Jumistik
E-ISSN: 2964-3953 P-ISSN: 2964-3888 Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer

Vol. 4 No.1 Juni 2025
Hal 392-400
Https://ojs.amiklps.ac.id

PERBANDINGAN ALGORITMA *NAIVE BAYES* DAN SVM DALAM ANALISIS SENTIMEN *TWITTER* TERHADAP ISU IJAZAH JOKOWI PALSU

Oca Meilika Wulandari[1*], Irvan Maulana[2], Fatih Syamsudin[3], Retno Waluyo[4]
[1,2,4]Sistem Informasi, Universitas Amikom Purwokerto
[3]Informatika, Universitas Amikom Purwokerto
E-mail: ocawulandari89@gmail.com[1*], irvanml474@gmail.com[2], fsyamsudin24@gmail.com[3], waluyo@amikompurwokerto.ac.id[4]

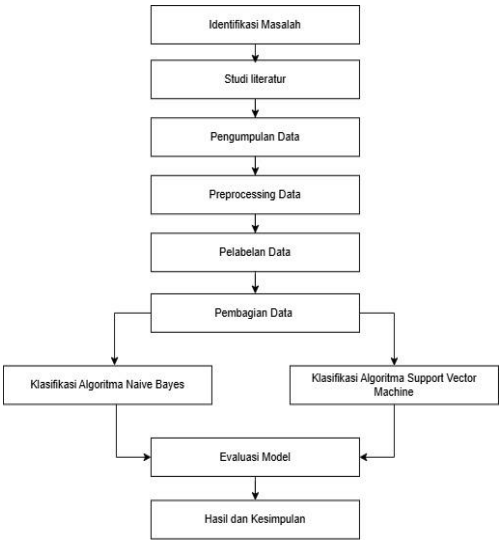
INFO ARTIKEL

Sejarah Artikel
Diterima : 05/06/2025
Direvisi : 10/06/2025
Diterbitkan : 11/06/2025

***Corresponding author**
ocawulandari89@gmail.com

DOI: 10.70247/jumistik.v4i1.145
Https://ojs.amiklps.ac.id

GRAPHICAL ABSTRACT



ABSTRACT

Social media, especially Twitter, has evolved into a vibrant public forum for discussing various social and political issues. One notable topic of debate is the alleged fake diploma of former President Joko Widodo. This study compares how well two classification algorithms Support Vector Machine and Naïve Bayes perform, in analzying user sentiment on Twitter regarding the alleged fake diploma issue involving former President Joko Widodo. The analysis begins with data collection using crawling techniques, followed by preprocessing, data labeling, algorithm implementation, and model evaluation. Out of a total of 3,055 Indonesian-language Twitter comments related to the alleged fake diploma of former President Jokowi, the analysis revealed 1,453 negative comments, 942 positive comments, and 660 neutral comments. The results indicated that the Naïve Bayes Classifier achieved an accuracy of 65%, while the Support Vector Machine reached a higher accuracy of 69.23%.

***Keywords**: Algorithms, Naïve Bayes, Support Vector Machine, Sentiment Analysis, Twitter*

ABSTRAK

Media sosial, khususnya Twitter, telah berkembang menjadi forum publik yang dinamis untuk membahas berbagai isu sosial dan politik. Salah satu topik yang banyak diperdebatkan adalah dugaan ijazah palsu milik mantan Presiden Joko Widodo. Penelitian ini membandingkan kinerja dua algoritma klasifikasi, yaitu *Support Vector Machine* dan *Naïve Bayes*, dalam menganalisis sentimen pengguna Twitter terhadap isu dugaan ijazah palsu tersebut. Analisis dimulai dengan pengumpulan data menggunakan teknik *crawling*, dilanjutkan dengan proses pra-pemrosesan, pelabelan data, implementasi algoritma, dan evaluasi model. Dari total 3.055 komentar Twitter berbahasa Indonesia yang berkaitan dengan dugaan ijazah palsu Presiden Jokowi, diperoleh 1.453 komentar negatif, 942 komentar positif, dan 660 komentar netral. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi sebesar 65%, sementara *Support Vector Machine* menghasilkan akurasi yang lebih tinggi, yaitu 69,23%.

***Kata kunci**: Algoritma, Naïve Bayes, Support Vector Machine, Analisis Sentimen, Twitter*

PENDAHULUAN

Media sosial menandai era media baru, yang mengubah cara masyarakat modern dalam berinteraksi antara satu sama lain. Media sosial menjadi wadah bagi individu untuk merepresentasikan diri, berkomunikasi, bekerja sama, berbagi informasi, dan membentuk hubungan secara virtual dengan pengguna lainnya [1]. Twitter merupakan platform mikroblogging yang bersifat cepat dan *real-time*, yang memberi ruang bagi pengguna untuk mengekspresikan pendapat, menyebarkan informasi, serta menanggapi berbagai isu secara luas dan terbuka [2].

Media sosial seperti Twitter menyajikan lautan data yang sangat besar dan beragam, yang mencerminkan persepsi, emosi, dan tanggapan masyarakat terhadap suatu peristiwa atau tokoh tertentu. Namun, data yang dihasilkan dari platform ini umumnya bersifat tidak terstruktur, singkat, dan penuh unsur khas media sosial seperti emotikon, singkatan, sarkasme, serta bahasa informal, sehingga memerlukan pendekatan analisis yang lebih kompleks. Dalam konteks isu dugaan ijazah palsu yang melibatkan mantan Presiden Joko Widodo, data yang terkumpul dari media sosial tidak hanya mencerminkan sikap individu terhadap figur publik tersebut, tetapi juga merepresentasikan tingkat kepercayaan masyarakat terhadap institusi pendidikan dan pemerintahan secara umum. Oleh karena itu, pemanfaatan analisis sentimen menjadi langkah strategis untuk mengidentifikasi kecenderungan persepsi publik secara lebih objektif dan berbasis data aktual. Selain itu, perbandingan antara algoritma klasifikasi seperti Naïve Bayes dan Support Vector Machine menjadi relevan untuk menentukan pendekatan mana yang paling sesuai dalam mengolah data opini publik di media sosial, khususnya dalam kasus dengan sensitivitas politik yang tinggi. Penelitian ini diharapkan dapat memberikan kontribusi teoritis dan praktis dalam pengembangan metode analisis sentimen di era digital.

