



Big Data-Based YouTube Feedback Analytics Architecture with Generative AI Integration

Sistem Analitik Umpan Balik YouTube Berbasis Big Data dan Generative AI

Asro^{1*}, Novi Rukhviyanti²

^{1,2}Department of Informatics, STMIK Indonesia Mandiri, Indonesia

E-Mail: ¹asroharun6@gmail.com, ²novi.rukhviyanti@stmik-im.ac.id

Received Mar 07th 2026; Revised Apr 04th 2026; Accepted Apr 11th 2026; Available Online Apr 26th 2026

Corresponding Author: Asro

Copyright ©2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The rapid growth of social media platforms generates large volumes of unstructured user feedback, particularly on YouTube. User comments contain valuable insights regarding perceptions of digital products; however, manual analysis remains time-consuming and inefficient. This study aims to develop a Big Data-based feedback analytics architecture integrating sentiment analysis and Generative Artificial Intelligence (Generative AI) within an end-to-end pipeline. A total of 10,625 YouTube comments were collected via the YouTube Data API and stored in MongoDB. The CRISP-DM framework was employed, including text preprocessing and TF-IDF feature representation. Sentiment classification was evaluated using Logistic Regression, Support Vector Machine (SVM), and Random Forest (RF), with SMOTE applied to the training dataset. Experimental results show that SVM without SMOTE achieved the best performance, with an accuracy of 83.97% and a macro F1-score of 80.29%. The integration of Generative AI enables automatic issue summarization and prioritized recommendation generation, supporting more efficient data-driven decision-making.

Keywords: Big Data, Generative Artificial Intelligence, Sentiment Analysis, YouTube Comments.

Abstrak

Perkembangan media sosial menghasilkan data opini pengguna dalam jumlah besar dan tidak terstruktur, khususnya pada platform YouTube. Komentar pengguna mengandung informasi penting mengenai persepsi terhadap produk digital, namun analisis manual masih menjadi kendala. Penelitian ini bertujuan mengembangkan arsitektur sistem analitik umpan balik berbasis Big Data yang mengintegrasikan analisis sentimen dan *Generative Artificial Intelligence* (AI) dalam satu *pipeline end-to-end*. Sebanyak 10.625 komentar dikumpulkan melalui *YouTube Data API* dan disimpan dalam *MongoDB*. Proses analisis menggunakan kerangka *CRISP-DM* dengan preprocessing teks dan representasi fitur *TF-IDF*. Klasifikasi sentimen dibandingkan menggunakan *Logistic Regression*, *Support Vector Machine* (SVM), dan *Random Forest*, dengan pengujian *SMOTE* pada data latih. Hasil menunjukkan bahwa *SVM* tanpa *SMOTE* memberikan performa terbaik dengan akurasi 83,97% dan F1-macro 80,29%. Integrasi *Generative AI* memungkinkan peringkasan isu dominan serta penyusunan rekomendasi perbaikan secara otomatis. Sistem yang dikembangkan mendukung pengambilan keputusan berbasis data secara lebih efisien.

Kata Kunci: Analisis Sentimen, Big Data, Generative Artificial Intelligence, YouTube.

1. PENDAHULUAN

Dalam beberapa tahun terakhir, pertumbuhan media sosial telah menghasilkan jumlah data opini pengguna yang sangat besar dengan karakteristik tidak terstruktur. Platform seperti YouTube tidak hanya digunakan untuk distribusi konten digital, tetapi juga menjadi sarana komunikasi dan interaksi pengguna yang menghasilkan volume komentar yang sangat besar secara real-time. Komentar pengguna mengandung informasi penting mengenai pengalaman dan persepsi kualitas layanan, sehingga memiliki potensi strategis apabila dianalisis secara sistematis. Karakteristik data yang memiliki volume tinggi, pertumbuhan cepat, serta variasi format menjadikan komentar media sosial sebagai bagian dari ekosistem Big Data berbasis teks. Data

media sosial umumnya memiliki karakteristik volume besar, pertumbuhan cepat, dan variasi format yang tinggi sehingga membutuhkan metode pengolahan dan pengelolaan data yang efisien [1].

Dalam praktiknya, komentar pada YouTube tidak hanya berjumlah besar, tetapi juga terus bertambah dalam waktu singkat, terutama ketika terjadi pembaruan aplikasi, gangguan layanan, atau isu yang menjadi perhatian publik. Kondisi ini menyebabkan organisasi sulit melakukan evaluasi secara manual karena proses pembacaan, pengelompokan, dan penarikan kesimpulan dari ribuan komentar membutuhkan waktu yang panjang serta hasil yang sering kali tidak konsisten. Selain itu, komentar pengguna pada platform digital umumnya mengandung bahasa informal, singkatan, kata tidak baku, campuran opini, serta ekspresi emosional yang beragam. Karakteristik tersebut membuat data komentar menjadi sumber informasi yang kaya, namun sekaligus menimbulkan tantangan dalam proses analisis. Untuk menangani data tidak terstruktur seperti komentar YouTube, basis data NoSQL seperti MongoDB sering digunakan karena lebih fleksibel dibandingkan basis data relasional konvensional serta mampu menangani struktur dokumen yang dinamis dan mendukung skalabilitas tinggi [2], [3].

Untuk menjawab kebutuhan tersebut, penelitian ini memanfaatkan Natural Language Processing (NLP) sebagai pendekatan utama dalam pengolahan data teks. NLP digunakan untuk membersihkan, menormalkan, dan mengubah komentar pengguna menjadi representasi yang dapat diproses oleh model klasifikasi sentimen. Melalui proses tersebut, sistem dapat mengidentifikasi kecenderungan opini pengguna ke dalam kategori sentimen positif, negatif, dan netral secara otomatis. Salah satu pendekatan utama dalam pengolahan opini pengguna adalah analisis sentimen, yaitu bagian dari NLP yang digunakan untuk mengidentifikasi dan mengelompokkan opini teks ke dalam kategori tertentu. Pada dataset dengan distribusi kelas yang tidak seimbang, teknik oversampling seperti *Synthetic Minority Oversampling Technique* (SMOTE) sering digunakan untuk meningkatkan representasi kelas minoritas sehingga proses klasifikasi menjadi lebih seimbang [4]. Hasil yang diharapkan dari proses ini bukan hanya berupa nilai akurasi model, tetapi juga tersedianya dasar analitik yang dapat digunakan untuk memahami persepsi pengguna, mengenali isu dominan, serta mendukung penyusunan prioritas perbaikan layanan atau produk digital secara lebih objektif.

