

ENSEMBLE MAJORITY VOTING UNTUK ANALISIS SENTIMEN DAN EMOSI PADA KOMENTAR YOUTUBE: STUDI KASUS RESIDENT EVIL 4 REMAKE

Ahmad Abdul Hadi ^{a,1}, Ressa Priskila^{b,2} *, Viktor Handrianus Pranatawijaya ^{c,3}, Novera Kristianti ^{d,4}

^{a,b,c} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya

¹ahmaad.abdul.hadi@gmail.com; ²viktorhp@it.upr.ac.id^{2*}; ³ressa@it.upr.ac.id³, ⁴noverakristianti@eng.upr.ac.id⁴

* Corresponding author

ARTICLE INFO

Keywords

Sentiment Analysis; Emotion Detection; Naive Bayes; Support Vector Machine; Bert; Majority Voting.

ABSTRACT

Currently, social media can be said to be one of the important things in the fields of marketing, broadcasting and entertainment, such as the gaming industry. In this case, Sentiment Analysis and Emotion Detection can be a tool for understanding the public's response and perception of the content presented. One of them is for the game Resident Evil 4 Remake, which was announced on March 24, 2023, and received a lot of public response on various social media platforms such as YouTube, one of which received responses in the form of 7177 comments between June 3 2022 and February 9, 2024. The research methodology used includes data collection methodology and simulation methodology, by combining the Naive Bayes algorithm, SVM and BERT using the Majority Voting method where these algorithms were previously trained using two different datasets which showed Naive Bayes performance with an accuracy of 84%, SVM with 89%, BERT with 93% and the Majority Voting Method with 90% accuracy with training using the Resident Evil 4 Remake dataset. And in training with the Steam Game Review dataset, Naive Bayes and SVM were obtained with an accuracy of 53%, BERT with 66%, and the Majority Voting Method with an accuracy of 57%. The Majority Voting classification model trained on the Resident Evil 4 dataset was used to perform Sentiment Analysis classification on comments from the YouTube video entitled "Resident Evil 4 Remake: Reveal Trailer" from the IGN Channel. The ratio of positive and negative sentiments was 60.2% and 39%. .8% with the frequency of emotions of anger, excitement and anticipation appearing most frequently.

1. Pendahuluan

Perkembangan teknologi informasi dan internet telah membawa dampak besar pada berbagai aspek kehidupan, termasuk dalam bidang pemasaran, penyiaran, dan hiburan, seperti industri game. Dalam era di mana informasi dengan cepat tersebar melalui platform online, menjadi sangat penting untuk memahami respons dan persepsi publik terhadap konten yang disajikan. Dalam hal ini, analisis sentimen menjadi alat yang sangat berharga.

Analisis sentimen sendiri menurut Turney, P. D. [1] adalah tugas untuk mengidentifikasi orientasi subjektif suatu teks, apakah positif, negatif, atau netral. Orientasi subjektif ini dapat diidentifikasi dengan mempertimbangkan berbagai aspek linguistik seperti kata-kata, frasa, dan struktur kalimat. Analisis sentimen sering kali digunakan untuk menganalisis pendapat dan perasaan pelanggan tentang produk atau layanan tertentu. Menurut Zendesk [2] analisis sentimen membantu perusahaan mengidentifikasi dan menyelesaikan masalah pelanggan dengan lebih cepat, yang dapat meningkatkan skor CSAT (Customer Satisfaction) hingga 10%.

Deteksi emosi merupakan hal yang mungkin dilakukan yang mana bertujuan untuk mengidentifikasi dan mengklasifikasikan emosi yang terkandung dalam pesan-pesan tertulis, mulai dari pesan singkat hingga teks panjang seperti misalnya komentar produk. Dengan menggunakan teknik-teknik seperti analisis sentimen dan pembelajaran mesin. Majority voting adalah salah satu metode yang dapat digunakan dalam analisis sentimen. Menurut Mohammad, Saif, & Turney [3] Majority Voting adalah metode klasifikasi teks yang sederhana dan efektif, di mana sentimen suatu teks ditentukan berdasarkan mayoritas suara dari beberapa klasifikasi.

Model Naive Bayes adalah model klasifikasi probabilistik yang berdasarkan pada teorema Bayes dengan asumsi bahwa fitur-fitur dalam data adalah independen satu sama lain. Dan juga Model Naive Bayes adalah model yang mudah diadaptasi dan dapat digunakan untuk berbagai jenis tugas klasifikasi [4]. Model SVM (Support Vector Machine) adalah salah satu model klasifikasi yang populer dalam analisis sentimen. Model ini didasarkan pada algoritma machine learning yang memisahkan data menjadi dua kelas, positif dan negatif, dengan margin terbesar. Model SVM juga merupakan model yang interpretable dan mudah dipahami [5].

BERT adalah model bahasa yang dikembangkan oleh Google yang berbasis pada arsitektur Transformer dan dilatih pada tugas-tugas pemahaman bahasa alami. BERT mampu memahami konteks kata dalam teks dengan lebih baik karena dapat memperhitungkan hubungan antara kata-kata sebelumnya dan sesudahnya dalam sebuah kalimat. Dari Yang et al. [6] Dalam penelitian mereka yang berjudul "BERT for Sentiment Analysis of Social Media Posts", mereka menunjukkan bahwa model BERT dapat mencapai kinerja yang baik dalam mengklasifikasikan sentimen unggahan media sosial.

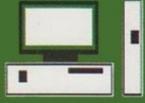
Pada tanggal 24 Maret 2023, industri game dikejutkan dengan perilisan game Resident Evil 4 Remake. Sejak diluncurkan, game ini telah terjual lebih dari 6 juta kopi di seluruh dunia. Angka ini menunjukkan antusiasme yang tinggi dari para penggemar terhadap game klasik yang dikemas ulang dengan grafis dan gameplay modern [7].

