Sentiment Analysis of YouTube Comments Using XLM-RoBERTa to Support the Understanding of Multilingual Public Opinion

Kevin R. K. Meliala[1], Dhafin W. Wicaksono [1], Jonathan E. Siregar[1], Naufal D. Prakintara[1], and Raafi A. Ramadhan [1]

[1] School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

kevin.meliala@binus.ac.id; dhafin.wicaksono@binus.ac.id; jonathan.siregar001@binus.ac.id; naufal.prakintara@binus.ac.id; raafi.ramadhan@binus.ac.id

Received: 13th July 2026/ Revised: 24th July 2026/ Published: 31st August 2026

Abstrak

YouTube merupakan salah satu platform media sosial terbesar yang menghasilkan jutaan komentar pengguna setiap harinya. Komentar-komentar tersebut mengandung berbagai opini, tanggapan, dan persepsi yang dapat dimanfaatkan untuk memahami respons publik terhadap suatu konten, produk, maupun layanan. Namun, jumlah komentar yang sangat besar membuat proses analisis secara manual menjadi tidak efektif dan memerlukan waktu yang lama. Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen komentar YouTube berbasis Natural Language Processing (NLP) menggunakan model XLM-RoBERTa yang mampu melakukan klasifikasi sentimen secara otomatis pada lingkungan multibahasa. Sistem dikembangkan menggunakan FastAPI sebagai backend, React sebagai frontend, dan YouTube Data API v3 sebagai sumber pengambilan data komentar. Proses analisis meliputi pengambilan komentar, preprocessing, filtering spam dan noise, inferensi model, heuristic post-processing, serta visualisasi hasil melalui dashboard interaktif. Hasil penelitian menunjukkan bahwa sistem mampu mengelompokkan komentar ke dalam kategori positif dan negatif dengan performa yang baik. Dashboard juga menyediakan berbagai fitur analisis seperti distribusi

sentimen, word cloud, pencarian komentar, serta top comments berdasarkan confidence score. Kebaruan penelitian ini terletak pada integrasi model multilingual XLM-RoBERTa dengan mekanisme heuristic post-processing untuk meningkatkan kualitas klasifikasi komentar berbahasa Indonesia dan campuran bahasa. Sistem yang dikembangkan diharapkan dapat membantu pengguna, peneliti, maupun organisasi dalam memahami opini publik secara lebih cepat, akurat, dan efisien.

Kata Kunci: Analisis Sentimen, YouTube, XLM-RoBERTa, Natural Language Processing, Multilingual

Abstract

YouTube is one of the largest social media platforms, generating millions of user comments every day. These comments contain various opinions, responses, and perceptions that can be utilized to understand public reactions toward content, products, and services. However, the massive volume of comments makes manual analysis inefficient and time-consuming. This study aims to develop a YouTube comment sentiment analysis system based on Natural Language Processing (NLP) using the XLM-RoBERTa model to automatically classify sentiment in a multilingual environment. The system is developed using FastAPI as the backend, React as the frontend, and the YouTube Data API v3 as the source for retrieving comment data. The analysis process includes comment retrieval, preprocessing, spam and noise filtering, model inference, heuristic post-processing, and result visualization through an interactive dashboard. The results demonstrate that the system is capable of classifying comments into positive and negative sentiment categories with strong performance. The dashboard also provides various analytical features, including sentiment distribution, word cloud visualization, comment search, and top comments ranked by confidence score. The novelty of this research lies in the integration of the multilingual XLM-RoBERTa model with a heuristic post-processing mechanism to improve the classification quality of Indonesian and code-mixed language comments. The developed system is expected to assist users, researchers, and organizations in understanding public opinion more quickly, accurately, and efficiently.

***Keywords**: Sentiment Analysis, YouTube, XLM-RoBERTa, Natural Language Processing, Multilingual*

Pendahuluan

Perkembangan teknologi digital telah mengubah cara masyarakat berkomunikasi, berbagi informasi, dan menyampaikan opini. Media sosial menjadi salah satu sarana utama yang digunakan masyarakat untuk mengekspresikan pandangan mereka terhadap berbagai isu, produk, layanan, maupun konten digital (Liu, 2020). Di antara berbagai platform media sosial yang tersedia saat ini, YouTube merupakan salah satu platform terbesar yang memiliki jutaan pengguna aktif di seluruh dunia. Selain berfungsi sebagai media berbagi video, YouTube juga menyediakan fitur komentar yang memungkinkan pengguna memberikan tanggapan secara langsung terhadap suatu video.

Komentar yang dihasilkan pengguna mengandung informasi yang sangat berharga karena mencerminkan persepsi, kepuasan, kritik, maupun dukungan terhadap suatu topik tertentu. Bagi perusahaan, kreator konten, peneliti, maupun organisasi, informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi dan pengambilan keputusan. Namun, jumlah komentar yang sangat besar membuat proses analisis secara manual menjadi sulit dilakukan. Oleh karena itu, diperlukan suatu sistem yang mampu melakukan analisis komentar secara otomatis sehingga informasi penting dapat diperoleh dengan lebih cepat dan efisien.

Salah satu pendekatan yang banyak digunakan untuk memahami opini pengguna adalah analisis sentimen (Liu, 2020). Analisis sentimen merupakan cabang dari Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen yang terkandung dalam suatu teks ke dalam kategori tertentu seperti positif atau negatif (Liu, 2020; Semary et al., 2023). Perkembangan teknologi NLP telah memungkinkan sistem komputer untuk memahami konteks bahasa manusia dengan tingkat akurasi yang semakin baik.

Dalam beberapa tahun terakhir, arsitektur transformer telah menjadi standar utama dalam berbagai tugas NLP karena kemampuannya memahami hubungan kontekstual antar kata secara lebih efektif dibandingkan metode tradisional (Vaswani et al., 2017). Model seperti BERT dan berbagai turunannya menunjukkan performa yang sangat baik dalam tugas klasifikasi teks, termasuk analisis sentimen (Devlin et al., 2019). Salah satu model yang banyak digunakan

adalah XLM-RoBERTa, yaitu model transformer multilingual yang dilatih menggunakan data dalam lebih dari seratus bahasa (Conneau et al., 2020). Penelitian terbaru menunjukkan bahwa model multilingual berbasis transformer memiliki performa yang sangat baik pada tugas klasifikasi sentimen lintas bahasa dan code-mixed text (Barbieri et al., 2021; Hashmi et al., 2024). Kemampuan tersebut menjadikan XLM-RoBERTa sangat cocok untuk menganalisis komentar YouTube yang sering kali mengandung campuran Bahasa Indonesia, Bahasa Inggris, maupun bentuk code-switching lainnya.

