bapak Dosen yang telah membantu dan mengarahkan saya dalam penyelesaian karya tulis ini.

REFERENCES

- [1] A. S. Putra and Ubaidullah, "Analisis Kegagalan Marketing Politik Partai NasDem Pada Pemilu Legislatif Tahun 2019 Di Aceh," *Jurnal Ilmiah Mahasiswa FISIP Unsyiah*, vol. 6, no. 2, May 2021.
- [2] W. Budiharto and M. Meiliana, "Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis," *J Big Data*, vol. 5, no. 1, Dec. 2018, doi: 10.1186/s40537-018-0164-1.
- [3] A. Putri Nardilarsari, A. Lia Hananto, S. Shofia Hilabi, and B. Priyatna, "Terakreditasi SINTA Peringkat 3 Analisis Sentimen Calon Presiden 2024 Menggunakan Algoritma SVM Pada Media Sosial Twitter," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 7, no. 1, pp. 11–18, Mar. 2023.
- [4] Fajri Rahman Jurdi, *Pengantar Hukum Pemilihan Umum*, Edisi Pertama. Jakarta: Kencana, 2018.



Hak Sipta Bimbingan dan Pengajaran

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

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

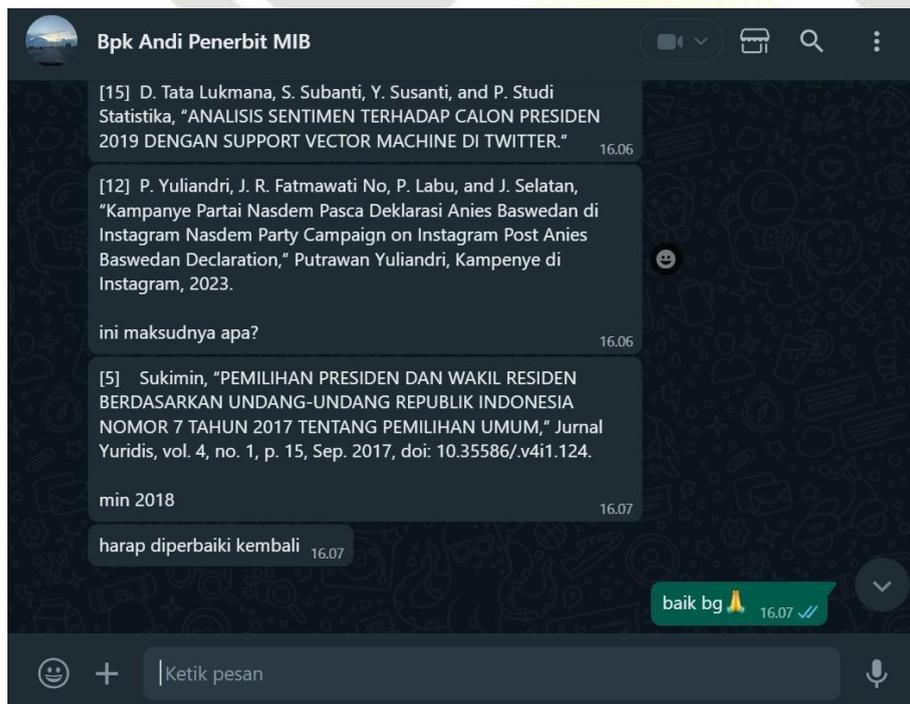
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Fathin, "Analisis Sentimen Artikel Berita Pemilu Berbasis Metode Klasifikasi," Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, 2023.
- D. Wiryany, S. Natasha, and R. Kurniawan, "PERKEMBANGAN TEKNOLOGI INFORMASI DAN KOMUNIKASI TERHADAP PERUBAHAN SISTEM KOMUNIKASI INDONESIA," *Jurnal Nomosleca*, Oktober, vol. 8, no. 2, pp. 242–252, Oct. 2022.
- R. Rutra, "Penerapan Metode SVM Pada Klasifikasi Sentimen Terhadap Anies Baswedan Sebagai Bakal Calon Presiden 2024," *Jurnal Informatika Universitas Pamulang*, vol. 8, no. 2, pp. 145–152, Jun. 2023.
- F. Agustinah and Widayati, "PEMANFAATAN MEDIA SOSIAL SEBAGAI SARANA PROMOSI MAKANAN RINGAN KRIPIK SINGKONG DI KABUPATEN SAMPANG," *Jurnal Dialektika*, vol. 4, no. 2, pp. 1–20, Sep. 2019.
- G. A. BUNTORO, R. ARIFIN, G. N. SYAIFUDDIIN, A. SELAMAT, O. KREJCAR, and H. FUJITA, "Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia's 2019 Presidential Election," *IJUM Engineering Journal*, vol. 22, no. 1, pp. 78–92, 2021, doi: 10.31436/IJUM.EJ.V22I1.1532.
- P. Yulhandri, J. R. Fatmawati No, P. Labu, and J. Selatan, "Kampanye Partai Nasdem Pasca Deklarasi Anies Baswedan di Instagram Nasdem Party Campaign on Instagram Post Anies Baswedan Declaration," *PROMEDIA, (PUBLIC RELATION DAN MEDIA KOMUNIKASI)*, vol. 9, no. 1, pp. 30–55, Jun. 2023.
- H. Tsaniya, R. Rosadi, and A. S. Abdullah, "Sentiment analysis towards Jokowi's government using twitter data with convolutional neural network method," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1722/1/012017.
- [12] A. Zruddin, K. Rahman, and M. A. Maulidi, *Merawat Negara Hukum*. Guepedia, 2023.
- [13] D. Tata Lukmana, S. Subanti, Y. Susanti, and P. Studi Statistika, "ANALISIS SENTIMEN TERHADAP CALON PRESIDEN 2019 DENGAN SUPPORT VECTOR MACHINE DI TWITTER," in *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, 2019, pp. 154–160.
- [14] I. F. Fanhar Nur, A. Herdiani, and W. Astuti, "Analisis Sentimen Berbasis Leksikon InSet Terhadap Partai Politik Peserta Pemilu 2019 Pada Media Sosial Twitter," in *e-Proceeding of Engineering*, Dec. 2019, pp. 10397–10407.