Sejumlah penelitian terdahulu telah menunjukkan bahwa analisis sentimen pada data media sosial dapat dimanfaatkan untuk mengevaluasi opini publik dan kualitas layanan digital [3], [5]. Beberapa studi juga membandingkan performa algoritma klasifikasi seperti Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), dan Random Forest pada data ulasan atau komentar pengguna [6], [7]. Hasil penelitian tersebut umumnya menunjukkan bahwa model berbasis *machine learning* cukup efektif untuk klasifikasi sentimen teks, terutama ketika dikombinasikan dengan *preprocessing* yang tepat, representasi fitur TF-IDF, serta penanganan ketidakseimbangan kelas menggunakan SMOTE [6], [8]. Di sisi lain, penelitian lain juga menyoroti pentingnya penyimpanan data tidak terstruktur menggunakan basis data NoSQL seperti MongoDB karena lebih fleksibel dalam menangani dokumen teks berskala besar [1], [7]. Temuan tersebut memperlihatkan bahwa aspek klasifikasi dan pengelolaan data telah banyak dibahas, namun belum seluruhnya terintegrasi dalam satu sistem analitik yang utuh.

Meskipun berbagai penelitian menunjukkan tingkat akurasi klasifikasi yang tinggi, sebagian besar studi masih berfokus pada evaluasi performa model tanpa mengembangkan mekanisme pemanfaatan hasil klasifikasi sebagai sistem pendukung keputusan secara operasional. Dalam banyak penelitian, analisis sentimen masih terbatas pada pelaporan kinerja model melalui metrik evaluasi seperti *accuracy* dan *F1-score*, tanpa mengintegrasikan hasil klasifikasi ke dalam sistem analitik yang mampu menghasilkan temuan analitik serta rekomendasi perbaikan secara otomatis. Selain itu, integrasi antara pengumpulan komentar, penyimpanan *Big Data*, klasifikasi sentimen komparatif, dan interpretasi hasil berbasis *Generative Artificial Intelligence* (AI) masih relatif terbatas. Berdasarkan celah tersebut, penelitian ini bertujuan mengembangkan arsitektur sistem analitik umpan balik YouTube berbasis *Big Data* yang mengintegrasikan analisis sentimen dan *Generative AI* dalam satu sistem *end-to-end*. Kontribusi utama penelitian ini terletak pada pembangunan sistem terintegrasi yang mencakup *data layer*, *analytics layer*, dan *interpretative layer*, sehingga mampu mengotomatisasi proses ekstraksi *insight* dari komentar pengguna serta mendukung proses pengambilan keputusan berbasis data secara lebih terstruktur dan efektif.

1.1 Tinjauan Literatur

1.1.1 Big Data dan Analitik Umpan Balik Media Sosial

Konsep *Big Data* ditandai oleh karakteristik *volume*, *velocity*, dan *variety* yang menuntut pendekatan pengolahan data berskala besar [8]. Data komentar pada media sosial seperti YouTube termasuk dalam kategori data tidak terstruktur yang tumbuh dengan cepat dan memiliki format yang beragam. Kondisi ini menuntut adanya infrastruktur penyimpanan yang adaptif sekaligus arsitektur analitik yang mampu memproses data secara terukur. Dalam konteks tersebut, basis data NoSQL seperti MongoDB dinilai lebih sesuai dibandingkan dengan sistem relasional tradisional karena mendukung penyimpanan dokumen berbasis JSON dengan struktur yang dinamis [8]. Fleksibilitas ini memungkinkan pengelolaan data opini dalam jumlah besar tanpa ketergantungan pada skema tetap. Namun, tantangan utama dalam pengelolaan data

media sosial tidak hanya terletak pada penyimpanan, tetapi juga pada kemampuan sistem untuk mengubah data mentah menjadi informasi yang bernilai. Berbagai studi menunjukkan bahwa implementasi *Big Data* pada media sosial masih sering berhenti pada tahap pengumpulan dan penyimpanan data, tanpa dilanjutkan dengan mekanisme analitik yang mampu menghasilkan *insight* strategis [9]. Oleh karena itu, pengembangan sistem analitik umpan balik membutuhkan pendekatan yang tidak hanya kuat pada sisi infrastruktur data, tetapi juga pada integrasi proses analisis secara menyeluruh.

1.1.2 Analisis Sentimen Berbasis Machine Learning

Analisis sentimen merupakan salah satu pendekatan dalam *Natural Language Processing* (NLP) yang digunakan untuk mengenali dan mengelompokkan opini dalam teks ke dalam kategori sentimen seperti positif, negatif, atau netral. Teknik ini banyak dimanfaatkan untuk mengevaluasi persepsi pengguna terhadap produk, layanan, maupun isu publik tertentu. Berbagai algoritma berbasis *machine learning* telah digunakan dalam penelitian analisis sentimen, di antaranya *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, hingga pendekatan berbasis *deep learning* [5]. Selain itu, *Decision Tree* juga sering dimanfaatkan dalam proses klasifikasi karena mampu menghasilkan model yang relatif mudah dipahami serta mendukung interpretasi hasil secara lebih transparan [10]. Di antara berbagai algoritma tersebut, SVM dikenal memiliki performa yang konsisten dalam menangani data teks berdimensi tinggi karena kemampuannya membangun *hyperplane* optimal untuk memisahkan kelas data secara efektif. Ketika dikombinasikan dengan pembobotan fitur seperti *Term Frequency-Inverse Document Frequency* (TF-IDF), pendekatan ini dapat menyederhanakan ruang fitur sekaligus mempertahankan informasi penting dari teks. Meskipun demikian, salah satu tantangan utama dalam analisis sentimen adalah distribusi kelas yang tidak seimbang. Pada data media sosial, komentar sering kali didominasi oleh polaritas tertentu sehingga proses pembelajaran model menjadi kurang optimal. Untuk mengatasi hal tersebut, teknik *Synthetic Minority Oversampling Technique* (SMOTE) sering digunakan guna meningkatkan representasi kelas minoritas pada data latih [11]. Namun, beberapa penelitian menunjukkan bahwa penerapan SMOTE tidak selalu memberikan peningkatan performa yang signifikan, terutama ketika distribusi data awal masih relatif seimbang. Selain itu, sebagian besar penelitian analisis sentimen masih menitikberatkan pada peningkatan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, tanpa mengintegrasikan hasil klasifikasi ke dalam sistem analitik yang dapat dimanfaatkan secara langsung sebagai pendukung pengambilan keputusan. Akibatnya, hasil analisis sering berhenti pada tahap evaluasi model dan belum sepenuhnya dimanfaatkan untuk menghasilkan rekomendasi berbasis data secara otomatis.

