

Analisis Sentimen Acara Clash of Champions dengan Algoritma Naïve Bayes dan Support Vector Machine

Putri Intan Purnama*, Ryan Randy Suryono

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ^{1,*}putri_intan_purnama@teknokrat.ac.id, ²ryan@teknokrat.ac.id

Email Penulis Korespondensi: putri_intan_purnama@teknokrat.ac.id

Submitted: 30/12/2024; Accepted: 26/02/2025; Published: 01/03/2025

Abstrak—Dengan kemajuan teknologi informasi dan komunikasi, masyarakat semakin mudah bertukar informasi dan mengakses konten pendidikan, termasuk melalui platform pembelajaran daring seperti Ruangguru. Salah satu program unggulan Ruangguru adalah *Clash of Champions*, yang menarik perhatian publik dan menghasilkan beragam sentimen di media sosial. Namun, menganalisis sentimen publik terhadap program ini menghadapi tantangan, terutama karena adanya ketidakseimbangan jumlah data antara sentimen *mayoritas* dan *minoritas*, yang dapat memengaruhi akurasi model analisis sentimen. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), dalam menganalisis sentimen publik terhadap program ini. Dengan menggunakan 5.226 *tweet* dari media sosial X, data diseimbangkan menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi masalah ketidakseimbangan data. Setelah data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, hasil menunjukkan bahwa sebelum menggunakan SMOTE, *Naïve Bayes* memiliki akurasi 78%, sedangkan SVM mencapai 82%. Setelah SMOTE diterapkan, akurasi *Naïve Bayes* meningkat menjadi 79%, sementara SVM naik menjadi 84%. Selain akurasi, peningkatan signifikan juga terlihat pada *precision*, *recall*, dan *f1-score*, terutama untuk sentimen positif. Hasil penelitian menunjukkan bahwa SVM lebih unggul dibandingkan *Naïve Bayes*, baik dalam akurasi maupun metrik evaluasi lainnya. Penelitian ini memberikan pemahaman mendalam tentang efektivitas algoritma dalam analisis sentimen pada program edukasi berbasis hiburan dan diharapkan menjadi referensi untuk pengembangan model serupa di masa depan.

Kata Kunci: Analisis Sentimen; Clash of Champions; Naive Bayes; SMOTE; Support Vector Machine

Abstract—With the advancement of information and communication technology, it has become easier for people to exchange information and access educational content, including through online learning platforms such as Ruangguru. One of Ruangguru's flagship programs is Clash of Champions, which attracts public attention and generates various sentiments on social media. However, analyzing public sentiment towards this program faces challenges, especially due to the imbalance in the amount of data between majority and minority sentiments, which may affect the accuracy of sentiment analysis models. This study aims to compare the performance of two algorithms, namely Naïve Bayes and Support Vector Machine (SVM), in analyzing public sentiment towards this program. Using 5,226 tweets from social media X, the data was balanced using the Synthetic Minority Oversampling Technique (SMOTE) method to overcome the data imbalance problem. After the data was divided into 80% for training and 20% for testing, the results showed that before using SMOTE, Naïve Bayes had an accuracy of 78%, while SVM reached 82%. After SMOTE was applied, Naïve Bayes' accuracy increased to 79%, while SVM rose to 84%. In addition to accuracy, significant improvements were also seen in precision, recall, and f1-score, especially for positive sentiments. The results show that SVM is superior to Naïve Bayes, both in accuracy and other evaluation metrics. This research provides an in-depth understanding of the effectiveness of algorithms in sentiment analysis on entertainment-based educational programs and is expected to be a reference for the development of similar models in the future.

Keywords: Sentiment Analysis; Clash of Champions; Naive Bayes; SMOTE; Support Vector Machine

1. PENDAHULUAN

Dengan berkembangnya teknologi informasi dan komunikasi, cara orang mengakses dan berbagi informasi telah berubah, termasuk melalui program-program edukasi yang disediakan oleh *platform* pembelajaran daring [1]. Perkembangan pesat teknologi ini memberikan dampak besar, terutama di bidang pendidikan[2]. Pendidikan memiliki peran krusial dalam membekali generasi muda untuk menghadapi berbagai tantangan di era global. Proses pendidikan yang berkesinambungan bertujuan membantu individu berkembang secara fisik, mental, dan spiritual, sehingga mampu beradaptasi dengan berbagai aspek kehidupan, baik intelektual, emosional, maupun sosial. Peningkatan kualitas pendidikan menjadi bagian penting dari pengembangan sumber daya manusia. Tanggung jawab ini tidak hanya berada pada lembaga pendidikan, tetapi juga membutuhkan kerja sama antara masyarakat dan pemerintah[3]. Di era digital, *platform* pembelajaran daring telah menjadi salah satu sarana utama dalam mendukung proses belajar mengajar. Sebuah *platform* yang cukup populer di Indonesia adalah Ruangguru. Dengan berbagai program inovatifnya, Ruangguru tidak hanya menarik minat siswa dan orang tua, Namun, juga berperan signifikan dalam mendukung peningkatan mutu pendidikan di Indonesia[4].

Sebagai perusahaan teknologi pendidikan terbesar di Indonesia, Ruangguru menyediakan berbagai layanan berbasis teknologi, seperti kelas virtual, *platform* ujian online, video pembelajaran berlangganan, *marketplace* les privat, dan berbagai konten edukasi lainnya yang dapat diakses melalui aplikasi dan situs web. Dengan lebih dari 22 juta pengguna dan dukungan 300.000 guru di lebih dari 100 bidang pelajaran, Ruangguru terus menghadirkan inovasi untuk mendukung pembelajaran. Selain itu, platform ini telah menjalin kemitraan strategis dengan 32 pemerintah provinsi dan 326 pemerintah kota dan kabupaten di Indonesia[5]. Tidak hanya berhenti pada layanan pendidikan, Ruangguru juga menghadirkan inovasi di bidang hiburan edukatif. Salah satunya adalah program *Clash of Champions*,

sebuah kompetisi akademik yang terinspirasi dari program edukasi sukses Korea Selatan, *University War*. Program ini dikemas dengan visual menarik, audio yang dinamis, dan strategi pemasaran yang efektif. Dengan format yang matang, *Clash of Champions* berhasil menjadi gebrakan baru dalam tayangan edukasi yang menarik perhatian generasi muda di media sosial. Program ini menampilkan mahasiswa berprestasi dari perguruan tinggi dalam dan luar negeri, seperti NTU, NUS, Oxford, dan KAIST[6].

