Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 5, No. 1, Januari 2025, Hal. 39-51
DOI: https://doi.org/10.52436/1.jpti.556
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Analisis Sentimen *Game Show Clash of Champions* Ruangguru dengan Algoritma KNN dan SVM

Elsa Apriani[1], Nunik Pratiwi[*2]

[1,2]Teknik Informatika, Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. DR. HAMKA, Indonesia
Email: [1]elsaapriani14@gmail.com, [2]npratiwi@uhamka.ac.id

Abstrak

Penelitian ini menganalisis sentimen publik terhadap acara *Clash of Champions (CoC)* oleh Ruangguru menggunakan algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)*. Data dianalisis melalui proses evaluasi menggunakan *Confusion Matrix* dengan metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-Score*. Hasil menunjukkan bahwa algoritma KNN memiliki *accuracy* tertinggi sebesar 74.01%, sedangkan SVM memiliki *accuracy* 73.68%. Dengan performa yang lebih stabil, KNN terbukti lebih unggul dalam mendeteksi sentimen positif dan negatif dibandingkan SVM. Penelitian ini menunjukkan bahwa algoritma KNN lebih efektif untuk analisis sentimen pada acara edukatif seperti CoC. Hasil ini diharapkan dapat memberikan masukan bagi pengembang program untuk meningkatkan kualitas acara di masa depan.

Kata kunci: *Analisis Sentimen, Clash of Champions, K-Nearest Neighbor, Ruangguru, Support Vector Machine*

Sentiment Analysis of the Clash of Champions Game Show by Ruangguru Using the KNN and SVM Algorithms

Abstract

This study analyzes public sentiment towards the Clash of Champions (CoC) event by Ruangguru using the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms. The data was analyzed through an evaluation process using the Confusion Matrix with evaluation metrics in the form of accuracy, precision, recall, and F1-Score. The results show that the KNN algorithm has the highest accuracy of 74.01%, while SVM has an accuracy of 73.68%. With more stable performance, KNN is proven to be superior in detecting positive and negative sentiments compared to SVM. This study shows that the KNN algorithm is more effective for sentiment analysis on educational events such as CoC. These results are expected to provide input for program developers to improve the quality of future events.

Keywords: *Clash of Champions, K-Nearest Neighbor, Ruangguru, Sentiment Analysis, Support Vector Machine*

1. PENDAHULUAN

Di era digital saat ini, perkembangan teknologi informasi telah berkembang pesat. Teknologi adalah kumpulan alat, metode, dan proses yang dirancang untuk memudahkan atau meningkatkan efektivitas berbagai aktivitas manusia [1]. Informasi adalah hasil pengolahan data yang memiliki makna dan dapat digunakan untuk membuat keputusan, menyelesaikan masalah, atau memahami suatu situasi [2]. Secara umum, ilmu ini dapat dipahami sebagai pengetahuan yang mencakup pengembangan dan penerapan berbagai temuan ilmiah untuk menciptakan teknologi yang memberikan manfaat bagi manusia.

Dunia pendidikan juga dipengaruhi oleh kemajuan teknologi informasi, seperti dengan munculnya *platform e-learning* yang memudahkan siswa dan pendidik untuk belajar secara daring. Pembelajaran jarak jauh sekarang lebih mudah daripada pembelajaran tatap muka karena *e-learning* telah menjadi bagian dari kurikulum. E-learning memungkinkan pelaksanaan pembelajaran jarak jauh, sehingga menjadi aspek yang sangat penting dalam dunia pendidikan, terutama selama masa pandemi COVID-19 [3]. Salah satu bentuk inovasi dalam pendidikan adalah pengembangan *platform e-learning*, seperti Ruangguru, yang menawarkan berbagai fitur untuk mendukung proses belajar mengajar. Salah satu fitur utamanya adalah video pembelajaran yang dibawakan oleh guru-guru yang dilatih dengan baik, dilengkapi dengan animasi untuk memperjelas topik yang dibahas [4].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Ruangguru merupakan salah satu *platform* belajar online terkemuka di Indonesia yang menawarkan berbagai fitur yang dirancang untuk mendukung proses belajar mengajar secara efektif dan efisien [5]. Fitur-fitur ini mencakup video pembelajaran interaktif, latihan soal, ujian online, serta sesi tanya jawab dengan tutor. Dengan memanfaatkan teknologi terkini, Ruangguru berupaya untuk menyediakan pengalaman belajar yang adaptif dan personal sesuai dengan kebutuhan masing-masing siswa. Salah satu inovasi terbaru dari Ruangguru adalah program *Game Show Clash of Champions (CoC)*.

Clash of Champions (CoC) adalah *game show* yang diadakan oleh Ruangguru yang melibatkan 40 mahasiswa berprestasi dari berbagai universitas ternama di Indonesia [6]. Peserta dari berbagai universitas berkompetisi secara individu dan kelompok dalam bidang matematika, deduksi, dan hafalan [7]. Program *Game Show Clash of Champions (CoC)* dirancang untuk menguji pengetahuan dan keterampilan mahasiswa dalam suasana yang menyenangkan dan kompetitif [8]. Acara ini tidak hanya menjadi sarana hiburan, tetapi juga diharapkan dapat meningkatkan motivasi belajar. Dengan tingginya tingkat partisipasi dan antusiasme masyarakat terhadap program ini, penting untuk memahami bagaimana opini masyarakat terhadap program tersebut. Analisis sentimen terhadap opini masyarakat dapat memberikan wawasan yang berharga bagi pengembang program dan pihak terkait untuk mengevaluasi dan meningkatkan kualitas program di masa mendatang.

Analisis sentimen adalah cara untuk mengolah data teks yang belum terstruktur secara otomatis. Tujuannya adalah memahami, mengambil, dan memproses informasi untuk mengetahui pendapat atau perilaku seseorang. Dalam analisis ini, sering digunakan pendekatan seperti algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)*.

Algoritma *K-Nearest Neighbor (KNN)* adalah algoritma yang bekerja dengan mengklasifikasikan data baru berdasarkan mayoritas tetangga terdekatnya. Proses ini dilakukan dengan menghitung jarak antara data baru dengan data yang ada di dalam dataset. Kelebihan *K-Nearest Neighbor (KNN)* adalah kesederhanaan dan fleksibilitasnya, tetapi algoritma ini kurang efisien untuk dataset besar karena memerlukan perhitungan jarak untuk setiap data baru [9]. Sementara itu, *Support Vector Machine (SVM)* adalah algoritma machine learning yang berfungsi untuk menemukan hyperplane terbaik yang memisahkan data ke dalam kelas-kelas berbeda. *Support Vector Machine (SVM)* sangat efektif digunakan pada data dengan dimensi tinggi, seperti teks, karena memanfaatkan fungsi kernel untuk memetakan data ke ruang dimensi yang lebih tinggi. Hal ini membuat *Support Vector Machine (SVM)* sering memberikan tingkat akurasi yang tinggi, terutama pada dataset yang kompleks dan beragam [10].