Proses pengumpulan data tidak hanya terbatas pada dataset yang telah diberi label untuk melatih model yang akan digunakan dalam analisis sentimen. Di era digital yang kaya akan platform media sosial seperti YouTube, di mana trailer game sering dipublikasikan, misalnya Trailer Resident Evil 4 Remake yang sedang ramai dibicarakan oleh para penggemar dan komunitas gamer saat ini, kita dapat memanfaatkan komentar-komentar dari video YouTube tersebut sebagai sumber data tambahan. Data ini akan menjadi landasan bagi analisis untuk memahami pandangan umum yang muncul dari komunitas tentang trailer tersebut.

Berdasarkan latar belakang yang telah diuraikan diatas, maka perumusan masalah yang muncul adalah "Bagaimana melakukan analisis sentimen dan emosi dengan metode majority voting terhadap komentar-komentar pada video youtube Resident Evil 4 Remake: Reveal Trailer untuk menemukan pola-pola sentiment dan emosi yang dominan terhadap pengumuman ini?".

Diharapkan bahwa hasil analisis ini akan memberikan wawasan yang berharga tentang bagaimana game tersebut diterima oleh penggemar, serta memperkuat pemahaman tentang efektivitas metode Majority Voting dalam konteks analisis sentimen pada komentar-komentar YouTube dengan penggunaan model Naive Bayes, SVM, dan Bert. Analisis sentimen (sentiment analysis atau disebut juga opinion mining) merupakan suatu bidang dalam Natural language processing (NLP) yang berfokus kepada penentuan sentimen atau muatan apakah data tersebut dapat dikatakan positif, negatif atau lain sebagainya [8]. Sentimen analisis juga dapat didefinisikan sebagai pemanfaatan NLP, analisis teks, linguistik komputer dan biometriks dalam mengidentifikasi, mengekstrak, dan mempelajari informasi subjektif [9].

Machine learning merupakan satu bidang dalam kecerdasan buatan (artificial intelligence atau AI) yang berfokus pada pengembangan mesin yang dapat belajar dengan sendirinya atau tanpa perintah [10]. Machine learning mempelajari dan menggunakan data yang ada sehingga dapat mengerjakan suatu tugas tertentu. Algoritma Naive Bayes merupakan algoritma yang mengimplementasikan teori Naive Bayes. Teorem Bayes membahas suatu kemungkinan munculnya atau terjadinya suatu fakta dalam beberapa kejadian yang terjadi. Teorem Bayes digunakan dalam memprediksi fakta dengan menggunakan fakta-fakta yang telah terjadi sebelumnya [11].



Support Vector Machine (SVM) adalah salah satu metode klasifikasi yang memisah kelas-kelas berdasar hyperplane pembatasnya. SVM berfokus untuk mencari hyperplane terbaik yang dapat memisah atau mengklasifikasikan data-data [12]. Bidirectional Encoder Representation from Transformer (BERT) merupakan teknik machine learning jaringan saraf (artificial neural network) yang dikembangkan oleh google untuk memahami konteks suatu kata dalam bahasa Inggris [13]. Ada dua teknik voting dalam machine learning, majority voting dan soft voting. Dalam soft voting, dilakukan penggabungan peluang tiap prediksi ke dalam model yang kemudian akan dipilih prediksi yang memiliki total peluang prediksi tertinggi. Sedangkan majority voting mengambil keputusan dengan melihat nilai atau prediksi yang muncul pada tiap metode dan memilih prediksi yang paling banyak muncul [14]. Web scraping merupakan suatu teknik ekstraksi data atau informasi dari satu atau banyak website dan memprosesnya ke dalam suatu struktur simpel seperti file spreadsheet, File CSV atau database yang akan digunakan di kemudian hari. [15].

2. Metodologi Penelitian

2.1 Objek Penelitian

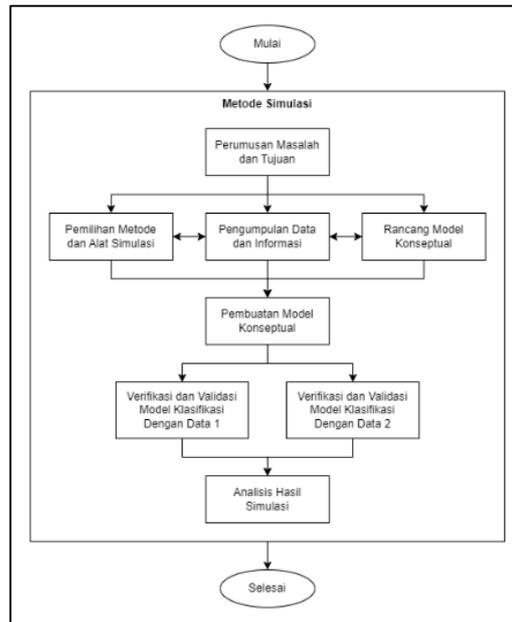
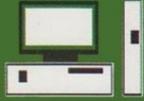
Objek penelitian disini yaitu dua model machine learning yaitu Naive Bayes, dan SVM. Kemudian satu model deep learning yaitu BERT yang akan dilatih dengan dua set data yang berbeda yaitu dataset kumpulan komentar yang diambil dari video YouTube "Resident Evil 4 Remake: Reveal Trailer" selama rentang waktu dari tanggal 3 Juni 2022 hingga 9 Februari 2024. Dan data dari "Steam Game Review Dataset".

2.2 Metode Pengumpulan Data

Pengumpulan data yang bersumber dari pustaka-pustaka baik literatur maupun sumber lainnya yang mengandung teori-teori yang digunakan sebagai dasar dari penelitian yaitu teori-teori terkait analisis sentimen, model klasifikasi Naive Bayes, SVM, BERT, dan Majority Voting serta konsep-konsep terkait penelitian. Data yang diolah bersumber pada video Youtube berjudul "Resident Evil 4 Remake: Reveal Trailer" yang diunggah oleh kanal IGN pada 3 Juni 2022 dengan scraping komentar yang dilakukan pada 9 Februari 2024. Komentar yang didapat berjumlah 7177 komentar dengan menggunakan App Script yang memanfaatkan fungsi Youtube API. komentar-komentar kemudian disimpan dalam file csv.

