

Analisis Sentimen Komentar pada Kanal Youtube Team RRQ Menggunakan Algoritma Support Vector Machine

Kelvin Sakun¹, Chris Rudianto²

^{1,2} Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jalan Dr. O. Notohamidjojo No. 1-10, Kel. Blotongan, Kec. Sidorejo, Kota Salatiga, Jawa Tengah 50715
Email: 682022027@student.uksw.edu, chris.rudianto@uksw.edu

ABSTRAK

Interaksi antara organisasi *e-sport* dan penggemar kerap berlangsung melalui kolom komentar pada platform YouTube. Aktivitas tersebut menghasilkan data teks dalam jumlah besar yang bersifat tidak terstruktur, sehingga analisis sentimen secara manual menjadi kurang efektif. Penelitian ini bertujuan untuk mengidentifikasi sentimen komentar pengguna pada kanal YouTube Team RRQ dengan menggunakan *algoritma Support Vector Machine* (SVM). Data yang digunakan sebanyak 300 komentar yang diperoleh melalui YouTube Data API dan dikategorikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Tahapan analisis meliputi preprocessing teks yang mencakup *case folding*, *cleansing*, *tokenisasi*, normalisasi, serta *stemming*. Selanjutnya, dilakukan pembobotan fitur menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebelum proses klasifikasi menggunakan algoritma SVM. Model dikembangkan dengan skema pembagian data sebesar 80% sebagai data latih dan 20% sebagai data uji. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa model mampu mencapai tingkat akurasi sebesar 75%, dengan performa terbaik pada kelas sentimen negatif. Temuan ini mengindikasikan bahwa kombinasi metode SVM dan TF-IDF cukup efektif dalam mengklasifikasikan sentimen pada data komentar YouTube yang tidak terstruktur. Penelitian ini memberikan kontribusi dalam pengembangan analisis sentimen pada domain *e-sport* di Indonesia, khususnya pada data yang memiliki karakteristik bahasa informal dan penggunaan *slang*.

Kata kunci: Analisis Sentimen, YouTube, Support Vector Machine, e-sport, TF-IDF

ABSTRACT

Interactions between e-sports organizations and their fans frequently occur through the comment sections on the YouTube platform. These interactions generate large volumes of unstructured textual data, making manual sentiment analysis inefficient. This study aims to identify the sentiment of user comments on the Team RRQ YouTube channel using the Support Vector Machine (SVM) algorithm. The dataset consists of 300 comments collected through the YouTube Data API and categorized into three sentiment classes: positive, negative, and neutral. The analysis process includes text preprocessing stages such as case folding, cleansing, tokenization, normalization, and stemming. Feature weighting is then performed using the Term Frequency-Inverse Document Frequency (TF-IDF) method prior to classification using the SVM algorithm. The model is developed using an 80:20 split for training and testing data. Model performance is evaluated using a confusion matrix with metrics including accuracy, precision, recall, and F1-score. The results indicate that the model achieves an accuracy of 75%, with the best performance observed in the negative sentiment class. These findings suggest that the combination of SVM and TF-IDF is effective for classifying sentiment in unstructured YouTube comment data. This study contributes to the development of sentiment analysis in the e-sports domain in Indonesia, particularly for data characterized by informal language and extensive use of slang.

Keywords: sentiment analysis, YouTube, Support Vector Machine, e-sport, TF-IDF

Pendahuluan

Dalam beberapa tahun terakhir, pemanfaatan media sosial mengalami peningkatan yang signifikan sebagai sarana komunikasi dan pertukaran informasi di ruang digital. Melalui platform tersebut, pengguna dapat menyampaikan opini maupun tanggapan secara terbuka, termasuk pada layanan berbagi video seperti YouTube [1]. Berdasarkan laporan terbaru, jumlah pengguna internet global telah melampaui 5,56 miliar orang, dengan sebagian besar aktivitas berfokus pada pencarian, produksi, serta distribusi informasi [2]. Fenomena ini menjadikan media sosial sebagai ruang interaksi yang dinamis bagi berbagai komunitas, termasuk sektor *e-sport* yang memanfaatkan platform digital untuk membangun keterlibatan dengan penggemarnya. Team RRQ

merupakan salah satu organisasi e-sport terkemuka di Indonesia dengan komunitas penggemar yang besar. Kanal YouTube yang dimiliki tidak hanya berfungsi sebagai media distribusi konten, tetapi juga sebagai sarana komunikasi antara tim dan penggemar melalui fitur komentar yang mencerminkan opini serta sentimen pengguna terhadap berbagai aspek, seperti performa pemain, strategi tim, maupun kebijakan manajemen. Sejumlah penelitian terdahulu telah memanfaatkan analisis sentimen pada media sosial untuk mengkaji opini publik yang bersumber dari data teks tidak terstruktur. Sucahyo et al. [3] menunjukkan bahwa penerapan teknik *machine learning* mampu mengidentifikasi persepsi masyarakat secara efektif. Dalam konteks klasifikasi, algoritma *Support Vector Machine* (SVM) termasuk algoritma yang paling efektif digunakan karena efektif dalam menangani data berdimensi tinggi [4]. Penelitian oleh Ma'rifudin dan Yudhistira [5] serta Br Sinulingga dan Sitorus [6] menunjukkan bahwa SVM dapat memberikan hasil yang baik melalui tahapan prapemrosesan dan pembobotan fitur. Selain itu, dalam konteks *e-sport*, Ichsan et al. [7] menyatakan bahwa media sosial berperan sebagai sarana komunikasi aktif antara organisasi dan komunitas, sedangkan Amal et al. [8] menekankan bahwa komentar pada platform YouTube mengandung respons spontan yang bernilai dalam mendukung pengambilan keputusan.