Meskipun berbagai penelitian mengenai analisis sentimen telah dilakukan sebelumnya, masih terdapat beberapa tantangan yang perlu diatasi. Komentar pada media sosial sering kali mengandung spam, humor, sarkasme, singkatan, emoji, dan penggunaan bahasa tidak baku yang dapat memengaruhi kualitas klasifikasi. Selain itu, banyak sistem analisis sentimen hanya berfokus pada proses klasifikasi tanpa menyediakan visualisasi yang memudahkan pengguna memahami hasil analisis secara menyeluruh. Akibatnya, hasil klasifikasi yang diperoleh sering kali sulit dimanfaatkan secara optimal oleh pengguna non-teknis.

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan sistem analisis sentimen komentar YouTube berbasis XLM-RoBERTa yang terintegrasi dengan dashboard visualisasi interaktif. Sistem tidak hanya melakukan klasifikasi sentimen secara otomatis, tetapi juga menerapkan proses filtering dan heuristic post-processing untuk mengurangi kesalahan klasifikasi yang umum terjadi pada komentar media sosial. Hasil analisis kemudian disajikan dalam bentuk visualisasi yang mudah dipahami, termasuk distribusi sentimen, word cloud, serta daftar komentar dengan tingkat keyakinan tertinggi.

Penelitian ini memiliki beberapa kontribusi utama. Pertama, penelitian mengimplementasikan model XLM-RoBERTa untuk melakukan analisis sentimen komentar YouTube dalam lingkungan multibahasa. Kedua, penelitian menerapkan mekanisme heuristic post-processing untuk meningkatkan kualitas prediksi pada komentar yang mengandung sarkasme, humor, maupun ekspresi informal. Ketiga, penelitian mengintegrasikan proses analisis dengan dashboard interaktif yang memungkinkan pengguna memahami hasil secara cepat dan intuitif. Dengan demikian, sistem yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi sentimen, tetapi juga sebagai platform analisis opini publik yang lebih komprehensif.

Berdasarkan latar belakang tersebut, tujuan penelitian ini adalah mengembangkan dan mengevaluasi sistem analisis sentimen komentar YouTube berbasis XLM-RoBERTa yang mampu melakukan klasifikasi sentimen secara otomatis, mendukung berbagai bahasa, serta menyajikan hasil analisis dalam bentuk visualisasi yang informatif.

Kebaruan penelitian ini terletak pada pengembangan sistem terintegrasi yang menggabungkan pengambilan data komentar YouTube secara otomatis, analisis sentimen multibahasa menggunakan XLM-RoBERTa, heuristic post-processing, serta dashboard visualisasi interaktif dalam satu platform. Berbeda dengan penelitian sebelumnya yang umumnya hanya berfokus pada proses klasifikasi sentimen, penelitian ini juga menekankan aspek visualisasi dan kemudahan interpretasi hasil bagi pengguna.

Tinjauan Pustaka

Analisis Sentimen

Analisis sentimen merupakan proses identifikasi dan klasifikasi opini, emosi, atau sikap yang terkandung dalam suatu teks (Liu, 2020). Analisis sentimen banyak digunakan untuk memahami persepsi masyarakat terhadap suatu produk, layanan, organisasi, maupun isu tertentu. Dalam praktiknya, analisis sentimen umumnya mengelompokkan teks ke dalam kategori positif atau negatif. Seiring meningkatnya jumlah data yang dihasilkan pengguna internet, analisis sentimen menjadi salah satu bidang penelitian yang berkembang pesat dalam NLP. Perkembangan terbaru menunjukkan bahwa model transformer berbasis RoBERTa dan turunannya secara konsisten menghasilkan performa yang lebih baik dibandingkan metode machine learning tradisional dalam tugas sentiment classification pada media sosial (Semary et al., 2023; Jahin et al., 2024).

Pendekatan awal analisis sentimen umumnya menggunakan metode berbasis kamus (lexicon-based) dan algoritma machine learning tradisional seperti Naïve Bayes, Support Vector Machine (SVM), serta Random Forest. Meskipun metode tersebut relatif sederhana dan mudah diimplementasikan, performanya sering kali terbatas dalam memahami konteks kalimat yang kompleks. Perkembangan deep learning kemudian memungkinkan terciptanya model yang mampu memahami representasi semantik teks secara lebih baik.

Evaluasi model analisis sentimen pada media sosial juga didukung oleh keberadaan benchmark standar seperti TweetEval yang menyediakan berbagai tugas klasifikasi teks untuk mengukur performa model NLP pada data media sosial (Barbieri et al., 2020).

Natural Language Processing

Natural Language Processing merupakan bidang ilmu yang menggabungkan linguistik, ilmu komputer, dan kecerdasan buatan untuk memungkinkan komputer memahami bahasa manusia. NLP digunakan dalam berbagai aplikasi seperti chatbot, machine translation, text summarization, question answering, dan sentiment analysis.

Tahapan NLP umumnya mencakup preprocessing data, tokenisasi, ekstraksi fitur, pemodelan, dan evaluasi. Pada penelitian ini, NLP digunakan untuk memproses komentar YouTube yang memiliki karakteristik bahasa tidak formal, penggunaan singkatan, emoji, serta campuran beberapa bahasa dalam satu komentar.

Perkembangan NLP pada media sosial juga didukung oleh framework TweetNLP yang menyediakan berbagai model dan utilitas untuk pemrosesan teks pada platform digital seperti Twitter dan media sosial lainnya (Camacho-Collados et al., 2022).

Transformer Architecture

Transformer merupakan arsitektur deep learning yang diperkenalkan untuk mengatasi keterbatasan model berbasis recurrent neural network (RNN) (Vaswani et al., 2017). Keunggulan utama transformer terletak pada mekanisme self-attention yang memungkinkan model memahami hubungan antar kata dalam suatu kalimat secara paralel (Vaswani et al., 2017). Berbagai penelitian terkini menunjukkan bahwa arsitektur transformer tetap menjadi pendekatan dominan dalam berbagai tugas NLP karena kemampuannya menangkap konteks bahasa secara mendalam dan mendukung transfer learning pada berbagai bahasa (Zhang & Shafiq, 2024). Pendekatan ini menghasilkan performa yang lebih baik dalam berbagai tugas NLP sekaligus meningkatkan efisiensi proses pelatihan.

Keberhasilan transformer melahirkan berbagai model modern seperti BERT, RoBERTa, ALBERT, DistilBERT, dan XLM-RoBERTa (Devlin et al., 2019; Conneau et al.,

2020). Model-model tersebut telah digunakan secara luas dalam berbagai penelitian karena mampu menghasilkan representasi bahasa yang lebih kaya dibandingkan metode sebelumnya.

XLM-RoBERTa

Model yang digunakan dalam penelitian ini adalah *cardiffnlp/twitter-xlm-roberta-base-sentiment* yang tersedia melalui platform Hugging Face dan telah dilatih untuk tugas klasifikasi sentimen multibahasa (Hugging Face, 2026).