2.3 Metode Simulasi

Bagian ini, akan menjelaskan alur metode simulasi yang dilakukan untuk mencapai tujuan penelitian yang telah ditetapkan. Untuk alur penelitian yang dilakukan dapat dilihat pada Gambar 3.1 berikut.



Gambar 1. Alur Dalam Metode Simulasi

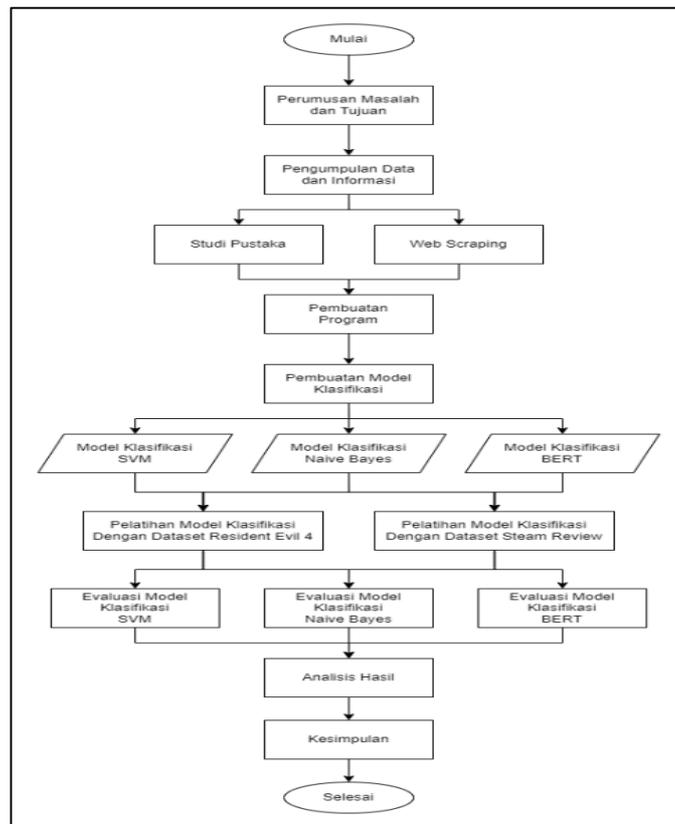
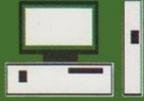
Perumusan Masalah dan Tujuan, Masalah dan tujuan dari penelitian akan dirumuskan sesuai dengan penjelasan pada Bab 1 dengan rumusan masalah "Bagaimana Melakukan Analisis Sentimen dan Emosi Dengan Metode Majority Voting Terhadap Komentar-Komentar Pada Video Youtube Resident Evil 4 Remake: Reveal Trailer Guna Menemukan Pola-Pola Sentiment Dan Emosi Yang Dominan Dalam Komunitas Gamer?" dengan batasan-batasan yang sebelumnya sudah disebutkan.

Rancang Model Konseptual, Pembuatan representasi dari sistem, konsep, yang akan diteliti dalam penelitian. Model ini membantu dalam pemahaman konsep secara lebih mendalam dan menyajikan struktur dasar dari berbagai elemen yang terlibat dalam studi ini. Pemilihan Metode dan Alat Simulasi, Metode simulasi dipilih berdasarkan batasan yang telah dijelaskan, yaitu dengan menggunakan metode majority voting yang didukung dengan metode Naive Bayes, SVM dan BERT. Sedangkan alat simulasi akan dibatasi sesuai dengan yang sudah dijelaskan pada bagian Analisis Kebutuhan Sistem sebelumnya baik perangkat keras maupun perangkat lunak.

Pengumpulan Data dan Informasi, Data-data beserta informasi-informasi yang terkait dengan penelitian dikumpulkan sesuai dengan yang sudah dijelaskan pada bagian Metode Pengumpulan Data Sebelumnya. Pembuatan Model Konseptual. Model konsep direalisasikan menggunakan alat-alat yang telah ditetapkan guna melakukan pembuatan model termasuk diantaranya penulisan kode, pelatihan dan uji model simulasi serta hal-hal lain yang mendukung jalannya model simulasi. Verifikasi dan Validasi Model Klasifikasi, Model klasifikasi yang dibuat akan diverifikasi dan divalidasi menggunakan metrik akurasi, presisi dan recall yang dapat menggambarkan kinerja dan performa model simulasi. Analisis Hasil Simulasi, Output dari simulasi dapat divisualisasi untuk memudahkan pemahaman akan performa model yang diteliti dan informasi yang dihasilkan sehingga dapat ditarik suatu kesimpulan.

2.4 Alur Penelitian

Dan untuk alur dari penelitian ini sendiri, alurnya secara terperinci dapat dilihat pada Gambar 3.2 dibawah ini.



Gambar 2. Alur-Alur Yang Dilakukan Pada Penelitian

3. Hasil dan Pembahasan

Analisis sentimen yang dilakukan memiliki input berupa data-data tekstual yang ada dalam kolom komentar youtube pada video "Resident Evil 4 Remake: Reveal Trailer " dari Channel IGN, yang akan diproses melalui model atau algoritma Naive Bayes, SVM, dan BERT yang menghasilkan sentiment "Negatif" atau "Positif" yang kemudian dilakukan majority voting untuk tiap hasilnya. Untuk melakukan validasi dari model yang telah dibuat digunakan pembagian data ke dalam data latih dan data uji dengan rasio 80:20 menggunakan dua dataset sebagai perbandingan.

3.1 Rancang Model Konseptual

Untuk tiap keperluan pada penelitian semua bagian-bagian yang diperlukan akan dibuat model konseptualnya sebelum direalisasikan untuk penggunaan untuk model konseptualnya sendiri antara lain adalah Model Konseptual Pengumpulan Data (Web Scraping), Model Konseptual Preprocessing, Model Konseptual Deteksi Emosi, Model Konseptual Pemberian Label, Model Konseptual Pembagian Data, Model Konseptual Klasifikasi.