1.1.3 Generative AI dalam Interpretasi Data Teks

Perkembangan *Generative AI*, khususnya model bahasa berskala besar atau *Large Language Models* (LLM), membuka peluang baru dalam pengolahan dan interpretasi data teks dalam skala besar [12]. Berbeda dari model klasifikasi tradisional yang berfokus pada pengelompokan label, model generatif mampu menghasilkan ringkasan, penjelasan, dan rekomendasi berbasis narasi dari kumpulan teks yang kompleks. Beberapa studi menunjukkan bahwa model bahasa skala besar memiliki kemampuan generalisasi yang baik, termasuk melalui pendekatan *few-shot learning* [13]. Selain itu, laporan teknis GPT-4 menunjukkan bahwa model generatif dapat digunakan secara efektif untuk tugas peringkasan dan ekstraksi informasi dari teks panjang dengan tingkat konsistensi yang tinggi [14], [15]. Dalam konteks analisis sentimen, kemampuan tersebut penting karena hasil klasifikasi yang semula berbentuk distribusi numerik dapat diterjemahkan menjadi ringkasan isu dominan dan rekomendasi berbasis konteks yang lebih mudah dipahami oleh pengambil keputusan. Namun demikian, pemanfaatan *Generative AI* dalam literatur masih sering berdiri terpisah dari proses klasifikasi sentimen, sehingga integrasi antara klasifikasi berbasis *machine learning* dan interpretasi berbasis model generatif masih belum banyak dikembangkan dalam bentuk sistem analitik yang benar-benar terpadu.

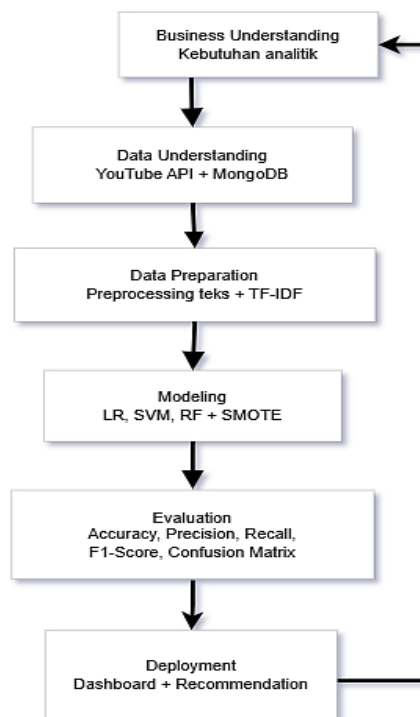
1.1.4 Research Gap dan Posisi Penelitian

Berdasarkan telaah literatur, terdapat beberapa celah penelitian yang masih terbuka. Pertama, penelitian analisis sentimen pada media sosial umumnya berfokus pada optimalisasi performa model klasifikasi, tetapi belum banyak yang mengintegrasikan hasil klasifikasi ke dalam sistem pendukung keputusan yang operasional. Akibatnya, keluaran penelitian sering terbatas pada metrik evaluasi tanpa menghasilkan rekomendasi strategis yang aplikatif. Kedua, studi *Big Data* pada umumnya lebih menekankan aspek infrastruktur penyimpanan dan manajemen data, tanpa mengintegrasikan lapisan interpretatif berbasis *Generative AI* secara menyeluruh dalam arsitektur sistem analitik. Ketiga, integrasi antara penyimpanan data tidak terstruktur berbasis MongoDB, perbandingan model klasifikasi sentimen seperti *Logistic Regression*, SVM, dan *Random Forest*, penanganan ketidakseimbangan kelas menggunakan SMOTE, serta pemanfaatan *Generative AI* sebagai lapisan peringkasan dan rekomendasi dalam satu arsitektur sistem terpadu masih relatif terbatas. Sebagian besar penelitian masih memisahkan proses pengumpulan data, klasifikasi, dan interpretasi, sehingga belum membentuk sistem analitik yang terintegrasi secara *end-to-end*. Berdasarkan

celah tersebut, penelitian ini memosisikan diri pada pengembangan arsitektur sistem analitik umpan balik YouTube berbasis *Big Data* yang mengintegrasikan tiga lapisan utama, yaitu *data layer* (YouTube API dan MongoDB), *analytics layer* (TF-IDF, *Logistic Regression*, SVM, *Random Forest*, dan SMOTE), serta *interpretative layer* berbasis *Generative AI*. Pendekatan ini tidak hanya melakukan klasifikasi sentimen dan evaluasi performa model, tetapi juga mengotomatisasi ekstraksi *insight* serta penyusunan rekomendasi prioritas perbaikan produk berbasis data komentar pengguna dalam skala besar. Dengan demikian, *novelty* penelitian ini terletak pada integrasi arsitektural yang menghubungkan penyimpanan *Big Data*, analisis sentimen komparatif, dan interpretasi berbasis *Generative AI* dalam satu sistem analitik terpadu. Berbeda dari penelitian sebelumnya yang umumnya masih melakukan evaluasi model klasifikasi secara parsial, penelitian ini mengembangkan sistem yang menghubungkan proses *crawling*, penyimpanan data, klasifikasi sentimen, dan interpretasi hasil dalam satu *pipeline* operasional.

2. MATERIALS AND METHOD

Penelitian ini menggunakan pendekatan kuantitatif berbasis rekayasa sistem (*system engineering approach*) untuk merancang dan mengimplementasikan sistem analitik umpan balik YouTube yang terintegrasi secara *end-to-end*. Pendekatan ini tidak hanya berfokus pada evaluasi performa model klasifikasi sentimen, tetapi juga pada pembangunan arsitektur sistem yang mengintegrasikan proses pengumpulan data, penyimpanan, pemrosesan, analisis, hingga interpretasi hasil dalam satu pipeline terpadu. Penelitian ini menggunakan kerangka kerja CRISP-DM yang mencakup enam tahapan proses analisis data [16]. Alur metodologi penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Menggunakan CRISP-DM

2.1 Business Understanding

Tahap *business understanding* bertujuan untuk mengidentifikasi kebutuhan analitik dalam pemanfaatan komentar YouTube sebagai sumber evaluasi produk digital. Permasalahan utama yang dihadapi adalah tingginya volume data opini pengguna yang bersifat tidak terstruktur, sehingga sulit dianalisis secara manual dan membutuhkan waktu yang lama. Kondisi ini menuntut pengembangan sistem analitik yang mampu melakukan klasifikasi sentimen secara otomatis, menganalisis distribusi sentimen untuk mengidentifikasi isu dominan, serta menghasilkan rekomendasi perbaikan produk berbasis data. Dengan demikian, tahap ini menjadi landasan konseptual dalam perancangan arsitektur sistem analitik yang terintegrasi dan berorientasi pada pengambilan keputusan.

2.2 Data Understanding

Pada tahap *data understanding*, proses pengumpulan komentar dilakukan menggunakan YouTube Data API dengan kata kunci yang relevan dengan topik penelitian. Data yang diperoleh meliputi informasi seperti *comment_id*, *author*, isi komentar, waktu publikasi, serta jumlah *like*. Seluruh data disimpan dalam

basis data *MongoDB* yang termasuk dalam kategori *NoSQL database* sehingga mampu menangani data teks tidak terstruktur dalam jumlah besar. Selain pengumpulan data, tahap ini juga mencakup eksplorasi awal terhadap *dataset*, seperti identifikasi duplikasi komentar.