Meskipun demikian, pemanfaatan komentar YouTube sebagai objek analisis sentimen dalam domain *e-sport* masih belum optimal. Sebagian besar penelitian sebelumnya cenderung menggunakan data teks yang bersifat formal, sementara komentar pada komunitas *e-sport* umumnya didominasi oleh bahasa informal, penggunaan *slang*, serta ekspresi emosional yang tinggi [9], [10]. Kondisi ini menunjukkan adanya kesenjangan penelitian dalam pemanfaatan data komentar YouTube pada domain *e-sport* di Indonesia. Pemilihan algoritma *Support Vector Machine* (SVM) dalam penelitian ini didasarkan pada kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan kinerja klasifikasi yang stabil. Selain itu, SVM memiliki kemampuan generalisasi yang baik sehingga dinilai sesuai untuk diterapkan pada data teks yang tidak terstruktur. Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk fokus dalam menganalisis sentimen komentar pengguna pada kanal YouTube Team RRQ menggunakan algoritma SVM. Penelitian ini memberikan kontribusi dalam mengevaluasi kinerja model klasifikasi sentimen pada data berbahasa informal serta mengkaji tantangan klasifikasi yang muncul akibat penggunaan bahasa *slang* dalam komunitas *e-sport*.

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang dikombinasikan dengan metode eksperimen komputasional. Pendekatan tersebut dipilih karena penelitian difokuskan pada pengolahan data komentar yang diperoleh dari kanal YouTube Team RRQ dan selanjutnya diubah ke dalam bentuk representasi numerik. Representasi numerik tersebut kemudian dianalisis untuk mengidentifikasi kecenderungan sentimen pengguna yang dikategorikan menjadi tiga kelas, yaitu positif, negatif, dan netral [12]. Metode eksperimen komputasional dalam penelitian yang dilakukan dalam studi ini melalui beberapa tahapan yang terstruktur, yang diawali dengan tahap pelatihan sampai pada tahap pengujian model klasifikasi berbasis *Support Vector Machine* (SVM). Dalam proses tersebut, algoritma SVM berupaya membentuk suatu *hyperplane* yang bertindak sebagai batas pemisah antar kelas pada ruang fitur dengan dimensi yang tinggi. Pembentukan *hyperplane* ini bertujuan untuk menghasilkan pemisahan data yang optimal sehingga kinerja klasifikasi dapat ditingkatkan berdasarkan parameter model serta tahapan *preprocessing* yang telah diterapkan sebelumnya [13].



Gambar 1. Alur Penelitian

Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari komentar pengguna pada kanal resmi YouTube Team RRQ. Proses pengambilan data dilakukan dengan memanfaatkan layanan YouTube Data API melalui bahasa pemrograman Python [14]. Dalam studi analisis sentimen, data umumnya berbentuk teks tidak terstruktur yang merepresentasikan opini atau tanggapan pengguna terhadap suatu topik tertentu. Teks tersebut kemudian dianalisis untuk mengetahui kecenderungan sentimen yang dikelompokkan menjadi tiga kategori kelas sentimen, yaitu positif, negatif, dan netral. Pada penelitian ini, *dataset* terdiri dari 300 komentar yang diambil dari beberapa video pada kanal tersebut. Seluruh komentar yang berhasil dikumpulkan selanjutnya disimpan dalam bentuk file CSV yang berisi atribut teks komentar beserta label sentimennya. Pelabelan data dilakukan secara manual oleh peneliti dengan berpedoman pada kriteria klasifikasi sentimen yang telah ditentukan. Setiap komentar dievaluasi berdasarkan konteks kalimat, makna yang terkandung, serta penggunaan kata yang mengindikasikan kecenderungan sentimen tertentu. Untuk menjaga konsistensi hasil pelabelan, dilakukan proses verifikasi ulang terhadap data yang telah diberi label guna mengurangi potensi subjektivitas dalam proses klasifikasi. Klasifikasi sentimen dalam penelitian ini dibagi menjadi tiga kategori utama, yaitu positif, negatif, dan netral, sebagaimana skema klasifikasi yang umum diterapkan dalam penelitian analisis sentimen. Untuk menjaga keseimbangan distribusi data pada tahap klasifikasi, jumlah data pada setiap kategori dibuat seimbang (*balanced*), masing-masing terdiri dari 100 komentar positif, 100 komentar negatif, dan 100 komentar netral. *Dataset* yang telah diberi label tersebut kemudian digunakan sebagai data latih dan data uji dalam proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Data penelitian dikumpulkan pada periode September hingga November 2025 yang diperoleh dari satu video yang diunggah pada kanal YouTube Team RRQ dengan tingkat interaksi yang tinggi, yang ditunjukkan oleh jumlah komentar yang signifikan. Pemilihan video tersebut didasarkan pada pertimbangan bahwa tingginya interaksi pengguna dapat merepresentasikan keragaman opini yang lebih luas. Proses pengambilan data dilakukan menggunakan YouTube Data API dengan memanfaatkan parameter berupa ID video serta filter komentar utama. Selanjutnya, data yang diperoleh melalui proses seleksi dengan mempertimbangkan relevansi konten serta penggunaan bahasa Indonesia.