Dalam beberapa tahun terakhir, analisis sentimen telah menjadi bidang penelitian yang sangat populer, terutama dalam memahami opini masyarakat terhadap produk atau layanan tertentu. Banyak pendekatan algoritmik yang telah digunakan untuk analisis sentimen, mulai dari algoritma klasik seperti *Naïve Bayes Classifier (NBC), K-Nearest Neighbor (KNN), Decision Tree* hingga algoritma yang lebih canggih seperti *Support Vector Machine (SVM)* dan model berbasis deep learning.

Adapun penelitian mengenai analisis sentimen dengan menggunakan *Naïve Bayes Classifier (NBC)* oleh Kevin, Margareta Enjeli, dan Andri Wijaya pada tahun 2024 terhadap ulasan aplikasi Kinemaster di *Google Play Store* menggunakan algoritma *Multinomial Naïve Bayes* menghasilkan *accuracy* 85%, *precision* 82%, *recall* 74%, dan *F1-score* 78%. Sentimen positif menunjukkan tingkat evaluasi lebih tinggi (88%) dibandingkan sentimen negatif (78%) [11].

Penelit Penelitian lain mengenai analisis sentimen dengan menggunakan *Decision Tree* oleh Maharani dan Fathoni pada tahun 2024 terharap ulasan PayPal di *Google Play Store*. Sentimen negatif mendominasi aspek produk (97,33%), harga (99,68%), dan tempat (96,72%), sedangkan aspek promosi mendapatkan sentimen positif sebesar 62,18%. Uji *5-Fold Cross Validation* dengan kriteria gain ratio menghasilkan *accuracy* tertinggi sebesar 90,69% [12].

Penelitian sebelumnya mengenai analisis sentimen dengan menggunakan Metode NLP dan Model BERT oleh Siti Helmiyah dan Arry Verdian pada tahun 2024 terhadap sentimen masyarakat terhadap *acara Clash of Champions (CoC)* dari Ruangguru menggunakan algoritma *TextBlob* dan model BERT pada 351 tweets di media sosial X. Hasilnya menunjukkan sentimen positif sebesar 85,27% *TextBlob* dan 86,12% BERT, mencerminkan antusiasme masyarakat dan keberhasilan acara dalam meningkatkan minat belajar. Sentimen negatif sangat kecil, yaitu 3,99% *TextBlob* dan 0% BERT, sebagian besar terkait tantangan teknis, memberikan masukan untuk perbaikan [13].

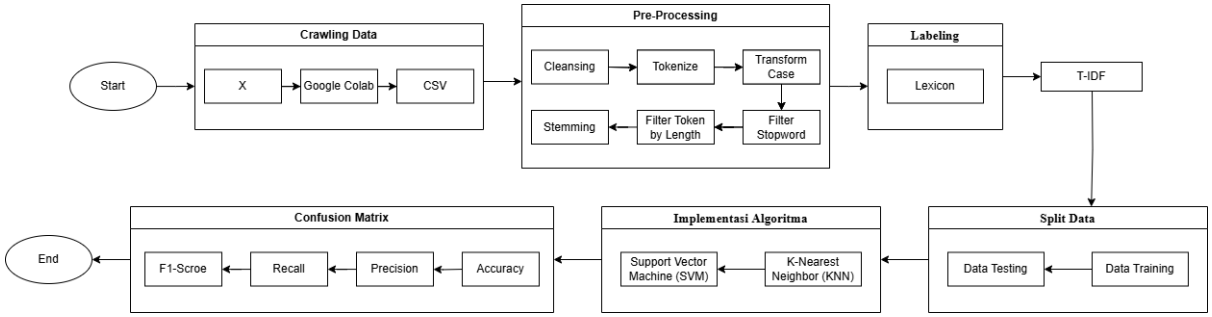
Penelitian ini berbeda dengan penelitian sebelumnya dalam hal pendekatan metodologis. Penelitian sebelumnya menggunakan algoritma BERT, sedangkan penelitian ini menggunakan algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* untuk menganalisis persepsi masyarakat terhadap program *Clash of Champions (CoC)*. Dengan demikian, penelitian ini menawarkan perspektif baru dalam analisis sentimen, menggantikan pendekatan berbasis model BERT dengan metode yang lebih sederhana namun efektif seperti KNN dan SVM. Analisis ini diharapkan dapat mengungkap kelebihan dan kekurangan program serta

p-ISSN: 2775-4227
e-ISSN: 2775-4219

memberikan masukan berharga bagi Ruangguru dalam mengembangkan dan menyempurnakan program mereka di masa depan agar lebih sesuai dengan kebutuhan dan harapan masyarakat. Dengan demikian, program edukatif seperti *Clash of Champions (CoC)* diharapkan mampu memberikan manfaat optimal bagi peningkatan kualitas pendidikan di Indonesia. Penelitian ini bertujuan untuk membandingkan efektivitas kedua algoritma dalam menganalisis sentimen publik terhadap acara tersebut. Hasil penelitian diharapkan dapat memberikan rekomendasi yang berharga bagi pengembang program untuk meningkatkan kualitas dan efektivitas acara di masa depan.

2. METODE PENELITIAN

Tahapan penelitian ini dimulai dengan pengumpulan data, di mana data diambil dari media sosial X, yang dipilih karena merupakan platform populer yang digunakan oleh banyak pengguna untuk berdiskusi dan berbagi pendapat tentang acara-acara publik, termasuk *Clash of Champions (CoC)* dari Ruangguru. Media sosial X menyediakan aliran data real-time yang sangat berguna untuk analisis sentimen, di mana pengguna sering kali menyatakan opini mereka secara langsung dalam bentuk teks yang mudah dianalisis. Oleh karena itu, X menjadi sumber data yang relevan untuk penelitian ini, karena mencerminkan persepsi publik yang dinamis terhadap acara tersebut. Data kemudian dicrawling menggunakan *Google Colab*, sebuah alat yang efisien untuk mengakses dan mengolah data dalam jumlah besar, dan disimpan dalam format CSV untuk memudahkan pengelolaan dan analisis selanjutnya. Selanjutnya, data tersebut diberi label dan diproses lebih lanjut dalam tahap *Pre-Processing* yang mencakup *Cleansing*, *Tokenize*, *Transform Cases*, *Filter Token by Length*, *Filter Stopword*, dan *Stemming*. Setelah *Pre-Processing*, data diimplementasikan menggunakan algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* dengan menggunakan fitur TF-IDF dan visualisasi *Wordcloud*. Terakhir, hasil klasifikasi dievaluasi menggunakan *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk mengukur kinerja model. Proses ini memberikan gambaran menyeluruh mengenai tahapan yang dilakukan dalam pengolahan dan klasifikasi data teks untuk menghasilkan model yang efektif. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Pada penelitian ini, dataset yang digunakan yaitu sebanyak 5000 data yang diambil dari platform X (sebelumnya dikenal sebagai Twitter) pada tanggal 16 Juni - 25 Juli 2024. Data ini dikumpulkan menggunakan *Google Collab* dan kemudian disimpan dalam format CSV. Pengumpulan data dapat dilihat pada Gambar 2.