2.3 Data Preparation

Tahap *data preparation* dilakukan untuk meningkatkan kualitas data melalui proses *text preprocessing*. Proses ini meliputi pembersihan teks dari URL, simbol, emoji, serta karakter non-alfabet. Selanjutnya dilakukan normalisasi huruf (*case folding*) dan tokenisasi untuk memisahkan teks menjadi unit kata. Setelah itu dilakukan penghapusan *stopword* dan proses *stemming* untuk mengubah kata ke bentuk dasar [17]. Teks yang telah diproses kemudian direpresentasikan dalam bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* dengan kombinasi fitur *n-gram*. Untuk menangani ketidakseimbangan distribusi kelas sentimen, penelitian ini menerapkan teknik *Synthetic Minority Oversampling Technique (SMOTE)* pada data latih setelah proses vektorisasi sehingga tidak terjadi *data leakage* pada data uji.

2.4 Modeling

Pada tahap *modeling*, penelitian ini membandingkan tiga algoritma klasifikasi berbasis *machine learning*, yaitu *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Random Forest*. *Dataset* dibagi menjadi 80% data latih dan 20% data uji dengan menggunakan metode *stratified sampling* untuk mempertahankan proporsi distribusi kelas pada setiap subset data. Untuk menangani ketidakseimbangan kelas pada data latih, diterapkan teknik *Synthetic Minority Oversampling Technique (SMOTE)*. Selanjutnya, masing-masing algoritma dilatih menggunakan data latih dan diuji menggunakan data uji yang sama sehingga perbandingan performa antarmodel dapat dilakukan secara konsisten.

2.5 Evaluation

Evaluasi performa model dilakukan menggunakan *confusion matrix* untuk menganalisis pola kesalahan klasifikasi pada setiap kategori sentimen (*positive*, *negative*, dan *neutral*). *Confusion matrix* menunjukkan jumlah prediksi yang benar maupun salah yang terdiri dari *True Positive (TP)*, *False Positive (FP)*, *False Negative (FN)*, dan *True Negative (TN)* pada masing-masing kelas dengan pendekatan *one-vs-rest* pada kasus multikelas [18]. Berdasarkan nilai pada *confusion matrix*, kinerja model kemudian diukur menggunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (4)$$

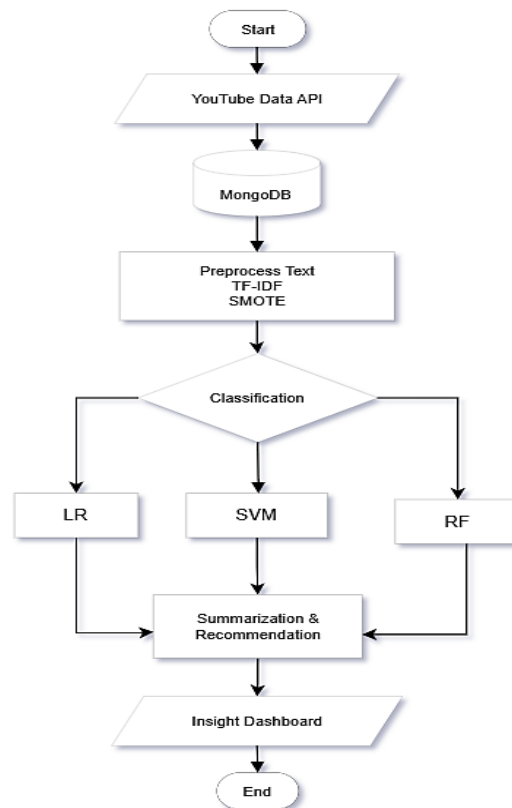
Selain itu, dilakukan perbandingan performa model tanpa *SMOTE* dan dengan *SMOTE* untuk menganalisis pengaruh oversampling terhadap kinerja klasifikasi. Penerapan *SMOTE* dilakukan hanya pada data latih setelah proses vektorisasi untuk mencegah *data leakage*, sedangkan evaluasi tetap menggunakan data uji yang sama agar perbandingan antar skenario tetap objektif [19].

2.6 Deployment

Tahap *deployment* dilakukan dengan mengintegrasikan model terbaik ke dalam sistem dashboard analitik. Distribusi sentimen yang dihasilkan dari proses klasifikasi selanjutnya menjadi input bagi modul *Generative AI*. Modul *Generative AI* berfungsi sebagai lapisan interpretatif yang merangkum isu dominan, khususnya dari komentar dengan sentimen negatif, serta menghasilkan rekomendasi prioritas perbaikan produk secara otomatis. Hasil akhir sistem disajikan dalam bentuk *Dashboard Insight* dan Rekomendasi yang menampilkan visualisasi distribusi sentimen, ringkasan isu utama, serta rekomendasi berbasis analisis data. Dengan demikian, sistem tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai pendukung pengambilan keputusan berbasis data.

2.7 Arsitektur Sistem Analitik

Arsitektur sistem analitik yang dikembangkan dalam penelitian ini ditunjukkan pada Gambar 2. Sistem dirancang menggunakan pendekatan tiga lapisan utama (*three-layer architecture*), yaitu *Data Layer*, *Analytics Layer*, dan *Generative AI Layer*, yang terintegrasi secara *end-to-end* mulai dari proses pengumpulan data hingga penyajian *insight*.



Gambar 2. Arsitektur Sistem Analitik Umpan Balik YouTube

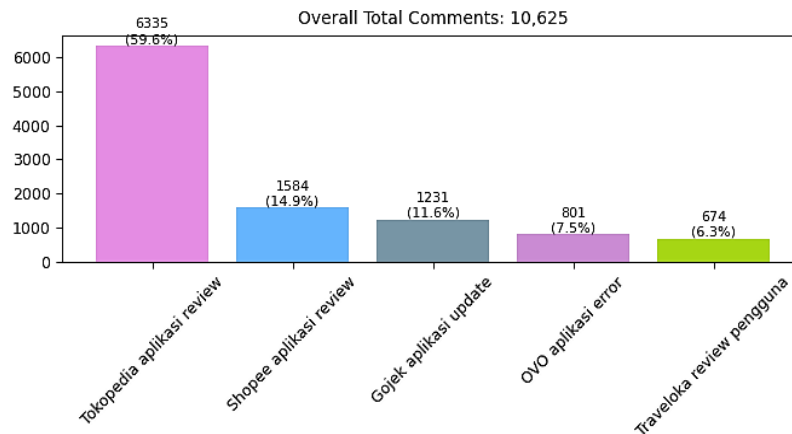
Pada *Data Layer*, sistem melakukan pengambilan komentar melalui *YouTube Data API* berdasarkan kata kunci tertentu. Data yang diperoleh berupa teks tidak terstruktur dan disimpan dalam *MongoDB* sebagai basis data *NoSQL* yang mendukung pengelolaan data dalam skala besar dengan struktur yang dinamis. Lapisan kedua, yaitu *Analytics Layer*, bertanggung jawab atas pemrosesan dan klasifikasi sentimen. Tahap ini meliputi preprocessing teks dan representasi fitur menggunakan TF-IDF. Proses klasifikasi dilakukan dengan membandingkan tiga algoritma machine learning, yaitu *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Random Forest* (RF). Untuk menangani ketidakseimbangan kelas, teknik SMOTE diterapkan pada data latih [20]. Hasil dari lapisan ini berupa distribusi sentimen dan metrik evaluasi model. Selanjutnya, pada *Generative AI Layer*, distribusi sentimen digunakan sebagai input untuk melakukan peringkasan isu dominan serta menghasilkan rekomendasi perbaikan produk secara otomatis. *Output* akhir sistem disajikan dalam bentuk *Insight Dashboard* yang menampilkan visualisasi sentimen dan rekomendasi berbasis data. Arsitektur ini menunjukkan integrasi antara penyimpanan *Big Data*, klasifikasi sentimen berbasis machine learning, dan interpretasi berbasis *Generative AI* dalam satu sistem analitik terpadu.