Preprocessing

Tahap *preprocessing* dilakukan untuk mempersiapkan data teks sebelum proses ekstraksi fitur sehingga data memiliki format yang lebih konsisten dan mudah dianalisis. Proses ini diimplementasikan menggunakan pustaka Sastrawi dan NLTK melalui bahasa pemrograman Python [15]. Tahapan *preprocessing* pada penelitian ini terdiri dari beberapa langkah. Proses awal yang dilakukan adalah case folding, yaitu mengonversi seluruh huruf dalam teks ke bentuk huruf kecil untuk menyamakan format penulisan. Setelah itu dilakukan proses *cleansing* yang bertujuan menghilangkan elemen yang tidak relevan, seperti URL, angka, dan tanda baca. Tahap berikutnya adalah *tokenizing*, yaitu memecah teks menjadi unit kata sehingga setiap kata dapat dianalisis secara terpisah. Selanjutnya diterapkan proses *stopword removal* untuk menyaring kata-kata dengan frekuensi tinggi yang tidak memberikan informasi penting dalam analisis sentimen [16]. Selain itu, dilakukan proses normalisasi terhadap kata-kata *slang* yang sering muncul dalam komentar komunitas Team RRQ. Pada tahap akhir, diterapkan proses *stemming* untuk mengubah kata menjadi bentuk dasarnya sehingga menghasilkan teks yang lebih bersih sehingga menghasilkan teks yang lebih bersih untuk selanjutnya diproses pada tahap ekstraksi fitur. Pada tahap normalisasi, istilah tidak baku yang umum ditemukan dalam komentar media sosial dikonversi ke dalam bentuk bahasa baku. Sebagai ilustrasi, kata “gk” diubah menjadi “tidak”, “mantul” menjadi “mantap”, serta “parah bgt” menjadi “parah banget”.

Pembobotan Fitur Menggunakan TF-IDF

Setelah melalui tahap *preprocessing*, data teks direpresentasikan ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini digunakan untuk menentukan tingkat kepentingan setiap kata dalam suatu dokumen berdasarkan frekuensi kemunculannya (*term frequency*) serta distribusinya pada seluruh kumpulan dokumen (*inverse document frequency*) [17]. Dalam mekanisme ini, kata yang muncul secara dominan pada banyak dokumen cenderung memiliki bobot yang lebih rendah karena dianggap kurang diskriminatif, sedangkan kata yang kemunculannya relatif jarang akan memperoleh bobot yang lebih tinggi karena dinilai lebih representatif terhadap karakteristik dokumen tertentu [18]. Dalam implementasinya, proses pembobotan TF-IDF dikonfigurasi menggunakan skema n-gram dalam rentang (1,2) untuk menangkap representasi kata tunggal maupun pasangan kata. Selain itu, diterapkan parameter minimum document frequency (*min_df*) sebesar 2 untuk menyaring kata-kata yang memiliki frekuensi kemunculan sangat rendah sehingga dapat meningkatkan kualitas fitur yang digunakan dalam proses klasifikasi. Melalui proses pembobotan tersebut, setiap komentar direpresentasikan dalam bentuk vektor fitur numerik yang menunjukkan tingkat kepentingan masing-masing kata, yang selanjutnya digunakan sebagai masukan bagi algoritma *Support Vector Machine* (SVM) dalam tahap klasifikasi sentimen.

Klasifikasi Menggunakan SVM

Representasi numerik yang dihasilkan dari proses pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) digunakan sebagai input bagi algoritma *Support Vector Machine* (SVM). Implementasi model dalam penelitian ini dilakukan menggunakan pustaka *scikit-learn* melalui bahasa pemrograman Python. Pembagian dataset dilakukan menggunakan metode *hold-out*, dengan proporsi sebesar 80% sebagai data latih dan 20% sebagai data uji. Proses pembagian dilakukan secara acak (*random splitting*) untuk memastikan distribusi data yang representatif pada masing-masing kelas sentimen, sehingga model dapat mempelajari pola dari data latih secara optimal dan dievaluasi secara objektif menggunakan data uji [19]. Secara konseptual, SVM melakukan klasifikasi dengan membentuk *hyperplane* sebagai batas pemisah antar kelas pada ruang fitur. *Hyperplane* tersebut dipilih untuk memaksimalkan margin antar kelas sehingga model mampu membedakan sentimen positif, negatif, dan netral secara optimal. Pendekatan ini efektif dalam menangani data teks berdimensi tinggi, seperti representasi fitur TF-IDF, karena memiliki kemampuan generalisasi yang baik terhadap data baru. Model SVM yang digunakan dalam penelitian ini menerapkan *kernel linear* dengan nilai parameter regularisasi $C=1.0$ sebagai nilai default dari pustaka *scikit-learn*. Pemilihan *kernel linear* didasarkan pada kemampuannya dalam memproses data teks berdimensi tinggi secara efisien serta menghasilkan kinerja klasifikasi yang stabil. Selain menggunakan metode *hold-out*, penelitian ini juga menerapkan teknik validasi tambahan berupa *k-fold cross validation* dengan nilai $k = 5$ untuk mengevaluasi kestabilan model. Pendekatan ini bertujuan untuk memastikan bahwa kinerja model tidak bergantung pada satu skenario pembagian data tertentu, sehingga hasil evaluasi yang diperoleh menjadi lebih reliabel dan memiliki kemampuan generalisasi yang lebih baik.