3.2 Pengumpulan Data dan Informasi

Pengumpulan Data disini ada data yang sudah diberi label sebelumnya yang didapat dari situs Kaggle berupa dataset Review Game pada Steam dengan jumlah data 17.494 yang nantinya akan digunakan untuk pelatihan model sebagai perbandingan. Berikut adalah beberapa contoh dari "Steam Game Review Dataset". Untuk pengumpulan data ke dua dilakukan dengan melalui scraping web dengan menggunakan App Script terhadap video dari Channel IGN yang berjudul "Resident Evil 4 Remake: Reveal Trailer" di YouTube pada 9 Februari 2024 dengan memanfaatkan youtube api untuk scraping komentar-komentar pada video nya, kemudian didapatlah data teks berjumlah 7177 Komentar.

<https://doi.org/10.47111/JTI>

Available online at <https://e-journal.upr.ac.id/index.php/JTI>

3.3 Pemilihan Metode dan Alat Simulasi

Berdasarkan apa yang telah dijelaskan mengenai rumusan masalah dan tujuan penelitian sebelumnya maka dipilihlah metode klasifikasi teks, perangkat lunak dan perangkat keras yang akan digunakan pada penelitian ini yaitu. Metode klasifikasi teks berdasarkan penjelasan pada bab sebelumnya digunakan metode Majority Voting dengan menggabungkan penggunaan algoritma:

- NB (Naive Bayes).
- SVM (Support Vector Machine).
- BERT (Bidirectional Encoder Representations from Transformers).

Sedangkan rincian perangkat keras dan perangkat lunak yaitu device Acer Nitro AN515-51 dan Google Chrome beserta alat-alat lainnya seperti Google Colab, App Script, dll.

3.4 Pembuatan Model Konseptual

Berdasarkan Model Konseptual yang telah dibuat sebelumnya sebagai rencana awal, pembuatannyapun direalisasikan dan dijalankan sesuai dengan bagiannya masing-masing yaitu sebagai berikut.

3.4.1 *Preprocessing*

Data-data mentah hasil dari Web Scraping kemudian dibersihkan, dan disaring dalam proses preprocessing mulai dari proses pembersihan pertama, menghapus Tag HTML, Menjemahkan Emoji, dan pembuangan nilai Null pada data untuk mempermudah proses pendeteksian bahasa. Dari proses Cleaning didapatkan 7172 data yang sebelumnya berjumlah 7177. Data ini kemudian diolah kembali dan dideteksi bahasanya. Data selain Bahasa Inggris kemudian dibuang untuk mengurangi jumlah data dan didapatlah data dengan jumlah 5338. Dan pada bagian terakhir proses ini yaitu penghapusan Stopword atau kata-kata umum yang sering muncul dalam teks, yang tidak memberikan kontribusi signifikan terhadap makna teks kemudian lematisasi mengubah kata-kata dalam teks menjadi bentuk dasarnya, pada bagian ini juga dilakukan penghapusan data dengan nilai Null dan jumlah akhir data yang didapat dari proses ini adalah 5302 Data. Untuk sample data sebelum dan sesudah proses dapat dilihat pada Tabel 4 dibawah.

3.4.2 Deteksi Emosi

Data-data yang sudah mengalami proses preprocessing sebelumnya kemudian digunakan untuk dilakukan deteksi emosi. Untuk setiap kata dalam teksnya, jika kata tersebut ditemukan dalam NRC Lexicon, nilai emosi yang sesuai dengan kata tersebut diakumulasi. Untuk Langkah-langkahnya sendiri pertama tiap emosi akan menyimpan bobot yang berbeda untuk skor emosinya antara lain yaitu anger: 0.80, anticipation: 0.95, disgust: 0.89, fear: 0.91, joy: 1, sadness: 0.83, surprise: 0.93, trust: 0.85. Kemudian Teks di-tokenisasi menggunakan fungsi `word_tokenize` dari pustaka NLTK. Setelah itu dilakukan pencocokan kata-kata dalam teks dengan kamus NRC Lexicon yang berisi kata-kata bersama dengan emosi yang terkait. Untuk setiap kata yang cocok, skor emosi ditambahkan dengan bobot yang sesuai. Akhirnya, emosi dengan skor tertinggi diambil sebagai hasil emosi utama dari teks. Untuk hasilnya dapat dilihat pada Tabel 5 dibawah ini.

3.4.3 Pemberian Label

Untuk penglabelan digunakan juga data-data yang sudah mengalami proses preprocessing sebelumnya. penglabelan bertujuan untuk mempermudah proses pelatihan untuk model-model yang akan digunakan nantinya. Pemberian label sentimen dilakukan dengan menggunakan metode VADER. Metode ini menghitung skor sentimen keseluruhan dari sebuah teks dengan cara menjumlahkan nilai valensi dari setiap kata di dalamnya. Dari hasil perhitungan ini, sebuah label sentimen (positif, atau negatif) dapat ditentukan. Pada proses ini juga data-data yang tidak memiliki nilai sentimen dihapus kemudian dari perhitungan menggunakan VADER ini dihasilkanlah data berjumlah 3688 dengan jumlah data positif 2511 dan jumlah data negatif 1177 berikut adalah sedikit sample data yang sudah diberi label.

3.4.4 Pembagian Data

Kedua dataset yaitu “Steam Game Review Dataset” dan dataset “Resident Evil 4 Remake: Reveal Trailer” yang sudah memiliki label kemudian, dibagi menjadi data latih dan data uji. Tapi sebelum itu, akan dilakukan downsampling kepada data-data tersebut guna mengurangi sampel yang bertujuan untuk menghasilkan data latih yang memiliki label atau kelas yang simetris antara label positif dan label negative untuk hasilnya dapat dilihat pada table 1 dan 2 berikut.

Tabel 1. Pembagian Data Resident Evil 4 Remake: Reveal Trailer.