- [15] R. A. Husen, R. Astuti, L. Marlia, R. Rahmadden, and L. Efrizoni, "Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 211–218, Oct. 2023, doi: 10.57152/malcom.v3i2.901.
- [16] C. Chazar and B. E. Widhiaputra, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 12, no. 1, pp. 67–80, May 2020.
- [17] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *JURNAL INFORMATIKA UPGRIS*, vol. 7, no. 1, pp. 36–40, 2021.
- [18] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, "Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023," *Jurnal Teknik Komputer AMIK BSI*, vol. 6, no. 1, pp. 121–129, Jan. 2020, doi: 10.31294/jtk.v4i2.
- [19] A. Maulana, Inayah Khasnaputri Afifah, Asghafi Mubarrak, Kiagus Rachmat Fauzan, Ardhan Dwintara, and B. P. Zen, "COMPARISON OF LOGISTIC REGRESSION, MULTINOMIALNB, SVM, AND K-NN METHODS ON SENTIMENT ANALYSIS OF GOJEK APP REVIEWS ON THE GOOGLE PLAY STORE," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1487–1494, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.863.
- [20] M. Wahyudin, "Analisis Sentimen pada Data Twitter untuk Memprediksi Hasil Pemilihan Umum," *JUCA (Jurnal Cyber Area)*, vol. 3, no. 3, pp. 1–18, 2023.
- [21] E. Nur, "STRATEGI KOMUNIKASI TIM SUKSES PADA KAMPANYE POLITIK UNTUK MEMENANGKAN CALON LEGISLATIF KOTA MAKASSAR The Communication Strategy of Success Team in Political Campaign to Win Legislative Candidates in Makassar," *Jurnal Diakom*, vol. 2, no. 1, pp. 120–128, Sep. 2019.
- [22] A. Hossain, M. Karimuzzaman, M. M. Hossain, and A. Rahman, "Text mining and sentiment analysis of newspaper headlines," *Information - MDPI*, vol. 12, no. 10, pp. 1–15, Oct. 2021, doi: 10.3390/info12100414.
- [23] D. Pramana, M. Afdal, and I. Permana, "Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier dan K-Nearest Neighbors," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, pp. 1–10, 2023, doi: 10.30865/mib.v5i1.2293.
- [24] W. Sihalahi and A. Hartanto, "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024," *JRST (Jurnal Riset Sains dan Teknologi)*, vol. 7, no. 2, p. 245, Sep. 2023, doi: 10.30595/jrst.v7i2.18133.
- [25] F. F. Abdulloh and I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19," *CSRID (Computer Science Research and Its Development Journal)*, vol. 13, no. 3, p. 141, Nov. 2021, doi: 10.22303/csrid.13.3.2021.141-148.
- [26] I. Adhal et al., "Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [27] A. R. Hidayati, A. S. Fitriani, M. A. Rosid, F. Sains, and D. Teknologi, "Analisa Sentimen Pemilu 2019 Pada Judul Berita Online Menggunakan Metode Logistic Regression," *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 4, no. 2, pp. 298–305, Apr. 2023.
- [28] A. R. Dewi, S. Diana, M. A. Fakhrezi, N. Awang, H. Ma'arif, and D. D. Saputra, "SENTIMEN ANALISIS TERHADAP PUAN MAHARANI SEBAGAI KANDIDAT CALON PRESIDEN 2024 BERDASARKAN OPINI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN ADABOOST," *JSii (Jurnal Sistem Informasi)*, vol. 10, no. 1, pp. 75–80, Mar. 2023, doi: 10.30656/jsii.v10i1.5785.