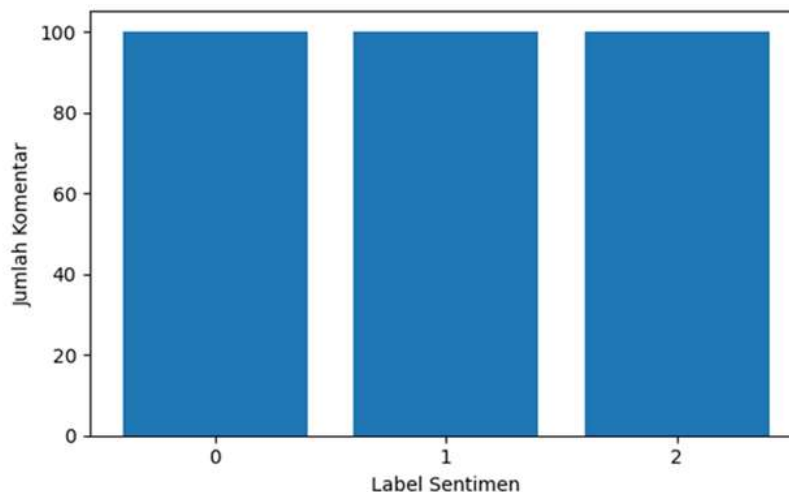
Evaluasi Hasil

Evaluasi model dilakukan untuk mengukur kinerja klasifikasi sentimen setelah proses pelatihan selesai. Penilaian yang dilakukan dengan menggunakan *confusion matrix* yang memperoleh sejumlah metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* [20]. Penggunaan berbagai metrik tersebut bertujuan agar evaluasi tidak hanya berfokus pada tingkat ketepatan prediksi secara keseluruhan, tetapi juga mampu menggambarkan kemampuan model dalam mengidentifikasi setiap kategori sentimen secara lebih rinci. Nilai F1-score digunakan untuk menunjukkan keseimbangan antara *precision* dan *recall*, sehingga memberikan gambaran performa model yang lebih komprehensif. Selain itu, analisis terhadap *confusion matrix* juga dilakukan untuk mengidentifikasi distribusi kesalahan klasifikasi serta pola kesalahan yang terjadi pada masing-masing kelas sentimen. Untuk memastikan kestabilan model, pembagian data dilakukan secara acak menggunakan skema *hold-out*. Pendekatan ini bertujuan untuk meminimalkan bias pada proses pelatihan dan pengujian, sehingga hasil evaluasi tidak bergantung pada satu konfigurasi pembagian data tertentu dan mampu melakukan generalisasi yang baik terhadap data baru. Seluruh proses penelitian diimplementasikan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka *pandas* dan *numpy* untuk pengolahan data, *NLTK* dan *Sastrawi* untuk *preprocessing* teks, *scikit-learn* untuk ekstraksi fitur dan klasifikasi, serta *matplotlib* untuk visualisasi data. Selain itu, proses pengambilan data dilakukan menggunakan *googleapiclient* melalui YouTube Data API.

Hasil Dan Pembahasan

Penerapan algoritma *Support Vector Machine* (SVM) pada dataset yang terdiri dari 300 komentar dari kanal YouTube Team RRQ memberikan gambaran mengenai kemampuan model dalam melakukan klasifikasi sentimen pada data teks media sosial. Berdasarkan hasil pengujian dengan skema pembagian data sebesar 80% sebagai data latih dan 20% sebagai data uji, model yang dibangun mampu memperoleh nilai *accuracy* sebesar 75%. Hasil pengujian tersebut mengindikasikan bahwa model SVM mampu mengidentifikasi pola sentimen yang muncul pada komentar pengguna dengan cukup baik. Dalam mengenali pola sentimen yang terkandung pada komentar pengguna. Hasil pengujian juga menunjukkan bahwa tahapan *preprocessing* teks, seperti pembersihan data, *tokenizing*, serta normalisasi kata, berperan penting dalam meningkatkan kualitas fitur yang digunakan oleh model [21]. Selain itu, penerapan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai teknik pembobotan fitur memungkinkan kata-kata pada komentar pengguna direpresentasikan dalam bentuk numerik sehingga dapat diproses oleh algoritma klasifikasi [22]. Dengan demikian, kombinasi antara proses *preprocessing* dan pembobotan fitur menggunakan *TF-IDF* membantu model SVM dalam mengenali pola bahasa yang umum digunakan oleh komunitas penggemar *e-sport* pada kanal YouTube Team RRQ.

Tahap awal pembahasan hasil penelitian dilakukan dengan meninjau karakteristik dataset yang digunakan. Dataset sebanyak 300 komentar digunakan sebagai dataset penelitian yang diperoleh dari beberapa video pada kanal YouTube Team RRQ. Untuk mendukung proses pelatihan model secara optimal, distribusi kelas dalam dataset dibuat relatif seimbang sehingga model tidak cenderung memprediksi salah satu kategori sentimen secara dominan. Setiap komentar kemudian diberi label ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Distribusi masing-masing kategori sentimen dalam dataset penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Label

Mengingat bahwa komentar pada kanal YouTube Team RRQ banyak menggunakan bahasa *slang* serta singkatan yang umum digunakan dalam komunitas *e-sport*, tahapan preprocessing menjadi langkah penting untuk menghasilkan data dengan kualitas yang lebih baik sebelum memasuki proses klasifikasi oleh algoritma. Untuk memperlihatkan karakteristik data yang digunakan dalam penelitian ini, beberapa contoh komentar pengguna beserta kategori sentimennya ditampilkan pada Tabel 1.