Klasifikasi	Jumlah	Data Latih (80%)	Data Uji (20%)
Positif	1000	801	199
Negatif	1000	799	201
Total	2000	1600	400

Tabel 2. Pembagian Data Steam Game Review Dataset.

Klasifikasi	Jumlah	Data Latih (80%)	Data Uji (20%)
Positif	7526	6052	1474
Negatif	7526	5989	1537
Total	15052	12041	3011

3.5 Klasifikasi

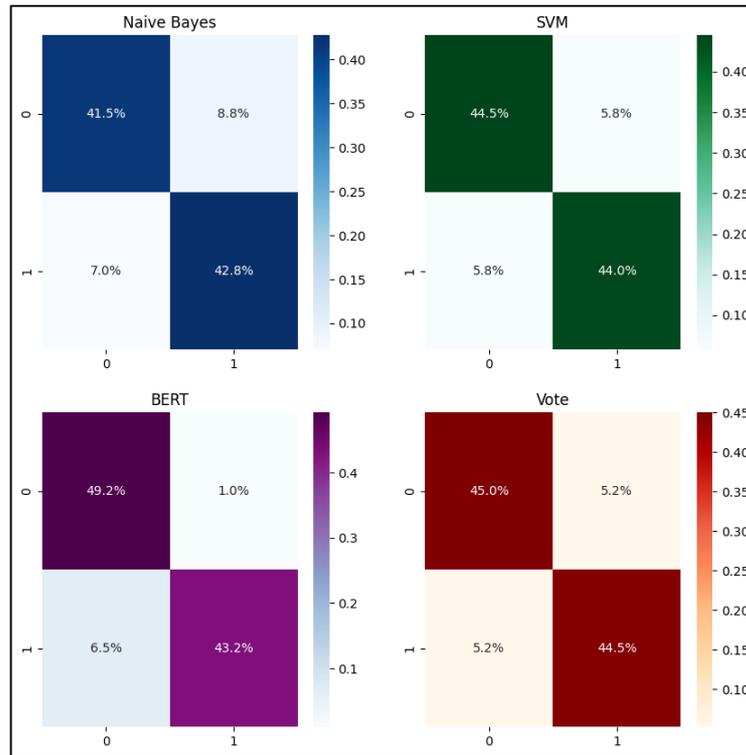
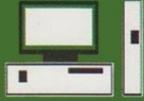
Pada pembuatan model untuk Klasifikasi digunakan dua dataset yang sudah disebutkan sebelumnya yaitu dataset “Steam Game Review Dataset” dan dataset “Resident Evil 4 Remake: Reveal Trailer” Kedua dataset ini digunakan untuk pelatihan klasifikasi pada model Naive Bayes, SVM, dan BERT. Kemudian hasil dari prediksi ketiga model dimasukkan kedalam Majority Voting guna menggabungkan prediksi dari tiga model berbeda dan melakukan perhitungan suara mayoritas untuk menentukan prediksi akhir untuk setiap data. Dari hasil metode ini didapatlah hasil sebagai berikut.

3.6 Verifikasi dan Validasi

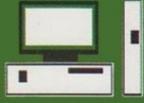
Model yang sudah dilatih untuk klasifikasi kemudian diverifikasi, untuk memastikan apakah dapat bekerja sesuai dengan apa yang diharapkan. Disini diharapkan model yang sudah dilatih bisa menghasilkan output berupa hasil analisis sentiment dengan label positif atau negatif yang dapat menggambarkan sentimen pengguna youtube terhadap pengumuman Resident Evil 4 Remake dari video "Resident Evil 4 Remake: Reveal Trailer". Dari hasil pelatihan menggunakan dua dataset yang sudah disebutkan sebelumnya, model menghasilkan output yang diinginkan sesuai seperti yang terlihat pada hasil pembuatan model klasifikasi sebelumnya. Untuk validasi model berdasarkan dataset yang digunakan untuk pelatihannya adalah sebagai berikut.

3.6.1 Dataset Resident Evil 4 Remake

Dataset ini berisi teks yang didapat dari kolom komentar video youtube "Resident Evil 4 Remake: Reveal Trailer" dari hasil validasi menggunakan metode evaluasi confusion matrix untuk menunjukkan tingkat performa model didapat hasil seperti pada gambar 4.7 dibawah.



Gambar 3. Confusion Matrix Dataset Resident Evil 4 Remake



Kemudian untuk hasil dari Classification Report tiap model yang digunakan untuk melihat tingkat akurasi, recall, presisi dan nilai F1 didapatkan nilai-nilai sebagai berikut.

Tabel 3. Classification Report Naive Bayes I.

	precision	recall	f1-score	support
negatif	0.86	0.83	0.84	201
positif	0.83	0.86	0.84	199
accuracy			0.84	400
macro avg	0.84	0.84	0.84	400
weighted avg	0.84	0.84	0.84	400

Tabel 4. Classification Report SVM I.

	precision	recall	f1-score	support
negatif	0.89	0.89	0.89	201
positif	0.88	0.88	0.88	199
accuracy			0.89	400
macro avg	0.88	0.88	0.88	400
weighted avg	0.89	0.89	0.89	400

Tabel 5. Classification Report BERT I.