Gambar 2. Pengumpulan Data

Google Collab mengambil data dari media sosial X dengan mengautentikasi akses ke media sosial X dan menggunakan *kode_auth* dari Twitter. Setelah terhubung, *Google Collab* dapat mengambil data tweet dengan keyword tertentu.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2.2. *Pre-Processing*

Pre-processing adalah tahap penting dalam proses klasifikasi yang dimana akan memperbaiki kata atau menghapus kata yang tidak penting. Data yang dimasukkan ke tahap ini adalah data yang asli dan belum melalui proses apa pun, sehingga hasilnya adalah data yang bersih yang memudahkan proses klasifikasi [14].

a. *Cleansing*

Cleansing merupakan proses menghilangkan karakter dari teks yang tidak termasuk alfabet, bertujuan untuk mengurangi karakter atau simbol yang tidak penting, seperti tanda baca, url, hastag (#), mention (@), emoticon, dan sebagainya [15].

b. *Tokenize*

Tokenize merupakan proses mengubah dokumen teks menjadi serangkaian token, seperti kata atau frasa, yang lebih mudah diolah oleh komputer. Tokenisasi mengubah teks menjadi kata-kata secara terpisah, membuatnya lebih mudah diproses [16].

c. *Transform Case*

Transform Case merupakan proses mengubah data menjadi huruf kecil dalam dataset untuk konsistensi format penelitian. Proses ini mengubah huruf kapital dalam dokumen menjadi huruf kecil sehingga semua data yang diperoleh berhuruf kecil [17].

d. *Filter Stopword*

Filter stopwords merupakan proses dalam analisis teks untuk mengeliminasi kata-kata yang kurang signifikan terhadap makna ulasan dalam dokumen. Kata-kata ini disebut stopword, seperti "dan", "di", "tetapi", "atau", yang tidak memiliki nilai emosional atau kontribusi penting [18].

e. *Filter Token by Length*

Filter Token by Length merupakan proses analisis teks untuk menghapus kata-kata dengan jumlah huruf minimal atau maksimal tertentu. Data yang sudah difilter berdasarkan Kaggle kemudian diimplementasikan ke model [19].

f. *Stemming*

Stemming merupakan proses mengonversi kata ke bentuk dasarnya sehingga kata-kata yang berbeda dapat dianggap sebagai kata yang sama [20].

2.3. Labeling

Pada tahap ini, proses pelabelan dilakukan secara otomatis menggunakan lexicon dengan memberikan label sentimen positif atau negatif pada setiap dataset. Metode berbasis lexicon membandingkan kata dalam teks dengan nilai polaritas di lexicon, menghitung jumlah kata positif dan negatif untuk mengklasifikasikan sentimen [21]. Pendekatan ini sederhana dan cepat tanpa memerlukan data pelatihan, tetapi memiliki keterbatasan dalam menangkap konteks, sarkasme, atau sentimen implisit.

2.4. *Term Frequency-Inverse Document Frequency (TF-IDF)*

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode pembobotan yang digunakan dalam pengolahan teks untuk mengevaluasi pentingnya suatu kata dalam sebuah dokumen relatif terhadap koleksi dokumen lainnya [22]. Nilai TF menghitung frekuensi kata dalam dokumen, sementara IDF menghitung seberapa unik kata tersebut di seluruh dokumen. Pembobotan ini efektif dalam mengurangi pengaruh kata-kata umum yang tidak relevan dalam proses analisis teks. TF-IDF sering digunakan sebagai representasi fitur dalam tugas klasifikasi teks, clustering, dan pencarian informasi.

2.5. *K-Nearest Neighbor (KNN)*

K-Nearest Neighbor (KNN) adalah algoritma klasifikasi berbasis instance yang bekerja dengan mengukur kedekatan suatu data dengan *k* data terdekat lainnya berdasarkan jarak tertentu, seperti jarak *Euclidean* [23]. Data baru diklasifikasikan berdasarkan mayoritas label dari tetangga terdekatnya. KNN bersifat non-parametrik, sehingga cocok untuk dataset dengan distribusi yang tidak diketahui. Meskipun sederhana, KNN dapat memberikan performa yang baik jika jumlah tetangga (k) dipilih dengan tepat dan data telah dinormalisasi. Pemilihan jumlah tetangga (k) yang optimal sangat penting karena terlalu sedikit tetangga dapat menyebabkan model terlalu sensitif terhadap noise, sedangkan terlalu banyak tetangga dapat mengurangi kemampuan model dalam mendeteksi pola yang lebih halus. Oleh karena itu, pemilihan k dilakukan berdasarkan eksperimen untuk mencari nilai k yang memberikan hasil terbaik dalam mengklasifikasikan sentimen. Selain itu, KNN juga menggunakan ukuran jarak *(distance metric)* untuk menentukan tetangga terdekat, dengan jarak *Euclidean* sering

p-ISSN: 2775-4227
e-ISSN: 2775-4219

digunakan, meskipun jarak *Manhattan* atau *Minkowski* juga dapat dipilih sesuai dengan karakteristik data. Rumus jarak Euclidean yang digunakan adalah sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(xi - yi)^2} \quad (1)$$

2.6. *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi dengan mencari *hyperplane* terbaik yang memisahkan kelas data [24]. *Hyperplane* dipilih untuk memaksimalkan margin, yaitu jarak antara hyperplane dan data terdekat dari masing-masing kelas. SVM efektif untuk data berdimensi tinggi karena memanfaatkan kernel trick untuk memproyeksikan data *non-linear* ke ruang yang lebih tinggi. Dalam penelitian ini, kernel yang digunakan adalah kernel linear karena sifat data yang diharapkan relatif dapat dipisahkan secara linier. Namun, kernel lain seperti polynomial atau radial basis *function* (RBF) dapat dipertimbangkan jika data membutuhkan pemisahan yang lebih kompleks. Selain kernel, parameter lain yang penting dalam SVM adalah parameter C, yang mengontrol *trade-off* antara memaksimalkan margin dan meminimalkan kesalahan klasifikasi. Nilai C yang lebih besar cenderung menghasilkan model yang lebih sensitif terhadap kesalahan, sedangkan nilai yang lebih kecil menghasilkan model yang lebih umum. Untuk kernel RBF, parameter gamma juga mempengaruhi seberapa besar pengaruh masing-masing titik data terhadap klasifikasi, dengan nilai gamma yang tinggi menyebabkan model lebih fokus pada data terdekat, berpotensi menyebabkan *overfitting*. Rumus klasifikasi SVM adalah sebagai Berikut:

$$f(x) = sgn\,(w.x + b) \quad (2)$$

2.7. *Confusion Matrix*

Confusion Matrix adalah tabel yang digunakan untuk mengevaluasi performa model klasifikasi dengan membandingkan hasil prediksi model terhadap data sebenarnya. Matriks ini terdiri dari empat komponen utama: *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)* [25]. TP adalah jumlah data positif yang terklasifikasi dengan benar, sementara FP adalah data negatif yang salah diklasifikasikan sebagai positif. TN merepresentasikan data negatif yang benar diklasifikasikan, dan FN adalah data positif yang salah diklasifikasikan sebagai negatif. Proses evaluasi model dalam penelitian ini menggunakan *Confusion Matrix* untuk menilai performa klasifikasi algoritma. Data yang digunakan dibagi menjadi dua bagian, yaitu 80% untuk pelatihan dan 20% untuk pengujian, guna memastikan model dapat belajar pola dari data pelatihan dan diuji pada data yang tidak terlihat sebelumnya. Rumus *confusion matrix* adalah sebagai berikut::

a. *Accuracy*
Accuracy merupakan rasio prediksi benar terhadap total prediksi, cocok untuk dataset seimbang.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

b. *Precision*
Precision merupakan keandalan prediksi positif, penting untuk meminimalkan kesalahan positif palsu.