Pada bulan Juni 2024, Ruangguru mengadakan *Clash of Champions*, yang mempertemukan mahasiswa dari universitas ternama Indonesia, seperti Universitas Airlangga, Universitas Gadjah Mada, Institut Teknologi Bandung, Universitas Indonesia, Universitas Udayana, dan Binus. Kompetisi ini juga melibatkan mahasiswa Indonesia yang berkuliah di kampus-kampus bergengsi dunia, menjadikannya ajang bergengsi untuk menunjukkan prestasi akademik di tingkat global[7]. Program *Clash of Champions* yang diadakan oleh Ruangguru sukses besar di berbagai platform media sosial. Episode pertama di YouTube mencapai 9 juta penonton dan 16.397 komentar dalam dua minggu, sementara episode kelima meraih hampir 6 juta penonton dalam enam hari. Acara ini juga viral di TikTok dan menjadi trending topic di *Twitter* (X) setiap kali episode baru tayang, menunjukkan antusiasme masyarakat yang tinggi[6].

Tingginya interaksi di platform media sosial X menunjukkan bahwa acara ini telah menjadi perbincangan yang signifikan. Platform ini, dengan format *tweet*-nya yang ringkas namun representatif, menyediakan ruang bagi masyarakat untuk mengekspresikan pendapat mereka secara langsung. Banyaknya tanggapan yang muncul di X menciptakan sumber data yang kaya untuk analisis sentimen. Analisis sentimen memegang peranan penting dalam memahami persepsi publik terhadap suatu acara atau program. Melalui analisis sentimen, dapat diungkap pola-pola opini dan kecenderungan sikap masyarakat, baik positif maupun negatif[8]. Melalui analisis ini, pandangan, kritik, dan apresiasi masyarakat terhadap *Clash of Champions* dapat diidentifikasi dengan lebih jelas, memberikan wawasan yang berguna bagi pengembangan acara di masa mendatang.

Kajian-kajian terdahulu telah mendemonstrasikan efektivitas implementasi analisis sentimen dalam ranah pendidikan, menyiratkan bahwa analisis ini mampu memberikan gambaran lebih mendalam tentang persepsi publik terhadap program pendidikan dan acara-acara terkait. Salah satunya Penelitian Yud Caesar Ivan Sabastian dkk. (2024) menunjukkan bahwa algoritma SVM mencapai akurasi tinggi 99,34% dalam menganalisis sentimen terhadap acara *Clash of Champions* Ruangguru, dengan 93,56% respons positif[7]. Efrico Fernando Baharsyah dkk. (2024) membandingkan SVM dan *Naïve Bayes*, di mana SVM lebih unggul dengan akurasi 89% dibandingkan 84,86% dari *Naïve Bayes*[9]. Indah Novitasari dkk. (2024) melaporkan akurasi *Naïve Bayes* sebesar 69% dalam menganalisis sentimen terhadap *tweet* Ruangguru[10]. Penelitian Ari Wibowo dkk. (2022) menunjukkan akurasi 88,5% untuk *Naïve Bayes* dalam menganalisis keefektifan pembelajaran daring selama pandemi COVID-19[11]. Nurdin dkk. (2024) melaporkan akurasi *Naïve Bayes* 72,73% dalam analisis sentimen dampak AI untuk pendidikan[12]. Sementara itu, Amira Aida Rashifa (2024) menunjukkan bahwa algoritma SVM memiliki akurasi 75%, lebih tinggi dibandingkan *Naïve Bayes* yang hanya 25% dalam mengukur minat mahasiswa untuk bergabung dengan MSIB[13].

Sebagian besar penelitian sebelumnya lebih berfokus pada analisis sentimen terhadap platform pembelajaran daring atau aplikasi pendidikan menggunakan algoritma *Naïve Bayes* atau SVM, tanpa perbandingan menyeluruh antara kedua algoritma tersebut dalam konteks acara edukasi dengan elemen hiburan seperti *Clash of Champions*. Selain itu, kajian analisis sentimen terhadap acara edukasi viral di media sosial, seperti X (*Twitter*), yang memiliki tingkat interaksi publik tinggi, masih jarang dilakukan. Faktor sosial dan hiburan yang memengaruhi respons masyarakat terhadap acara semacam ini juga belum banyak diteliti.

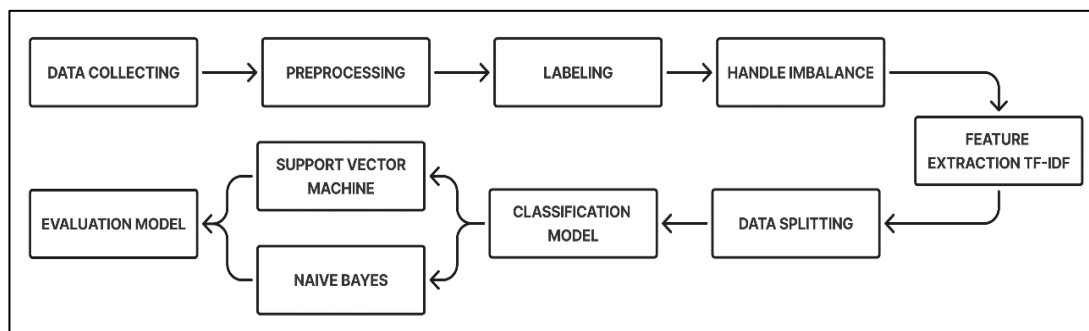
Tingginya interaksi publik di media sosial X terhadap program *Clash of Champions* yang diselenggarakan oleh Ruangguru menunjukkan kebutuhan untuk memahami persepsi masyarakat secara lebih mendalam. Namun, proses analisis sentimen menghadapi tantangan ketidakseimbangan data antara sentimen positif dan negatif, yang dapat memengaruhi kinerja algoritma dalam mengklasifikasi sentimen. Oleh karena itu, penelitian ini dilakukan untuk membandingkan akurasi dan efektivitas algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM), yang keduanya sering digunakan dalam analisis sentimen namun menunjukkan hasil yang beragam pada berbagai konteks.