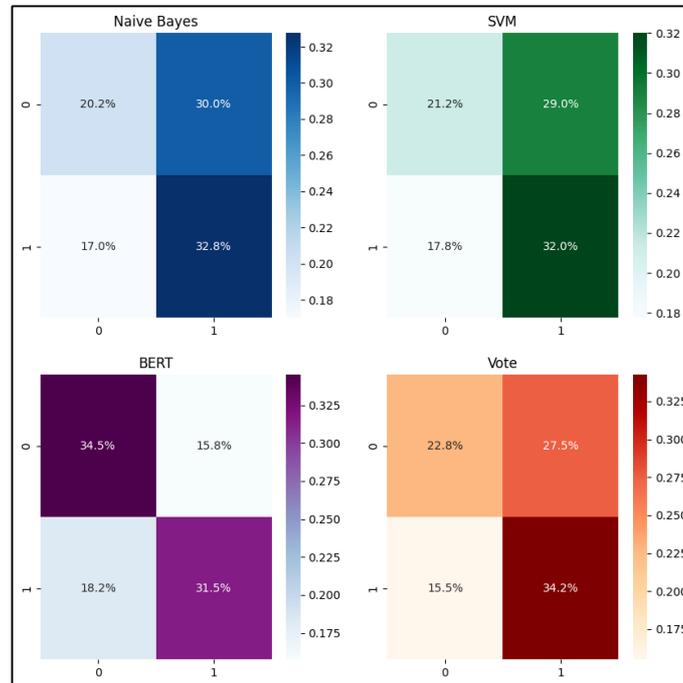
	precision	recall	f1-score	support
negatif	0.88	0.98	0.93	201
positif	0.98	0.87	0.82	199
accuracy			0.93	400
macro avg	0.93	0.92	0.92	400
weighted avg	0.93	0.93	0.92	400

Tabel 6. Classification Report *Majority Voting* I.

	precision	recall	f1-score	support
negatif	0.90	0.90	0.90	201
positif	0.89	0.89	0.89	199
accuracy			0.90	400
macro avg	0.89	0.89	0.89	400
weighted avg	0.90	0.90	0.90	400

3.6.2 Dataset Steam Game Review

Data yang dipakai untuk pelatihan selanjutnya yaitu “Steam Game Review Dataset” yang berisi data teks dari review game di steam, hasil validasi menggunakan metode evaluasi confusion matrix untuk menunjukkan tingkat performa model didapat hasil seperti pada gambar 4.8 berikut.



Gambar 4. Confusion Matrix Dataset Steam Game Review

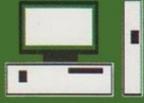
Kemudian untuk hasil dari Classification Report tiap model yang digunakan untuk melihat tingkat akurasi, recall, presisi dan nilai F1 didapatlah nilai-nilai sebagai berikut.

Tabel 7. Classification Report Naive Bayes II.

	precision	recall	f1-score	support
negatif	0.54	0.40	0.46	201
positif	0.52	0.66	0.58	199
accuracy			0.53	400
macro avg	0.53	0.53	0.52	400
weighted avg	0.53	0.53	0.52	400

Tabel 8. Classification Report SVM II.

	precision	recall	f1-score	support
negatif	0.54	0.42	0.48	201
positif	0.52	0.64	0.58	199
accuracy			0.53	400
macro avg	0.53	0.53	0.53	400



weighted avg	0.53	0.53	0.53	400
--------------	------	------	------	-----

Table 9. Classification Report BERT II.

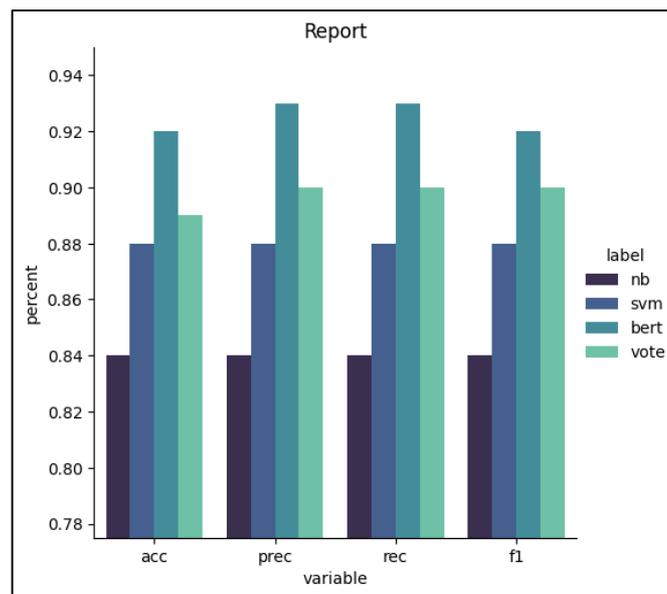
	precision	recall	f1-score	support
negatif	0.65	0.69	0.67	201
positif	0.67	0.63	0.65	199
accuracy			0.66	400
macro avg	0.66	0.66	0.66	400
weighted avg	0.66	0.66	0.66	400

Tabel 10. Classification Report Majority Voting II.

	precision	recall	f1-score	support
negatif	0.59	0.45	0.51	201
positif	0.55	0.69	0.61	199
accuracy			0.57	400
macro avg	0.57	0.57	0.56	400
weighted avg	0.57	0.57	0.56	400

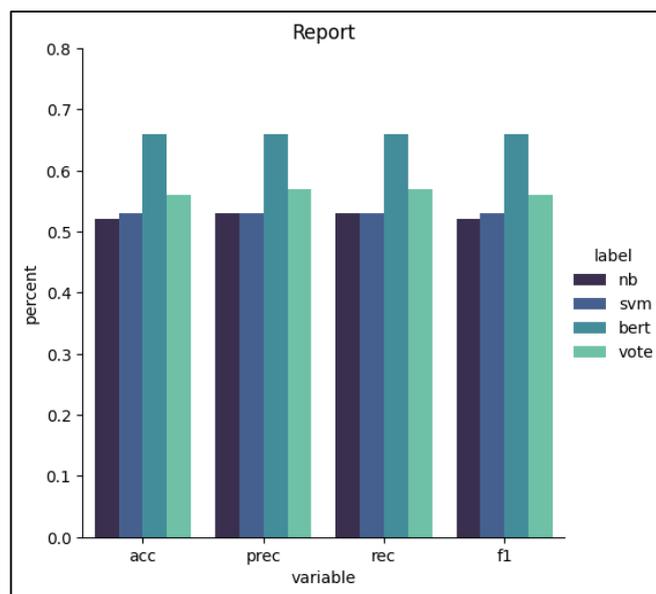
3.7 Analisis Hasil Simulasi

Setelah semua langkah-langkah dalam metode simulasi telah dijalankan didapatkan tingkat akurasi masing-masing metode klasifikasi yang dilatih dengan kedua dataset “Dataset Resident Evil 4 Remake” dan “Dataset Steam Game Review”, dari metode klasifikasi dengan model Naive Bayes, SVM, BERT dan majority voting didapat hasil yang dapat dilihat pada Gambar 5 dan 6 berikut ini.