Isu mengenai ijazah palsu yang menyeret nama Mantan Presiden Joko Widodo menjadi sorotan besar di tengah masyarakat dan media, khususnya media sosial. Meskipun pihak universitas dan pemerintah telah memberikan klarifikasi resmi untuk membantah tuduhan tersebut, isu ini tetap berkembang luas di ruang digital. Banyak pengguna media sosial yang menyuarakan pendapatnya baik mendukung maupun meragukan keaslian dokumen tersebut yang mencerminkan polarisasi opini publik. Isu Ijazah Jokowi Palsu tidak hanya menyentuh ranah personal seorang tokoh publik, tetapi juga berpotensi menggoyahkan legitimasi institusi pendidikan dan kepercayaan terhadap sistem politik di Indonesia. Hal ini menunjukkan bagaimana isu-isu politik yang menyangkut figur penting dapat dengan cepat meluas dan membentuk narasi tersendiri di tengah masyarakat digital [3].

Isu politik seperti dugaan pemalsuan ijazah tidak hanya menimbulkan kegaduhan di ruang publik digital, tetapi juga dapat mempengaruhi tingkat kepercayaan masyarakat terhadap kredibilitas figur publik yang bersangkutan maupun institusi pemerintahan [3]. Analisis sentimen berperan penting untuk memetakan pola persepsi masyarakat secara objektif berdasarkan data nyata yang bersumber dari media sosial. Namun, karena karakteristik data Twitter yang bersifat tidak terstruktur dan dinamis, dibutuhkan algoritma yang efektif dan efisien dalam proses klasifikasinya. Oleh karena itu, penting untuk mengevaluasi dan memilih algoritma mana yang paling relevan untuk menganalisis sentimen masyarakat dalam konteks isu yang sensitif seperti ini.
Analisis sentimen merupakan bagian komponen *text mining* yang berfokus pada pemrosesan teks untuk mengekstraksi emosi, opini dan persepsi seseorang tentang suatu subjek dengan menggunakan teknik pemrosesan bahasa alami [4]. Analisis sentimen di media sosial mampu menghasilkan informasi berharga bagi pihak yang menjadi objek opini dengan mengklasifikasikan pernyataan pengguna ke dalam kategori positif, negatif, atau netral [5].

Teknik dan algoritma data mining dapat digunakan untuk mengidentifikasi pola, mengklasifikasikan data, dan membuat prediksi yang mendukung analisis sentimen [6]. Diantarannya yaitu algoritma *Naïve Bayes* dan *Support Vector Machine*.
Naive Bayes merupakan algoritma probabilistik yang mengasumsikan bahwa setiap komponen dalam sebuah data adalah independen satu dengan yang lainnya. Meskipun anggapan independensi ini tidak selalu benar dalam dunia nyata, *Naive Bayes* tetap menunjukkan hasil yang sangat baik dalam banyak studi klasifikasi, termasuk analisis sentimen [7]. Algoritma ini cepat, sederhana, dan sering digunakan dalam klasifikasi teks karena kemampuannya untuk mengatasi data besar dengan efisiensi yang tinggi [8].

Support Vector Machine merupakan algoritma klasifikasi yang menggunakan pemrosesan bahasa alami yang efektif, dan pada penelitian ini digunakan untuk mengkategorikan komentar pengguna aplikasi twitter ke dalam kategori komentar positif, negatif dan netral [4]. Meskipun *Support Vector Machine* cenderung memerlukan lebih banyak waktu komputasi dibandingkan f, namun menghasilkan hasil yang lebih baik dalam tugas-tugas klasifikasi yang kompleks [9].

Penelitian sebelumnya yang dilakukan untuk menganalisis sentimen aplikasi ruang guru pada pengguna twitter yang menemukan hasil bahwa dengan perbandingan algoritma yang digunakan, algoritma *Support Vector Machine* mendapatkan hasil akurasi terbaik dengan nilai akurasi sebesar 78,55%. Hasil penelitian lain (sumber) menemukan bahwa algoritma *Support Vector Machine* memiliki tingkat akurasi sebesar 96.2704% untuk klasifikasi penyakit diabetes menggunakan software WEKA [9]. penelitian lain (sumber) menganalisis sentimen mobil listrik di indonesia menghasilkan tingkat akurasi yang lebih

tinggi pada algoritma *Support Vector Machine* dengan tingkat akurasi 70,82% dibandingkan dengan algoritma *Naïve Bayes* yang memiliki tingkat akurasi 63.02% [10].

Tujuan penelitian ini adalah membandingkan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* dalam analisis sentimen data Twitter terkait isu ijazah palsu Mantan Presiden Joko Widodo, untuk mengetahui algoritma mana yang lebih unggul dalam mengolah data media sosial yang tidak terstruktur dan dinamis. Dari penelitian ini, diharapkan akan membantu peneliti lain dalam menentukan teknik analisis sentimen yang lebih efektif dan akurat untuk mengolah data media sosial yang tidak terstruktur.

TINJAUAN PUSTAKA

Penelitian analisis sentimen di media sosial telah menjadi topik penting dalam bidang *text mining* dan *data mining*, terutama dalam memahami persepsi publik terhadap isu-isu sosial dan politik. Data yang dihasilkan oleh pengguna media sosial seperti Twitter bersifat dinamis, masif, dan tidak terstruktur, sehingga memerlukan pendekatan yang tepat dalam pengolahannya. Salah satu pendekatan yang paling umum digunakan adalah klasifikasi sentimen menggunakan algoritma *machine learning*.

Penelitian oleh Pratiwi et al. [11] mengenai analisis sentimen pengguna Twitter terhadap kebijakan PPKM menunjukkan bahwa data Twitter dapat dimanfaatkan secara efektif untuk menangkap opini publik. Dalam penelitian tersebut, algoritma Support Vector Machine (SVM) digunakan untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Hasilnya menunjukkan bahwa SVM menghasilkan akurasi sebesar 76,2%, yang mengindikasikan performa yang cukup baik dalam menangani data tidak terstruktur dari media sosial.