Penelitian ini menggunakan dataset dari 5.226 *tweet* di media sosial X, yang mencakup sentimen mayoritas dan minoritas. Untuk mengatasi ketidakseimbangan data, metode *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan. Perbandingan kedua algoritma dilakukan untuk mengidentifikasi mana yang lebih efektif dalam mengatasi tantangan data tidak seimbang, sekaligus memberikan wawasan tentang respons masyarakat terhadap program edukasi berbasis hiburan. Hasil penelitian ini diharapkan tidak hanya memberikan rekomendasi untuk pengembangan acara serupa, tetapi juga memperkaya metodologi analisis sentimen dalam konteks media sosial dengan tingkat interaksi tinggi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan mengikuti beberapa tahapan yang sistematis untuk memastikan hasil yang valid dan dapat dipertanggungjawabkan. Setiap langkah penelitian dirancang untuk mendalami topik secara menyeluruh dan memberikan pemahaman yang lebih baik mengenai permasalahan yang diangkat. Keseluruhan tahapan penelitian ini divisualisasikan dalam bentuk skema pada Gambar 1.



Gambar 1. Skema Penelitian

Berdasarkan gambar 1 tahap penelitian dimulai dengan proses pengumpulan data adalah awalan dari penelitian ini, dengan menggunakan metode *crawling* melalui *library Tweet Harvest* di *Google Colab* dengan memanfaatkan *API Token Twitter*. Data berupa *tweet* dengan kata kunci tertentu dikumpulkan dalam rentang waktu yang telah ditentukan, selanjutnya data disimpan dalam format *CSV (Comma Separated Values)* untuk diproses lebih lanjut. Tahap berikutnya adalah pra-pemrosesan data elemen yang tidak relevan, akan di bersihkan dan dihilangkan teksnya. Setelah itu, data diberi label sesuai kategori tertentu, seperti sentimen positif dan negatif. Untuk mengatasi ketidakseimbangan data, dilakukan teknik *Synthetic Minority Oversampling Technique (SMOTE)* agar distribusi kelas lebih merata. Selanjutnya, fitur teks diekstraksi menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* agar menghasilkan gambaran dalam bentuk angka. Dataset kemudian dipisahkan antara data latih dan data uji untuk membangun model klasifikasi dengan algoritma seperti *Support Vector Machine (SVM)* dan *Naive Bayes*. Terakhir, performa model dievaluasi menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengevaluasi kinerja model dalam pengklasifikasian data yang tepat.

2.2 Data Collection

Proses pengumpulan data menggunakan metode *crawling*, yaitu pendekatan yang memungkinkan pengambilan informasi dari situs web secara otomatis dengan bantuan perangkat lunak[14]. Data dikumpulkan menggunakan teknik *crawling* melalui *library tweet-harvest* dari *Node.js*[15]. Salah satu keunggulan *tweet-harvest* adalah kemampuannya untuk mengumpulkan data dalam jumlah besar, yang dapat dijalankan melalui *Command Line Interface (CLI)* hanya dengan menggunakan *auth_token*[16]. Data dikumpulkan dalam *tweet* berbahasa Indonesia dengan kata kunci "*clash of champions*". Sebanyak 7.903 *tweet* berhasil diperoleh dalam rentang waktu 1 Agustus 2024 hingga 30 November 2024.

2.3 Data Preprocessing

Preprocessing data adalah tahap persiapan yang bertujuan untuk memudahkan pengolahan dan analisis data. Ini merupakan proses untuk membersihkan data dari gangguan, mengurangi dimensi data, dan membuat data lebih terstruktur[17]. Tahap *preprocessing* diperlukan untuk meningkatkan hasil analisis. Tahap ini, *preprocessing* dilakukan melalui, *data cleaning*, *case folding*, *tokenizing*, *stopwords removal*, dan *stemming*[18]. Berikut adalah penjelasannya:

- Cleaning* Tahap ini melibatkan penghapusan karakter *non-alfabet* untuk mengurangi *noise*. Karakter yang dihilangkan meliputi tanda baca seperti titik (.), koma (,), tanda tanya (?), tanda seru (!), serta simbol seperti '@' untuk *username*, *hashtag* (#), emotikon, dan alamat situs web[19].
- Case Folding* Semua huruf di dataset diubah menjadi huruf kecil untuk memastikan konsistensi dan menghindari perbedaan yang tidak diperlukan antara huruf kapital dan huruf kecil[20]
- Tokenizing* Tahap ini memecah sekumpulan karakter dalam teks diubah menjadi unit kata. Pada proses ini juga dilakukan penghapusan karakter pembatas, angka, serta tanda baca atau karakter lain yang bukan huruf[21].
- Filtering* Langkah ini bertujuan menghapus *stopwords*, yaitu kata-kata seperti "atau," "yang," "dengan," "di," "ke," dan "tetapi," yang tidak memiliki makna penting dalam analisis sentimen. Langkah ini menyederhanakan teks dan mempertahankan hanya kata-kata yang relevan dengan konteks sentimen[22].
- Stemming* Tahap ini menyederhanakan kata menjadi bentuk dasarnya, contoh kata "berlari," "berlari-lari," dan "lari" disederhanakan jadi "lari" dengan menghapus sufiks, prefiks, dan konfiks [23]. Penelitian ini menggunakan metode *stemming Sastrawi*[24].

2.4 Labeling

Proses pelabelan dilakukan secara otomatis menggunakan *TextBlob*, yang menganalisis teks berdasarkan dua metrik utama, yaitu *polarity* dan *subjectivity*. *Polarity* mengukur arah sentimen dalam teks, apakah condong ke sentimen positif atau negatif. Sementara itu, *subjectivity* mengindikasikan sejauh mana teks mengandung opini atau fakta. Teks dengan nilai *subjectivity* tinggi cenderung berupa opini, sedangkan nilai *polarity* yang tinggi menunjukkan sentimen yang lebih positif. Berdasarkan pengukuran *polarity*, teks lalu diklasifikasi menjadi kategori positif dan negatif[25]. Proses ini menghasilkan 3.349 data *tweet* yang memiliki sentimen negatif dan 2.177 data *tweet* yang termasuk

sentimen positif, membentuk dataset yang kuat untuk analisis lebih lanjut. Rumus yang digunakan dalam pelabelan ini disajikan pada Persamaan 1.

$$S_{sentiment} = \begin{cases} \text{positive if } S_{positive} > S_{negative} \\ \text{negative if } S_{positive} < S_{negative} \end{cases} \quad (1)$$

Rumus ini digunakan untuk mengklasifikasikan sentimen teks berdasarkan perbandingan skor positif ($S_{positive}$) dan negatif ($S_{negative}$). Jika $S_{positive}$ lebih besar dari $S_{negative}$, teks dianggap "positif." Sebaliknya, jika $S_{negative}$ lebih besar, teks dianggap "negatif." Rumus ini secara otomatis menentukan sentimen teks berdasarkan nilai kedua skor tersebut.