3. HASIL DAN PEMBAHASAN

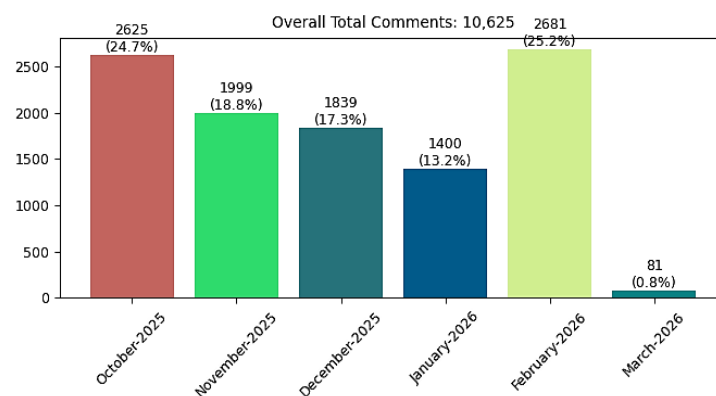
3.1 Arsitektur Karakteristik dan Distribusi Data

Berdasarkan proses *crawling* menggunakan *YouTube Data API*, diperoleh total 10.625 komentar dari periode Oktober 2025 hingga Maret 2026. Distribusi temporal menunjukkan variasi volume komentar yang cukup signifikan. Februari 2026 memiliki proporsi tertinggi sebesar 25,2%, diikuti Oktober 2025 sebesar 24,7%, sedangkan Maret 2026 menunjukkan penurunan drastis menjadi 0,8%. Pola ini mengindikasikan adanya momentum tertentu seperti pembaruan sistem, rilis versi baru, atau isu viral yang memicu peningkatan interaksi pengguna. Distribusi komentar berdasarkan kata kunci ditunjukkan pada Gambar 3. Terlihat bahwa pencarian dengan kata kunci “Tokopedia aplikasi review” mendominasi dataset sebesar 59,6%, diikuti Shopee sebesar 14,9%, Gojek sebesar 11,6%, OVO sebesar 7,5%, dan Traveloka sebesar 6,3%. Dominasi satu entitas dalam *dataset* ini perlu diperhatikan karena berpotensi memengaruhi distribusi sentimen dan bias model klasifikasi sebagaimana dijelaskan dalam penelitian analitik *Big Data* sebelumnya [1].

Selanjutnya, distribusi komentar berdasarkan periode waktu ditampilkan pada Gambar 4. Fluktuasi jumlah komentar menunjukkan bahwa interaksi pengguna sangat dipengaruhi oleh dinamika layanan digital. Kenaikan tajam pada periode tertentu dapat berkorelasi dengan pembaruan sistem atau isu teknis yang memicu respons pengguna.



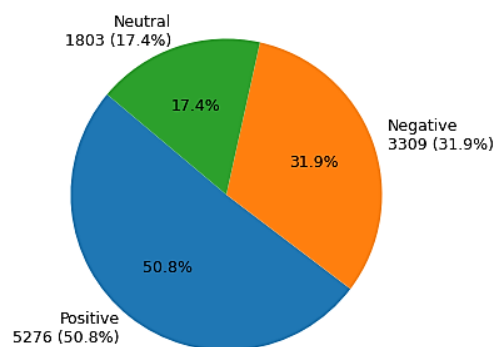
Gambar 3. Distribusi Komentar Berdasarkan Kata Kunci



Gambar 4. Distribusi Komentar Berdasarkan Periode

3.2 Distribusi Sentimen

Hasil klasifikasi sentimen terhadap seluruh komentar ditunjukkan pada Gambar 5. Distribusi sentimen menunjukkan bahwa 50,8% komentar bersifat positif, 31,9% negatif, dan 17,4% netral.



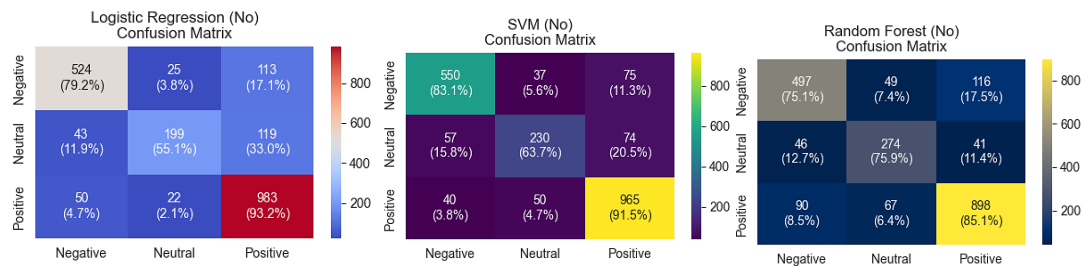
Gambar 5. Distribusi Sentimen Komentar

Meskipun sentimen positif mendominasi, proporsi sentimen negatif yang mendekati sepertiga menunjukkan adanya isu yang signifikan. Temuan ini konsisten dengan studi analisis sentimen media sosial yang menunjukkan bahwa opini publik sering terdistribusi tidak merata dan mengandung polaritas dominan tertentu [5]. Proporsi sentimen negatif ini menjadi dasar analisis lebih lanjut pada tahap evaluasi model dan identifikasi isu prioritas.

3.3 Evaluasi Kinerja Model Klasifikasi

Evaluasi dilakukan terhadap tiga algoritma klasifikasi, yaitu *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Random Forest* (RF), menggunakan representasi fitur TF-IDF. Pengujian dilakukan dalam dua skenario, yaitu tanpa SMOTE dan dengan SMOTE. Metrik evaluasi yang digunakan adalah *accuracy* dan F1-macro untuk menjaga keseimbangan performa antarkelas.