Tabel 1. Contoh Komentar Berdasarkan Kelas Sentimen

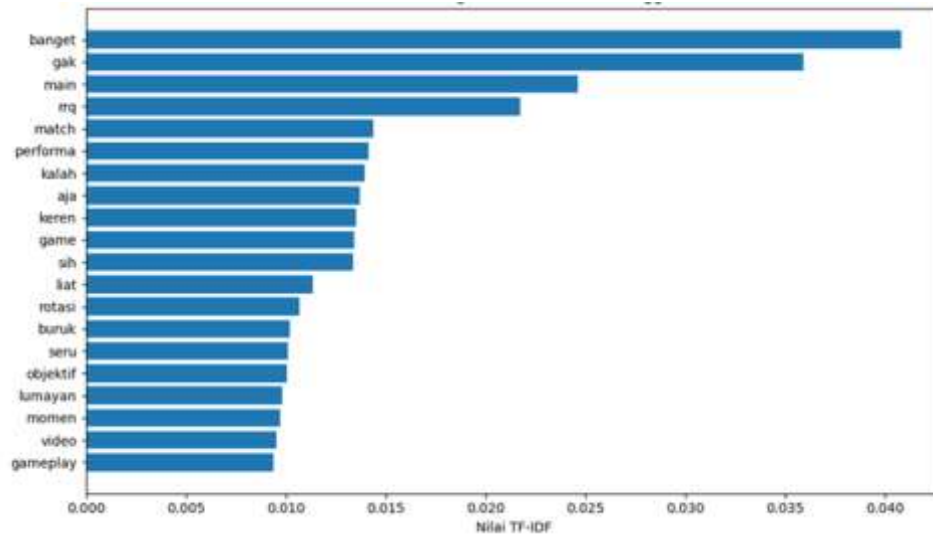
No	Komentar Youtube	Label	Keterangan
1	RRQ mainnya parah banget hari ini	0	Negatif
2	Kok mainnya nggak siap sama sekali	0	Negatif
3	Bangga sama permainan RRQ hari ini	1	Positif
4	Akhirnya menang, mainnya bagus	1	Positif
5	Draft pick-nya cukup menarik	2	Netral
6	Game-nya cukup oke untuk ditonton	2	Netral

Tabel 1 menampilkan beberapa contoh komentar pengguna pada platform *YouTube* yang digunakan sebagai dataset dalam penelitian ini beserta label sentimen yang menyertainya. Kolom komentar berisi teks asli yang ditulis oleh pengguna, sedangkan kolom label menunjukkan kategori tiga kelas sentimen, yaitu positif, negatif, dan netral. Selain itu, kolom keterangan disediakan untuk memberikan penjelasan tambahan mengenai konteks dari masing-masing komentar.

Tabel 2. Contoh Data Mentah & Preprocessing

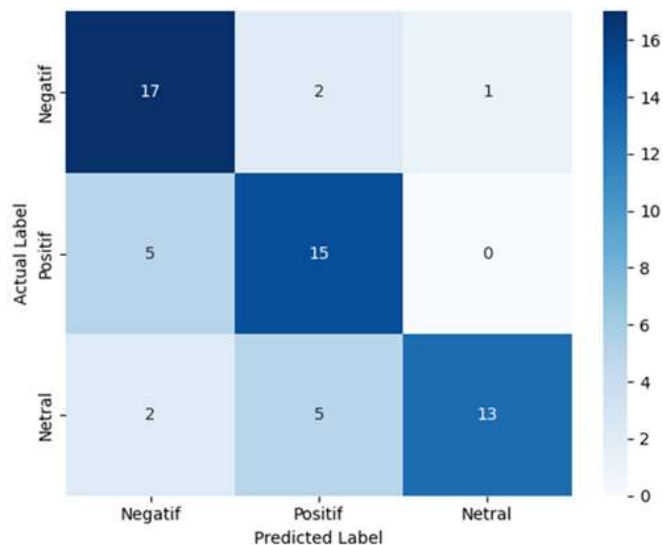
No	Komentar Mentah YouTube	Hasil Preprocessing
1	dah kalah draft, kalah gameplay juga	kalah draft kalah gameplay
2	Coach dan player kompak, mantap.	coach player kompak mantap
3	Early game bagus, late game hancur	early game bagus late game hancur
4	Udah unggul tapi malah blunder	unggul tapi malah blunder
5	RRQ lagi kurang beruntung hari ini	rrq lagi kurang untung hari

Distribusi kelas dalam dataset menjadi aspek penting yang perlu diperhatikan karena ketidakseimbangan jumlah sampel antar kategori dapat memengaruhi kinerja model klasifikasi. Ketidakseimbangan data berpotensi mengurangi kemampuan algoritma dalam mengidentifikasi kelas yang memiliki jumlah sampel lebih sedikit serta dapat berdampak pada tingkat akurasi prediksi yang dihasilkan [23]. Setelah melalui tahap *preprocessing* dan analisis distribusi data, proses pembobotan fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF) dilakukan untuk mengidentifikasi kata-kata yang paling representatif dalam dataset. Kata-kata yang memiliki bobot tertinggi dapat memberikan gambaran mengenai topik utama yang sering dibicarakan oleh para penggemar, sebagaimana divisualisasikan pada Gambar 3.