Hak Cipta Dilindungi Undang-Undang

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



Hak Cipta Dilindungi Undang-Undang

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

LAMPIRAN B

ACCEPTED SUBMISSION



JURNAL MEDIA INFORMATIKA BUDIDARMA

eISSN 2548-8368 / pISSN 2614-5278

Sekretariat : UNIVERSITAS BUDI DARMA | Jl. Sisingamangaraja No. 338, Medan, Sumatera Utara
Website: <https://ejournal.stmik-budidarma.ac.id/index.php/mib>
Email: mib.stmikbd@gmail.com

Medan, 21 Maret 2024

No : 975/MIB/LOA/III/2024
Lamp :-
Hal : Surat Penerimaan Naskah Publikasi Jurnal

Kepada Yth,
Bapak/Ibu **Zairy Cindy Dwinnie**
Di Tempat

Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada **Jurnal Media Informatika Budidarma** (eISSN 2548-8368 / pISSN 2614-5278), dengan judul:

Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum dan Sesudah Debat Calon Presiden dan Wakil Presiden Tahun 2024

Penulis: **Zairy Cindy Dwinnie(*)**, Rice Novita

Berdasarkan hasil review dari reviewer, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Volume 8, Nomor 2, April 2024**.

Sebagai informasi QR-Code digunakan untuk melihat link LOA Jurnal Media Informatika Budidarma, **Volume 8, Nomor 2, April 2024** yang telah dikeluarkan. Mohon segera untuk mengirimkan Copyright Transfer Form ke Email Jurnal MIB.

Demikian informasi yang kami sampaikan, atas perhatiannya kami ucapkan terimakasih.



Hormat Kami,

Surya Darma Nasution, M.Kom
Ketua Editor Jurnal MIB

Tembusan:
1. Author
2. Files

UIN SUSKA RIAU

LAMPIRAN C

Published

© Hak

Hak Cipta Dilindungi Undang-Undang

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

Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum dan Sesudah Debat Calon Presiden dan Wakil Presiden Tahun 2024

 **Zairy Cindy Dwinnie** (Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia) 

 **Rice Novita** (Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia)

(*) Corresponding Author

Submitted: **February 22, 2024**; Published: **April 23, 2024**

 Like  Share [Sign Up to see what your friends like](#)
DOI: <https://dx.doi.org/10.30865/mb.v8i2.7504>



	?	Total citations
	?	Recent citations
	n/a	Field Citation Ratio
	n/a	Relative Citation Ratio

Abstract

The 2024 Presidential Election has become the hottest topic in the past two years. The KPU has confirmed that there are 3 candidates for President and Vice President. For this reason, as a momentum for voters to assess the 2024 Presidential and Vice Presidential candidates, the KPU is holding the 2024 Presidential Choice Debate which is based on Law Number 7 of 2017 concerning General Elections. Based on the information presented on the kpu.go.id page, the debate will be held 5 times with 3 presidential candidate debates and 2 vice presidential candidate debates. For this reason, it is necessary to carry out an analysis to find out how public sentiment is positive, negative, and neutral on Twitter towards the three candidates for President and Vice President in 2024 before and after the debate was held. The aim is to estimate public support or disapproval of the three candidate pairs. This research uses three algorithms as a comparison of classification accuracy, namely the Support Vector Machine algorithm, Random Forest, and Logistic Regression. Where the data used is tweet data on Twitter related to before and after the debate as many as 30 datasets with a total of 9000 data. From the classification results, the average accuracy obtained for the three algorithms, namely SVM and Random Forest, was 79%, and Logistic Regression was 79%. The highest polarity obtained from the classification of the three algorithms is in the positive class. This indicates that the Logistic Regression algorithm provides better performance in classifying Twitter sentiment regarding the 2024 presidential and vice presidential candidate pairs.