3.3.1 Model Tanpa SMOTE

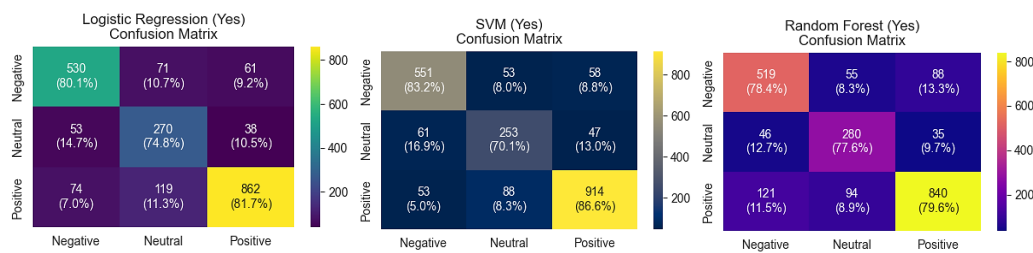


Gambar 6. Confusion Matrix Model Tanpa SMOTE (LR, SVM, dan RF)

Berdasarkan hasil evaluasi pada Gambar 6, SVM tanpa SMOTE menunjukkan performa terbaik dengan *accuracy* sebesar 83,97% dan F1-macro sebesar 80,29%. *Logistic Regression* mencapai *accuracy* 82,10%, sedangkan *Random Forest* berada pada 80,32%. Dari *confusion matrix* terlihat bahwa ketiga model mampu mengklasifikasikan kelas Positive dengan tingkat *recall* tertinggi, khususnya pada SVM yang mencapai lebih dari 91% pada kelas tersebut. Hal ini menunjukkan bahwa pola sentimen positif relatif lebih konsisten dan memiliki karakteristik fitur yang lebih jelas dalam ruang representasi TF-IDF. Sebaliknya, kelas Neutral merupakan kelas yang paling sulit diklasifikasikan. Terlihat adanya kesalahan klasifikasi yang cukup tinggi antara kelas Neutral dan Positive. Fenomena ini konsisten dengan karakteristik teks netral yang cenderung ambigu secara emosional dan sering kali mengandung kata-kata yang juga muncul pada sentimen positif ringan. Performa SVM yang lebih unggul dapat dijelaskan oleh kemampuannya membangun hyperplane optimal pada ruang fitur berdimensi tinggi, yang sangat sesuai dengan karakteristik data teks berbasis TF-IDF.

3.3.2 Model Dengan SMOTE

Penerapan SMOTE tidak memberikan peningkatan performa yang signifikan. SVM dengan SMOTE menghasilkan *accuracy* 82,68% dan F1-macro 79,40%, sedikit lebih rendah dibandingkan dengan tanpa SMOTE. *Logistic Regression* dan *Random Forest* juga mengalami penurunan performa. Dari *confusion matrix* pada Gambar 7 terlihat bahwa distribusi prediksi menjadi lebih merata, namun tidak meningkatkan kemampuan model dalam membedakan kelas secara signifikan. Hal ini menunjukkan bahwa dataset tidak berada dalam kondisi ketidakseimbangan ekstrem sehingga teknik oversampling sintesis tidak memberikan keuntungan yang berarti. Beberapa penelitian menyatakan bahwa penerapan SMOTE pada data teks berbasis TF-IDF yang bersifat sparse tidak selalu meningkatkan performa karena pembentukan data sintesis dapat menghasilkan vektor fitur yang kurang representatif.



Gambar 7. Confusion Matrix Model Dengan SMOTE (LR, SVM, dan RF)

3.4 Perbandingan Kinerja Model

Setelah evaluasi individual, dilakukan perbandingan agregat performa model seperti pada Tabel 1.

Tabel 1. Perbandingan Accuracy dan F1-macro Model

No	Model	SMOTE	Accuracy (%)	F1-Macro (%)
1	Logistic Regression	No	82.10	78.04
2	SVM	No	83.97	80.29
3	Random Forest	No	80.32	78.28
4	Logistic Regression	Yes	79.98	77.22
5	SVM	Yes	82.68	79.40
6	Random Forest	Yes	78.87	77.05

Hasil perbandingan pada Tabel 1 menunjukkan bahwa SVM tanpa SMOTE secara konsisten menghasilkan nilai *accuracy* dan F1-macro tertinggi dibandingkan dengan konfigurasi model lainnya.

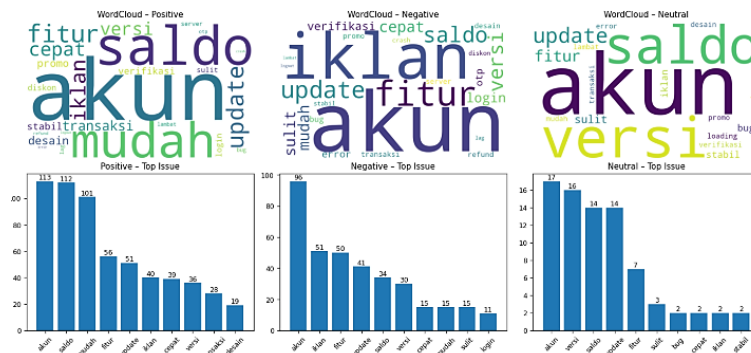


Gambar 8. Perbandingan Accuracy dan F1-macro Model

Grafik komparasi pada Gambar 8 memperjelas bahwa penerapan *SMOTE* tidak memberikan peningkatan signifikan terhadap performa model. Secara konsisten, *SVM* tanpa *SMOTE* berada pada posisi terbaik.

3.5 Analisis Konten dan Isu Dominan

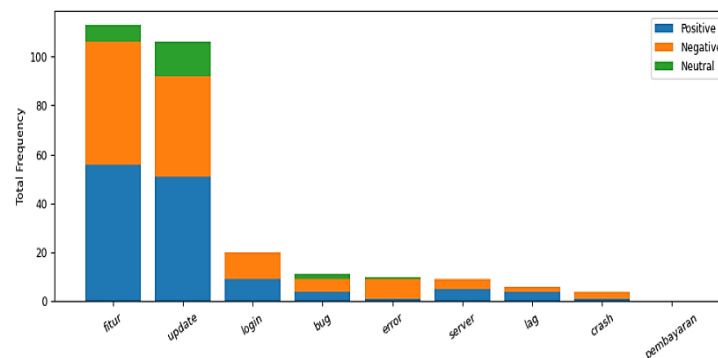
Analisis *wordcloud* pada Gambar 9 menunjukkan bahwa Sentimen Positif didominasi oleh kata seperti saldo, akun, mudah, fitur, dan update. Hal ini menunjukkan bahwa pengguna mengapresiasi kemudahan penggunaan dan fitur sistem. Sentimen Negatif didominasi oleh kata akun, login, error, server, dan bug. Isu teknis menjadi sumber utama ketidakpuasan pengguna. Sentimen Netral didominasi oleh kata versi, update, akun, dan saldo, yang menunjukkan sifat informatif tanpa ekspresi emosional kuat. Meskipun secara kuantitatif sentimen positif mendominasi, isu teknis pada sentimen negatif memiliki implikasi yang lebih strategis terhadap perbaikan sistem.