Gambar 3. Visualisasi Kata dengan Nilai TF-IDF Tertinggi pada Komentar Team RRQ

Setelah proses ekstraksi fitur selesai dilakukan, tahap berikutnya adalah pelatihan model klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). Representasi fitur dalam penelitian ini menerapkan skema n-gram yang mencakup unigram dan bigram untuk membantu model memahami hubungan antar kata dalam suatu kalimat. Pendekatan ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa kombinasi n-gram dapat meningkatkan kemampuan model dalam menangkap struktur bahasa pada data teks [24]. Penerapan algoritma SVM dengan skema pembagian dataset dimana 80% digunakan untuk proses pelatihan model dan 20% digunakan sebagai data pengujian untuk menghasilkan nilai accuracy sebesar 75%. Selanjutnya untuk menilai performa model secara lebih rinci pada setiap kategori sentimen, digunakan metode Confusion Matrix yang ditampilkan pada Gambar 4. Berdasarkan visualisasi tersebut, model menunjukkan mampu mengidentifikasi komentar positif dengan cukup baik, model masih mengalami kendala dalam membedakan komentar netral dan negatif, khususnya pada komentar yang mengandung unsur sarkasme.



Gambar 4. *Confusion Matrix*

Sebagai tahap validasi akhir terhadap kinerja sistem, dilakukan pengukuran metrik evaluasi yang meliputi precision, recall, dan F1-score. Rincian performa model untuk setiap kategori sentimen disajikan secara lengkap pada Tabel 2. Berdasarkan hasil evaluasi tersebut, dapat diidentifikasi beberapa pola kesalahan klasifikasi yang masih terjadi pada model. Pola kesalahan ini menunjukkan bahwa model masih menghadapi kesulitan dalam membedakan ekspresi yang bersifat netral dengan sentimen positif yang memiliki tingkat intensitas emosional yang rendah. Temuan ini sejalan dengan hasil penelitian sebelumnya yang melaporkan adanya tantangan serupa dalam proses klasifikasi sentimen berbasis teks [25].

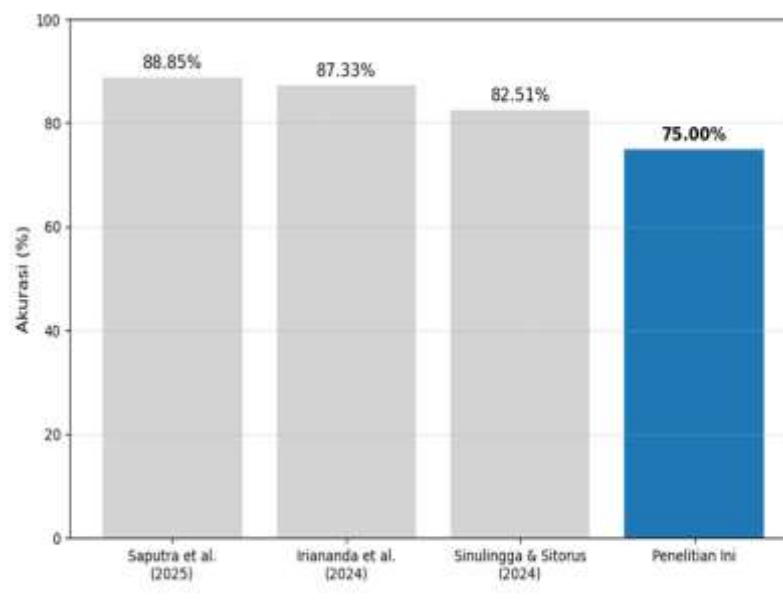
```
Classification Report:
              precision    recall  f1-score   support

   Negatif      0.71      0.85      0.77      20
   Positif      0.68      0.75      0.71      20
   Netral       0.93      0.65      0.76      20

 accuracy              0.75      60
 macro avg              0.77      0.75      0.75      60
 weighted avg          0.77      0.75      0.75      60
```

Gambar 5. Akurasi *support vector machine* (SVM)

Berdasarkan rangkaian eksperimen yang telah dilakukan, model klasifikasi dengan algoritma *Support Vector Machine* (SVM) mampu menghasilkan tingkat accuracy sebesar 75%. Nilai tersebut mendefinisikan bahwa metode yang digunakan pada penelitian ini cukup efektif untuk mengidentifikasi sentimen pada komentar pengguna di kanal YouTube Team RRQ. Selain itu, penerapan tahapan prapemrosesan teks, termasuk proses normalisasi terhadap kata tidak baku maupun bahasa *slang* yang umum digunakan dalam komunitas *e-sport*, turut membantu meningkatkan kualitas data sebelum diproses oleh model klasifikasi. Meskipun demikian, masih terdapat tingkat kesalahan klasifikasi sebesar 25%, yang menunjukkan bahwa sebagian komentar belum dapat diidentifikasi secara tepat oleh sistem. Hal ini umumnya dipengaruhi oleh karakteristik bahasa pada media sosial yang cenderung tidak terstruktur, seperti penggunaan singkatan, kata tidak baku, serta adanya konteks sarkasme atau ironi yang sulit ditangkap oleh model berbasis machine learning. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa penggunaan algoritma *Support Vector Machine* (SVM) yang dikombinasikan dengan menggunakan metode pembobotan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) serta proses normalisasi bahasa *slang* dapat menjadi pendekatan yang cukup efektif dalam memetakan opini publik pada komunitas *e-sport*.