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

c. *Recall*
Recall merupakan kemampuan model mendeteksi semua data positif, penting untuk mengurangi kesalahan negatif palsu.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

d. *F1-Score*
F1-Score merupakan rata-rata harmonis Precison dan Recall, ideal untuk dataset tidak seimbang.

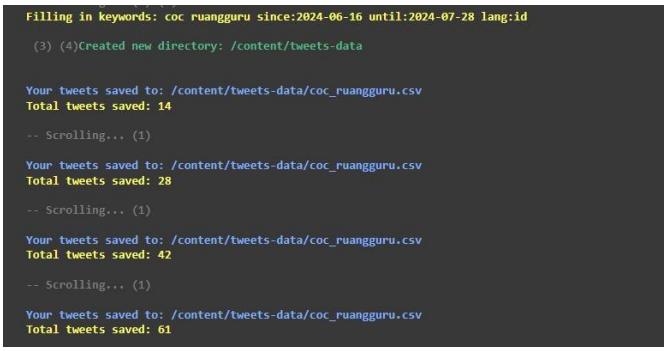
$$F1 = 2.\frac{Precision\,.\,Recall}{Precision + Recall} \quad (6)$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data dikumpulkan dengan menggunakan *Google Collab* dan bahasa pemrograman Python. Proses ini memanfaatkan library tweet-harvest untuk mengumpulkan data selama periode 16 Juni 2024 hingga 28 Juli 2024. Kata kunci yang digunakan untuk pencarian adalah "coc ruangguru." Hasil pencarian tersebut kemudian disimpan dalam sebuah file dengan nama "coc_ruangguru" dalam format CSV. Total data yang berhasil dikumpulkan yaitu berjumlah 5000 dataset. Proses pengumpulan data dapat dilihat pada Gambar 3.



Gambar 3. Proses Pengumpulan Data

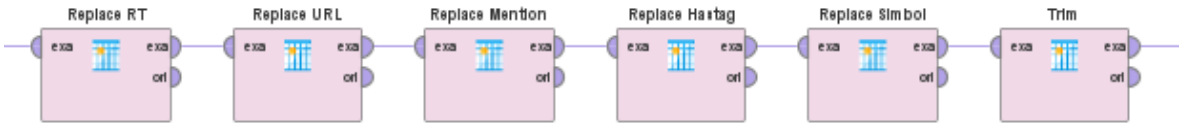
Proses pengambilan dataset pada *Tweet-Harvest* yang menggunakan *Google Collab* dan API X dimulai dengan menyiapkan dan memastikan akses ke API X. Pertama, library *Tweepy*, yang berinteraksi dengan API Twitter dan Pandas untuk mengolah data, diinstal. Setelah itu, langkah selanjutnya adalah mengautentikasi aplikasi dengan API X dengan menggunakan kunci API yang diperoleh dari akun *developer* X. Metode ini mengumpulkan data tweet dan menyimpannya dalam format CSV. Jumlah total data yang dikumpulkan dari dataset ini, yang dikumpulkan per-tanggal dari 16 Juni 2024 hingga 28 Juli 2024, berjumlah 5000 data. Setelah melakukan proses cleansing data, dataset yang berhasil dikumpulkan yaitu sebanyak 3023 data. Hasil pengumpulan data dapat dilihat pada Gambar 4.

	conversation_id_str	created_a	favorite_c	full_text	id_str	image_url	in_reply_t	lang	location	quote_co	reply_cou	retweet_c	tweet_url	user_id_s	username
2	1809609659776659784	Sat Jul 06	0	Gua lulus	1,81E+18			in		0	0	0	https://x.(	1,36E+18	leviosafa
3	1809609652151414959	Sat Jul 06	0	Bateku to	1,81E+18	https://pbs.twimg.c		in	Makassar,	0	0	0	https://x.(	8,47E+17	farhanputra___
4	1809609056895791487	Sat Jul 06	0	Rapat per	1,81E+18	https://pbs.twimg.c		in	bangkok	0	0	0	https://x.(	1,41E+18	IbrahimNiar
5	1809532154172359144	Sat Jul 06	0	@ayyacia	1,81E+18		ayyacia	in	221b Bake	0	0	0	https://x.(	1,67E+18	thizzmoriarty
6	1809608316743741476	Sat Jul 06	0	Eh demi c	1,81E+18			in	menolak1	0	0	0	https://x.(	1,3E+18	liv1us
7	1809606443966034057	Sat Jul 06	0	@ruanggu	1,81E+18		hwangwa	in	butterwol	0	0	0	https://x.(	1,34E+18	hwangwatanabe_2
8	1809607906926698771	Sat Jul 06	0	REAAAAL	1,81E+18			in	svtbz	0	1	0	https://x.(	9,86E+17	warmestwaves
9	1809606674954743850	Sat Jul 06	0	@ruanggu	1,81E+18		ruangguru	in	she	0	0	0	https://x.(	3,3E+09	acnomali
10	1809606443966034057	Sat Jul 06	11	Ini maxwe	1,81E+18	https://pbs.twimg.c		in	butterwol	0	3	1	https://x.(	1,34E+18	hwangwatanabe_2
11	1809606431718666583	Sat Jul 06	5	sama kaya	1,81E+18	https://pbs.twimg.c		in		0	0	0	https://x.(	1,51E+18	anotherchoco
12	1809606197391290817	Sat Jul 06	0	waw max(	1,81E+18			in	she/her, S	0	0	0	https://x.(	1,19E+18	anneliany
13	1809606071213777081	Sat Jul 06	0	ini tuh coc	1,81E+18			in	17! ðŸ	0	0	0	https://x.(	1,68E+18	strwbewrrys
14	1809605904830197918	Sat Jul 06	0	Pertama k	1,81E+18			in	Palestine	0	0	0	https://x.(	1,53E+18	wonwhales
15	1809605024961687768	Sat Jul 06	0	@glfamor	1,81E+18		glfamor	in	in my del(	0	1	0	https://x.(	1,41E+18	tiearsz
16	1809605464956760242	Sat Jul 06	0	ruangguru	1,81E+18			in	Bandar La(	0	0	0	https://x.(	81041917	eternallenel
17	1809603970266194022	Sat Jul 06	0	aaa tamba	1,81E+18	https://pbs.twimg.c		in		0	0	0	https://x.(	1,14E+18	cylaf7
18	1809603770642493449	Sat Jul 06	0	Pas nonto	1,81E+18			in		0	0	0	https://x.(	1,08E+18	alipsih
19	1809602859111182590	Sat Jul 06	1	@ruanggu	1,81E+18		ruangguru	in		0	0	0	https://x.(	7,43E+17	huzkies
20	1809602308478730534	Sat Jul 06	0	makasiii b	1,81E+18			in		0	0	0	https://x.(	1,54E+18	chocoflavoor
21	1809601823646577082	Sat Jul 06	0	abis nonton coc ruangguru eps 3 &amp											
22	1809601381977997420	Sat Jul 06	0	moo nang	1,81E+18	https://pbs.twimg.c		in	s/herâ"ã	0	1	0	https://x.(	1,52E+18	prodbymarsha
23	1809601082152173611	Sat Jul 06	0	Reminder	1,81E+18			in		1	1	0	https://x.(	1,65E+18	seceyess

Gambar 4. Hasil Pengumpulan Data

3.2. *Pre-Processing*

Tahapan berikutnya adalah melakukan *pre-processing* dataset dengan menggunakan *software RapidMiner*. Salah satu proses penting dalam *pre-processing* adalah data cleansing, yang bertujuan untuk menghilangkan kesalahan yang mungkin terdapat dalam data yang dikumpulkan. Pada penelitian ini, proses data cleansing dilakukan dengan menggunakan sub-operator yang terdapat di dalamnya, termasuk operator *replace* dan trim. Proses *cleansing* data dapat dilihat pada Gambar 5.