Gambar 9. Wordcloud dan Frekuensi Kata Dominan Berdasarkan Sentimen

3.6 Analisis Prioritas Perbaikan

Dari Gambar 10, berdasarkan agregasi frekuensi kata negatif, prioritas utama perbaikan adalah: sistem login dan autentikasi, stabilitas server, bug pasca-update, error pada fitur tertentu. Temuan ini menunjukkan bahwa perbaikan teknis memiliki urgensi lebih tinggi dibandingkan dengan penambahan fitur baru.



Gambar 10. Prioritas Perbaikan Berdasarkan Dominasi Sentimen

3.7 Diskusi

Hasil penelitian ini menunjukkan bahwa kombinasi *TF-IDF* dan *Support Vector Machine (SVM)* memberikan performa terbaik dibandingkan dengan Logistic Regression dan Random Forest pada klasifikasi

sentimen komentar *YouTube*. Temuan ini menunjukkan bahwa *SVM* masih efektif untuk menangani data teks berdimensi tinggi, terutama ketika data telah melalui tahap *preprocessing* yang memadai dan direpresentasikan dengan fitur *TF-IDF*. Dalam penelitian ini, *SVM* menghasilkan keseimbangan performa antarkelas yang lebih baik dibandingkan dengan model lain, sehingga memperoleh akurasi dan nilai *F1-macro* tertinggi. Hasil tersebut sejalan dengan penelitian sebelumnya yang menunjukkan bahwa *SVM* tetap kompetitif pada tugas klasifikasi sentimen berbasis teks dan studi komparatif model klasifikasi [6], [21].

Di sisi lain, hasil penelitian ini juga menunjukkan bahwa penerapan *SMOTE* tidak selalu meningkatkan performa model. Pada data teks berbasis *TF-IDF* yang bersifat *sparse*, pembentukan sampel sintesis dapat menghasilkan representasi fitur yang tidak selalu lebih baik daripada distribusi data asli. Dalam penelitian ini, model tanpa *SMOTE* justru memberikan hasil yang lebih stabil dibandingkan dengan model dengan *SMOTE*. Temuan tersebut sejalan dengan penelitian sebelumnya yang menegaskan bahwa efektivitas *SMOTE* sangat dipengaruhi oleh distribusi kelas awal, karakteristik fitur, dan metode klasifikasi yang digunakan [19], [20].

Dari sisi substansi hasil, dominasi sentimen positif tidak menghilangkan arti penting proporsi sentimen negatif yang cukup besar. Analisis kata dominan menunjukkan bahwa istilah seperti *login*, *server*, *bug*, dan *error* menjadi sumber utama keluhan pengguna. Hal ini menunjukkan bahwa analisis sentimen tidak hanya berguna untuk menggambarkan kecenderungan opini secara umum, tetapi juga dapat dimanfaatkan untuk mengidentifikasi area perbaikan layanan yang paling mendesak. Dengan demikian, hasil klasifikasi dapat digunakan sebagai dasar penyusunan prioritas tindakan yang lebih terarah dalam evaluasi produk digital berbasis umpan balik pengguna.

Kontribusi akademik penelitian ini terletak pada integrasi antara penyimpanan *Big Data* berbasis *MongoDB*, klasifikasi sentimen komparatif, dan interpretasi hasil menggunakan *Generative AI* dalam satu arsitektur sistem analitik yang terpadu. Berbeda dengan penelitian sebelumnya yang umumnya berhenti pada tahap evaluasi performa model, penelitian ini menempatkan hasil klasifikasi sebagai dasar untuk proses interpretasi lanjutan berupa peringkasan isu dominan dan rekomendasi prioritas perbaikan. Pendekatan ini memperluas fungsi analisis sentimen dari sekadar pengukuran performa model menjadi sistem pendukung keputusan yang lebih operasional. Posisi ini juga sejalan dengan perkembangan penelitian terkini yang mendorong integrasi *pipeline NLP* secara *end-to-end* [22], [23], [24].

4. KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan arsitektur sistem analitik umpan balik *YouTube* berbasis *Big Data* yang mengintegrasikan data *layer*, *analytics layer*, dan *interpretative layer* dalam satu pipeline terpadu. Sistem memanfaatkan *YouTube Data API* untuk pengumpulan data, *MongoDB* sebagai penyimpanan dokumen tidak terstruktur, serta pendekatan analisis sentimen berbasis *TF-IDF* dan machine learning untuk klasifikasi opini pengguna. Hasil eksperimen menunjukkan bahwa *Support Vector Machine* (*SVM*) tanpa *SMOTE* memberikan performa terbaik dengan accuracy sebesar 83,97% dan *F1-macro* sebesar 80,29%, dibandingkan dengan *Logistic Regression* dan *Random Forest*. Penerapan *SMOTE* tidak memberikan peningkatan signifikan terhadap performa model, yang mengindikasikan bahwa distribusi kelas dalam dataset tidak berada pada kondisi ketidakseimbangan ekstrem. Temuan ini menunjukkan bahwa perbaikan teknis memiliki urgensi yang tinggi berdasarkan dominasi sentimen negatif. Selain itu, hasil penelitian juga menunjukkan bahwa penerapan *SMOTE* pada data teks berbasis *TF-IDF* yang bersifat *sparse* tidak selalu meningkatkan kinerja klasifikasi. Dari sisi analisis konten, meskipun sentimen positif mendominasi secara kuantitatif, isu teknis seperti *login*, *server*, *bug*, dan *error* menjadi faktor utama yang mendorong sentimen negatif. Integrasi *Generative AI* dalam sistem memungkinkan proses klasifikasi tidak berhenti pada pelaporan metrik evaluasi, tetapi dilanjutkan dengan peringkasan isu dominan dan penyusunan rekomendasi prioritas perbaikan produk secara otomatis. Kontribusi utama penelitian ini terletak pada pendekatan arsitektural yang mengintegrasikan penyimpanan *Big Data* berbasis *MongoDB*, klasifikasi sentimen komparatif, dan interpretasi berbasis *Generative AI* dalam satu sistem analitik *end-to-end*. Sistem yang dikembangkan mampu mengotomatisasi ekstraksi insight dari komentar pengguna serta mendukung pengambilan keputusan berbasis data secara lebih sistematis dan efisien. Sebagai pengembangan selanjutnya, penelitian dapat diperluas dengan memanfaatkan model berbasis deep learning atau transformer untuk meningkatkan performa klasifikasi, serta menguji arsitektur sistem pada domain produk digital yang lebih beragam guna meningkatkan generalisasi dan skalabilitas sistem [24].