Gambar 6. Perbandingan Nilai Akurasi SVM dengan Penelitian Terdahulu

Hasil evaluasi menunjukkan bahwa model yang dikembangkan mampu mencapai tingkat akurasi sebesar 75% dalam mengklasifikasikan sentimen pada komentar pengguna YouTube dalam konteks e-sport. Untuk mengkaji kontribusi penelitian, hasil tersebut dibandingkan dengan beberapa studi terdahulu sebagaimana disajikan pada Gambar 6. Perbandingan tersebut mengindikasikan bahwa performa akurasi dalam penelitian ini masih berada di bawah beberapa penelitian sebelumnya. Perbedaan tersebut diduga dipengaruhi oleh karakteristik dataset yang digunakan, di mana komentar YouTube pada domain *e-sport* cenderung bersifat informal serta mengandung bahasa *slang* dan istilah teknis permainan yang meningkatkan kompleksitas proses klasifikasi. Sebagai pembanding, beberapa penelitian terdahulu menunjukkan bahwa algoritma *Naive Bayes* sering digunakan sebagai model baseline dalam analisis sentimen. Dibandingkan dengan pendekatan tersebut, algoritma *Support Vector Machine (SVM)* pada penelitian ini menunjukkan performa yang cukup kompetitif, khususnya dalam menangani data berdimensi tinggi seperti representasi fitur berbasis *TF-IDF*. Analisis lebih lanjut terhadap setiap kategori sentimen menunjukkan bahwa model memiliki kinerja terbaik pada kelas negatif dengan nilai *recall* sebesar 0,85. Hal ini dipengaruhi oleh keberadaan kata-kata eksplisit bernada kritik, seperti “kalah”, “parah”, dan “jelek”, yang memiliki bobot fitur tinggi serta muncul secara konsisten dalam data. Sebaliknya, performa model pada kelas positif dan netral relatif lebih rendah, yang disebabkan oleh adanya ambiguitas linguistik serta penggunaan kalimat majemuk yang sering menggabungkan unsur positif dan negatif dalam satu pernyataan. Selain itu, kesalahan klasifikasi juga ditemukan pada komentar yang mengandung sarkasme maupun makna ambigu. Model cenderung mengalami kesulitan dalam membedakan antara sentimen netral dan positif dengan intensitas emosi yang rendah. Faktor lain yang memengaruhi kesalahan klasifikasi adalah penggunaan bahasa *slang* yang belum sepenuhnya terakomodasi dalam proses normalisasi. Temuan ini menunjukkan bahwa pendekatan berbasis frekuensi kata seperti *TF-IDF* masih belum mampu dalam menjelaskan konteks semantik yang lebih kompleks.

Simpulan

Hasil penelitian menunjukkan bahwa tujuan utama dari analisis sentimen komentar pengguna pada kanal YouTube Team RRQ menggunakan algoritma *Support Vector Machine (SVM)* telah berhasil dicapai. Model yang dibangun mampu mengklasifikasikan komentar ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral, dengan tingkat *accuracy* sebesar 75%. Penerapan tahapan prapemrosesan teks serta pembobotan fitur Term Frequency–Inverse Document Frequency (*TF-IDF*) berperan dalam mengubah data komentar yang bersifat tidak terstruktur menjadi representasi numerik yang dapat diproses oleh model klasifikasi. Selama proses penelitian, beberapa kendala juga ditemukan, terutama terkait karakteristik bahasa pada komentar media sosial yang cenderung informal, seperti penggunaan bahasa *slang*, singkatan, serta adanya unsur sarkasme yang dapat memengaruhi hasil klasifikasi. Meskipun demikian, pendekatan menggunakan algoritma *Support Vector Machine* terbukti memiliki kemampuan yang cukup efektif dalam mengidentifikasi pola sentimen pada komentar komunitas *e-sport*. Secara umum, penelitian ini memberikan gambaran mengenai potensi penerapan teknik analisis sentimen dalam memahami opini penggemar terhadap tim *e-sport* melalui data komentar pada media sosial. Selain itu, hasil penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya dalam mengembangkan metode klasifikasi sentimen yang lebih mampu menangkap konteks bahasa yang kompleks.

Daftar Pustaka

- [1] Adiawaty, S., & Daso, M. A. (2023). Dampak Pemberdayaan dan Komunikasi Interpersonal dalam Meningkatkan Organizational Citizenship Behavior Karyawan Generasi Y. *JURNAL LENTERA BISNIS*, 12(3), 771. <https://doi.org/10.34127/jrlab.v12i3.900>
- [2] We Are Social. (2025, February 5). Digital 2025: The essential guide to the global state of digital. <https://wearesocial.com/id/blog/2025/02/digital-2025-the-essential-guide-to-the-global-state-of-digital/>
- [3] Suahyo, N., Kurniati, I., & Harvit, K. (2022). Analisis Sentimen Masyarakat terhadap UU Cipta Kerja pada Media Sosial Twitter. *Jurnal Rekayasa Informasi Swadharma*, 2(1), 63–70.
- [4] Farahdinna, F., & Prabawati, P. (2026). Perbandingan Algoritma Support Vector Machine dan K-Nearest Neighbor terhadap Efektivitas Program Makan Siang Gratis. *Sistemasi: Jurnal Sistem Informasi*, 15(2), 699–709. <https://doi.org/10.32520>
- [5] Ma'rufudin, M., & Yudhistira, A. (2025). Analisis Sentimen Petani Milenial Pada Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(3), 845–857. <https://doi.org/10.52436/1.jpti.717>
- [6] Br Sinulingga, J. E., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, 14(1), 42–53. <https://doi.org/10.34010/jamika.v14i1.11946>
- [7] Ichsan, M. C., Setyawati, H., & Awang, F. (2022). Social Media and Sports: Engagement For Sports Fans. *Journal of Physical Education and Sports*, 11(4), 456–463. <https://journal.unnes.ac.id/sju/index.php/jpes>