Gambar 5. Hasil Performa Dengan Dataset Resident Evil

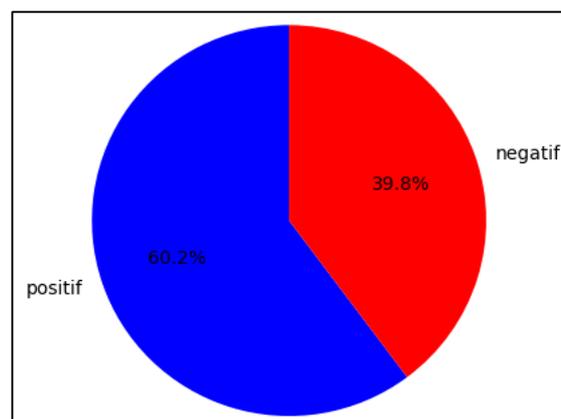
Pada klasifikasi dengan dataset Resident Evil 4 Remake, Naive Bayes memiliki performa terkecil dengan akurasi sebesar 84% dengan nilai presisi dan recall masing-masing mencapai 84%. SVM memiliki nilai akurasi sebesar 89% dengan nilai presisi dan recall masing-masing mencapai 88%. Kemudian BERT dengan performa terbesar dengan nilai akurasi mencapai 93%, dengan nilai presisi dan recall mencapai 90% dan untuk hasil klasifikasi dari penggabungan ketiga model dengan Metode Majority Voting didapat akurasi sebesar 90% dengan nilai presisi dan recall masing-masing mencapai 89%.



Gambar 6. Hasil Dengan Dataset Steam Game Review

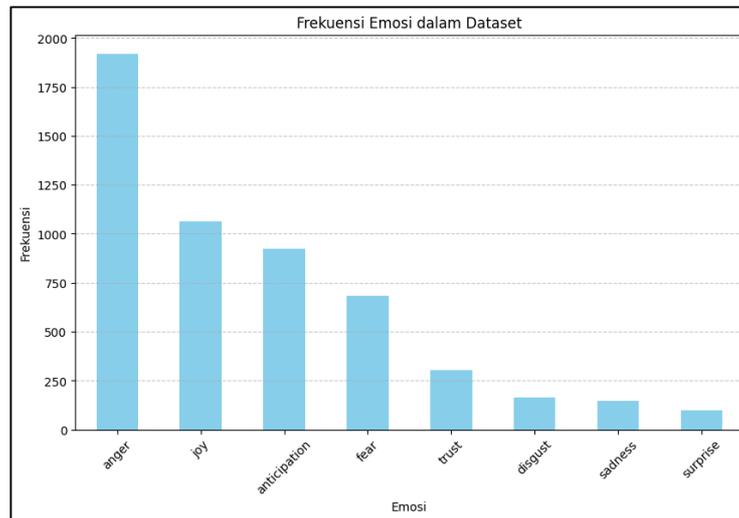
Pada klasifikasi dengan dataset Steam Game Review didapat hasil performa yang kecil, Naive Bayes dan SVM memiliki performa yang hampir sama dengan akurasi sebesar 53% dengan nilai presisi dan recall masing-masing mencapai 53% juga. Sedangkan BERT dengan performa terbesar dengan nilai akurasi mencapai 66%, dengan nilai presisi dan recall mencapai 66% dan untuk hasil klasifikasi dari penggabungan ketiga model dengan Metode Majority Voting didapat akurasi sebesar 57% dengan nilai presisi dan recall masing-masing mencapai 57%.

Dari hal ini maka diambilah model klasifikasi Majority Voting yang dilatih dengan dataset Resident Evil 4 Remake yang memiliki tingkat akurasi mencapai 90% untuk melakukan klasifikasi Analisis Sentimen pada komentar-komentar dari video YouTube berjudul "Resident Evil 4 Remake: Reveal Trailer" dari Channel IGN dari hasil klasifikasinya didapatlah mayoritas sentiment pada video tersebut adalah Positif dengan rasio sentiment positif dan negatif 60,2% dan 39,8% seperti pada Gambar 7.



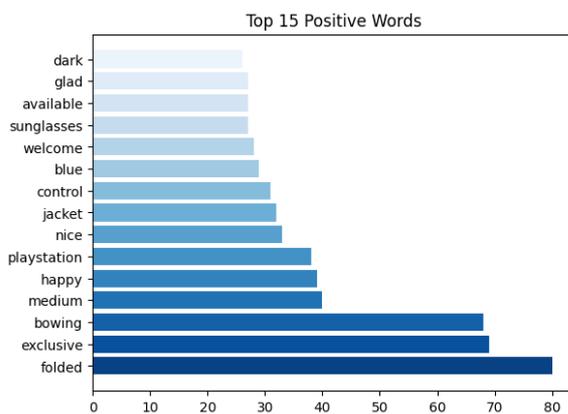
Gambar 7. Rasio Dari Hasil Analisis Sentimen

Selain itu juga dilakukan perhitungan untuk frekuensi emosi yang muncul dari hasil deteksi emosi yang dilakukan diawal, didapat emosi marah (anger) yang paling banyak muncul disusul dengan Joy, Anticipation, dan Fear.

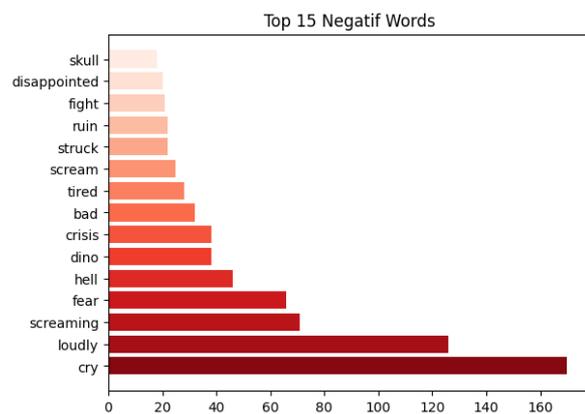


Gambar 8. Frekuensi Emosi Yang Muncul Dari Deteksi Emosi

Pada Analisis Sentimen ini didapat daftar kata-kata yang paling sering muncul pada komentar yang merepresentasikan kata positif dan negatif pada kalimat yang di analisis kata-kata ini dikelompokkan dan dilist menjadi 15 kata positif dan negatif yang paling sering muncul.



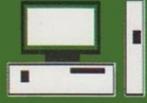
Gambar 9. Kata Positif Yang Sering Muncul



Gambar 10. Kata Negatif Yang Sering Muncul

4. Kesimpulan

Berdasarkan hasil penelitian pada Skripsi/Tugas Akhir dengan judul “Analisis Sentimen Dan Emosi Dengan Metode Majority Voting Terhadap Komentar Video Youtube Resident Evil 4 Remake: Reveal Trailer”, dapat ditarik kesimpulan. Model yang dilatih dengan dataset Resident Evil 4 Remake memiliki performa yang jauh lebih baik daripada pelatihan dengan dataset Steam Review ini dikarenakan mulai dari dataset Resident Evil 4 Remake yang fokus pada game itu sendiri sedangkan Steam Review memiliki data dengan bermacam game didalamnya. Kemudian hasil pelatihan dengan



model BERT mempunyai nilai akurasi mencapai 93% namun hasil dari Metode Majority Voting didapat akurasi yang sedikit lebih kecil dari BERT yaitu sebesar 90% dikarenakan terpengaruh dari performa model lain yang memiliki akurasi lebih kecil seperti 84% untuk Naive Bayes dan 89% untuk SVM.

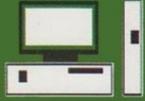
Dari 5302 komentar di video YouTube berjudul "Resident Evil 4 Remake: Reveal Trailer" dari Channel IGN yang sudah dibersihkan didapat mayoritas orang-orang menanggapi pengumuman ini dengan positif dengan rasio sentiment positif dan negatif 60,2% dan 39,8% meskipun untuk frekuensi emosi yang paling sering muncul adalah marah (anger) disusul dengan Joy, Anticipation, dan Fear besar kemungkinan hal ini disebabkan karena pendekatan yang digunakan untuk deteksi emosi menggunakan NRC Emotion Lexicon yang mana melakukan pencocokkan kata-kata dalam teks dengan emosi tertentu yang terdaftar dalam leksikon dan dapat diketahui Resident Evil 4 adalah permainan aksi horor yang terkenal dengan suasana yang intens dan lumayan dipenuhi dengan kekerasan dan seringkali menegangkan.

Daftar Pustaka

- [1] Turney, P. D., "Sentiment analysis: A survey and framework," arXiv preprint arXiv:2001.08598, 2020.
- [2] Zendesk, "The Power Of Sentiment Analysis For Customer Support," 2021.
- [3] Mohammad, S. M., Saif, H., & Turney, P. D., Sentiment analysis: A survey. *ACM Transactions on Information Systems (TOIS)*, 36(4), 1-37.K, 2018.
- [4] Feldman, R., Sentiment analysis: A practical guide, Cambridge University Press, 2022.
- [5] Kim, S. M., & Hovy, E., Determining the sentiment of text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 136-143). Association for Computational Linguistics, 2020.
- [6] Yang, S., Wang, X., & Liu, S., BERT for sentiment analysis of social media posts, In *Proceedings of the 16th International Conference on Advanced Communications Technology (ICT)* (pp. 1-5). IEEE. <https://ieeexplore.ieee.org/document/9545261>, 2023.
- [7] Capcom, Resident Evil 4, <https://www.residentevil.com/re4/>. 2023.
- [8] MonkeyLearn, Sentiment analysis Guide, MonkeyLearn, <https://monkeylearn.com/sentiment-analysis/>, 2022.
- [9] Hamborg, F., & Donnay, K., NewsMTSC: A Dataset for (Multi-)Target- dependent Sentiment Classification in Political News Articles, *ACL Anthology, Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*: 10.18653/v1/2021.eacl-main.142 Main Volume, 1663-1675, 2021.
- [10] Dicoding Intern, Apa itu Machine learning? Beserta Pengertian dan Cara Kerjanya, Dicoding, <https://www.dicoding.com/blog/machine-learning-adalah/>, 2020.
- [11] Ranganathan, S., Nakai, K., & Schonbach, C. (Eds.), *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, Elsevier Science, 2018.
- [12] Kasim, A. A., & Sudarsono, M, Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di SEMNASTIK, Kecamatan Simpang Raya Sulawesi Tengah, <http://publikasi.dinus.ac.id/index.php/semnastik/article/view/2922>, 2019.
- [13] Alaydrus, A. A., Penerapan Algoritma BERT dalam Search Engine Google, MTI, <https://mti.binus.ac.id/2020/09/03/penerapan-algoritma-bert-dalam-search-engine-google/>, 2020.
- [14] Nair, A., Combine Your Machine learning Models With Voting | by Aashish Nair, Towards Data Science, <https://towardsdatascience.com/combine-your-machine-learning-models-with-voting-fa1b42790d84>, 2021.
- [15] Diouf, R., Sarr, A. N., Sall, O., Birregah, B., Bouso, M., & Mbaye, S. N., Web scraping: State-of-the-Art and Areas of Application. *IEEE Xplore, 2019 IEE International Conference* <https://ieeexplore.ieee.org/abstract/document/9005594/citations>, 2019.

<https://doi.org/10.47111/JTI>

Available online at <https://e-journal.upr.ac.id/index.php/JTI>



JURNAL TEKNOLOGI INFORMASI

Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika

[E-ISSN 2656-0321]

[Vol 19 No 2]

[Agustus 2025]