Gambar 5. Proses Cleansing Data

p-ISSN: 2775-4227
e-ISSN: 2775-4219

a. *Replace* RT

Replace RT merupakan proses yang bertujuan untuk menghilangkan kalimat yang dimulai dengan "RT@" menggunakan pola "RT @.*". Proses ini dilakukan untuk menghapus bagian teks yang diawali dengan "RT" diikuti oleh nama pengguna tertentu.

b. *Replace* URL

Replace URL merupakan proses menghapus kalimat yang dimulai dengan "https" menggunakan pola "https.*?". Ini dilakukan untuk menghapus tautan URL dari teks.

c. *Replace Mention*

Replace Mention merupakan proses menghapus kalimat yang dimulai dengan "@" menggunakan pola "@.*?". Tujuannya adalah untuk menghapus nama pengguna dari teks.

d. *Replace Hastag*

Replace Hastag merupakan proses yang menghapus kalimat yang dimulai dengan tanda pagar "#" dengan pola "#.*?". Tujuannya adalah untuk menghapus seluruh teks yang muncul setelah tanda pagar.

e. *Replace Simbol*

Replace Simbol merupakan proses yang menggunakan pola "[-!"#$%&'()*+,./:;<=>?@_`{|}~]" untuk menghapus simbol-simbol khusus dari teks. Ini dilakukan untuk menghilangkan simbol-simbol yang tidak relevan dari teks.

f. *Trim*

Trim merupakan proses yang memastikan bahwa tidak ada spasi tambahan di awal dan akhir teks, menjaga agar teks bebas dari karakter kosong yang tidak diperlukan.

Langkah pertama yang dilakukan pada tahapan pre-processing yaitu cleansing data, dan kemudian dilanjutkan proses tokenize transform cases, filter stopword, filter token by length, dan terakhir stemming. Proses pre-processing dapat dilihat pada Gambar 6.

Gambar 6. Proses Pre-Processing

a. *Cleansing*

Penghapusan karakter yang tidak relevan, seperti tanda baca dan mention (@ruangguru), sehingga teks menjadi lebih bersih dan hanya berisi kata-kata yang penting untuk analisis lebih lanjut. Hasil *cleansing* dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Cleansing*

Sebelum *Cleansing*	**Sesudah *Cleansing***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	Kasian ya jadi talent coc Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu Sekarang kalian dituntut harus jadi orang paling sempurna didunia Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya ruangguru

b. *Tokenize*

Proses pemisahan teks yang belum *ditokenize* masih berupa kalimat utuh, sementara setelah ditokenize, setiap kata dipisahkan menjadi token-token individual yang terpisah, memudahkan proses analisis dan pemrosesan lebih lanjut. Hasil *tokenize* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Tokenize*

Sebelum *Tokenize*	**Sesudah *Tokenize***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	‘Kasian’, ‘ya’, ‘jadi’, ‘talent’, ‘coc’, ‘Selamat’, ‘anda’, ‘bukan’, ‘jadi’, ‘manusia’, ‘biasa’, ‘lagi’, ‘gabakal’, ‘bisa’, ‘nafas’, ‘sebebas’, ‘dlu’, ‘ngomong’, ‘sebebas’, ‘dlu, ‘Sekarang’, ‘kalian’, ‘dituntut’, ‘harus’, ‘jadi’, ‘orang’, ‘paling’, ‘sempurna’, ‘didunia’, ‘Karna’, ‘mau’, ‘gamau’, ‘netizen’, ‘Indonesia’, ‘itu’, ‘wajib’, ‘harus’, ‘kudu’, ‘dipuaskan’, ‘egonya’, ‘ruangguru’

p-ISSN: 2775-4227
e-ISSN: 2775-4219

c. ***Transform Case***

Proses mengubah semua huruf kapital dalam teks menjadi huruf kecil untuk konsistensi format data, sehingga analisis tidak dipengaruhi oleh perbedaan kapitalisasi. Hasil *transform case* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Transform Case*

Sebelum *Transform Case*	**Sesudah *Transform Case***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	kasian ya jadi talent coc selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu sekarang kalian dituntut harus jadi orang paling sempurna didunia karna mau gamau netizen indonesia itu wajib harus kudu dipuaskan egonya ruangguru

d. ***Filter Stopword***

Penghapusan kata-kata umum yang tidak membawa makna penting (seperti "ya", "anda", "lagi", "itu", dll.), sehingga teks menjadi lebih ringkas dan fokus pada kata-kata yang relevan untuk analisis. Hasil *filter stopword* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Filter Stopword*

Sebelum *Filter Stopword*	**Sesudah *Filter Stopword***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	kasian ya talent coc selamat manusia gabakal nafas sebebas dlu ngomong sebebas dlu dituntut orang sempurna didunia karna gamau netizen indonesia wajib kudu dipuaskan egonya ruangguru

e. ***Token by Length***

Proses penghilangan kata-kata yang dianggap tidak penting atau memiliki makna kurang signifikan, seperti "ya," "jadi," "anda," "lagi," "dlu," dan "itu," sehingga hanya menyisakan kata-kata yang lebih relevan untuk analisis. Hasil *token by length* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Token by Length*

Sebelum *Token by Length*	**Sesudah *Token by Length***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	kasian talent selamat manusia gabakal nafas sebebas ngomong sebebas dituntut orang sempurna didunia karna gamau netizen indonesia wajib kudu dipuaskan egonya ruangguru

f. ***Stemming***

Proses penghilangan imbuhan dan bentuk kata yang kembali ke bentuk dasar atau akar kata. Misalnya, kata "jadi" menjadi "jadi", "ngomong" menjadi "bicara", dan kata-kata lain disederhanakan menjadi bentuk yang lebih sederhana untuk memudahkan analisis. Hasil *stemming* dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Stemming*

Sebelum *Filter Stopword*	**Sesudah *Filter Stopword***
Kasian ya jadi talent coc. Selamat anda bukan jadi manusia biasa lagi gabakal bisa nafas sebebas dlu ngomong sebebas dlu. Sekarang kalian dituntut harus jadi orang paling sempurna didunia. Karna mau gamau netizen Indonesia itu wajib harus kudu dipuaskan egonya. @ruangguru	kasihan talent selamat manusia gabakal nafas sebebas bicara sebebas dituntut orang sempurna dunia karna gamau netizen indonesia wajib kudu dipuaskan egonya ruangguru