Studi lain yang membandingkan algoritma Naïve Bayes dan SVM dalam menganalisis sentimen terhadap layanan tiga marketplace populer di Indonesia—Tokopedia, Shopee, dan Lazada—menunjukkan bahwa algoritma SVM secara konsisten memberikan hasil akurasi yang lebih tinggi dibandingkan Naïve Bayes. Untuk Tokopedia, akurasi Naïve Bayes mencapai 85,34%, sementara SVM menghasilkan akurasi sebesar 86,82%. Pada Shopee, Naïve Bayes mencatat akurasi 80,04%, sedangkan SVM sedikit lebih tinggi dengan 80,91%. Sementara itu, pada data Lazada, perbedaan yang cukup signifikan terlihat, di mana Naïve Bayes memperoleh akurasi 83,52% dan SVM mencapai 88,93%. Temuan ini mengindikasikan bahwa algoritma SVM memiliki performa klasifikasi yang lebih andal dalam konteks analisis sentimen pada data e-commerce [12].

Penelitian oleh Ryandi et al. [13] memperkuat temuan tersebut dengan membandingkan algoritma Naïve Bayes dan SVM dalam menganalisis sentimen masyarakat terhadap kinerja Kementerian Kesehatan di media sosial X (dulu Twitter). Hasil analisis menunjukkan bahwa SVM menghasilkan akurasi yang lebih tinggi dibandingkan Naïve Bayes, dengan selisih akurasi sekitar 5%. Penelitian ini juga menekankan pentingnya tahap preprocessing dalam meningkatkan kualitas data yang dianalisis, seperti proses tokenisasi, stopword removal, dan stemming.

Selanjutnya, Indriyani et al. [14] melakukan analisis sentimen terhadap aplikasi TikTok dengan membandingkan kinerja dua algoritma tersebut. Penelitian ini menggunakan 2.000 data ulasan pengguna yang telah melalui proses praproses teks. Hasilnya menunjukkan bahwa metode SVM mencapai akurasi sebesar 84%, mengungguli Naïve Bayes yang memperoleh akurasi 79%. Hal ini menguatkan anggapan bahwa SVM lebih efektif dalam mengklasifikasikan opini yang terkandung dalam teks, khususnya jika teks tersebut bersifat kompleks, informal, dan penuh variasi seperti dalam komentar media sosial.

Dari berbagai studi tersebut dapat dilihat bahwa SVM memiliki beberapa keunggulan utama, yaitu kemampuannya dalam menangani data berdimensi tinggi dan kemampuan untuk membentuk *decision boundary* yang optimal antara kelas sentimen yang berbeda. Meskipun algoritma Naïve Bayes dikenal dengan kesederhanaannya dan efisiensi dalam proses pelatihan model, algoritma ini memiliki keterbatasan ketika dihadapkan pada korelasi antar fitur yang tinggi, yang umum ditemukan dalam data teks.

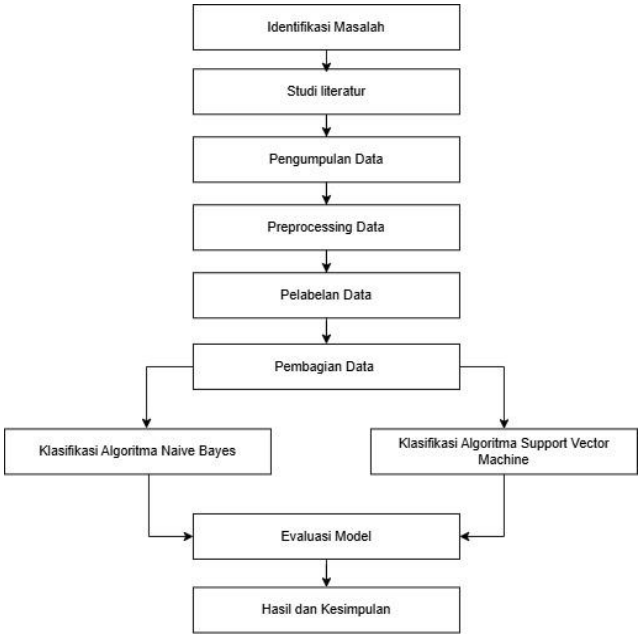
Secara umum, algoritma SVM lebih cocok untuk data teks dari media sosial yang tidak terstruktur, beragam, dan penuh nuansa emosional, karena pendekatan margin maksimum-nya mampu memisahkan kelas sentimen dengan lebih baik. Oleh karena itu, banyak penelitian cenderung memilih SVM ketika ketepatan klasifikasi menjadi prioritas utama, meskipun dengan biaya komputasi yang sedikit lebih tinggi.

Berdasarkan berbagai studi sebelumnya, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) secara konsisten menunjukkan performa yang lebih baik dibandingkan Naïve Bayes dalam tugas analisis sentimen terhadap data media sosial yang bersifat tidak terstruktur. Keunggulan SVM terlihat dari tingkat akurasi yang lebih tinggi dalam berbagai konteks, baik isu sosial-politik, layanan publik, hingga e-commerce dan aplikasi hiburan.

Oleh karena itu, penelitian ini menghadirkan kebaruan (*state of the art*) dengan menerapkan kedua algoritma tersebut secara langsung dalam menganalisis sentimen masyarakat terhadap isu dugaan ijazah palsu Presiden Joko Widodo—sebuah topik politik aktual yang sangat sensitif dan masih menjadi perdebatan di berbagai platform media sosial. Kajian ini penting dilakukan karena sampai saat ini belum banyak studi yang membahas isu ini secara ilmiah melalui pendekatan *machine learning*, terutama dengan fokus pada perbandingan performa algoritma klasifikasi dalam konteks opini publik digital.

METODOLOGI PENELITIAN

Gambar 1 menunjukkan bagan alur tahapan penelitian yang akan membantu peneliti melakukan penelitian dari awal hingga akhir. Dengan mengikuti bagan ini, peneliti dapat memastikan bahwa setiap komponen penelitian berjalan secara terintegrasi dan sesuai dengan rencana.