XLM-RoBERTa merupakan model transformer multilingual yang dikembangkan untuk mendukung pemrosesan berbagai bahasa dalam satu model (Conneau et al., 2020). Model ini dilatih menggunakan dataset multibahasa berskala besar sehingga mampu memahami konteks bahasa yang beragam (Conneau et al., 2020). Keunggulan utama XLM-RoBERTa adalah kemampuannya mempertahankan performa tinggi pada berbagai tugas NLP tanpa memerlukan model terpisah untuk setiap bahasa.

Dalam konteks analisis sentimen YouTube, kemampuan multilingual menjadi sangat penting karena komentar pengguna sering kali mengandung campuran Bahasa Indonesia, Bahasa Inggris, serta istilah informal lainnya. Oleh karena itu, penggunaan XLM-RoBERTa diharapkan mampu meningkatkan kualitas klasifikasi sentimen dibandingkan model monolingual.

XLM-RoBERTa dan model turunannya telah digunakan secara luas dalam penelitian sentiment analysis multilingual. Studi oleh Barbieri et al. (2021) menunjukkan bahwa model multilingual mampu mempertahankan performa yang baik pada berbagai bahasa sekaligus. Penelitian lain juga menunjukkan efektivitas transformer multilingual dalam menangani code-mixed text yang umum ditemukan pada media sosial (Hashmi et al., 2024).

Penelitian Terdahulu

Berbagai penelitian sebelumnya telah menerapkan analisis sentimen pada media sosial menggunakan pendekatan yang berbeda-beda. Sebagian penelitian menggunakan algoritma machine learning tradisional seperti SVM dan Naïve Bayes, sementara penelitian yang lebih

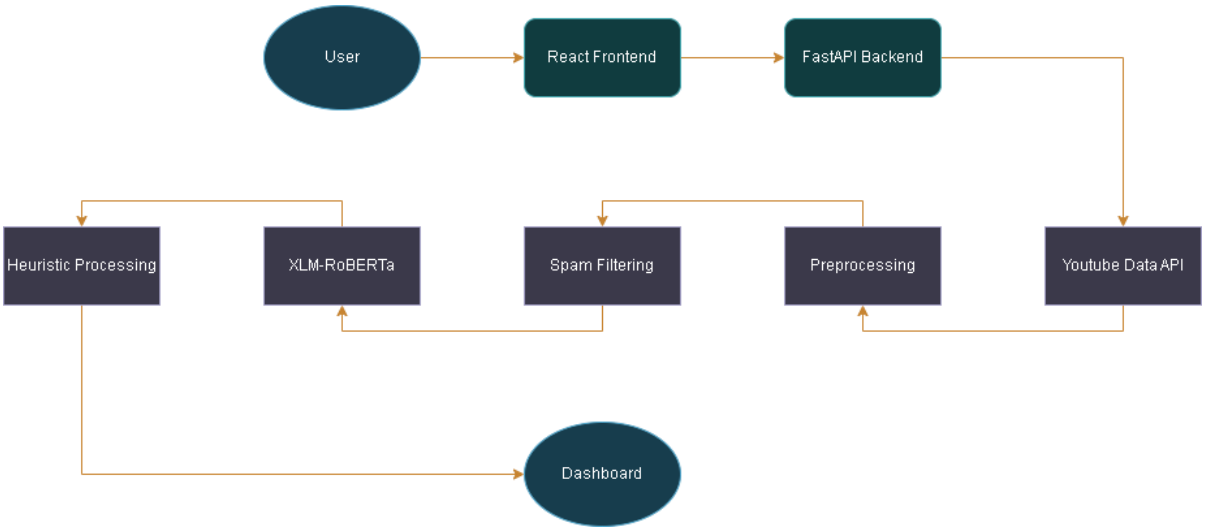
baru mulai memanfaatkan model transformer seperti BERT dan RoBERTa. Hasil penelitian menunjukkan bahwa model transformer umumnya menghasilkan tingkat akurasi yang lebih tinggi karena mampu memahami konteks bahasa secara lebih mendalam.

Penelitian terdahulu menunjukkan bahwa transformer multilingual memberikan hasil yang lebih baik dibandingkan pendekatan tradisional dalam berbagai tugas sentiment analysis (Mercha & Benbrahim, 2023). Namun, sebagian besar penelitian masih berfokus pada dataset Twitter dan belum banyak yang mengintegrasikan proses pengambilan data YouTube, analisis sentimen multibahasa, serta dashboard visualisasi interaktif dalam satu sistem yang terintegrasi. Selain itu, penelitian mengenai pemanfaatan XLM-RoBERTa pada komentar YouTube berbahasa Indonesia dan campuran bahasa masih relatif terbatas. Oleh karena itu, penelitian ini berupaya mengisi kesenjangan tersebut melalui pengembangan sistem analisis sentimen YouTube berbasis XLM-RoBERTa yang dilengkapi dengan fitur visualisasi dan heuristic post-processing.

Metodologi Penelitian

Desain Penelitian

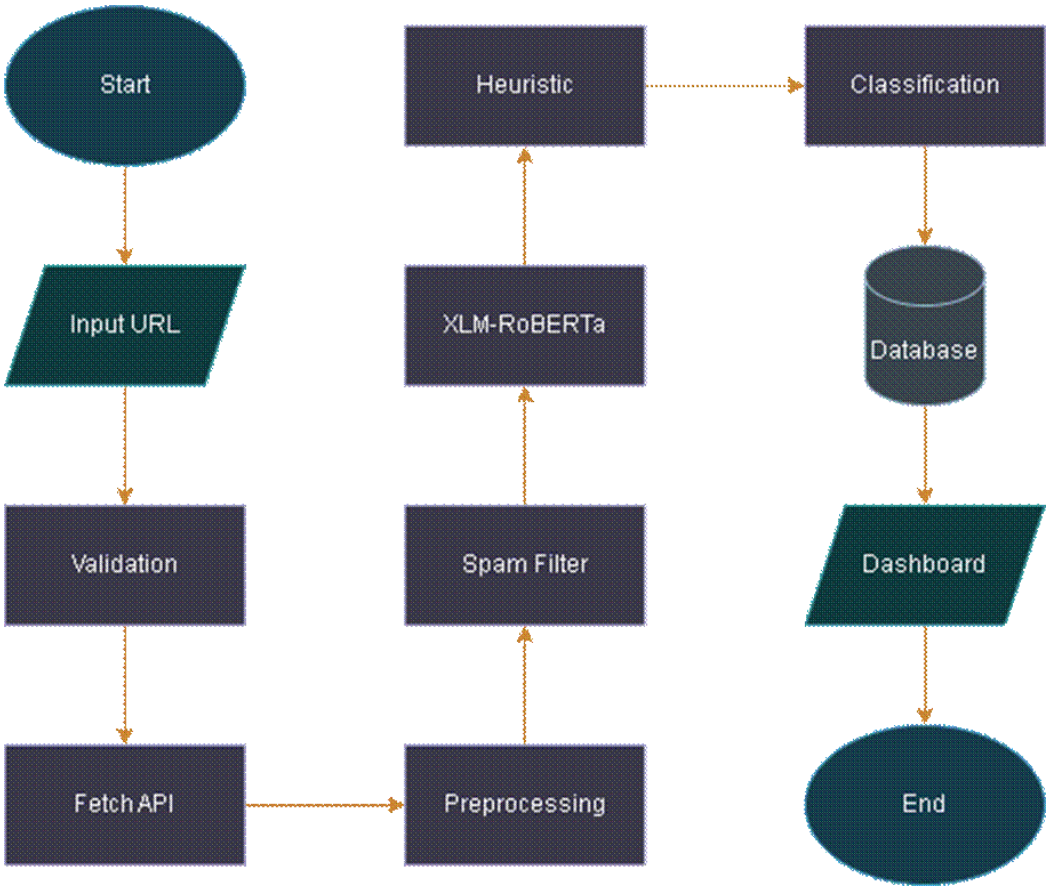
Penelitian ini menggunakan pendekatan pengembangan sistem (system development research) yang mengintegrasikan teknologi Natural Language Processing (NLP), machine learning, dan web application untuk melakukan analisis sentimen komentar YouTube secara otomatis. Tujuan utama penelitian adalah membangun sistem yang mampu mengambil komentar secara real-time, melakukan klasifikasi sentimen, serta menyajikan hasil analisis dalam bentuk dashboard interaktif yang mudah dipahami pengguna.