2.5 Handle Imbalance

Hasil dari proses pelabelan menunjukkan adanya ketidakseimbangan data, sehingga diperlukan penanganan lebih lanjut menggunakan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Metode ini sebuah pembelajaran mesin yang dirancang untuk mengatasi ketidakseimbangan kelas pada data[26]. SMOTE bekerja menggabungkan sampel baru dari kelas minoritas untuk menyeimbangkan dataset. Dengan menyertakan contoh-contoh baru dari kelas minoritas, data pelatihan menjadi lebih seimbang, dan data tambahan ini digunakan untuk melatih model pengklasifikasi[27].

2.6 Feature Extraction

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode pembobotan kata yang digunakan untuk menilai sejauh mana pentingnya suatu kata dalam sebuah dokumen, berdasarkan frekuensi kemunculannya.. Pada penelitian ini, implementasi TF-IDF dilakukan dengan menggunakan *library scikit-learn* pada bahasa *Python*[28]. Data yang telah melalui tahap prapemrosesan kemudian diberi bobot menggunakan metode ini. Proses pembobotan terdiri dari dua langkah utama: pertama, memperkirakan frekuensi kemunculan suatu kata dalam dokumen dibandingkan dengan total jumlah kata dalam dokumen tersebut (*Term Frequency*, TF); kedua, menghitung seberapa jarang kata tersebut muncul di seluruh dokumen (*Inverse Document Frequency*, IDF)[29]. Rumus TF-IDF disajikan pada persamaan 2.

$$TF * IDF(d, t) = TF(d, t) * \log \frac{N}{df(t)} \quad (2)$$

TF(d, t) - *Term Frequency* (Frekuensi Kemunculan *Term*), TF mengukur seberapa sering *term t* muncul dalam dokumen *d.*, IDF(t) - *Inverse Document Frequency* (Frekuensi Dokumen Terbalik), IDF mengukur seberapa jarang *term t* muncul dalam semua dokumen di korpus, N Jumlah total dokumen dalam korpus. df(t) Jumlah dokumen yang mengandung *term t*.

2.7 Data Splitting

Langkah selanjutnya dalam penelitian ini adalah membagi data menjadi data latih (*training data*) dan data uji (*testing data*), dengan proporsi 80% pelatihan dan 20% uji. Data latih digunakan untuk melatih model dengan menyesuaikan parameter agar dapat membuat prediksi atau klasifikasi secara akurat, Sementara itu, data uji digunakan untuk menilai kinerja model pada data yang belum pernah dilihat sebelumnya[30]. Strategi bertujuan untuk memastikan model dapat menggeneralisasi data dengan baik, menghindari overfitting, juga tetap efektif saat menangani data baru.

2.8 Classification Model

2.8.1 Naïve Bayes

Naive Bayes classifier yaitu algoritma klasifikasi yang berlandaskan kepada *teorema Bayes* dan mengasumsikan independensi antar fitur. Algoritma ini terkenal karena kesederhanaannya, kecepatan eksekusinya, dan sering digunakan dalam berbagai aplikasi seperti analisis teks, deteksi spam, dan sistem rekomendasi. Dalam proses klasifikasinya, *Naive Bayes* menghitung *probabilitas* kelas berdasarkan distribusi fitur yang ada[31] Algoritma ini menggunakan *teorema Bayes* dengan asumsi bahwa semua atribut bersifat independen, yang berarti setiap fitur dianggap tidak saling berinteraksi, meskipun dalam kenyataannya mungkin ada hubungan antar fitur. Meskipun sederhana, metode ini sangat efektif dan sering digunakan sebagai tolok ukur untuk membandingkan performa model klasifikasi lainnya. Dalam implementasinya, *Naive Bayes* mengestimasi *probabilitas* awal (*prior probability*) dari data pelatihan dan menentukan *probabilitas* untuk setiap parameter berdasarkan nilai *prior* tersebut[32]. Secara matematis algoritma ini dapat dituliskan seperti pada persamaan 3.

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

P(H|X) merupakan *probabilitas hipotesis H* (kelas) diberikan data X (fitur), P(X|H) merupakan *probabilitas likelihood* dari fitur X diberikan *hipotesis H*. P(H) merupakan probabilitas apriori dari *hipotesis H*. P(X) adalah *probabilitas* total dari data X, yang biasanya dianggap konstan dan dapat diabaikan dalam perhitungan jika kita membandingkan beberapa kelas.

2.8.2 Support Vector Machine (SVM)

Algoritma *Support Vector Machine* (SVM) yaitu metode *supervised* yang mengklasifikasikan data ke dalam kelas tertentu. SVM bekerja dengan memetakan data pelatihan yang telah dilabeli ke dalam ruang vektor, sehingga jarak antar kelas menjadi lebih besar. Penerapan model SVM memerlukan pemilihan parameter yang tepat, karena keakuratan model sangat bergantung pada nilai parameter tersebut. Oleh karena itu, untuk mendapatkan nilai optimal, teknik *grid search* sering digunakan[33]. Tiga jenis kernel dapat digunakan oleh SVM: *radial*, *linier*, dan *sigmoid*. Dalam penelitian ini Persamaan 4 dan 5 digunakan untuk kernel linier.

$$(w \cdot x_i + b) \leq 1, y_i = -1 \tag{4}$$

$$(w \cdot x_i + b) \geq 1, y_i = 1 \tag{5}$$

Persamaan tersebut merupakan bagian dari algoritma *Support Vector Machine* (SVM) yang digunakan untuk memisahkan dua kelas data dengan margin terbesar. Persamaan $(w \cdot x_i + b) \leq 1, y_i = -1$ mengatur data kelas negatif ($y_i = -1$) untuk berada di satu sisi *hyperplane*, sementara persamaan $(w \cdot x_i + b) \geq 1, y_i = 1$ mengatur data kelas positif ($y_i = 1$) di sisi lainnya. Tujuan SVM adalah menemukan *hyperplane* yang memaksimalkan jarak (*margin*) antara kedua kelas tersebut.

2.9 Evaluation Model

Pengujian model diperlukan untuk menilai kinerja metode *Naive Bayes classifier* dan *Support Vector Machine* (SVM). Penelitian ini menggunakan metrik akurasi untuk menilai sejauh mana model klasifikasi atau prediksi mampu menghasilkan hasil yang benar berdasarkan data yang tersedia. Metrik akurasi menunjukkan persentase keberhasilan model dalam memprediksi kelas atau label yang tepat[34]. Rumus-rumus berikut digunakan untuk menghitung *accuracy*, *precision*, dan *recall* berdasarkan nilai-nilai *confusion matrix* :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{6}$$

$$Presisi = \frac{TP}{TP+FP} \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$f1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{9}$$

Dimana *True Positive (TP)* jumlah data yang berhasil diidentifikasi dengan baik sebagai positif oleh model, *False Positive (FP)* Jumlah data yang model salah mengklasifikasikan sebagai positif, *True Negative (TN)* jumlah data yang berhasil diidentifikasi dengan benar sebagai negative oleh model, *False Negative (FN)* Jumlah data yang model salah mengklasifikasikan sebagai *negative*.