3.3. Labeling

Data dikumpulkan dengan menggunakan *Google Collab* dan bahasa pemrograman Python. Proses ini memanfaatkan library tweet-harvest untuk mengumpulkan data selama periode 16 Juni 2024 hingga 28 Juli 2024. Kata kunci yang digunakan untuk pencarian adalah "coc ruangguru." Hasil pencarian tersebut kemudian

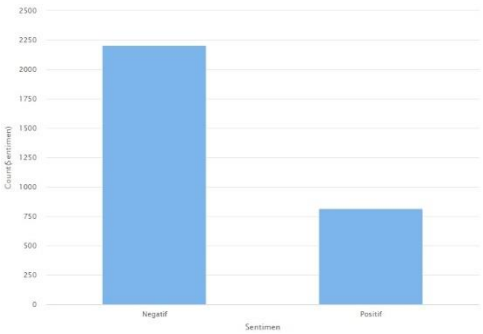
p-ISSN: 2775-4227
e-ISSN: 2775-4219

disimpan dalam sebuah file dengan nama "coc_ruangguru" dalam format CSV. Total data yang berhasil dikumpulkan yaitu berjumlah 3023 dataset. Hasil labeling dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Labeling*

Opini	Sentimen
baru nonton coc ruang guru ep 1 masya Allah keren bgt tiap kali dispill prestasi2 pesertanya gue ikut bangga bgt dan bertanya2 gimana cara parenting ortunyaaa	Positif
Miris ya skrg lagi booming acara COC ruang guru yg pesertanya anakanak jenius tapi di satu sisi yg diusung jadi pejabat daerah malah yg kyk gitu	Negatif

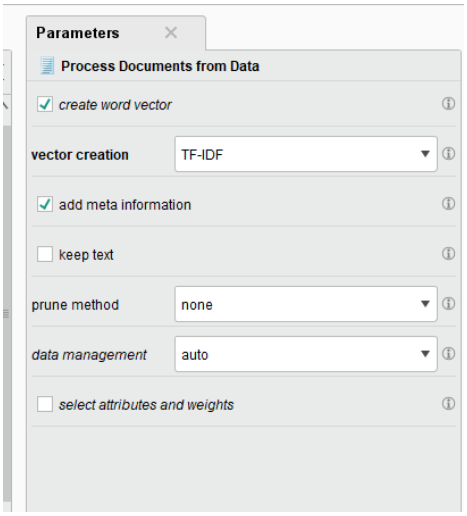
Setelah melakukan labeling, proses berikutnya menentukan presentasi sentimen positif dan negatif dari 3023 dataset. Kinerja *K-Nearest Neighbor (KNN)* yang lebih unggul dibandingkan *Support Vector Machine (SVM)* pada dataset ini disebabkan oleh sifat KNN yang adaptif terhadap distribusi data yang tidak seimbang. Dataset ini didominasi oleh sentimen negatif yaitu sebanyak 2207 data dibandingkan sentimen positif sebanyak 816 data, yang menyebabkan SVM cenderung bias terhadap kelas mayoritas, terutama dengan kernel linear yang kurang optimal untuk pola non-linear. Sebaliknya, KNN mengklasifikasikan data berdasarkan kedekatan dengan tetangga terdekat, sehingga lebih efektif dalam menangkap pola lokal. Dominasi sentimen negatif juga memengaruhi evaluasi model, di mana KNN menunjukkan performa yang lebih stabil. Presentasi sentimen positif dan negatif dapat dilihat pada Gambar 7.



Gambar 7. Presentasi Sentimen Positif dan Negatif.

3.4. *Frequency-Inverse Document Frequency (TF-IDF)*

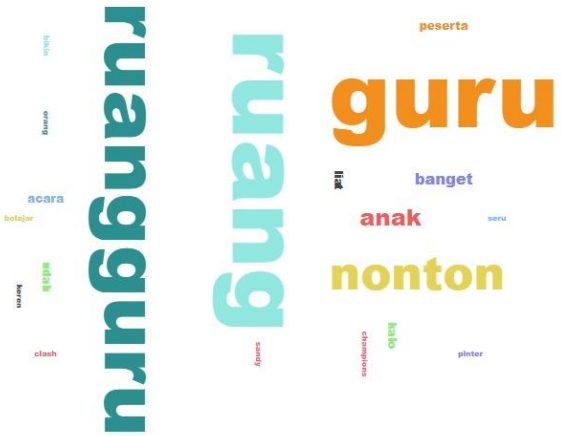
Langkah berikutnya melakukan pembobotan kata menggunakan metode TF-IDF. Semakin banyak dokumen yang diproses, semakin banyak pula fitur yang dihasilkan dalam pembobotan kata ini. Dengan TF-IDF, dapat menentukan kata-kata yang memiliki bobot tinggi dalam sebuah adalah dokumen, yang menunjukkan tingkat kepentingan kata tersebut dalam konteks dokumen yang dianalisis. Presentasi TF-IDF dapat dilihat pada Gambar 8.



Gambar 8. Proses TF-IDF

p-ISSN: 2775-4227
e-ISSN: 2775-4219

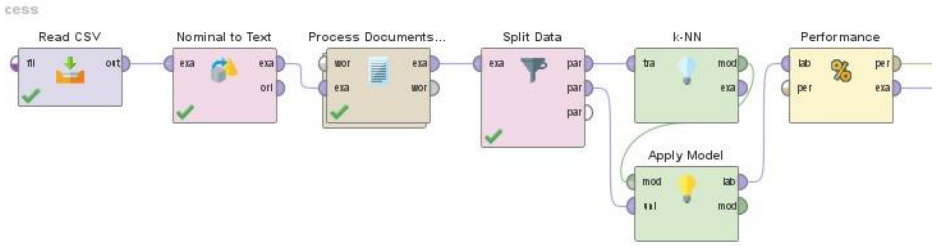
Untuk memberikan visualisasi kata-kata yang paling sering muncul dalam penelitian, menggunakan teknik wordcloud berdasarkan 20 kata yang paling sering muncul dalam data penelitian. Visualisasi wordcloud dapat dilihat pada Gambar 9.



Gambar 9. Visualisasi Wordcloud.

3.5. *K-Nearest Neighbor (KNN)*

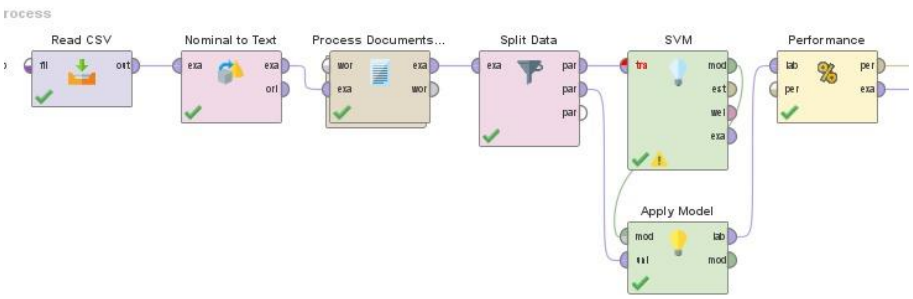
Algoritma *K-Nearest Neighbor (KNN)* digunakan untuk menganalisis sentimen dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Hasil implementasi menunjukkan performa klasifikasi berdasarkan jarak tetangga terdekat, dengan mayoritas tetangga menentukan hasil prediksi. Tingkat *accuracy*, *precision*, *recall*, dan *f1-score* dihitung untuk mengevaluasi efektivitas algoritma ini dalam mengklasifikasikan data sentimen positif dan negatif. Implementasi algoritma *K-Nearest Neighbor (KNN)* dapat dilihat pada Gambar 10.