- [8] Amal, I., Muhyidin, A., & Rinsi, E. S. (2025). Implikatur Percakapan Dalam Tuturan Pemain Mobile Legends Dari Tim RRQ Pada Pertandingan Esport. *Jurnal Ilmiah Wahana Pendidikan*, 11(C), 57–64.
- [9] Saputra, J., Maryani, L., Wulandari, D., & Eka, W. (2025). Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0). 10(1). <https://journal.uin-alaudhin.ac.id/index.php/instek>
- [10] Mursidah, I., Sanjaya, R., Yulianto, B., Sweetania, D., & Sularsih, P. (2025). Analisis Sentimen Pengguna pada Ulasan Game Honkai Star Rail Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Minfo Polgan*, 14(1), 1178–1188. <https://doi.org/10.33395/jmp.v14i1.14982>
- [11] Arisenja, N. L. B., & Mulyana, D. I. (2025). Analisis Sentimen Pemecatan Shin Tae-yong pada Media Sosial X untuk Monitoring Opini Publik Menggunakan Naive Bayes dan Support Vector Machines. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi (JIMIK)*, 6(3), 1649–1705. <https://journal.stmiki.ac.id>
- [12] Sari, N., Jazman, M., Ahsyar, T. K., Marsal, A., Syaifullah, & Marsal, A. (2024). Penerapan Algoritma Klasifikasi Naive Bayes dan Support Vector Machine untuk Analisis Sentimen Cyberbullying Bilingual di Aplikasi X. 14(1), 211–971924. <http://sistemasi.ftik.unisi.ac.id>
- [13] Alqowiy, M. Q., & Utami, E. (2025). Sistemasi: Jurnal Sistem Informasi Optimasi Parameter Support Vector Machine menggunakan Algoritma Genetika untuk Deteksi Cyberbullying Parameter Optimisation of Support Vector Machine using Genetic Algorithm for Cyberbullying Detection. 14(1), 332–341. <http://sistemasi.ftik.unisi.ac.id>
- [14] Gunawan, A., Rohman, S. R., & Kartikasari, R. (2024). Analisis Sentimen Terhadap Pelayanan Samsat Di Tasikmalaya Menggunakan Metode Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, 8(3), 2884–2889.
- [15] Aesyi, U. S., & Cahyo, P. W. (2023). Peningkatan Penjualan Produk Berdasarkan Analisis Komentar Pelanggan di Marketplace: Shopee. *Jurnal Sains Dan Informatika*. <https://doi.org/10.34128/jsi.v9i1.539>
- [16] Miranda, E., Gabriella, V., Wahyudi, S. A., & Chai, J. (2023). Text Classification untuk Menganalisis Sentimen Pendapat Masyarakat Indonesia terhadap Vaksinasi Covid-19. *Sistemasi: Jurnal Sistem Informasi*, (2), 438–451. <http://sistemasi.ftik.unisi.ac.id>
- [17] Widodo, A., Herlambang, B. A., & Renaldy, R. (2025). Optimizing Support Vector Machine (SVM) for Sentiment Analysis of Blu by BCA Reviews with Chi-Square. *Journal of Applied Informatics and Computing (JAIC)*, 9(5), 2588. <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [18] Dani, A. H., Puspaningrum, E. Y., & Mumpuni, R. (2024). Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying. (2), 94–106. <https://doi.org/10.62951/router.v2i2.76>
- [19] Prastyo, D., Irawan, D., & Mursyidin, I. H. (2024). Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024. *Bit-Tech*, 7(2), 413–421. <https://doi.org/10.32877/bt.v7i2.1833>
- [20] Yahya, R. D., Wibowo, S. A., & Vendyansyah, N. (2024). Analisis Sentimen untuk Deteksi Ujaran Kebencian pada Media Sosial Terkait Pemilu 2024 Menggunakan Metode Support Vector Machine. *Jurnal Mahasiswa Teknik Informatika*, 8(2), 1182–1189.
- [21] Hermawan, M. A., Faqih, A., & Dwilestari, G. (2025). Implementasi Akurasi Model Naive Bayes Menggunakan Smote Dalam Analisis Sentimen Pengguna Aplikasi Brimo. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1), 855–862. <https://doi.org/10.23960/jitet.v13i1.5748>
- [22] Yahya, R. D., Wibowo, S. A., & Vendyansyah, N. (2024). Analisis Sentimen untuk Deteksi Ujaran Kebencian pada Media Sosial Terkait Pemilu 2024 Menggunakan Metode Support Vector Machine. *Jurnal Mahasiswa Teknik Informatika*, 8(2), 1182–1189.
- [23] Ikhsani, R. N., & Abdulloh, F. F. (2023). Optimasi SVM dan Decision Tree Menggunakan SMOTE Untuk Mengklasifikasi Sentimen Masyarakat Mengenai Pinjaman Online. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(4), 1667–1677. <https://doi.org/10.30865/mib.v7i4.6809>
- [24] Iriananda, S. W., Budiawan, R. W., Rahman, A. Y., & Istiadi, I. (2024). Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(4), 743–752. <https://doi.org/10.25126/jtiik.1148244>
- [25] Fazri, M., & Voutama, A. (2025). Analisis Sentimen Publik terhadap Danantara di Media Sosial X Menggunakan NLP dan Pembelajaran Mesin. *JOISIE (Journal Of Information Systems And Informatics Engineering)*, 9(1), 197–206. <https://doi.org/10.35145/joisie.v9i1.4924>