Gambar 1. Architecture of the YouTube Sentiment Analysis System

Pada gambar 1 ditunjukkan arsitektur sistem terdiri atas tiga komponen utama, yaitu frontend, backend, dan machine learning engine. Frontend dikembangkan menggunakan React dan Tailwind CSS untuk menyediakan antarmuka pengguna yang responsif. Backend dibangun menggunakan FastAPI yang berfungsi sebagai penghubung antara antarmuka pengguna, database, dan model analisis sentimen. Sementara itu, machine learning engine menggunakan model XLM-RoBERTa yang diimplementasikan melalui framework HuggingFace Transformers (Wolf et al., 2020).

Alur Proses Analisis Sentimen



Gambar 2. Sentiment Analysis Process Flow

Gambar 2 menunjukkan alur proses analisis sentimen yang diterapkan pada penelitian ini. Proses dimulai dengan pengguna memasukkan URL video YouTube yang akan dianalisis. Sistem kemudian mengambil komentar menggunakan YouTube Data API v3 dan melakukan tahap preprocessing untuk membersihkan data dari karakter yang tidak diperlukan, tautan, emoji tertentu, serta komentar yang terindikasi sebagai spam atau noise. Selanjutnya, komentar yang telah diproses dianalisis menggunakan model XLM-RoBERTa untuk mengklasifikasikan sentimen ke dalam kategori positif atau negatif. Hasil klasifikasi kemudian melalui tahap heuristic post-processing untuk memperbaiki prediksi pada komentar yang mengandung sarkasme, humor, atau pola bahasa tertentu. Setelah proses analisis selesai, hasil sentimen divisualisasikan melalui dashboard interaktif yang menampilkan distribusi sentimen, statistik analisis, dan informasi komentar sehingga pengguna dapat memahami hasil analisis dengan lebih mudah.

Sumber dan Pengumpulan Data

Data penelitian berupa komentar publik yang diperoleh dari video YouTube menggunakan YouTube Data API v3 (Google Developers, 2026). API tersebut memungkinkan sistem mengambil komentar berdasarkan URL video yang dimasukkan pengguna secara real-time (Google Developers, 2026).

Proses pengumpulan data dilakukan melalui beberapa tahapan sebagai berikut

1. Pengguna memasukkan URL video YouTube.
2. Sistem melakukan validasi URL.
3. Sistem mengakses YouTube Data API v3.
4. Komentar diambil secara bertahap sesuai batas API.
5. Data komentar disimpan sementara untuk diproses.

Komentar yang diperoleh terdiri atas berbagai bahasa, termasuk Bahasa Indonesia, Bahasa Inggris, dan campuran keduanya. Karakteristik ini menjadikan pendekatan multilingual sangat penting dalam penelitian.

Penelitian ini menggunakan dua sumber data. Dataset pertama merupakan dataset berlabel sebanyak 6.759 data yang digunakan untuk mengevaluasi performa model XLM-RoBERTa menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Dataset kedua berupa 5.000 komentar YouTube yang diperoleh melalui YouTube Data API v3 dan digunakan sebagai data implementasi sistem untuk analisis sentimen.

Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses klasifikasi sentimen. Menurut Liu (2020), kualitas preprocessing memiliki pengaruh yang signifikan terhadap performa sistem analisis sentimen.

Tahapan preprocessing yang diterapkan meliputi:

Tabel 1. Tahapan Preprocessing Data

Tahap	Deskripsi
URL Removal	Menghapus tautan yang terdapat dalam komentar
Duplicate Removal	Menghapus komentar duplikat

Spam Filtering	Menghapus komentar spam
Empty Comment Removal	Menghapus komentar kosong
Emoji-only Removal	Menghapus komentar yang hanya berisi emoji
Text Normalization	Menyeragamkan format teks

Berdasarkan hasil pengujian pada tabel 1, proses filtering berhasil menghilangkan sekitar 20–40% komentar yang teridentifikasi sebagai spam dan noise sehingga meningkatkan kualitas data yang digunakan dalam proses klasifikasi.

Model Analisis Sentimen

Model analisis sentimen pada penelitian ini menggunakan XLM-RoBERTa (Cross-lingual Language Model-RoBERTa) yang merupakan model transformer multibahasa berbasis arsitektur RoBERTa yang dirancang untuk mendukung berbagai bahasa melalui proses *cross-lingual representation learning* (Conneau et al., 2020). Model ini dipilih karena memiliki kemampuan memahami konteks semantik pada berbagai bahasa, termasuk Bahasa Indonesia, sehingga mampu menghasilkan klasifikasi sentimen yang lebih akurat dibandingkan pendekatan berbasis machine learning konvensional (Conneau et al., 2020). Selain itu, model XLM-T yang merupakan pengembangan XLM-RoBERTa untuk data media sosial juga menunjukkan performa yang baik pada tugas analisis sentimen di berbagai bahasa (Barbieri et al., 2021).

Sebelum diterapkan pada komentar YouTube, model terlebih dahulu dievaluasi menggunakan dataset berlabel sebanyak 6.759 data yang terdiri atas dua kelas sentimen, yaitu positif dan negatif. Dataset tersebut digunakan untuk mengukur kemampuan model dalam melakukan klasifikasi sentimen sebelum diimplementasikan pada data komentar YouTube yang diperoleh melalui YouTube Data API v3.