Keywords

U

ADN and Scope
Indexing & Abstracting
Author Guidelines
Publication Ethics
Access Submission
Submission Guidelines
Contact Us
Visitor Statistic
Author Fees
Statement of Originality
Copyright Notice
Copyright Transfer Form



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

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

LAMPIRAN D

SURAT PERNYATAAN PELABELAN

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini:

Nama : Elvina, M.Pd
Tempat Tanggal Lahir : Peranap, 02 Januari 1990
Pekerjaan : Dosen Bahasa Indonesia
Alamat : Perumahan Pranata Cluster Blok C4 Jl. Swakarya Tuah Karya,
Tampan, Pekanbaru

Menyatakan dengan sesungguhnya bahwa benar memvalidasi untuk pelabelan sentimen yang bersifat positif, negatif dan netral dalam komentar *Twitter (X)* sebelum dan pasca debat pilpres 2024 Indonesia sebanyak 9000 data dalam Tugas Akhir dari:

Nama : Zairy Cindy Dwinnie
Nim : 12050324081
Jurusan : Sistem Informasi
Judul Tugas Akhir : Penerapan Algoritma *Machine Learning* Pada Analisis Sentimen *Twitter* Terhadap Debat Calon Presiden Dan Wakil Presiden Tahun 2024.

Demikian surat ini saya buat dengan sesungguhnya tanpa paksaan dari pihak manapun. Atas perhatiannya saya ucapkan terima kasih.

Pekanbaru, 08 Februari 2024

Yang membuat pernyataan



Elvina, M.Pd



DAFTAR RIWAYAT HIDUP



© UIN Suska Riau
State Islamic University of Sultan Syarif Kasim

Zairy Cindy Dwinnie lahir di Kota Dumai, pada tanggal 08 Mei 2002. Penulis merupakan anak dari Bapak Ardi dan Ibu Betty Yuliane. Penulis adalah anak ketiga dari empat bersaudara. Pada tahun 2007 penulis memulai pendidikan di TK Melati Ujung Batu, Kabupaten Rokan Hulu dan lulus pada tahun 2008. Lalu melanjutkan pendidikan Sekolah Dasar di SD Negeri 001 Ujung Batu, kemudian pindah ke kota Pekanbaru dan melanjutkan kembali pendidikan Sekolah Dasar pada tahun 2011 di SD Negeri 048 Pekanbaru. Penulis menyelesaikan pendidikan Sekolah Dasar pada tahun 2014. Setelah menyelesaikan pendidikan Sekolah Dasar penulis melanjutkan pendidikan tingkat SLTP di SMP Negeri 35 Pekanbaru yang selesai pada tahun 2017. Penulis melanjutkan pendidikan tingkat SLTA di SMA Negeri 14 Pekanbaru. Setelah menyelesaikan pendidikan di SMA Negeri 14 Pekanbaru pada tahun 2020, penulis pun melanjutkan pendidikan dengan menjadi mahasiswa Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau dan penulis menyelesaikan kuliah Strata satu (S1) tersebut pada tahun 2024. Selama perkuliahan penulis aktif dalam mengikuti berbagai kegiatan yang diadakan oleh kampus maupun luar kampus. Selain itu penulis juga pernah menjadi anggota *Puzzle Research Data Technology* (PREDATECH) mulai dari tahun 2020 hingga 2024. Penulis juga pernah menjadi staff dinas advokasi, sosial dan politik Dewan Eksekutif Mahasiswa Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau periode 2022. Penulis juga pernah mengikuti Kuliah Kerja Nyata (KKN) di Desa Teluk Bano 1, Kecamatan Bangko Pusako, Kabupaten Rokan Hilir pada tahun 2023. Pada penelitian Tugas Akhir ini penulis mengambil judul penelitian Tugas Akhir yaitu "Penerapan Machine Learning Pada Analisis Sentimen Twitter Sebelum dan Sesudah Debat Calon Presiden dan Wakil Presiden Tahun 2024".

Hak Cipta Dilindungi Undang-Undang

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