Gambar 1. Bagan Alur Penelitian

Sumber: Hasil Olah Data

a. *Identifikasi Masalah*

Perdebatan tentang dugaan ijazah palsu mantan Presiden Joko Widodo sedang berlangsung di banyak platform media sosial khususnya Twitter. Banyaknya komentar yang tersebar dengan berbagai pandangan dan opini membuat pengelompokan sentimen publik menjadi tantangan tersendiri. Penelitian ini membandingkan performa algoritma *Naïve Bayes* dan *Support Vector Machine* dalam mengklasifikasikan sentimen publik terhadap isu tersebut.

b. *Studi Literatur*

Penulis melakukan studi literatur melalui berbagai jurnal ilmiah dari penelitian sebelumnya yang berkaitan dengan analisis sentimen, guna memperoleh gambaran umum mengenai penelitian yang akan dilaksanakan.

c. *Pengumpulan Data*

Data Twitter terbaru tahun 2025 digunakan sebagai data primer dalam penelitian ini.

d. *Preprocessing Data*

Preprocessing data merupakan proses membersihkan, mengubah format, dan mempersiapkan data untuk menjadi lebih mudah, akurat, dan siap untuk digunakan dalam analisis. Tahapan *preprocessing* dalam penelitian ini dilakukan melalui beberapa langkah, yaitu filter komentar, *cleaning text*, *normalization text*, *punctuation removal*, tokenisasi, *stopword removal*, dan *stemming*.

e. *Pelabelan Data*

Proses ini merupakan proses yang dilakukan untuk mengkategorikan komentar pengguna twitter terhadap data komentar Twitter mengenai isu Ijazah Jokowi Palsu ke dalam kelas positif, kelas negatif, dan kelas netral.

f. *Pembagian Data*

Pada proses ini, data yang sudah dibersihkan kemudian dibagi menjadi dua jenis, yaitu *data training* dan *data testing*. *Data training* merupakan data yang akan diolah menggunakan algoritma yang dipilih, dan hasil pengolahannya akan menjadi dasar dalam melakukan prediksi. Sementara itu, *data testing* digunakan untuk menguji dan memprediksi hasil dari model yang berasal dari data training yang telah dibentuk.

g. *Klasifikasi Algoritma Naïve Bayes*

Algoritma Naïve Bayes bekerja dengan memprediksi kemungkinan sebuah data termasuk ke dalam kelas tertentu berdasarkan perhitungan peluang [15]. *Naïve bayes* digunakan untuk memperoleh hasil akurasi dari data komentar yang diambil dari komentar postingan twitter dan digunakan untuk mengklasifikasi komentar positif, negatif, dan netral dari data komentar yang sudah diperoleh.

h. *Klasifikasi Algoritma Support Vector Machine (SVM)*

Support Vector Machine didasarkan pada tekniknya yang bertujuan untuk mencari cara atau aturan terbaik untuk membedakan dua kelompok data yang berasal dari dua kelas yang berbeda [16]. *Support Vector Machine* digunakan juga untuk memperoleh hasil akurasi dari komentar postingan twitter tentang objek yang diteliti dan digunakan untuk mengklasifikasi komentar positif, negatif, dan netral.

i. *Evaluasi Model*

Evaluasi model dilakukan dengan menggunakan *tools confusion matrix* sebagai dasar pengukuran performa, yang mencakup perhitungan nilai akurasi *(accuracy)*, presisi *(precision)*, *recall*, dan *f1-score*. Pengujian ini diterapkan pada dua algoritma yang digunakan dalam penelitian, yaitu *Naïve Bayes* dan *Support Vector Machine*, guna membandingkan sejauh mana masing-masing algoritma mampu mengklasifikasikan data ulasan secara akurat dalam konteks analisis sentimen.

j. Hasil dan Kesimpulan

Tahap selanjutnya adalah penarikan kesimpulan hasil akhir. Pada tahap ini, penulis menyimpulkan hasil penelitian berdasarkan proses pelabelan data sentimen yang menghasilkan jumlah kelas positif, negatif, dan netral terhadap data ulasan yang di analisis. Di tahap ini juga dilakukan perbandingan hasil evaluasi model dari dua algoritma, yaitu *Naïve Bayes* dan *Support Vector Machine*, dengan mempertimbangkan nilai akurasi *(accuracy)*, presisi *(precision)*, *recall*, dan *f1-score*. Dari hasil evaluasi tersebut, ditentukan algoritma yang memberikan performa terbaik dalam analisis sentimen penelitian kali ini.

HASIL DAN PEMBAHASAN

a. Proses Pengumpulan Data

Proses pengumpulan data dilakukan dengan memanfaatkan teknik crawling menggunakan pustaka atau tools yang mampu mengakses *Application Programming Interface* (API) dari Twitter. Teknik ini memungkinkan peneliti untuk mengunduh sejumlah besar data berdasarkan kata kunci spesifik, dalam hal ini "ijazah jokowi palsu", yang relevan dengan isu yang sedang diteliti. Pemilihan kata kunci tersebut dilakukan secara cermat agar data yang dikumpulkan benar-benar mencerminkan opini dan reaksi masyarakat terhadap isu yang berkembang. Selain itu, proses crawling juga disertai dengan penyaringan waktu untuk memastikan bahwa data yang diperoleh berasal dari periode tertentu, yaitu bulan Mei 2025, guna menjaga relevansi dan kekinian data terhadap konteks isu. Dataset yang digunakan dalam penelitian kali ini memiliki 4985 record data yang kemudian akan dilakukan preprocessing data terlebih dahulu sebelum kemudian diolah untuk melakukan analisis sentimen. Tabel 1 merupakan contoh data komentar yang diperoleh dari twitter.