Tahap analisis dimulai dengan menerima data komentar yang telah melalui proses preprocessing. Selanjutnya, setiap komentar diproses menggunakan tokenizer bawaan XLM-RoBERTa untuk mengubah teks menjadi token yang dapat dipahami oleh model. Token tersebut kemudian diproses oleh model transformer untuk menghasilkan representasi kontekstual dari setiap kalimat. Berdasarkan representasi tersebut, model menghasilkan

probabilitas pada masing-masing kelas sentimen, kemudian kelas dengan probabilitas tertinggi dipilih sebagai hasil prediksi.

Hasil prediksi berupa label sentimen positif atau negatif kemudian diteruskan ke tahap heuristic post-processing. Tahap ini bertujuan untuk memperbaiki beberapa hasil prediksi yang berpotensi keliru akibat penggunaan kata negasi, pengulangan karakter, maupun pola bahasa informal yang umum ditemukan pada komentar media sosial. Setelah seluruh proses selesai, hasil klasifikasi disimpan ke dalam basis data dan ditampilkan pada dashboard aplikasi dalam bentuk distribusi sentimen serta informasi pendukung lainnya.

Heuristic Post-Processing

Selain menggunakan model machine learning, penelitian ini menerapkan heuristic post-processing untuk meningkatkan kualitas prediksi. Pendekatan ini digunakan karena komentar media sosial sering kali mengandung humor, sarkasme, dan ekspresi informal yang sulit dipahami model secara sempurna.

Aturan heuristik diterapkan pada beberapa kondisi tertentu, seperti:

- Penggunaan emoji positif dan negatif.
- Frasa sarkastik yang umum digunakan.
- Komentar sangat pendek.
- Komentar yang mengandung kata-kata ambigu.

Pendekatan hybrid antara transformer dan aturan heuristik terbukti mampu mengurangi beberapa kesalahan klasifikasi yang muncul pada data media sosial.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat kinerja XLM-RoBERTa dalam mengklasifikasikan sentimen komentar ke dalam dua kelas, yaitu positif dan negatif. Proses evaluasi menggunakan dataset berlabel sebanyak 6.759 data yang telah memiliki label sentimen sebagai ground truth. Hasil prediksi model kemudian dibandingkan dengan label sebenarnya sehingga diperoleh nilai evaluasi berdasarkan confusion matrix.

Metrik evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score karena metrik tersebut merupakan standar evaluasi pada permasalahan klasifikasi teks dan analisis sentimen (Semary et al., 2023; Hashmi et al., 2024).

Accuracy digunakan untuk mengukur proporsi keseluruhan prediksi yang benar dibandingkan dengan seluruh data uji. Semakin tinggi nilai accuracy, semakin baik kemampuan model dalam melakukan klasifikasi sentimen.

Precision digunakan untuk mengukur tingkat ketepatan model dalam memberikan prediksi pada suatu kelas. Nilai precision yang tinggi menunjukkan bahwa sebagian besar hasil prediksi yang diberikan model merupakan prediksi yang benar.

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang termasuk ke dalam suatu kelas sentimen. Nilai recall yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data yang sebenarnya termasuk dalam kelas tersebut.

F1-score merupakan rata-rata harmonik antara precision dan recall sehingga memberikan gambaran keseimbangan antara kedua metrik tersebut. Nilai F1-score yang tinggi menunjukkan bahwa model memiliki performa klasifikasi yang baik tanpa terlalu bias terhadap precision maupun recall.

Selain menggunakan metrik evaluasi tersebut, penelitian ini juga memanfaatkan confusion matrix untuk menggambarkan jumlah prediksi yang benar maupun prediksi yang salah pada masing-masing kelas sentimen. Confusion matrix memberikan informasi mengenai distribusi True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) sehingga memudahkan analisis terhadap kesalahan klasifikasi yang dilakukan oleh model.

Hasil evaluasi model yang diperoleh menggunakan metrik accuracy, precision, recall, F1-score, serta confusion matrix disajikan dan dibahas lebih lanjut pada bagian Hasil dan Pembahasan sebagai dasar dalam menilai kelayakan model sebelum diimplementasikan pada analisis sentimen komentar YouTube.

Hasil dan Pembahasan

Evaluasi Model

Tabel 2. Hasil Evaluasi Model

Kelas	Precision	Recall	F1-score	Support	Kelas
Negatif	0,97	0,91	0,94	4176	Negatif

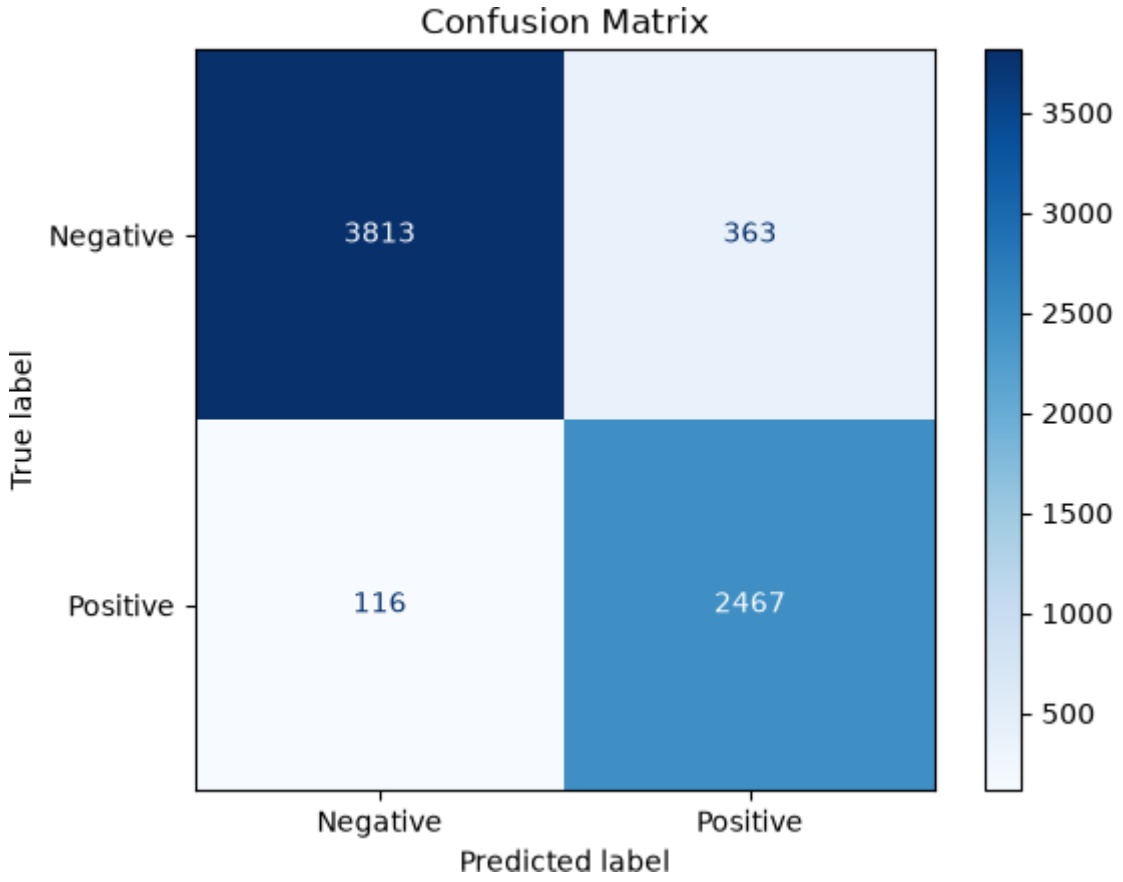
Positif	0,87	0,96	0,91	2583	Positif
Macro Avg	0,92	0,93	0,93	6759	Macro Avg
Weighted Avg	0,93	0,93	0,93	6759	Weighted Avg
Accuracy	0,93	-	-	6759	Accuracy