Gambar 10. Implementasi Algoritma K-Nearest Neighbor (KNN)

3.6. *Support Vector Machine (SVM)*

Algoritma *Support Vector Machine (SVM)* digunakan untuk memisahkan data sentimen ke dalam dua kelas dengan hyperplane terbaik, menggunakan kernel untuk menangani data *non-linear*. Dengan pembagian data yang sama 80% pelatihan, 20% pengujian, model dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *dan f1-score*. Implementasi algoritma *Support Vector Machine (SVM)* dapat dilihat pada Gambar 11.



Gambar 11. Implementasi Algoritma Support Vector Machine (SVM).

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.7. *Confusion Matrix*

Hasil perhitungan menunjukkan bahwa *Algoritma K-Nearest Neighbor (KNN)* memiliki *accuracy* sebesar 74,01%, menandakan 74,01% prediksi model benar secara keseluruhan. *Precision* sebesar 51,83% menunjukkan bahwa sekitar separuh prediksi positif model adalah benar. *Recall* sebesar 52,15% menunjukkan kemampuan model mendeteksi lebih dari separuh kasus positif sebenarnya. Dengan *F1-Score* sebesar 51,99%, model menunjukkan keseimbangan yang cukup baik antara precision dan *recall* meskipun performanya masih dapat ditingkatkan. Hasil *Confusion Matrix K-Nearest Neighbor (KNN)* dapat dilihat pada Gambar 12.

accuracy: 74.01%

	true Negatif	true Positif	class precision
pred. Negatif	362	78	82.27%
pred. Positif	79	85	51.83%
class recall	82.09%	52.15%	

Gambar 12. Hasil *Confusion Matrix K-Nearest Neighbor (KNN)*

Confusion Matrix K-Nearest Neighbor (KNN):
True Positive (TP): 85
False Positive (FP): 79
True Negative (TN): 362
False Negative (FN): 78

a. *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{85+362}{85+79+362+78} = \frac{447}{604} = 0.7401\ atau\ 74.01\%$$

b. *Precision*

$$Precision = \frac{TP}{TP+FP} = \frac{85}{85+79} = \frac{85}{164} = 0.5183\ atau\ 51.83\%$$

c. *Recall*

$$Recall = \frac{TP}{TP+FN} = \frac{85}{85+78} = \frac{85}{163} = 0.5215\ atau\ 52.15\%$$

d. *F1-Score*

$$F1 = 2 . \frac{Precision\ .\ Recall}{Precision+Recall} = 2 . \frac{0.5183\ .\ 0.5215}{0.5183 + 0.5215} = 2 . \frac{0.2703}{1.0398} = 0.5199\ atau\ 51.99\%$$

Hasil perhitungan menunjukkan bahwa Algoritma *Support Vector Machine (SVM)* memiliki *accuracy* sebesar 73,68%, artinya 73,68% prediksi model benar secara keseluruhan. Precision mencapai 75%, menunjukkan bahwa sebagian besar prediksi positif model adalah benar. Namun, *recall* hanya 3,68%, yang berarti model gagal mendeteksi banyak kasus positif sebenarnya. Akibatnya, *F1-Score* hanya 7,02%, mencerminkan ketidakseimbangan performa antara *precision* dan *recall*. Hasil *Confusion Matrix Support Vector Machine (SVM)* dapat dilihat pada Gambar 13.

accuracy: 73.68%

	true Negatif	true Positif	class precision
pred. Negatif	439	157	73.66%
pred. Positif	2	6	75.00%
class recall	99.55%	3.68%	

Gambar 13. Hasil *Confusion Matrix Support Vector Machine (SVM)*

Confusion Matrix Support Vector Machine (SVM):
True Positive (TP): 6
False Positive (FP): 2
True Negative (TN): 439
False Negative (FN): 157

a. *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} = \frac{6+439}{6+2+439+157} = \frac{445}{604} = 0.7368\ atau\ 73.68\%$$

b. *Precision*

$$Precision = \frac{TP}{TP+FP} = \frac{6}{6+2} = \frac{6}{8} = 0.75\ atau\ 75.00\%$$

p-ISSN: 2775-4227
e-ISSN: 2775-4219

c. *Recall*

$$Recall = \frac{TP}{TP+FN} = \frac{6}{6+157} = \frac{6}{163} = 0.0368\ atau\ 3.68\%$$

d. *F1-Score*

$$F1 = 2 . \frac{Precision\ .\ Recall}{Precision+Recall} = 2 . \frac{0.75\ .\ 0.0368}{0.75 + 0.0368} = 2 . \frac{0.0276}{0.7868} = 0.0702\ atau\ 7.02\%$$

Penelitian sebelumnya menggunakan TextBlob dan BERT untuk menganalisis sentimen pada 351 tweets terkait acara Clash of Champions (CoC), mencatat sentimen positif yang sangat tinggi, yaitu 85,27% (TextBlob) dan 86,12% (BERT), dengan sentimen negatif minimal (3,99% TextBlob, 0% BERT), menunjukkan antusiasme masyarakat dan keberhasilan acara. Sebaliknya, penelitian ini menggunakan algoritma *K-Nearest Neighbor (KNN)* dan *Support Vector Machine (SVM)* dengan hasil yang lebih kompleks. KNN menunjukkan *accuracy* sebesar 74,01%, *precision* 51,83%, *recall* 52,15%, dan *F1-Score* 51,99%,. Sedangkan SVM menunjukkan *accuracy* sebesar 73,68 precision 75,00%, *recall* hanya 3,68% dan *F1-Score* 7,02% Perbedaan ini mencerminkan bahwa TextBlob dan BERT lebih unggul dalam mendeteksi sentimen positif, sementara KNN dan SVM menunjukkan kekuatan pada dataset dengan dominasi sentimen negatif, seperti dalam penelitian ini.

4. KESIMPULAN

KNN menunjukkan performa terbaik dalam analisis sentimen *Clash of Champions (CoC)* dengan *accuracy* sebesar 74,01%, *precision* 51,83%, *recall* 52,15%, dan *F1-Score* 51,99%, serta kemampuan unggul dalam mengklasifikasikan data negatif. Sebaliknya, SVM mencatat *accuracy* sebesar 73,68 precision 75,00%, *recall* hanya 3,68% dan *F1-Score* 7,02%, mengindikasikan bias terhadap kelas negatif. Hasil ini menunjukkan bahwa KNN lebih efektif dan seimbang untuk analisis sentimen publik terhadap acara edukatif seperti CoC. Penelitian selanjutnya dapat mempertimbangkan metode pembobotan atau balancing data untuk meningkatkan performa SVM dalam menangani distribusi data yang tidak seimbang.