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

Data dikumpulkan melalui metode *crawling* dengan memanfaatkan *library Tweet Harvest* di *Google Colab* dan *API Token Twitter*. *API Token* digunakan untuk mengautentikasi aplikasi atau skrip yang mengakses data dari *Twitter* melalui *API* resmi. Hasil pengumpulan data dirangkum pada Tabel 1.

Tabel 1. Hasil pengumpulan data.

No	Username	Tweet
1	flynh__	Clash Of Champions bukan hanya ajang kompetisi antara mahasiswa tetapi juga menghibur memotivasi dan inspirasi untuk banyak orang dan juga mengajarkan apa itu arti persahabatan Terimakasih untuk para cast CoC dan Tim Ruang Guru kalian luar biasa #ruanguru #ThankYouCoC https://t.co/A5xbhrU1aF
7903	grosyeux_	Sumpah dah itu yg komenin class of champions kok gak gini kok gak gitu diem dahhh Kalo pengen sesuai yg kamu mau bikin sendiri aja njirr Mbok dinikmati aja gitu loh acaranyaaa hihhh Wong yg university war aja org korea juga gaada tuh yg komen Hobi bgt org indo nyinyir

Bedasarkan Tabel 1 yang menunjukkan hasil pengumpulan data, yang diakumulasikan berupa *tweet* berbahasa Indonesia yang memiliki kata kunci "clash of champions" selama periode 1 Agustus 2024 hingga 30 November 2024. Sebanyak 7.903 *tweet* berhasil dikumpulkan dan disimpan dalam format CSV untuk tahap *preprocessing* selanjutnya.

3.2 Preprocessing

Dari periode Agustus hingga November 2024, sebanyak 7.903 *tweet* berhasil dikumpulkan. Selanjutnya, data melalui serangkaian proses *preprocessing* untuk mempersiapkannya sebelum dianalisis. Pada tahap ini, data akan menjadi terorganisir, bersih, juga sudah bisa untuk dianalisis setelah melewati seluruh tahapan. Dapat dilihat pada tabel 2.

Tabel 2. Hasil *Preprocessing*

Tahapan	Tweet
Data Tweet	Clash Of Champions bukan hanya ajang kompetisi antara mahasiswa tetapi juga menghibur memotivasi dan inspirasi untuk banyak orang dan juga mengajarkan apa itu arti persahabatan Terimakasih untuk para cast CoC dan Tim Ruang Guru kalian luar biasa #ruanguru #ThankYouCoC https://t.co/A5xbhrU1aF
Data Cleaning	Clash OF Champions bukan hanya ajang kompetisi antara mahasiswa tetapi juga menghibur memotivasi dan inspirasi untuk banyak orang dan juga mengajarkan apa itu arti persahabatan Terimakasih untuk para cast CoC dan Tim Ruang Guru kalian luar biasa
Case Folding	clash of champions bukan hanya ajang kompetisi antara mahasiswa tetapi juga menghibur memotivasi dan inspirasi untuk banyak orang dan juga mengajarkan apa itu arti persahabatan terimakasih untuk para cast coc dan tim ruang guru kalian luar biasa
Tokenizing	['clash', 'of', 'champions', 'bukan', 'hanya', 'ajang', 'kompetisi', 'antara', 'mahasiswa', 'tetapi', 'juga', 'menghibur', 'memotivasi', 'dan', 'inspirasi', 'untuk', 'banyak', 'orang', 'dan', 'juga', 'mengajarkan', 'apa', 'itu', 'arti', 'persahabatan', 'terimakasih', 'untuk', 'para', 'cast', 'coc', 'dan', 'tim', 'ruang', 'guru', 'kalian', 'luar', 'biasa']
Filtering	['clash', 'of', 'champions', 'ajang', 'kompetisi', 'mahasiswa', 'menghibur', 'memotivasi', 'inspirasi', 'banyak', 'mengajarkan', 'apa', 'arti', 'persahabatan', 'terimakasih', 'cast', 'coc', 'tim', 'ruang', 'guru', 'kalian', 'luar', 'biasa']
Stemming	clash of champions ajang kompetisi mahasiswa hibur motivasi inspirasi banyak ajar apa arti sahabat terimakasih cast coc tim ruang guru kalian luar biasa

Berdasarkan Tabel 2 yang menunjukkan hasil proses *preprocessing* data *tweet* menjadi lebih bersih dan siap untuk analisis. Data *Tweet* awalnya berupa teks mentah dengan link dan *hashtag* seperti: "*Clash Of Champions* bukan hanya...#ruanguru #ThankYouCoC <https://t.co/A5xbhrU1aF>." Pada tahap *Data Cleaning*, *link* dan *hashtag* dihapus, menghasilkan teks: "*Clash Of Champions* bukan hanya...luar biasa." Selanjutnya, *Case Folding* mengubah teks menjadi huruf kecil: "*clash of champions*...luar biasa." Pada *Tokenizing*, teks dipecah menjadi kata-kata: ['*clash*', '*of*', '*champions*', ... 'luar', 'biasa']. Tahap *Filtering* menghapus *stopwords* seperti "dan" dan "itu," menyisakan kata-kata penting: ['*clash*', '*champions*', 'ajang', 'kompetisi', ... 'luar', 'biasa']. Terakhir, *Stemming* mengubah kata berimbuhan ke bentuk dasar, seperti "menghibur" menjadi "hibur," menghasilkan teks ringkas: "*clash of champions* ajang kompetisi mahasiswa hibur motivasi inspirasi banyak ajar arti sahabat terimakasih cast coc tim ruang guru kalian luar biasa." Proses ini membuat teks lebih fokus dan siap untuk analisis lebih lanjut.