Tabel 2 menunjukkan hasil evaluasi model XLM-RoBERTa menggunakan dataset berlabel sebanyak 6.759 data. Berdasarkan hasil pengujian, model memperoleh accuracy sebesar 93%, yang menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar. Pada kelas negatif, model memperoleh precision sebesar 0,97, recall sebesar 0,91, dan F1-score sebesar 0,94, sedangkan pada kelas positif diperoleh precision sebesar 0,87, recall sebesar 0,96, dan F1-score sebesar 0,91. Nilai macro average dan weighted average yang sama-sama mencapai 0,93 menunjukkan bahwa model memiliki performa yang konsisten dalam mengklasifikasikan kedua kelas sentimen. Hasil tersebut menunjukkan bahwa model XLM-RoBERTa memiliki kemampuan yang baik untuk diterapkan pada proses analisis sentimen komentar YouTube.

Tabel 3. Confusion Matrix

Aktual	Negatif	Positif
Negatif	3813	363
Positif	116	2467

Tabel 3 menampilkan confusion matrix yang menggambarkan jumlah prediksi benar maupun salah yang dihasilkan oleh model. Berdasarkan hasil pengujian, sebanyak 3.813 komentar negatif berhasil diklasifikasikan dengan benar sebagai negatif, sedangkan 363 komentar negatif diprediksi sebagai positif. Pada kelas positif, model berhasil mengklasifikasikan 2.467 komentar secara benar, sementara 116 komentar positif masih salah diklasifikasikan sebagai negatif. Jumlah kesalahan klasifikasi yang relatif kecil dibandingkan jumlah prediksi yang benar menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik serta mampu membedakan kedua kelas sentimen dengan tingkat akurasi yang tinggi.



Gambar 3. Sentiment Distribution Dashboard

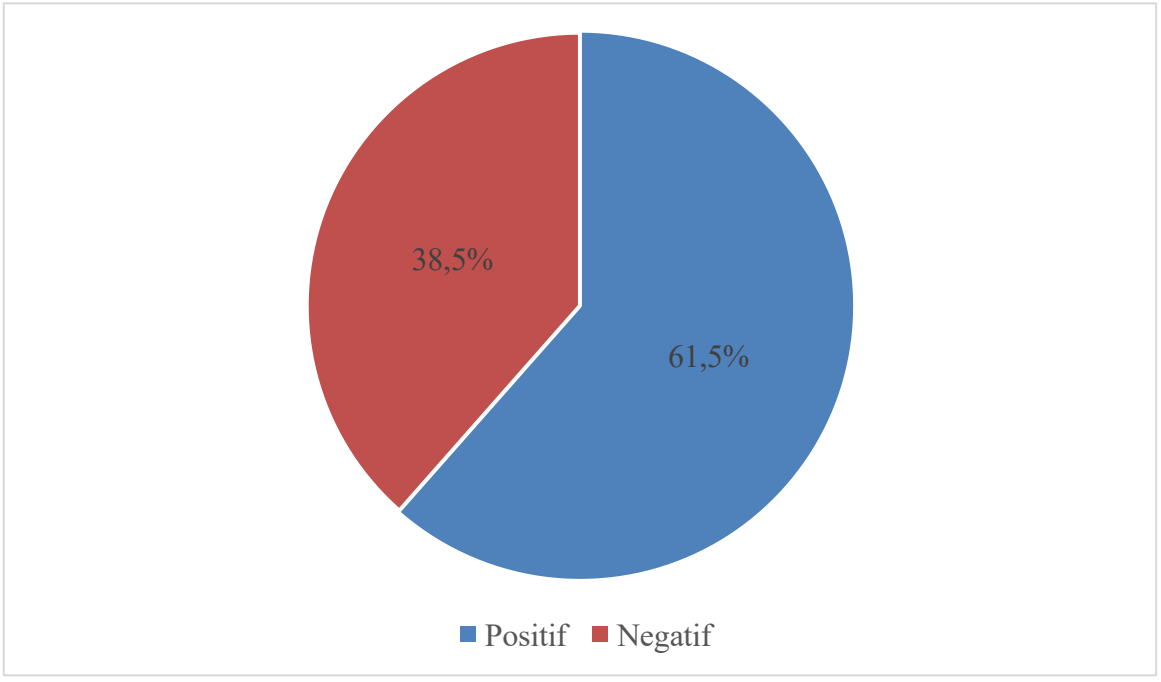
Gambar 3 merupakan visualisasi confusion matrix dalam bentuk heatmap untuk mempermudah interpretasi hasil klasifikasi model. Intensitas warna yang lebih gelap menunjukkan jumlah prediksi yang lebih besar. Terlihat bahwa sebagian besar data berada pada diagonal utama, yaitu sebanyak 3.813 komentar negatif dan 2.467 komentar positif berhasil diklasifikasikan dengan benar. Hal ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik, sedangkan jumlah kesalahan prediksi relatif kecil, yaitu 363 komentar negatif yang diprediksi sebagai positif dan 116 komentar positif yang diprediksi sebagai negatif.

Hasil Analisis Sentimen

Hasil penelitian menunjukkan bahwa sistem berhasil melakukan analisis sentimen komentar YouTube secara otomatis menggunakan model XLM-RoBER Komentar yang telah melalui tahap preprocessing dan filtering kemudian diklasifikasikan ke dalam tiga kategori utama, yaitu positif dan negatif.

Tabel 4. Distribusi Sentimen Komentar YouTube

Kategori Sentimen	Jumlah Komentar	Persentase
Positif	2.075	61,5%
Negatif	1.925	38,5%
Total	5.000	100%



Gambar 4. Sentiment Distribution Pie Chart

Berdasarkan Tabel 4 dan gambar 4, mayoritas komentar termasuk dalam kategori positif dengan persentase sebesar 61,5%. Hasil tersebut menunjukkan bahwa sebagian besar pengguna memberikan respons yang baik terhadap video yang dianalisis. Sementara itu, komentar negatif mencapai 38,5%.

Temuan ini menunjukkan bahwa sistem mampu memberikan gambaran umum mengenai persepsi pengguna secara cepat dan terstruktur tanpa memerlukan analisis manual.