DAFTAR PUSTAKA

[1] N. Y. S. Munti and D. A. Syaifuddin, "Analisa Dampak Perkembangan Teknologi Informasi Dan Komunikasi Dalam Bidang Pendidikan," *JPT*, vol. 4, no. 2, pp. 1799–1805, Oct. 2020, doi: https://doi.org/10.31004/jptam.v4i2.

[2] I. W. F. Fangestu and H. Syahrizal, "Digitalisasi Lembaga Pendidikandalam Menghadapi Perkembangan dan Kemajuan Teknologi Informasi Dunia Pendidikan," *Jurnal Ilmu Sosial dan Hukum*, vol. 1, no. 2, pp. 26–38, Dec. 2023, doi: https://doi.org/10.61104/alz.v1i2.89.

[3] J. A. Dewantara and T. H. Nurgiansah, "Efektivitas Pembelajaran Daring di Masa Pandemi COVID 19 Bagi Mahasiswa Universitas PGRI Yogyakarta," *JURNAL BASICEDU*, vol. 5, no. 1, pp. 367–375, Jan. 2021, doi: https://doi.org/10.31004/basicedu.v5i1.669.

[4] G. THABRONI, "Ruang Guru: Apa? Mengapa? Kelebihan & Kekurangan+Promo," serupa.id. Accessed: Jul. 06, 2024. [Online]. Available: https://serupa.id/ruang-guru-apa-mengapa-kelebihan-kekurangan-promo-diskon/

[5] D. Amanda, K. Safitri, V. Ferdiansyah, and Nurbaiti, "Media Pembelajaran Ruang Guru Berbasis Teknologi Sebagai Inovasi Pembelajaran Era Revolusi Industri 4.0," *EBISMAN : eBisnis Manajemen*, vol. 1, no. 4, pp. 23–29, Dec. 2023, doi: https://doi.org/10.59603/ebisman.v1i4.225.

[6] Y. C. 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*, vol. 7, pp. 820–838, Oct. 2024.

[7] Wikipedia, "Ruangguru Clash of Champions," Wikipedia. Accessed: Jul. 06, 2024. [Online]. Available: https://id.wikipedia.org/wiki/Ruangguru_Clash_of_Champions

[8] Ruangguru, "Daftar Peserta Clash of Champions & Jadwal Tayangnya," Ruangguru. Accessed: Jul. 06, 2024. [Online]. Available: https://www.ruangguru.com/blog/clash-of-champion-ruangguru

[9] M. R. R. Lillah, D. S. Maylawati, W. B. Zulfikar, W. Uriawan, and A. Wahana, "Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Analisis Sentimen Pengguna Aplikasi Tokopedia," *Intellect : Indonesian Journal of Innovation Learning and Technology*, vol. 2, no. 2, pp. 171–184, Dec. 2023, doi: https://doi.org/10.57255/intellect.v2i02.

[10] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *TEKNIKA*, vol. 1, no. 1, pp. 18–26, Feb.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

2021, doi: https://doi.org/10.34148/teknika.v10i1.311.

[11] Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *JICS*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: https://doi.org/10.58602/jics.v2i2.24.

[12] Maharan and Fathoni, "Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode Decision Tree," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 18, no. 1, pp. 71–83, May 2024, doi: https://doi.org/10.32815/jitika.v18i1.1002.

[13] S. Helmiyah and A. Verdian, "Analisis Sentimen Terhadap Minat Belajar pada Tayangan Acara CoC by Ruangguru Berdasarkan Tweets Menggunakan Metode NLP dan Model BERT," *Jurnal Pendidikan Rosalia*, vol. 7, no. 2, pp. 138–149, Aug. 2024.

[14] E. S. Basryah, A. Erfina, and C. Warman, "ANALISIS SENTIMEN APLIKASI DOMPET DIGITAL DI ERA 4.0 PADA MASA PENDEMI COVID-19 DI PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER," *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)*, pp. 189–196, Aug. 2021.

[15] S. G. Kaparang, D. R. Kaparang, and V. P. Rantung, "Analisis Sentimen New Normal Pada Masa Covid-19 Menggunakan Algoritma Naïve Bayes," *JOINTER–JOURNAL OF INFORMATICS ENGINEERING*, vol. 2, no. 1, pp. 16–23, Jun. 2021, doi: https://doi.org/10.53682/jointer.v2i01.33.

[16] V. Zuliana, Garno, and I. Maulana, "ANALISIS SENTIMEN PROGRAM MIGRASI TV DIGITAL MENGGUNAKAN ALGORITMA NAIVE BAYES DENGAN CHI SQUARE," *Jurnal informasi dan Komputer*, vol. 10, no. 2, pp. 90–95, 2022, doi: https://doi.org/10.35959/jik.v10i2.366.

[17] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI)* , vol. 5, no. 2, pp. 807–814, Sep. 2021, doi: http://dx.doi.org/10.30645/j-sakti.v5i2.

[18] A. H. Anshor and A. Safuwan, "ANALISIS SENTIMEN OPINI WARGANET TWITTER TERHADAP TES SCREENING GENOSE PENDETEKSI VIRUS COVID-19 MENGGUNAKAN METODE NAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION," *JINTEKS (Jurnal Informatika Teknologi dan Sains)*, vol. 5, no. 1, pp. 170–178, Feb. 2023, doi: https://doi.org/10.51401/jinteks.v5i1.2229.

[19] A. Erfina and M. F. Al-shufi, "ANALISIS SENTIMEN APLIKASI JASA KURIR DI PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES," *Jurnal Sistem Informasi dan Informatika (SIMIKA)*, vol. 5, no. 2, pp. 103–110, Jul. 2022.

[20] R. Saputra and F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 3, pp. 411–419, Jul. 2024, doi: https://doi.org/10.47233/jteksis.v6i3.1378.

[21] N. Habibah, E. Budianita, M. Fikry, and I. Iskandar, "Analisis Sentimen Mengenai Penggunaan E-Wallet Pada Google Play Menggunakan Lexicon Based dan K-Nearest Neighbor," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 1, pp. 192–200, Feb. 2023, doi: https://doi.org/10.30865/jurikom.v10i1.5429.

[22] T. I. Alfawas, A. Rahim, and Rudiman, "Penerapan Fitur Ekstraksi TF-IDF untuk Analisis Sentimen Ulasan Game Bus Simulator Indonesia dengan Algoritma Naive Bayes," *INNOVATIVE*, vol. 4, no. 5, pp. 3177–3193, Sep. 2024, doi: https://doi.org/10.31004/innovative.v4i5.13975.

[23] R. S. Al Fathir, T. R. Agus, A. A. Suyono, and F. Ibrahim, "Analisis Sentimen Haramnya Musik Secara Umum Menggunakan Metode KNN," *METIK*, vol. 5, no. 2, pp. 66–70, Dec. 2021, doi: https://doi.org/10.47002/metik.v5i2.284.

[24] R. W. Pratiwi, S. F. H, Dairoh, D. I. Af'idah, Q. R. A, and A. G. F, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *INISTA*, vol. 1, no. 1, pp. 41–46, Nov. 2021, doi: https://doi.org/10.20895/inista.v4i1.387.

[25] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI*, vol. 5, no. 2, pp. 697–771, Sep. 2021, doi: http://dx.doi.org/10.30645/j-sakti.v5i2.