3.3 Labeling

Setelah melalui tahap pemrosesan data, sebanyak 5.226 *tweet* berhasil disaring menjadi data yang bersih dan siap untuk dianalisis. Selanjutnya, setiap *tweet* dilabeli menggunakan *library TextBlob* berdasarkan analisis sentimen Dapat dilihat pada Tabel 3 berikut untuk hasil pelabelan:

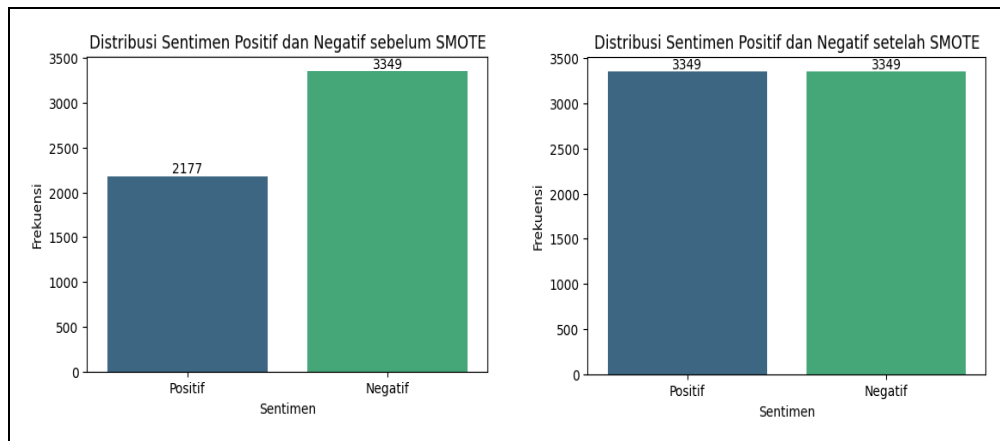
Tabel 3. Hasil Labeling

No	Username	Tweet	Sentimen Score	label
1	flynh__	clash of champions ajang kompetisi mahasiswa hibur motivasi inspirasi banyak ajar apa arti sahabat terimakasih cast coc tim ruang guru kalian luar biasa.	5	Positif
5226	grosyeux_	sumpah deh komenin class of champions kok begini kok begitu diem dahhh kalau sesuai kamu mau buat sendiri njir mbok dinikmati begitu loh acaranyaaaa hihhh wong yang university perang orang korea tidak uang yang komen hobi banget orang indo nyinyir	-4	Negatif

Berdasarkan Tabel 3, proses pelabelan dilakukan dengan menghitung nilai *polaritas* dalam rentang -1 hingga 1. Nilai di atas 0 dianggap positif, sementara di bawah 0 dianggap negatif. *Tweet* pertama dari pengguna flynh__ memiliki skor 5, menunjukkan sentimen Positif, dengan apresiasi terhadap *Clash of Champions*. *Tweet* kedua dari pengguna grosyeux_ memiliki skor -4, menunjukkan sentimen Negatif, berisi kritik terhadap komentar negatif tentang acara tersebut. Hasil pelabelan ini menghasilkan 2.177 *tweet* yang berlabel positif dan 3.349 *tweet* dengan label negatif.

3.4 SMOTE

Hasil awal analisis menunjukkan bahwa terdapat 2.177 *tweet* positif dan 3.349 *tweet* negatif. Yang menjukan ketidakseimbangan data *mayoritas* dan *minoritas*, Untuk mengatasi ketidakseimbangan data ini, digunakan teknik SMOTE. Hasil perbandingan distribusi sentimen dan dilihat pada gambar 2.



Gambar 2. Perbandingan Sentimen penggunaan teknik SMOTE

Gambar 2 menunjukkan perbandingan sebelum dan setelah penerapan SMOTE. Sebelum penerapan metode SMOTE, diagram batang memperlihatkan perbedaan signifikan dalam distribusi sentimen, dengan *tweet* sentimen *minoritas* sebanyak 2.177 dan *tweet* sentimen *mayoritas* sebanyak 3.349. Ketidakseimbangan ini dapat menyebabkan bias pada model klasifikasi, di mana model lebih cenderung memprediksi kelas *mayoritas*, dengan akurasi yang lebih tinggi, namun kurang efektif dalam memprediksi kelas *minoritas*. Setelah penerapan SMOTE, jumlah *tweet* dalam kedua kategori menjadi seimbang, masing-masing 3.349. Penyeimbangan ini penting untuk meningkatkan akurasi analisis sentimen secara keseluruhan, karena model dapat belajar secara adil dari kedua kelas, mengurangi bias, dan meningkatkan kinerja prediksi pada data *minoritas*. Penyesuaian ini memastikan bahwa model klasifikasi memberikan hasil yang lebih andal dan representatif.

3.5 Hasil Klasifikasi Model

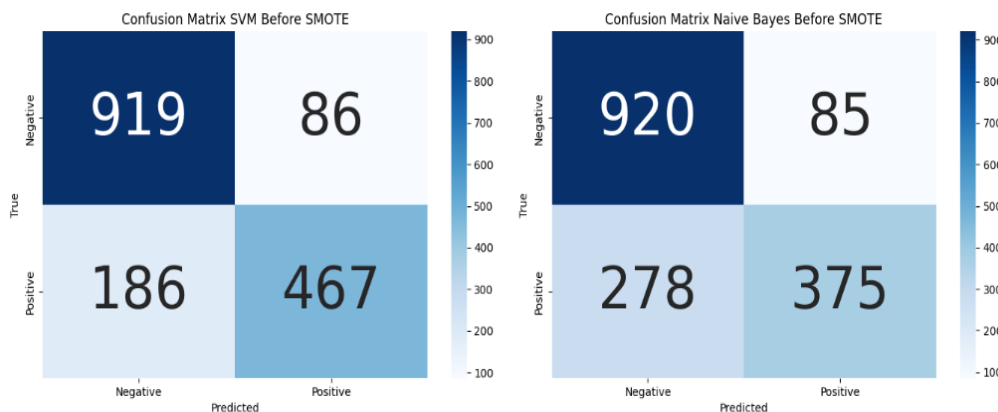
Perbandingan dilakukan antara algoritma SVM dan *Naive Bayes* dengan pembagian data training sebesar 80% dan data testing sebesar 20%. Evaluasi ini dilaksanakan dalam dua kondisi, yaitu sebelum dan setelah penerapan SMOTE. Proses evaluasi tersebut didasarkan pada data yang telah dikumpulkan dan siap untuk dianalisis. Tabel 4 berikut ini menunjukkan skor *precision*, *accuracy*, *recall*, dan *F1*.

Tabel 4. Perbandingan Klasifikasi Sebelum dan Setelah SMOTE

Matriks	NB	NB Setelah SMOTE	SVM	SVM Setelah SMOTE
Accuracy	78%	79%	82%	84%
Sentimen Negatif				
Precision	77%	83%	83%	85%
Recall	92%	82%	91%	88%
F1-Score	84%	82%	87%	87%
Sentimen Positif				
Precision	82%	72%	84%	81%
Recall	57%	73%	72%	77%
F1-Score	67%	73%	77%	79%

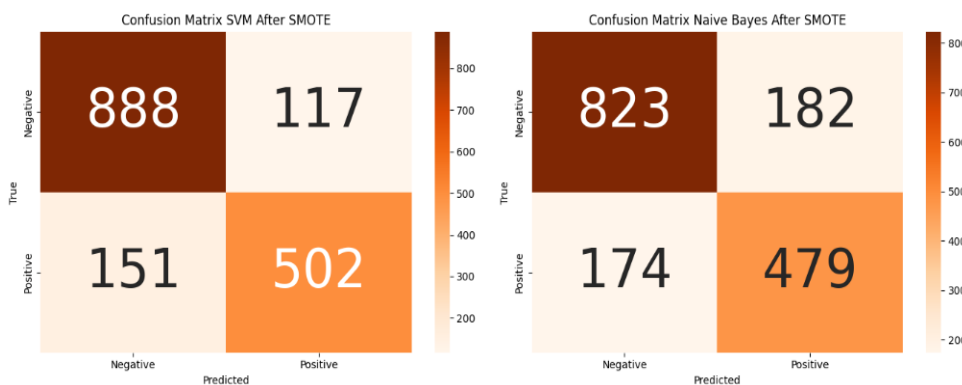
Hasil pengujian menunjukkan bahwa kedua model menghasilkan performa yang memuaskan. Sebelum optimasi SMOTE, algoritma *Naive Bayes* mencatat akurasi 78% dengan nilai *precision* 77% untuk sentimen negatif dan 82% untuk sentimen positif, *recall* 92% untuk negatif dan 57% untuk positif, serta *f1-score* 84% untuk negatif dan 67% untuk positif. Sementara itu, SVM mencapai akurasi 82% dengan nilai *precision* 83% untuk sentimen negatif dan 84% untuk sentimen positif, *recall* 91% untuk negatif dan 72% untuk positif, serta *f1-score* 87% untuk negatif dan 77% untuk positif. Setelah implementasi SMOTE, kedua model menunjukkan peningkatan performa yang signifikan. Algoritma *Naive Bayes* menunjukkan perbaikan setelah optimasi SMOTE dengan peningkatan akurasi menjadi 79%. Nilai *precision* menjadi 83% untuk sentimen negatif dan 72% untuk sentimen positif, *recall* mencapai 82% untuk negatif dan 73% untuk positif, serta *f1-score* sebesar 82% untuk negatif dan 73% untuk positif. Di sisi lain, algoritma SVM mengalami peningkatan akurasi menjadi 84%, dengan peningkatan nilai *precision* menjadi 85% untuk sentimen negatif dan 81% untuk sentimen positif. Nilai *recall* juga meningkat menjadi 88% untuk negatif dan 77% untuk positif,

diikuti dengan peningkatan *f1-score* menjadi 87% untuk negatif dan 79% untuk positif. Selain itu, penelitian ini akan melakukan perbandingan nilai *confusion matrix* untuk mengetahui seberapa baik kedua algoritma mengklasifikasikan data dan mengidentifikasi kesalahan yang terjadi. Dapat dilihat pada gambar 3 dan 4.



Gambar 3. Confusion matrix sebelum SMOTE

Pada Gambar 3 terlihat bahwa analisis *confusion matrix* memberikan wawasan lebih mendalam tentang kemampuan klasifikasi kedua model. Kedua algoritma menunjukkan kecenderungan yang kuat dalam mengidentifikasi sentimen negatif hal ini dilihat sebelum optimalisasi SMOTE dilakukan. SVM mencatat *True Negatif* sebanyak 919, *False Negatif* 186, sementara *True Positif* 467 dan *False Positif* 86. Naive Bayes memiliki distribusi serupa dengan *True Negatif* 920 dan *False Negatif* 278, serta *True Positif* 375 dan *False Positif* 85.



Gambar 4. Confusion matrix sesudah SMOTE

Sesudah penerapan SMOTE yang ditunjukkan pada gambar 4, terjadi peningkatan substansial dalam kemampuan klasifikasi kedua model. SVM menunjukkan peningkatan performa dengan *True Negatif* 888 dan *False Negatif* 151, serta *True Positif* 502 dan *False Positif* 117. Naive Bayes juga menampilkan perbaikan dengan *True Negatif* 823 dan *False Negatif* 174, diikuti *True Positif* 479 dan *False Positif* 182. Hasil ini mengindikasikan bahwa optimasi SMOTE berhasil meningkatkan keseimbangan dan akurasi klasifikasi untuk kedua algoritma, dengan SVM menunjukkan performa yang sedikit lebih unggul secara keseluruhan. Hasil optimasi SMOTE terbukti efektif dalam meningkatkan kemampuan kedua model untuk mengenali kelas positif. SVM menunjukkan performa yang terlihat unggul jika membandingkannya dengan Naive Bayes, terlihat dari tingkat keseksamaan SVM yang menjangkau hingga 83.83% dibandingkan Naive Bayes 78.53%. SVM juga menunjukkan kesanggupan yang lebih baik dalam menyeimbangkan prediksi antara kelas positif dan negatif, serta peningkatan yang lebih signifikan dalam *True Positive* setelah implementasi SMOTE.

3.6 Hasil Visualisasi WordCloud

Dalam eksperimen yang telah dilakukan, visualisasi *wordcloud* juga digunakan untuk menganalisis data teks. Teks ditampilkan dalam bentuk *cloud* kata, Dengan istilah yang paling sering muncul ditampilkan dengan ukuran visual yang lebih besar, sementara kata-kata yang jarang muncul tampil lebih kecil. Kata-kata seperti "champion" dan "clashofchampion", istilah yang paling sering muncul dalam teks akan ditampilkan dengan ukuran dan warna yang disesuaikan dengan frekuensi kemunculannya.



- [7] Y. Caesar, I. Sabastian, A. Kindarto, and A. Fathurrohman, “Analisis Sentiment Masyarakat Terhadap Clash of Champions Ruang Guru Menggunakan Metode Support Vector Machine (SVM),” *Prosiding, Seminar Nasional Unimus*, pp. 820–838, 2024.
- [8] A. Fauzi and A. H. Yunial, “Analisis Sentimen US Airline Pada Media Sosial Twitter/X Menggunakan Perbandingan Algoritma Data Mining,” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 10, no. 2, p. 277, 2024, doi: 10.26418/jp.v10i2.76024.
- [9] E. F. Baharsyah, A. Armanto, T. H. B. Aviani, and C. Wulandari, “Analisis Sentimen Pengguna Terhadap Aplikasi Belajar Online Ruang Guru Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes dan Support Vector Machine,” *Innovative: Journal Of Social Science Research*, vol. 4, no. 3, pp. 2965–2979, 2024, [Online]. Available: <https://j-innovative.org/index.php/Innovative/article/view/10769>
- [10] I. Novitasari, T. B. Kurniawan, D. A. Dewi, and Misinem, “Analisis sentimen masyarakat terhadap tweet ruang guru menggunakan algoritma naive bayes classifier (NBC) [Analysis of public sentiment towards ruang guru’s tweets using the Naive Bayes Classifier (NBC) algorithm],” *Jurnal Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022.
- [11] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *Jurnal Asimetrik: Jurnal Ilmiah Rekayasa & Inovasi*, vol. 4, pp. 239–248, 2022, doi: 10.35814/asimetrik.v4i1.3577.
- [12] N. Nurdin, L. Jama, T. Z. Magnus, R. Priskila, and V. H. Pranatawijaya, “Analisis Sentimen Dampak Artificial Intelligence (AI) Untuk Pendidikan Pada X Menggunakan Naïve Bayes,” *Jurnal Informatika Upgris*, vol. 10, no. 1, pp. 15–19, 2024, doi: 10.26877/jiu.v10i1.18867.
- [13] A. A. Rashifa, H. Marcos, P. Subarkah, and S. A. Sholikhatin, “Comparison of Svm and Naïve Bayes Classifier Algorithms on Student Interest in Joining Msib,” *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 10, no. 1, pp. 116–123, 2024, doi: 10.33480/jitk.v10i1.5270.
- [14] S. Pokhrel, “Analisis Sentimen Terhadap Iphone 16 Pada Data Twitter Menggunakan Orange Data Mining,” *Kohesi: Jurnal Multidisiplin Saintek*, vol. 15, no. 1, pp. 37–48, 2024.
- [15] S. Z. Rozaan, M. R. Andrianto, R. S. Purnama, N. I. Ramadhan, and W. Putra, “Analisis Sentimen terhadap Kenaikan UKT di Indonesia pasca Terpilihnya Capres 02 menggunakan VADER,” *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 4, pp. 88–92, 2024.
- [16] J. Sosial, B. Di, and K. Pertanahan, “Analisis Sentimen Respons Twitter Terhadap Persyaratan Badan Penyelenggara Jaminan Sosial (Bpjs) Di Kantor Pertanahan,” *Jurnal widya bhumi*, vol. 3, no. 2, pp. 113–136, 2023.
- [17] A. A. Syam, G. H. M. A. Salim, D. F. Surianto, and M. F. B, “Analisis teknik preprocessing pada sentimen masyarakat terkait konflik israel-palestina menggunakan support vector machine,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* vol. 9, no. 3, pp. 1464–1472, 2024, doi.org/10.29100/jupi.v9i3.5527
- [18] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [19] H. Syah and A. Witanti, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm),” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 5, no. 1, pp. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.
- [20] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, “Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily,” *eProceedings of Engineering*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [21] A. D. Adhi Putra, “Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 636–646, 2021.
- [22] O. I. Gifari, Muh. Adha, F. Freddy, and F. F. S. Durrand, “Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine,” *Journal of Information Technology*, vol. 2, no. 1, pp. 36–40, 2022.
- [23] Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, and Michael Indrawan, “Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z,” *Journal of Applied Computer Science and Technology*, vol. 5, no. 1, pp. 16–25, 2024, doi: 10.52158/jacost.v5i1.715.
- [24] E. A. Lisangan, A. Gormantara, and R. Y. Carolus, “Implementasi Naive Bayes pada Analisis Sentimen Opini Masyarakat di Twitter Terhadap Kondisi New Normal di Indonesia,” *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, pp. 23–32, 2022, doi: 10.24002/konstelasi.v2i1.5609.
- [25] A. Baita, Y. Pristyanto, and N. Cahyono, “Analisis Sentimen Mengenai Vaksin Sinovac Menggunakan Algoritma Support Vector Machine (Svm) Dan K-Nearest Neighbor (Knn),” *Infos*, vol. 4, no. 2, pp. 42–42, 2021.
- [26] A. Setiawan and R. R. Suryono, “Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 183–192, 2024, doi: 10.29408/edumatic.v8i1.25667.
- [27] R. R. S. Putri Kumala Sari, “Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse,” *Jurnal MNEMONIC*, vol. 7, no. 1, pp. 31–39, 2024.
- [28] A. Nofandi, N. Y. Setiawan, and D. W. Brata, “Analisis sentimen ulasan pelanggan dengan Metode Support Vector Machine (SVM) untuk peningkatan kualitas layanan pada Restoran Warung Wareg,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 1, pp. 458–466, 2023, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12218>
- [29] M. F. N. Ilham, K. D. Annurrahma, P. Wirayuda, and R. Rudiman, “Analisis Kepuasan Pengguna Aplikasi Donorku Dengan Pendekatan Metode Random Forest Dengan Smote”, *Jurnal Informatika Teknologi Dan Sains (JINTEKS)*, vol. 6, no. 3, pp. 508–513, Aug. 2024, doi.org/10.51401/jinteks.v6i3.4229.
- [30] E. A. Putri, “Penerapan Algoritma Naïve Bayes pada Analisis Sentimen Aplikasi Traveloka pada Platform Playstore,” vol. 6, no. 3, pp. 1467–1476, 2024, doi: 10.47065/bits.v6i3.6130.
- [31] Eskiyaturrofikoh and R. R. ’Suryono, “Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm),” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9,



- no. 3, pp. 1408–1419, 2024, [Online]. Available:
<https://www.jurnal.stkipggritulungagung.ac.id/index.php/jipi/article/view/5392>
- [32] A. Y. P. Yusuf and R. Sari, “Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Pemahaman Program MBKM Bagi Mahasiswa,” *Journal of Informatic and Information Security*, vol. 3, no. 2, pp. 171–180, 2022, doi: 10.31599/jiforty.v3i2.1713.
- [33] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, “Comparison of Random Forest, Naïve Bayes, and Support Vector Machine Algorithms in Analyzing Twitter Sentiment Regarding Public Opinion on the Removal of Honorary Employees,” *Seminar Nasional Official Statistics*, vol. 2022, no. 1, pp. 563–572, 2022.
- [34] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, “Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.