Performa Sistem

Selain menghasilkan klasifikasi sentimen, penelitian juga mengevaluasi performa sistem dalam memproses komentar YouTube.

Tabel 5. Hasil Pengujian Performa Sistem

Komponen	Hasil
Kecepatan Inferensi	60–120 komentar/detik
Jumlah Komentar Uji	5.000 komentar
Waktu Analisis Total	45–60 detik
Spam yang Dihapus	20–40%
Cache Response	< 1 detik

Hasil pengujian pada tabel 5 menunjukkan bahwa model mampu memproses komentar dengan kecepatan antara 60 hingga 120 komentar per detik menggunakan GPU. Kecepatan tersebut menunjukkan bahwa pendekatan transformer masih memungkinkan untuk digunakan pada sistem analisis sentimen real-time.

Tabel 6. Hasil Validasi Data

Parameter	Nilai
Total Komentar yang Dikumpulkan	6.235
Komentar Setelah Filtering	5.000
Komentar Spam/Noise yang Dihapus	1.235
Persentase Data yang Dihapus	19,8%
Tingkat Konsistensi Data	100%

Berdasarkan hasil validasi pada Tabel 6, sistem berhasil melakukan penyaringan data secara efektif sebelum proses klasifikasi sentimen dilakukan. Dari 6.235 komentar yang dikumpulkan, sebanyak 1.235 komentar berhasil dihapus karena mengandung spam, duplikasi, atau noise. Dengan demikian, sebanyak 5.000 komentar yang telah melalui proses validasi digunakan sebagai data akhir untuk analisis sentimen. Proses ini berkontribusi pada peningkatan kualitas data dan membantu menghasilkan prediksi sentimen yang lebih akurat dan konsisten.

Selain itu, implementasi caching berhasil mempercepat proses analisis ulang terhadap video yang sama. Ketika pengguna melakukan analisis terhadap video yang sebelumnya telah diproses, sistem tidak perlu menjalankan inferensi ulang sehingga waktu respons menjadi jauh lebih cepat.

Visualisasi Dashboard

Salah satu kontribusi utama penelitian ini adalah penyediaan dashboard interaktif yang memungkinkan pengguna memahami hasil analisis secara lebih mudah. Dashboard menampilkan beberapa komponen utama, yaitu:

1. Donut chart distribusi sentimen.
2. Word cloud kata yang sering muncul.
3. Top comments berdasarkan confidence score.
4. Fitur pencarian komentar.
5. Filter sentimen.

Visualisasi tersebut membantu pengguna mengidentifikasi pola opini publik dengan lebih cepat dibandingkan hanya membaca hasil klasifikasi dalam bentuk teks.

Untuk memudahkan pengguna dalam memahami hasil analisis, sistem menyediakan dashboard interaktif yang menampilkan distribusi sentimen, statistik analisis, jumlah komentar yang diproses, serta informasi performa sistem. Tampilan dashboard yang dikembangkan dapat dilihat pada Gambar 4.



Gambar 5. User Interface of the Sentiment Analysis Dashboard

Gambar 5 menunjukkan tampilan dashboard hasil analisis sentimen yang dikembangkan. Dashboard menyajikan informasi utama seperti distribusi sentimen, jumlah komentar yang

dianalisis, jumlah komentar yang difilter, serta waktu pemrosesan sehingga pengguna dapat memperoleh ringkasan hasil analisis secara cepat dan mudah dipahami.

Pembahasan

Hasil penelitian menunjukkan bahwa penggunaan XLM-RoBERTa memberikan performa yang baik dalam analisis sentimen multibahasa. Temuan ini sejalan dengan penelitian Barbieri et al. (2021) dan Hashmi et al. (2024) yang menunjukkan bahwa model multilingual transformer mampu menangani variasi bahasa dan code-mixed text secara efektif.

Dibandingkan pendekatan machine learning tradisional seperti Naïve Bayes atau Support Vector Machine, model transformer mampu memahami konteks kalimat secara lebih baik karena memanfaatkan mekanisme self-attention (Vaswani et al., 2017). Hal tersebut memungkinkan model mengenali hubungan antar kata yang sebelumnya sulit dipahami oleh metode konvensional.

Meskipun demikian, penelitian juga menemukan bahwa komentar yang mengandung humor dan sarkasme masih menjadi tantangan. Untuk mengatasi permasalahan tersebut, penelitian menerapkan heuristic post-processing yang membantu memperbaiki sebagian kesalahan klasifikasi. Secara keseluruhan, hasil evaluasi menunjukkan bahwa model XLM-RoBERTa memperoleh accuracy sebesar 93%, yang menunjukkan kemampuan klasifikasi yang baik pada data uji. Pendekatan ini menunjukkan bahwa kombinasi antara model transformer dan aturan berbasis domain berpotensi meningkatkan kualitas hasil analisis, khususnya pada komentar yang menggunakan bahasa informal.Selain aspek teknis, dashboard interaktif yang dikembangkan memberikan nilai tambah bagi pengguna. Tidak seperti banyak penelitian sebelumnya yang hanya menghasilkan label sentimen, sistem ini menyediakan visualisasi yang membantu pengguna memahami opini publik secara lebih komprehensif. Dengan demikian, hasil penelitian tidak hanya memberikan kontribusi akademik dalam bidang NLP, tetapi juga memiliki manfaat praktis bagi content creator, organisasi, maupun perusahaan yang ingin memahami respons audiens terhadap suatu konten.

Dari perspektif bisnis dan pengambilan keputusan, sistem yang dikembangkan dapat membantu organisasi memahami opini pelanggan secara lebih cepat sehingga mendukung strategi pemasaran, evaluasi layanan, dan pengembangan produk berbasis data. Temuan ini konsisten dengan penelitian terbaru yang menunjukkan bahwa analisis sentimen media sosial

memiliki peran penting dalam business intelligence dan customer insight analytics (Agnihotri et al., 2023).

Kesimpulan

Penelitian ini berhasil mengembangkan sistem analisis sentimen komentar YouTube berbasis web menggunakan model XLM-RoBERTa untuk mengklasifikasikan komentar ke dalam dua kelas sentimen, yaitu positif dan negatif. Sistem memanfaatkan YouTube Data API v3 untuk mengambil data komentar secara otomatis, kemudian melakukan proses preprocessing, klasifikasi sentimen, heuristic post-processing, dan visualisasi hasil analisis melalui dashboard interaktif.

Berdasarkan hasil evaluasi menggunakan dataset berlabel sebanyak 6.759 data, model XLM-RoBERTa memperoleh accuracy sebesar 93%, dengan nilai precision, recall, dan F1-score yang menunjukkan performa klasifikasi yang baik pada kedua kelas sentimen. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang andal dalam mengidentifikasi sentimen komentar sehingga layak diterapkan pada sistem analisis sentimen.