REFERENSI

- [1] H. L. A. Asro, A. Wicaksono, and P. Herwanto, "Strategi Dinamis Menggunakan MongoDB untuk Analisis Sentimen terhadap Komentar YouTube Pilkada Gubernur Indonesia 2024," 2024. [Online]. Available: <https://jurnaldrpm.budiluhur.ac.id/index.php/Kresna/>
- [2] H. Henderi, R. Irawatia, I. Indra, D. A. Dewi, and T. B. Kurniawan, "Big Data Analysis using Elasticsearch and Kibana: A Rating Correlation to Sustainable Sales of Electronic Goods," *HighTech and Innovation Journal*, vol. 4, no. 3, pp. 583–591, Sep. 2023, doi: 10.28991/HIJ-2023-04-03-09.

- [3] M. Jannatul Firdousi Zoti, A. Mohammad Akib, M. Rahman, S. Sadik Khan, and S. Shafin Ahmed, "Sentiment Analysis of YouTube Comments: A Comprehensive Study of Machine Learning Models," 2025.
- [4] Asro, A. Sulaiman, Henderi, and Sudaryono, "Performance Comparison of Naive Bayes and Logistic Regression for Sentiment Analysis of YouTube Comments on Indonesia's Education System," *Institute of Electrical and Electronics Engineers (IEEE)*, Dec. 2025, pp. 1–6. doi: 10.1109/icast68191.2025.11300032.
- [5] Henderi, Asro, A. Sulaiman, T. B. Kurniawan, D. A. Dewi, and M. Alqudah, "Utilizing Sentiment Analysis for Reflect and Improve Education in Indonesia," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 189–200, Jan. 2025, doi: 10.47738/jads.v6i1.527.
- [6] H. Basri, M. B. S. Junianto, and I. Kusyadi, "Enhancing Usability Testing Through Sentiment Analysis: A Comparative Study Using SVM, Naive Bayes, Decision Trees and Random Forest," *Jurnal Teknologi Sistem Informasi dan Aplikasi*, vol. 7, no. 4, pp. 1603–1610, Oct. 2024, doi: 10.32493/jtsi.v7i4.45117.
- [7] A. Sholekhah and M. Muntahanah, "Perbandingan Naïve Bayes dan Support Vector Machine Dalam Analisa Sentimen Tentang Penyitaan Aset Koruptor di Twitter," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 981–989, Jul. 2025, doi: 10.57152/malcom.v5i3.2068.
- [8] F. Suandi *et al.*, "Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning Algorithms," 2024, pp. 126–138. doi: 10.2991/978-94-6463-620-8_10.
- [9] I. G. B. A. Budaya and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1077–1086, Jul. 2024, doi: 10.57152/malcom.v4i3.1459.
- [10] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J. Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.
- [11] N. Kurniati and N. Rukhviyanti, "Minimarket Sales Optimization: Implementation Of Fp-Growth Algorithm And Mongodb With Python Optimalisasi Penjualan Minimarket: Implementasi Algoritma Fp-Growth Dan Mongodb Dengan Python," vol. 10, no. 1, p. 2025, 2025.
- [12] Susan Juli Safitri, Gelar Alam Ramdhaniawan, Asro Asro, and Novi Rukhviyanti, "Analisis Literatur Review Perencanaan Strategi Sistem Informasi Menggunakan Metode Metode Five Competitive Force Pada CV. Bio Chitosan Indonesia," *Bridge : Jurnal publikasi Sistem Informasi dan Telekomunikasi*, vol. 2, no. 4, pp. 319–327, Sep. 2024, doi: 10.62951/bridge.v2i4.263.
- [13] Novi Rukhviyanti, "Balanced Scorecard: Performance Measurement Of University Mediated By Student Loyalty," *Jurnal Manajemen*, vol. 29, no. 2, pp. 400–420, Jun. 2025, doi: 10.24912/jm.v29i2.2694.
- [14] F. Sufi, "A New Social Media-Driven Cyber Threat Intelligence," *Electronics (Switzerland)*, vol. 12, no. 5, Mar. 2023, doi: 10.3390/electronics12051242.
- [15] X. Liang, M. Zhang, G. Feng, D. Wang, Y. Xu, and F. Gu, "Few-Shot Learning Approaches for Fault Diagnosis Using Vibration Data: A Comprehensive Review," *Sustainability*, vol. 15, no. 20, p. 14975, Oct. 2023, doi: 10.3390/su152014975.
- [16] C. Suhaeni and H. S. Yong, "Enhancing Imbalanced Sentiment Analysis: A GPT-3-Based Sentence-by-Sentence Generation Approach," *Applied Sciences (Switzerland)*, vol. 14, no. 2, 2024, doi: 10.3390/app14020622.
- [17] OpenAI *et al.*, "GPT-4 Technical Report," Mar. 2024, [Online]. Available: <http://arxiv.org/abs/2303.08774>
- [18] F. Ayu Shefia *et al.*, "Sistemasi: Jurnal Sistem Informasi Analisis Sentimen Ulasan Aplikasi CapCut pada Google Play Store menggunakan Support Vector Machine dengan Teknik SMOTE Sentiment Analysis of CapCut Application Reviews using Support Vector Machine with the SMOTE Technique." [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [19] N. A. Smary, W. Ahmed, K. Amin, P. Plawiak, and M. Hammad, "Enhancing machine learning-based sentiment analysis through feature extraction techniques," *PLoS One*, vol. 19, no. 2 February, Feb. 2024, doi: 10.1371/journal.pone.0294968.
- [20] A. Robi Padri, A. Asro, and I. Indra, "Classification of Traffic Congestion in Indonesia Using the Naive Bayes Classification Method," *Journal of World Science*, vol. 2, no. 6, pp. 877–888, Jun. 2023, doi: 10.58344/jws.v2i6.285.
- [21] A. B. Putra Negara, "The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 14, no. 3, p. 172, Dec. 2023, doi: 10.24843/lkjiti.2023.v14.i03.p05.

- [22] C. Setia and N. Rukhviyanti, "Pipeline NLP End-to-End untuk Peringkasan Abstraktif dan Ekstraksi Entitas Berita Berbahasa Indonesia Berbasis Model Transformer," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 11, no. 1, pp. 1–11, Feb. 2026, doi: 10.30591/jpit.v11i1.10030.
- [23] Asro and Solihin, "Comparative Evaluation of Preprocessing Techniques in Twitter Sentiment Analysis for Indonesia's 2024 Regional Elections," *INOVTEK Polbeng-Seri Informatika*, vol. 11, no. 1, 2026.
- [24] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, "Transformer-based models for sentiment analysis of YouTube video comments," *Sci. Afr.*, vol. 29, Sep. 2025, doi: 10.1016/j.sciaf.2025.e02836.