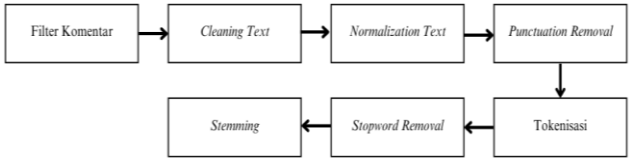
Tabel 1. Contoh Komentar Pengguna Twitter

Nomor	Text Komentar
1	@fendy_070L @GreatOfKratos @jokowi Nggak simple. Lu dituduh dan difitnah ijazah lu palsu trus nuntut orangnya nunjukin biar kelar.
2	Jokowi Laporkan 5 Orang ke Polisi Soal Tudingan Ijazah Palsu https://t.co/yUpDKhrP3r

Sumber: Hasil Olah Data

b. Preprocessing Data

Preprocessing merupakan tahapan krusial dalam analisis data teks karena bertujuan untuk membersihkan dan menyiapkan data mentah agar siap diproses oleh algoritma klasifikasi. Dalam konteks data media sosial seperti Twitter, proses ini menjadi semakin penting mengingat data yang dikumpulkan seringkali mengandung berbagai elemen tidak relevan seperti simbol, tagar, tautan, mention, serta kata-kata slang atau tidak baku. Tahapan ini tidak hanya membantu mereduksi kompleksitas data, tetapi juga meningkatkan akurasi model klasifikasi karena algoritma akan lebih mudah mengenali pola dan makna dalam teks yang telah dibersihkan. Dengan kata lain, preprocessing adalah fondasi penting yang menentukan kualitas hasil analisis sentimen di tahap berikutnya. Preprocessing pada penelitian kali ini merupakan tahap dimana data yang diperoleh dengan cara *crawling* sebanyak 4985 data selanjutnya dilakukan pra-proses data. Tahapan dalam melakukan *preprocessing* data dapat dilihat pada gambar dibawah ini.



Gambar 2. Tahapan Preprocessing
Sumber: Hasil Olah Data

- *Filter Komentar*

Pada tahap ini, dilakukan *filtering data* untuk menghapus postingan twitter yang bukan berasal dari masyarakat umum. Postingan yang terindikasi postingan dari akun pemerintah, akun marketing, akun berita atau akun-akun lain yang bukan milik masyarakat umum dihapus record datanya, supaya komentar pengguna twitter yang ada pada dataset merupakan opini yang benar-benar disuarakan oleh masyarakat terkait topik yang sedang dibahas yaitu mengenai isu ijazah Jokowi palsu. Setelah dilakukan filtering data, data yang tersisa untuk berikutnya diolah berjumlah 3055 data.

- *Cleaning Text*

Kemudian dari data yang sudah di filter pada tahap sebelumnya, dilakukan pembersihan *noise* dari suatu kalimat. Tahap ini dilakukan dengan menghilangkan karakter-karakter yang mengganggu seperti angka, simbol, *URL*, dan *hashtag*, kemudian seluruh teks diubah menjadi huruf kecil pada tahap *lowercase* untuk penyeragaman [17]. Hasil data setelah dilakukan proses *cleaning* dapat dilihat pada tabel 2.

Tabel 2. Hasil *Cleaning Text*

Nomor	Data Mentah	Hasil Cleaning
1	@__PASMANTAP @wiroi_ Doktor Rismon sudah mengatakan skripsi dan ijazah @UGMYogyakarta jokowi palsu di salah satu stasiun tv swasta dalam acara RAKYAT BERSUARA.	doktor rismon sudah mengatakan skripsi dan ijazah jokowi palsu di salah satu stasiun tv swasta dalam acara rakyat bersuara.

2	@Pepatah_dr_Chin @CNNIndonesia @jokowi Ok. 1. Mrk menggugat keaslian ijazah jokowi. Mrk sudah menunjukkan bukti dan membuktikan bhw ijazah Jokowi tdk asli.	ok. 1. mrk menggugat keaslian ijazah jokowi. mrk sudah menunjukkan bukti dan membuktikan bhw ijazah jokowi tdk asli.

Sumber: Hasil Olah Data

- *Normalization Text*

Setelah teks sudah dibersihkan dari *noise*, berikutnya teks diolah lagi masuk kedalam tahap normalisasi. Normalisasi teks diperlukan untuk memperbaiki kesalahan penulisan atau untuk memperbaiki *slang*, atau bahasa gaul, yang tidak ada dalam bahasa baku [18]. Penelitian kali ini dilakukan normalisasi komentar pengguna twitter yang terdapat pada tabel 3.

Tabel 3. Hasil *Normalization Text*

Nomor	Hasil Cleaning	Hasil Normalization
1	klau sy dalam posisi pak jokowi. akan melakukan hal serupa jika ijazah saya palsu.	kalau saya dalam posisi pak jokowi. akan melakukan hal serupa jika ijazah saya palsu.
2	mrk menggugat keaslian ijazah jokowi. mrk sudah menunjukkan bukti dan membuktikan bhw ijazah jokowi tdk asli.	mereka menggugat keaslian ijazah jokowi. mereka sudah menunjukkan bukti dan membuktikan bahwa ijazah jokowi tidak asli.

Sumber: Hasil Olah Data

- *Punctuation Removal*

Punctuation removal bertujuan untuk menghapus tanda baca yang ada pada dataset seperti tanda tanya (?), seru (!), titik (.), koma (,) dan tanda baca lainnya agar mesin hanya memahami teks yang terdapat pada data. Oleh karena tiu, pada penelitian kali ini dilakukan *punctuation removal*.

Tabel 4. Hasil *Punctuation Removal*

Nomor	Hasil Cleaning	Hasil Punctuation Removal
1	sangat memalukan lawyer ini.. sekolah sampai di luar negeri menuntut ilmu.. pulang2 membela ijazah palsu..	sangat memalukan lawyer ini sekolah sampai di luar negeri menuntut ilmu pulang membela ijazah palsu
2	dimaksud kemungkinan besar merujuk pada joko widodo (jokowi).	yang dimaksud kemungkinan besar merujuk pada joko widodo jokowi

Sumber: Hasil Olah Data

- *Tokenisasi*

Kemudian dilakukan tokenisasi pada teks komentar pengguna twitter. Tokenisasi adalah proses untuk membagi kalimat menjadi bagian kata yang lebih kecil yang disebut "*sub-word*" atau "bagian token" [19]. Hasil data setelah dilakukan tokenisasi dapat dilihat pada tabel 5.

Tabel 5. Hasil Tokenisasi

Nomor	Hasil Punctuation Removal	Hasil Tokenisasi
1	sangat memalukan lawyer ini sekolah sampai di luar negeri menuntut ilmu pulang membela ijazah palsu	['sangat', 'memalukan', 'lawyer', 'ini', 'sekolah', 'sampai', 'di', 'luar', 'negeri', 'menuntut', 'ilmu', 'pulang', 'membela', 'ijazah', 'palsu'
2	prediksi diamnya prabowo dalam kasus dugaan ijazah palsu jokowi mungkin beliau juga punya kecurigaan yg sama	['prediksi', 'diamnya', 'prabowo', 'dalam', 'kasus', 'dugaan', 'ijazah', 'palsu', 'jokowi', 'mungkin', 'beliau', 'juga', 'punya', 'kecurigaan', 'yg', 'sama'

Sumber: Hasil Olah Data

- *Stopword Removal*

Proses menghilangkan kata tidak penting yang sering muncul dalam sebuah data. Ini dilakukan untuk mengurangi dimensi data dan meningkatkan efisiensi komputasi [20]. Tahap *preprocessing* perlu dilakukan *stopword removal* supaya pada text kalimat yang diolah tidak ada kata kata yang tidak diperlukan. Hasil data setelah dilakukan *stopword removal* dapat dilihat pada tabel 5.

Tabel 6. Hasil *Stopword Removal*

Nomor	Hasil Tokenisasi	Hasil Stopword Removal
1	'sulit', 'memprediksi', 'pasti', 'siapa', 'yang', 'akan', 'jadi', 'tersangka', 'tersangka', 'dalam', 'kasus', 'tuduhan', 'ijazah', 'palsu', 'jokowi'	'sulit', 'memprediksi', 'tersangka', 'tuduhan', 'ijazah', 'palsu', 'jokowi'

2	'keyakinan', 'saya', 'dalam', 'persidangan', 'ijazah', 'palsu', 'seluruh', 'masyarakat', 'akan', 'mendukung', 'drtifa', 'dan', 'kepolisian'	'keyakinan', 'persidangan', 'ijazah', 'palsu', 'masyarakat', 'mendukung', 'drtifa', 'kepolisian'

Sumber: Hasil Olah Data

- *Stemming*

Proses *stemming* merupakan tahap akhir dari *preprocessing* data penelitian ini. Proses *stemming* digunakan untuk menemukan kata dasar dari kata berimbuhan [21].

Tabel 7. *Hasil Stopword Removal*

Nomor	Hasil Stopword Removal	Hasil Stemming
1	['sulit', 'memprediksi', 'tersangka', 'tuduhan', 'ijazah', 'palsu', 'jokowi'	'sulit', 'prediksi', 'sangka', 'tuduh', 'ijazah', 'palsu', 'jokowi'
2	'keyakinan', 'persidangan', 'ijazah', 'palsu', 'masyarakat', 'mendukung', 'drtifa', 'kepolisian'	'yakin', 'sidang', 'ijazah', 'palsu', 'masyarakat', 'dukung', 'drtifa', 'polisi'

Sumber: Hasil Olah Data

c. Pelabelan Data

Komentar pengguna twitter di berikan label dengan menggunakan teknik *lexicon based*. Dan kamus yang digunakan pada pelabelan ini yaitu menggunakan kamus dari *ID-OpinionWords* yang merupakan kumpulan kata-kata positif dan juga negatif berbahasa Indonesia. Dari hasil proses ini mengelompokan data menjadi label positif negatif dan netral dengan rincian sebanyak 942 data positif, 1453 negatif, dan, 660 data tergolong netral. Contoh hasil dari pelabelan komentar positif, negatif, dan netral dapat dilihat pada tabel 8.

Tabel 8. Contoh Hasil Pelabelan

Kategori	Text Komentar
Positif	menyala pak kami mendukungmu untuk melaporkan penyebar isu ijazah palsu dll
Negatif	mata mu picek ya entar kalau tidak lapor di bilang ijazah pak jokowi palsu pak jokowi lapor buar pengadilan yang membuktikan di bilang jahat otak mu dimana di pantat
Netral	bapakku serius banget menonton debat ijazah palsu jokowi

Sumber: Hasil Olah Data

d. Pembagian Data

Data kemudian dibagi menjadi dua yaitu *data training* dan *data testing*. Sebanyak 2444 untuk *training data* dan 611 untuk *testing data*. Membentuk model yang ideal untuk mengklasifikasikan sentimen berdasarkan dataset yang tersedia adalah tujuan dari pembagian ini.

e. Implementasi Algoritma Naïve Bayes

Dalam penelitian ini, bahasa pemrograman Python digunakan untuk menerapkan *Algoritma Naive Bayes*. Python memiliki *library Sklearn*, yang memungkinkan implementasi *Naive Bayes* untuk berbagai metode dan algoritma yang digunakan untuk pengajaran mesin. Baris code implementasi *naive bayes* dapat dilihat pada gambar dibawah.

```
from sklearn.naive_bayes import MultinomialNB
nb_model = MultinomialNB()
nb_model.fit(X_train_tfidf, y_train)
y_pred_nb = nb_model.predict(X_test_tfidf)
```

Gambar 3. *Kode implementasi algoritma Naïve Bayes*
Sumber: Hasil Olah Data

Mulai dari baris kode pertama yaitu "*from sklearn.naive_bayes import MultinomialNB*" yang digunakan untuk mengimport *library sklearn* untuk model *multinomial naive bayes*, dimana model *multinomial* ini cocok digunakan untuk data diskrit seperti teks. Baris kode nb_model = MultinomialNB() digunakan untuk membuat model *naive bayes* dan mengubahnya menjadi variabel string bernama nb_model. Kemudian baris kode nb_model.fit(X_train_tfidf, y_train) digunakan untuk melatih model dengan data latih. Lalu baris code y_pred_nb = nb_model.predict(X_test_tfidf) digunakan untuk melakukan prediksi dengan data uji menggunakan model yang sebelumnya sudah dilatih menggunakan data latih.

f. Implementasi Algoritma Support Vector Machine (SVM)

Implementasi algoritma *Support Vector Machine* masih dilakukan dengan menggunakan bahasa pemrograman yang sama dengan implementasi *naive bayes*. Baris code implementasi *Support Vector Machine* dapat dilihat pada gambar 4 dibawah ini.

```
from sklearn.svm import SVC
svm_model = SVC(kernel='linear')
svm_model.fit(X_train_tfidf, y_train)
y_pred_svm = svm_model.predict(X_test_tfidf)
```

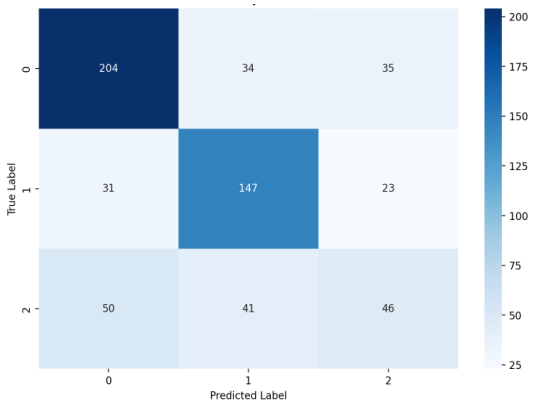
Gambar 4. *Kode Implementasi Algoritma SVM*
Sumber: Hasil Olah Data

Mulai dari baris kode pertama yaitu "*from sklearn.svm import SVC*" yang digunakan untuk mengimport *library sklearn* untuk model *Supoort Vector Machine*. Baris kode svm_model = SVC (kernel='linear') digunakan untuk membuat model *Support Vector Machine*

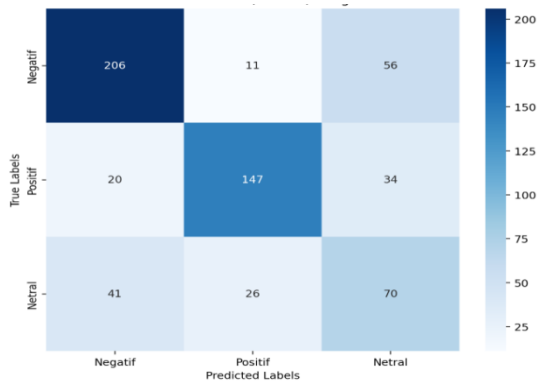
dengan kernel linear dan mengubahnya menjadi variabel string bernama svm_model. Kemudian baris kode svm_model.fit (X_train_tfidf, y_train) digunakan untuk melatih model dengan data latih. Lalu baris code y_pred_svm = svm_model. predict (X_test_tfidf) digunakan untuk melakukan prediksi dengan data uji menggunakan model yang sebelumnya sudah dilatih menggunakan data latih.

g. *Evaluasi Model*

Evaluasi model baik *Naive Bayes* maupun *Support Vector Machine* dilakukan dengan menggunakan *confussion matrix*. Menurut penelitian yang sudah dilakukan, *Confusion Matrix* adalah metode evaluasi model klasifikasi dengan tabel matriks untuk membandingkan hasil prediksi model dengan data sebenarnya. Berdasarkan hasil *Confussion Matrix*, Model *naive bayes* pada gambar 5, berhasil memprediksi 147 data positif, 204 data negatif, dan 46 data netral dengan benar dari total keseluruhan data. Namun, masih terjadi kesalahan prediksi pada 112 data negatif, 54 data positif, dan 91 data netral. Sedangkan hasil model dari *Support Vector Machine* di gambar 6 berhasil memprediksi benar 204 data Negatif, 147 data Positif, dan 46 data Netral. Namun, masih terjadi kesalahan prediksi pada 112 data Negatif, 54 data Positif, dan 91 data Netral. Berikut dibawah hasil *confusion matrix*.



Gambar 5. *Visualisasi confusion matrix Naïve Bayes*
Sumber: Hasil Olah Data



Gambar 6. Bagan Alur Penelitian
Sumber: Hasil Olah Data

Selain data di atas, *Confusion matrix* juga menghasilkan empat parameter utama yaitu akurasi, *precision*, *recall*, dan *f1-score*. Hasil dari confusion matrix pada model *naive bayes* mencapai akurasi sebesar 65%, dengan *precision* 64%, *recall* 65%, dan *f1-score* 64%. Sedangkan Hasil dari *confusion matrix* pada model support vector machine mencapai akurasi sebesar 69,23%, dengan nilai weighed average precision sebesar 70,56%, recall sebesar 69,23%, dan *f1-score* sebesar 69,78%. Hasil yang diperoleh dari pengujian model *naive bayes* dan support vector machine dapat dilihat pada tabel 9.

Tabel 9. Contoh Hasil Pelabelan

No	Algoritma	Accuracy	Precisi-on	Recal	F1-Score
1.	Naïve Bayes	65,00%	64,00%	65,00%	64,00%
2.	SVM	69,23%	70,56%	69,78%	69,78%

Sumber: Hasil Olah Data

KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* lebih unggul daripada Naïve Bayes dalam menilai sentimen pengguna Twitter tentang isu ijazah palsu milik mantan presiden Joko Widodo. SVM memiliki lebih banyak akurasi daripada Naïve Bayes, dengan 69,23% daripada 65%. Selain itu, SVM memiliki nilai *precision*, *recall*, dan *f1-score* yang lebih baik, menjadikannya algoritma yang lebih efektif dalam penelitian ini.

Perbedaan yang cukup mencolok terlihat pada nilai precision dan f1-score, di mana SVM memperoleh nilai precision sebesar 70,56% dan f1-score sebesar 69,78%, sedangkan Naïve Bayes hanya mencapai precision 64% dan f1-score 64%. Hal ini menunjukkan bahwa SVM lebih mampu dalam mengidentifikasi sentimen dengan tingkat kesalahan yang lebih rendah, terutama dalam klasifikasi data yang bersifat kompleks dan tidak seimbang. Selain itu, nilai recall yang lebih tinggi pada SVM juga mengindikasikan bahwa model ini lebih baik dalam mengenali seluruh data aktual yang termasuk dalam masing-masing kelas. Dengan demikian, dalam konteks penelitian ini, SVM merupakan algoritma yang lebih unggul dan direkomendasikan untuk digunakan pada analisis sentimen data media sosial, khususnya untuk isu-isu yang rawan polarisasi opini seperti dugaan ijazah palsu.

Untuk meningkatkan performa, penelitian selanjutnya dapat mengeksplorasi dan menguji algoritma lain yang dapat digunakan dalam analisis sentimen. Peneliti juga dapat menggunakan topik yang lebih netral agar data tidak hanya terbagi ke dalam data negatif terlalu banyak, yang berdampak pada kinerja mesin dalam pembuatan model. Selain itu, penggunaan teknik ekstraksi fitur yang lebih mendalam seperti TF-IDF, word embeddings (misalnya Word2Vec atau GloVe), serta pemrosesan data latih yang lebih berimbang dapat membantu

meningkatkan akurasi klasifikasi. Penelitian lanjutan juga dapat mengkaji perubahan sentimen secara temporal dan memvisualisasikan hasil analisis agar lebih mudah dipahami oleh pengguna akhir atau pengambil kebijakan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Amikom Purwokerto yang telah memberikan dukungan baik materil dan non-materil.

DAFTAR PUSTAKA

[1] A. Mustapa, R. Machmud, and D. L. Radji, "Pengaruh Penggunaan Media Sosial Terhadap Keputusan Pembelian Pada Umkm Jiksau Food," *J. Ilm. Manaj. dan Bisnis*, vol. 5, no. 1, p. 2022, 2022, [Online]. Available: http://ejurnal.ung.ac.id/index.php/JIMB

[2] B. R. T. Rizqi and H. Heriyanto, "Penyebaran Informasi melalui Thread Berita di Twitter oleh Mahasiswa S-1 Program Studi Ilmu Perpustakaan Universitas Diponegoro," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 7, no. 3, pp. 515–528, 2023, doi: 10.14710/anuva.7.3.515-528.

[3] F. Yusuf and N. Achmad, "Kontroversi Penggantian Calon Wakil Wali Kota Dalam Pemilihan Kepala Daerah Kota Baubau 2024," *J. Publicuho*, vol. 8, no. 1, pp. 91–105, 2025.

[4] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjeee.v5i1.16830.

[5] P. A. Permatasari, L. Linawati, and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," *Maj. Ilm. Teknol. Elektro*, vol. 20, no. 2, pp. 177–185, 2021, doi: 10.24843/mite.2021.v20i02.p01.

[6] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26

[7] D. M. Alvina Felicia Watratan, Arwini Puspita. B, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 7–14, 2020, doi: 10.55606/jurritek.v1i1.127.

[8] M. R. Fatturrahman and A. Kurniasih, "Penggunaan Metode NearMiss, SMOTE, dan Naïve Bayes untuk Klasifikasi Gangguan Tidur Berdasarkan Kualitas Tidur dan Gaya Hidup," in *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, Jakarta: Universitas Pembangunan Nasional Veteran Jakarta, 2023, pp. 567–576.

[9] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.isssue3.year2020.page133-143.

[10] W. Ningsih, B. Alfianda, R. Rahmaddeni, and D. Wulandari, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Twitter pada Penggunaan Mobil Listrik di Indonesia," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 556–562, 2024, doi: 10.57152/malcom.v4i2.1253.

[11] A. Putra, D. Haeirudin, H. Khairunnisa, and R. Latifah, "Analisis Sentimen Masyarakat Terhadap Kebijakan PPKM Pada Media Sosial Twitter Menggunakan Algoritma SVM," *Semin. Nas. Sains dan Teknol. 2021*, no. November, pp. 1–6, 2021.

[12] I. Kurniawan, A. Lia Hananto, S. Shofia Hilabi, A. Hananto, B. Priyatna, and A. Yuniar Rahman, "Perbandingan Algoritma Naive Bayes Dan SVM Dalam Sentimen Analisis Marketplace Pada Twitter," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 1, pp. 731–740, 2023, [Online]. Available: http://jurnal.mdp.ac.id

[13] J. Sains, F. A. Ryandi, D. Pratiwi, S. Sari, and J. Sains, "Analisis Sentimen Masyarakat Di Media Sosial X Terhadap Kemenkes Dengan Naive Bayes dan SVM," *J. Sains dan Teknol.*, vol. 7, no. 1, pp. 1–6, 2025.

[14] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.

[15] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.

[16] S. Suryani and M. Mustakim, "Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine," *Bull. Informatics Data Sci.*, vol. 1, no. 2, p. 81, 2022, doi: 10.61944/bids.v1i2.36.

[17] S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, "Pengaruh Kombinasi dan Urutan Pre-Processing pada Tweets Bahasa Indonesia," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 3, no. 2, p. 1074, 2022, doi: 10.24843/jtrti.2022.v03.i02.p06.

[18] S. Tuhpatussania, "Normalisasi Teks Komentar Instagram Masyarakat Makassar Menggunakan Metode Levenshtein Distence," *Explore*, vol. 12, no. 1, p. 29, 2022, doi: 10.35200/explore.v12i1.515.

[19] A. A. Mudding, "Mengungkap Opini Publik: Pendekatan BERT-based-caused untuk Analisis Sentimen pada Komentar Film," *J. Syst. Comput. Eng.*, vol. 5, no. 1, pp. 36–43, 2024, doi: 10.61628/jsce.v5i1.1060.

[20] A. H. Dani, E. Y. Puspaningrum, and R. Mumpuni, "Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying," *Router J. Tek. Inform. dan Terap.*, vol. 2, no. 2, pp. 94–106, 2024, [Online]. Available: https://doi.org/10.62951/router.v2i2.76

[21] N. W. Wardani and P. G. S. Cipta Nugraha, "Stemming Dokumen Teks Bahasa Bali Dengan Metode Rule Base Approach," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 510–521, 2020, doi: 10.35957/jatisi.v7i3.538.