Pada tahap implementasi, sistem berhasil menganalisis 5.000 komentar YouTube secara otomatis dan menyajikan hasil klasifikasi dalam bentuk dashboard yang menampilkan distribusi sentimen, pencarian komentar, serta informasi pendukung lainnya. Integrasi antara model XLM-RoBERTa dan aplikasi berbasis web memberikan kemudahan bagi pengguna dalam memahami opini publik secara lebih cepat dan efisien.

Untuk penelitian selanjutnya, sistem dapat dikembangkan dengan menambahkan klasifikasi sentimen multikelas, mendukung lebih banyak bahasa, menerapkan model transformer yang lebih mutakhir, serta melakukan evaluasi menggunakan dataset yang lebih beragam agar kemampuan generalisasi model semakin meningkat.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pembimbing dan Universitas Bina Nusantara yang telah memberikan dukungan selama pelaksanaan penelitian ini. Apresiasi juga diberikan kepada seluruh pihak yang telah memberikan masukan dan bantuan dalam proses pengembangan sistem serta penyusunan artikel. Penelitian ini tidak memperoleh pendanaan eksternal dari pihak mana pun.

Tentang Penulis

Penulis 1

Dhafin Wahyu Wicaksono merupakan mahasiswa Department of Computer Science, Bina Nusantara University, Indonesia. Minat penelitiannya meliputi pengembangan perangkat lunak, kecerdasan buatan, machine learning, dan Natural Language Processing. Dalam penelitian ini, penulis berkontribusi pada perancangan sistem, pengembangan backend, serta penyusunan metodologi penelitian.

Penulis 2

Jonathan Eliazer Siregar merupakan mahasiswa Department of Computer Science, Bina Nusantara University, Indonesia. Bidang minatnya mencakup pengembangan aplikasi web, desain antarmuka pengguna, dan visualisasi data. Dalam penelitian ini, penulis berkontribusi pada pengembangan frontend dan perancangan dashboard analisis sentimen.

Penulis 3

Kevin Rudolf Kennard Meliala merupakan mahasiswa Department of Computer Science, Bina Nusantara University, Indonesia. Ketertarikannya meliputi data analytics, data processing, dan data mining. Dalam penelitian ini, penulis bertanggung jawab atas pengumpulan data, preprocessing, serta validasi kualitas data yang digunakan dalam analisis.

Penulis 4

Naufal Dzakwan Prakintara merupakan mahasiswa Department of Computer Science, Bina Nusantara University, Indonesia. Bidang minatnya meliputi machine learning, deep learning, dan Natural Language Processing. Dalam penelitian ini, penulis berkontribusi pada implementasi model XLM-RoBERTa, proses inferensi, dan optimasi performa sistem.

Penulis 5

Raafi Achmad Ramadhan merupakan mahasiswa Department of Computer Science, Bina Nusantara University, Indonesia. Minat akademiknya meliputi software testing, software quality assurance, dan system analysis. Dalam penelitian ini, penulis berkontribusi pada pengujian sistem, analisis hasil penelitian, dokumentasi proyek, serta revisi akhir artikel.

Kontribusi Penulis

Penulis	Kontribusi
Dhafin Wahyu Wicaksono	Menginisiasi ide penelitian, menyusun metodologi penelitian, mengembangkan backend menggunakan FastAPI, serta menyusun dan merevisi naskah artikel.
Jonathan Eliazer Siregar	Mengembangkan antarmuka pengguna (frontend) menggunakan React, merancang dashboard visualisasi, serta melakukan pengujian fungsional antarmuka sistem.
Kevin Rudolf Kennard Meliala	Melakukan pengumpulan data melalui YouTube Data API, preprocessing data, filtering spam dan noise, serta validasi kualitas data yang digunakan dalam penelitian.
Naufal Dzakwan Prakintara	Mengimplementasikan model XLM-RoBERTa, melakukan proses inferensi dan heuristic post-processing, serta mengoptimalkan performa sistem analisis sentimen.
Raafi Achmad Ramadhan	Melaksanakan pengujian sistem, menganalisis hasil penelitian, menyusun dokumentasi proyek, serta melakukan pemeriksaan dan revisi akhir artikel

Ketersediaan Data dan Material

Data yang digunakan dalam penelitian ini berasal dari komentar publik yang diperoleh melalui YouTube Data API v3 sesuai dengan ketentuan dan kebijakan penggunaan yang berlaku. Dataset yang digunakan bersifat dinamis karena bergantung pada komentar yang tersedia pada video YouTube saat proses pengambilan data dilakukan. Kode sumber dan

dokumentasi sistem dapat disediakan oleh penulis untuk keperluan akademik dan penelitian lebih lanjut sesuai permintaan yang wajar.

Daftar Pustaka

Barbieri, F., Camacho-Collados, J., Espinosa-Anke, L., & Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1644–1650. https://doi.org/10.18653/v1/2020.findings-emnlp.148

Barbieri, F., Espinosa Anke, L., & Camacho-Collados, J. (2022). *XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond*. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 258–266). European Language Resources Association. https://aclanthology.org/2022.lrec-1.27/

Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., Boisson, J., Espinosa-Anke, L., Liu, F., Martinez-Camara, E., & others. (2022). TweetNLP: Cutting-edge natural language processing for social media. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–49. https://doi.org/10.18653/v1/2022.emnlp-demos.5

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Hashmi, E., Yayilgan, S. Y., & Shaikh, S. (2024). *Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers*. *Social Network Analysis and Mining, 14*(86). https://doi.org/10.1007/s13278-024-01245-6

Zhang, H., & Shafiq, M. O. (2024). *Survey of transformers and towards ensemble learning using transformers for natural language processing*. *Journal of Big Data, 11*(25). https://doi.org/10.1186/s40537-023-00842-0

Jahin, M. A., Shovon, M. S. H., Mridha, M. F., Islam, M. R., & Watanobe, Y. (2024). *A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets*. *Scientific Reports, 14*, 24882. https://doi.org/10.1038/s41598-024-76079-5

Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108639286

Mercha, E. M., & Benbrahim, H. (2023). *Machine learning and deep learning for sentiment analysis across languages: A survey*. *Neurocomputing, 531*, 195–216. https://doi.org/10.1016/j.neucom.2023.02.015

Agnihotri, R., Afshar Bakeshloo, K., & Mani, S. (2023). *Social media analytics for business-to-business marketing*. *Industrial Marketing Management, 115*, 110–126. https://doi.org/10.1016/j.indmarman.2023.09.012

Semary, N. A., Ahmed, W., Amin, K., Pławiak, P., & Hammad, M. (2023). *Improving sentiment classification using a RoBERTa-based hybrid model*. *Frontiers in Human Neuroscience, 17*, 1292010. https://doi.org/10.3389/fnhum.2023.1292010

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://papers.nips.cc/paper/7181-attention-is-all-you-need

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Google Developers. (2026). *YouTube Data API v3 documentation*. https://developers.google.com/youtube/v3

Hugging Face. (2026). *cardiffnlp/twitter-xlm-roberta-base-sentiment*. https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment