



Comparative Performance Evaluation of Naïve Bayes and Logistic Regression for Indonesian YouTube Comment Sentiment Classification

Evaluasi Komparatif Kinerja Algoritma *Naïve Bayes* dan *Logistic Regression* dalam Klasifikasi Sentimen Komentar YouTube Berbahasa Indonesia

Meisa Monica^{1*}, Asto Purwanto²

^{1,2}Department of Information Systems, STMIK IM Bandung, Indonesia

E-Mail: ¹meii.user90@gmail.com, ²astopurwanto@stmik-im.ac.id

Received Mar 13th 2026; Revised Apr 20th 2026; Accepted Apr 27th 2026; Available Online Apr 30th 2026

Corresponding Author: Meisa Monica

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

Abstract

User comments on YouTube represent a valuable source of information for understanding public opinion toward digital services. However, these comments are typically written in unstructured text, making manual analysis inefficient as data volumes grow. Therefore, a machine learning approach is required to automatically process and classify user opinions. This study performs sentiment analysis on Indonesian YouTube comments collected using several keywords related to digital services. The collected data were processed through text-cleaning and normalization, followed by feature representation using the TF-IDF technique. To improve classification performance, parameter optimization and SMOTE-based data balancing were applied. The evaluation results for accuracy, precision, recall, and F1-score indicate that the classification model achieves good performance, with a maximum accuracy of 84.88%. Furthermore, the sentiment distribution shows that positive sentiment occurs more frequently than neutral or negative sentiment in the analyzed comments.

Keywords: Logistic Regression, Naïve Bayes, Sentiment Analysis, Text Classification, TF-IDF.

Abstrak

Komentar pengguna pada platform YouTube merupakan sumber informasi yang dapat dimanfaatkan untuk memahami kecenderungan opini publik terhadap berbagai layanan digital. Namun, komentar tersebut umumnya berbentuk teks tidak terstruktur sehingga proses analisis secara manual menjadi kurang efektif ketika jumlah data sangat besar. Oleh karena itu, diperlukan pendekatan berbasis *machine learning* untuk mengolah serta mengklasifikasikan opini pengguna secara otomatis. Penelitian ini melakukan analisis sentimen terhadap komentar YouTube berbahasa Indonesia yang dikumpulkan menggunakan beberapa kata kunci terkait layanan digital. Data yang diperoleh kemudian diproses melalui tahapan pembersihan teks, normalisasi, serta pembentukan representasi fitur menggunakan metode TF-IDF. Untuk meningkatkan kualitas model klasifikasi, dilakukan optimasi parameter serta penyeimbangan distribusi data menggunakan teknik SMOTE. Hasil evaluasi berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score* menunjukkan bahwa model klasifikasi mampu menghasilkan performa yang baik dengan tingkat akurasi tertinggi mencapai 84,88%. Selain itu, distribusi hasil klasifikasi memperlihatkan bahwa kategori sentimen positif lebih dominan dibandingkan dengan kategori netral dan negatif pada data komentar yang dianalisis.

Kata kunci: Analisis Sentimen, Klasifikasi Teks, *Naïve Bayes*, Regresi Logistik, TF-IDF.

1. PENDAHULUAN

Media sosial telah berkembang menjadi ruang utama bagi masyarakat untuk menyampaikan opini terhadap berbagai isu, termasuk pengalaman mereka dalam menggunakan layanan digital. Dalam kehidupan sehari-hari, layanan digital seperti aplikasi *mobile*, platform berbasis web, layanan transaksi *online*, *e-commerce*, *mobile banking*, transportasi daring, pendidikan digital, serta sistem pelayanan publik berbasis internet semakin sering digunakan oleh masyarakat. Kondisi tersebut menjadikan layanan digital sebagai konteks yang penting untuk dikaji karena kualitas layanan yang diberikan secara langsung memengaruhi



kenyamanan, kepercayaan, kepuasan, dan loyalitas pengguna. Opini pengguna terhadap layanan digital dapat memberikan gambaran mengenai pengalaman mereka, baik dalam bentuk apresiasi, keluhan, kritik, maupun kebutuhan perbaikan terhadap layanan yang digunakan [1].

Pemilihan konteks layanan digital dalam penelitian ini didasarkan pada meningkatnya ketergantungan masyarakat terhadap teknologi digital dalam berbagai aktivitas. Semakin banyak pengguna yang berinteraksi dengan layanan digital, semakin besar pula kemungkinan munculnya komentar, tanggapan, dan keluhan di media sosial. Dalam hal ini, komentar pengguna dapat menjadi sumber informasi penting bagi pengembang layanan, pengambil keputusan, maupun pihak organisasi untuk memahami persepsi publik secara lebih cepat. Oleh karena itu, analisis terhadap opini pengguna pada layanan digital menjadi relevan untuk mengetahui kecenderungan sentimen masyarakat dan mendukung perbaikan kualitas layanan berbasis data.

YouTube dipilih sebagai sumber data karena *platform* ini memiliki tingkat interaksi yang tinggi dan menyediakan kolom komentar terbuka yang merekam respons pengguna secara langsung. Banyak konten YouTube yang membahas penggunaan aplikasi, layanan digital, sistem *online*, tutorial teknologi, pengalaman pengguna, maupun isu pelayanan digital tertentu. Komentar yang muncul pada konten tersebut sering kali memuat pengalaman nyata pengguna, seperti kepuasan terhadap kemudahan layanan, keluhan terhadap gangguan sistem, kritik terhadap fitur aplikasi, atau harapan terhadap peningkatan kualitas layanan. Berbeda dengan survei yang bersifat terstruktur, komentar YouTube cenderung muncul secara spontan sehingga dapat merepresentasikan opini pengguna secara lebih alami [2]. Namun, komentar YouTube umumnya berbentuk teks tidak terstruktur, menggunakan bahasa tidak baku, singkatan, campuran bahasa, serta memiliki jumlah yang besar. Kondisi ini membuat analisis manual menjadi kurang efisien dan berpotensi menghasilkan penilaian yang subjektif. Oleh karena itu, diperlukan pendekatan komputasional untuk mengidentifikasi pola sentimen secara otomatis. Dalam kajian *text mining*, analisis sentimen digunakan untuk mengelompokkan opini pengguna ke dalam kategori positif, netral, dan negatif. Pendekatan *machine learning* banyak digunakan karena mampu melakukan klasifikasi teks secara lebih konsisten pada data berukuran besar.

Pada penelitian ini, dua algoritma dipilih untuk dibandingkan, yaitu *Naïve Bayes* (NB) dan *Logistic Regression* (LR). NB dikenal sebagai algoritma yang sederhana, cepat, dan efisien untuk klasifikasi teks, sedangkan LR umumnya memiliki kestabilan yang baik dalam memodelkan hubungan antara fitur teks dan kelas sentiment [3],[4]. Perbandingan kedua algoritma ini penting dilakukan untuk mengetahui model yang lebih efektif dalam mengklasifikasikan sentimen komentar YouTube berbahasa Indonesia, khususnya pada konteks layanan digital. Penelitian ini juga menggunakan representasi fitur TF-IDF karena metode ini mampu memberikan bobot lebih besar pada kata-kata yang penting dalam suatu dokumen, tetapi tidak terlalu sering muncul di seluruh korpus. Dibandingkan representasi yang lebih kompleks seperti *word embedding* atau model transformer, TF-IDF lebih ringan secara komputasi, mudah diinterpretasikan, dan sesuai untuk skenario klasifikasi teks dengan algoritma klasik seperti NB dan LR [5].

Dengan demikian, penelitian ini bertujuan untuk melakukan evaluasi komparatif terhadap kinerja NB dan LR dalam klasifikasi sentimen komentar YouTube berbahasa Indonesia pada konteks layanan digital. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai algoritma yang lebih efektif untuk mengidentifikasi opini publik secara otomatis, sekaligus menunjukkan bahwa komentar YouTube dapat dimanfaatkan sebagai sumber data alternatif dalam memahami persepsi pengguna terhadap layanan digital[6].

1.1 Tinjauan Literatur

Kajian analisis sentimen pada media sosial terus mengalami peningkatan seiring bertambahnya konten opini yang diproduksi pengguna di berbagai *platform*, termasuk Twitter, Facebook, dan YouTube. Komentar yang muncul pada *platform* tersebut dapat diolah melalui pendekatan *text mining*, kemudian dimanfaatkan dalam model *machine learning* untuk melakukan klasifikasi opini secara otomatis ke dalam kelas sentimen tertentu [7]. Dalam konteks klasifikasi teks, *Naïve Bayes* menjadi salah satu metode yang sering digunakan karena berbasis probabilitas serta relatif ringan dari sisi komputasi, terutama ketika fitur yang digunakan berasal dari representasi kata [8]. Selain itu, *Logistic Regression* juga banyak dipakai pada tugas analisis sentimen karena mampu mengestimasi peluang kelas dari fitur teks secara lebih stabil, sehingga performanya cenderung konsisten pada variasi dataset tertentu [9]. Kinerja model klasifikasi tidak hanya bergantung pada pemilihan algoritma, tetapi juga dipengaruhi oleh tahapan *preprocessing* dan strategi representasi fitur. Pada penelitian ini, TF-IDF digunakan sebagai metode representasi teks karena mampu memberikan bobot pada kata berdasarkan tingkat kemunculannya dalam dokumen dan kekhasannya dalam seluruh korpus. Dengan cara ini, kata yang sering muncul pada satu komentar tetapi tidak terlalu umum pada seluruh data dapat memiliki bobot yang lebih bermakna untuk proses klasifikasi [10].

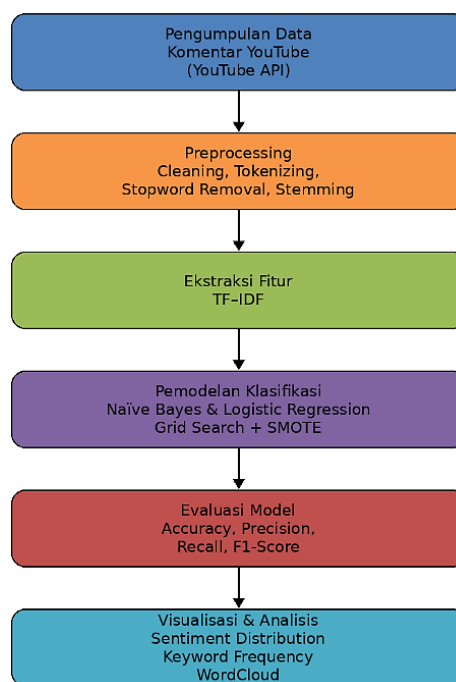
Pemilihan TF-IDF juga didasarkan pada kesesuaiannya dengan algoritma klasik seperti *Naïve Bayes* dan *Logistic Regression*. Dibandingkan metode sederhana seperti *Bag of Words*, TF-IDF lebih baik karena tidak hanya menghitung frekuensi kata, tetapi juga memperhatikan tingkat kepentingan kata dalam korpus. Sementara itu, dibandingkan dengan metode yang lebih kompleks seperti *word embedding* atau *transformer*, TF-IDF lebih ringan secara komputasi, lebih mudah diinterpretasikan, dan cukup sesuai untuk penelitian

yang berfokus pada evaluasi komparatif algoritma *machine learning* klasik. Oleh karena itu, TF-IDF dipilih agar proses klasifikasi tetap efisien, terukur, dan relevan dengan tujuan penelitian ini.

Permasalahan lain yang kerap dijumpai pada data sentimen adalah ketimpangan distribusi kelas, misalnya ketika salah satu kelas memiliki jumlah data jauh lebih sedikit dibandingkan dengan kelas lainnya. Untuk mengurangi bias model terhadap kelas mayoritas, teknik seperti SMOTE dapat diterapkan dengan menghasilkan sampel sintesis pada kelas minoritas agar distribusi data menjadi lebih proporsional [11]. Berdasarkan sejumlah penelitian terdahulu, pemilihan algoritma yang tepat, metode ekstraksi fitur, serta strategi penanganan data tidak seimbang dapat berkontribusi pada peningkatan performa analisis sentimen. Oleh sebab itu, penelitian ini mengkaji dan membandingkan kinerja *Naïve Bayes* dan *Logistic Regression* pada komentar YouTube berbahasa Indonesia dengan memanfaatkan TF-IDF, optimasi parameter, serta pendekatan penyeimbangan data sebagai bagian dari proses evaluasi model [12].

2 METODE PENELITIAN

Penelitian ini memanfaatkan *machine learning* untuk klasifikasi sentimen komentar YouTube berbahasa Indonesia. Proses penelitian mencakup pengumpulan data, *preprocessing*, ekstraksi fitur, pemodelan, serta evaluasi, sehingga analisis dapat dilakukan secara sistematis dan menghasilkan model klasifikasi otomatis[13]. Alur tahapan ditampilkan pada Gambar 1.



Gambar 1. Tahapan Penelitian Analisis

2.1 Pengumpulan Data

Sumber data penelitian berasal dari komentar pengguna di *platform* YouTube yang diekstraksi menggunakan YouTube Data API. Proses ekstraksi dilakukan berdasarkan sejumlah kata kunci terkait layanan digital, lalu hasilnya disimpan dalam basis data untuk diproses pada tahapan berikutnya.

2.2 Preprocessing Teks

Sebelum proses klasifikasi dilakukan, komentar YouTube terlebih dahulu melalui tahap *preprocessing* untuk meningkatkan kualitas data teks. Tahapan yang dilakukan meliputi *text cleaning*, yaitu menghapus URL, tanda baca, angka, simbol, emoji, dan karakter yang tidak relevan. Selanjutnya dilakukan *case folding* untuk mengubah seluruh huruf menjadi *lowercase*, kemudian normalisasi untuk memperbaiki kata tidak baku atau singkatan menjadi bentuk baku. Setelah itu, teks diproses melalui tokenisasi, *stopword removal*, dan *stemming* agar kata-kata yang tidak terlalu berpengaruh dapat dikurangi serta bentuk kata menjadi lebih konsisten. Tahapan ini dilakukan agar data komentar lebih bersih dan siap digunakan pada proses pembentukan fitur TF-IDF serta klasifikasi menggunakan *Naïve Bayes* dan *Logistic Regression*.

2.3 Ekstraksi

Pada tahap ekstraksi fitur, data komentar yang telah melalui proses *preprocessing* diubah menjadi representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen serta tingkat kekhasan kata tersebut dalam keseluruhan korpus. Dengan demikian, kata yang lebih relevan dan tidak terlalu umum muncul pada seluruh data akan memperoleh bobot yang lebih tinggi. Perhitungan TF-IDF dapat dinyatakan sebagai berikut:

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \quad (2)$$

TF (t,d) adalah frekuensi kemunculan term t pada dokumen d , N adalah jumlah seluruh dokumen, sedangkan $DF(t)$ adalah jumlah dokumen yang mengandung term t . Hasil pembobotan TF-IDF selanjutnya digunakan sebagai input untuk proses klasifikasi menggunakan algoritma *Naïve Bayes* dan *Logistic Regression*.

2.4 Pemodelan Klasifikasi

Penelitian ini menerapkan dua algoritma klasifikasi, yaitu *Naïve Bayes* (NB) dan *Logistic Regression* (LR) sebagai model pembandingan. Setelah kedua metode tersebut diperkenalkan, penulisan pada bagian selanjutnya menggunakan singkatan NB dan LR secara konsisten. Penyetelan parameter dilakukan secara sistematis menggunakan *Grid Search*, sedangkan permasalahan kelas tidak seimbang ditangani dengan *Synthetic Minority Over-sampling Technique* (SMOTE) agar distribusi data antarkelas menjadi lebih proporsional [14].

2.5 Evaluasi Model

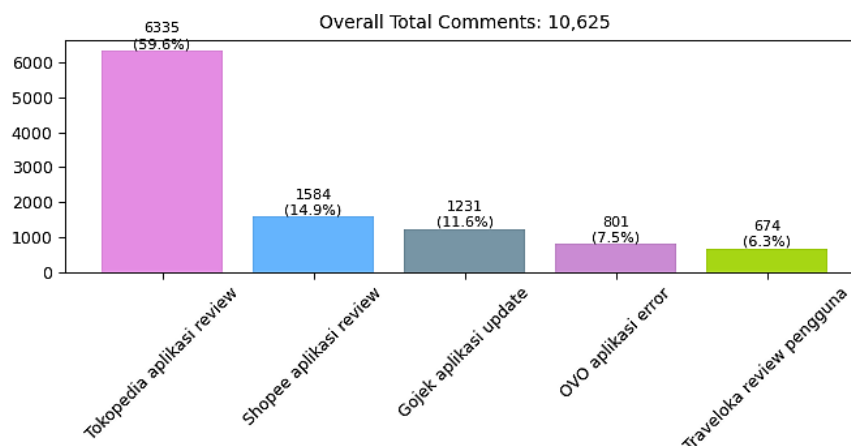
Evaluasi model dilakukan untuk mengetahui kemampuan NB dan LR dalam mengklasifikasikan sentimen komentar YouTube ke dalam kelas positif, netral, dan negatif. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk melihat ketepatan prediksi secara umum, *precision* untuk mengukur ketepatan hasil prediksi pada tiap kelas, *recall* untuk melihat kemampuan model mengenali data pada kelas yang benar, sedangkan *F1-score* digunakan untuk menilai keseimbangan antara *precision* dan *recall*. Selain itu, *confusion matrix* digunakan untuk melihat perbandingan antara label aktual dan hasil prediksi model. Penentuan parameter terbaik dilakukan menggunakan *Grid Search*, yaitu pencarian kombinasi parameter secara sistematis untuk memperoleh hasil model yang lebih optimal.

3. HASIL DAN PEMBAHASAN

Bab ini menguraikan temuan penelitian berdasarkan komentar YouTube yang dikumpulkan melalui proses *crawling* dengan beberapa kata kunci terkait layanan digital. Pembahasan mencakup gambaran jumlah dan distribusi komentar, serta hasil pengelompokan sentimen dari data yang telah diperoleh dan diproses pada tahapan sebelumnya.

2.1 Analisis Frekuensi Kata Kunci *Crawling*

Tahap awal dilakukan dengan menghitung jumlah komentar berdasarkan kata kunci yang digunakan dalam proses *crawling* YouTube. Kata kunci yang digunakan dalam penelitian ini meliputi “Tokopedia aplikasi review”, “Shopee aplikasi review”, “Gojek aplikasi update”, “Ovo aplikasi error”, dan “Traveloka review pengguna”. Dari proses *crawling* tersebut diperoleh total sebanyak 10.625 komentar.

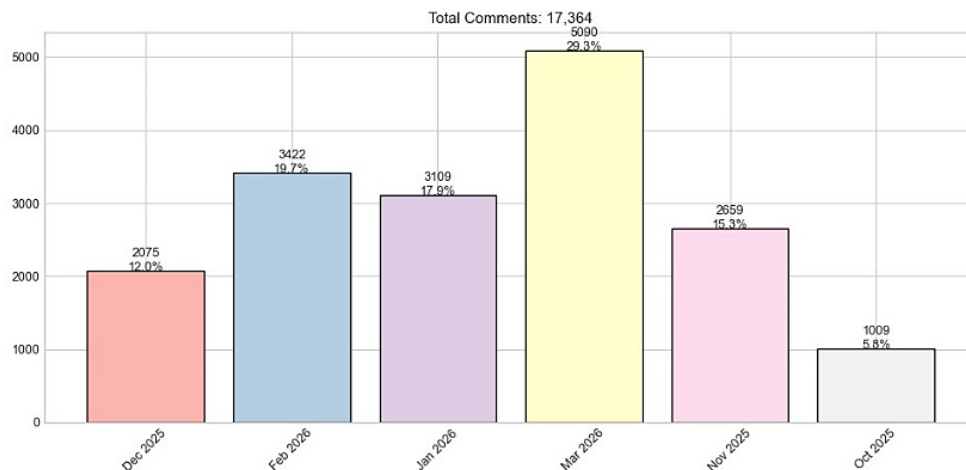


Gambar 2. Frekuensi Kata Kunci *Crawling* Komentar YouTube

Gambar 2 menunjukkan distribusi jumlah komentar berdasarkan kata kunci pencarian. Kata kunci “Tokopedia aplikasi *review*” menghasilkan komentar terbanyak, yaitu 6.335 komentar atau 59,6% dari total data. Selanjutnya, “Shopee aplikasi *review*” memperoleh 1.584 komentar atau 14,9%, “Gojek aplikasi update” sebanyak 1.231 komentar atau 11,6%, “Ovo aplikasi error” sebanyak 801 komentar atau 7,5%, dan “Traveloka *review* pengguna” sebanyak 674 komentar atau 6,3%. Hasil ini menunjukkan bahwa kata kunci Tokopedia memberikan kontribusi data paling besar dibandingkan dengan kata kunci lainnya.

2.2 Distribusi Komentar Berdasarkan Waktu

Setelah proses *crawling* data dilakukan, tahap berikutnya adalah menganalisis distribusi komentar berdasarkan waktu publikasi. Analisis ini bertujuan untuk melihat pola aktivitas pengguna dalam memberikan komentar pada periode tertentu.

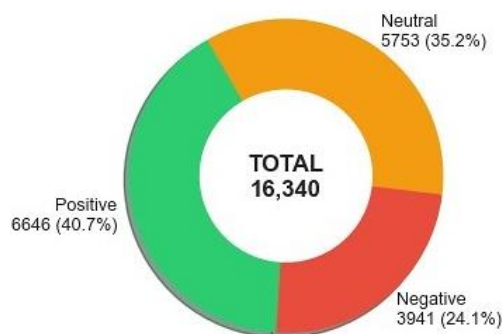


Gambar 3. Distribusi Komentar Berdasarkan Bulan

Gambar 3 menunjukkan distribusi komentar berdasarkan bulan dengan total 17.364 komentar. Jumlah komentar tertinggi terdapat pada Maret 2026, yaitu 5.090 komentar atau 29,3% dari total data. Sementara itu, jumlah komentar terendah terdapat pada Oktober 2025, yaitu 1.009 komentar atau 5,8%. Secara umum, jumlah komentar mengalami peningkatan dari Oktober 2025 hingga Maret 2026, meskipun terdapat sedikit variasi pada beberapa bulan. Hal ini menunjukkan bahwa aktivitas pengguna YouTube dalam memberikan komentar cenderung meningkat pada periode tertentu.

2.3 Distribusi Sentimen Komentar

Selanjutnya, komentar yang telah terkumpul diproses untuk menentukan kategori sentimen. Pada penelitian ini, pelabelan sentimen dilakukan dengan pendekatan berbasis lexicon/kamus sentimen Bahasa Indonesia, yaitu dengan melihat kecenderungan kata positif dan negatif pada setiap komentar. Komentar dengan skor positif lebih tinggi diberi label positif, komentar dengan skor negatif lebih tinggi diberi label negatif, sedangkan komentar yang tidak menunjukkan kecenderungan kuat dimasukkan ke dalam kategori netral. Proses pelabelan ini mengacu pada korpus atau kamus sentimen yang telah digunakan pada penelitian sebelumnya.



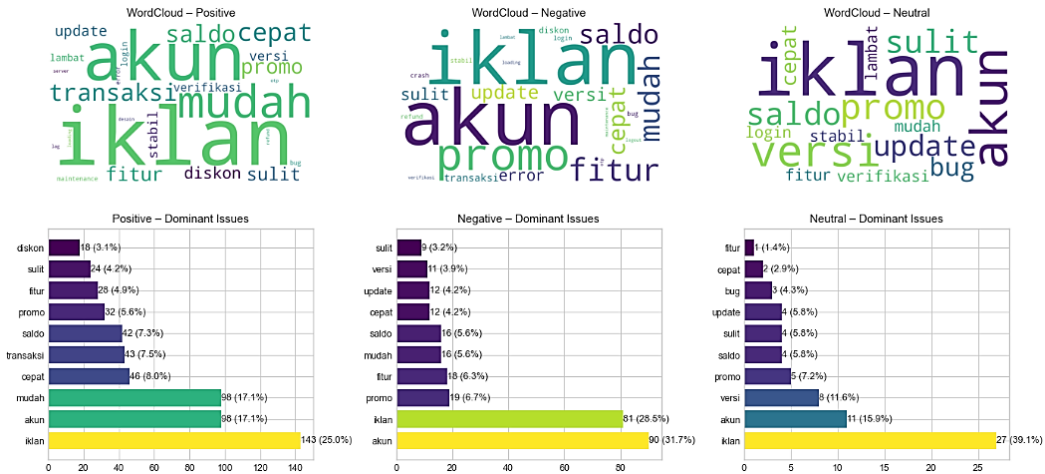
Gambar 4. Total sentiment analysis

Hasil klasifikasi mengelompokkan komentar pengguna ke dalam tiga kategori sentimen yaitu positif, netral, dan negatif. Distribusi hasil klasifikasi menunjukkan bahwa kategori sentimen positif memiliki

proporsi terbesar dibandingkan kategori lainnya (40,7%), diikuti netral (35,2%) dan negatif (24,1%), yang menunjukkan kecenderungan tanggapan pengguna lebih bersifat positif [14]

2.4 Perbandingan Kinerja Model

Selain distribusi sentimen, penelitian ini juga menampilkan *word cloud* untuk melihat kata yang paling sering muncul pada komentar pengguna di setiap kategori sentimen.

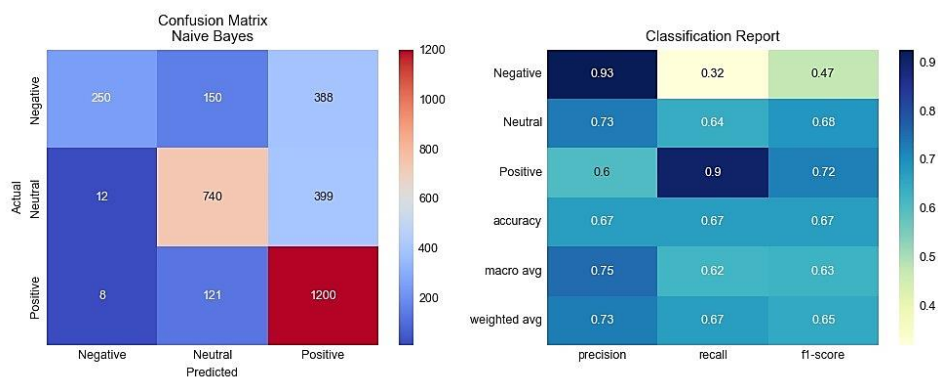


Gambar 5. Visualisasi Word Cloud

Gambar 5 menunjukkan kata-kata yang dominan pada komentar positif, negatif, dan netral. Pada sentimen positif, kata yang sering muncul antara lain “iklan”, “akun”, dan “mudah”, yang menunjukkan bahwa sebagian pengguna memberikan tanggapan baik terhadap kemudahan penggunaan aplikasi. Pada sentimen negatif, kata yang dominan seperti “akun”, “iklan”, dan “promo” mengindikasikan masih adanya keluhan atau ketidakpuasan pengguna pada aspek tertentu. Sementara itu, pada sentimen netral, kata “iklan”, “akun”, dan “versi” lebih banyak muncul sebagai bentuk komentar informatif atau tanggapan umum. Secara keseluruhan, visualisasi ini menunjukkan bahwa isu terkait akun, iklan, promo, dan versi aplikasi menjadi topik yang paling sering dibahas oleh pengguna.

2.5 Analisis Konten dan Isu Dominan

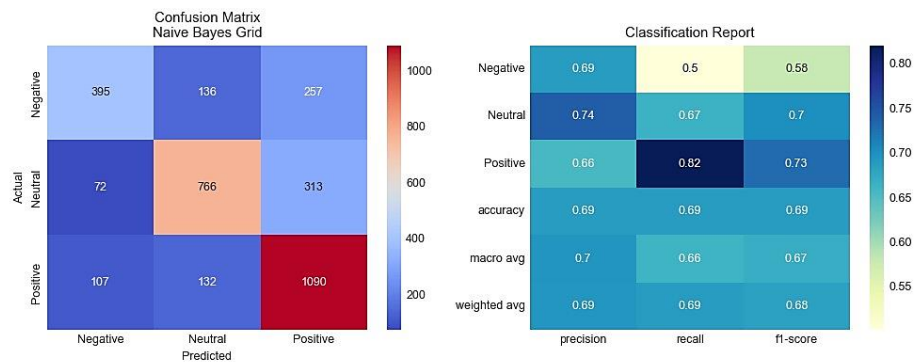
Kinerja *Naïve Bayes* (NB) dievaluasi menggunakan *confusion matrix* dan *classification report* untuk melihat kemampuan model dalam mengklasifikasikan sentimen negatif, netral, dan positif.



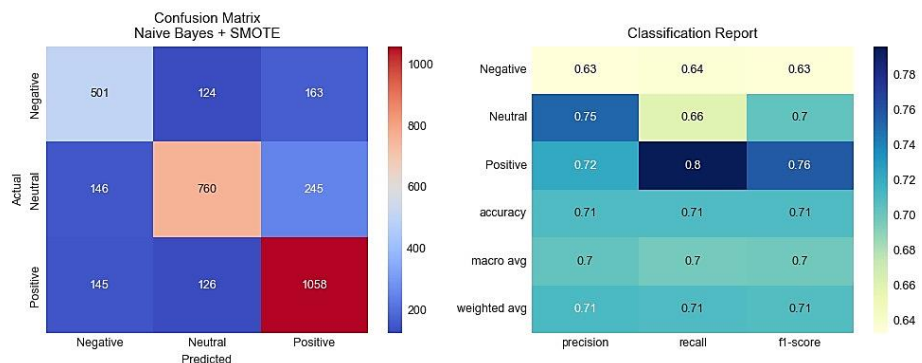
Gambar 6. Evaluasi Naïve Bayes Menggunakan Confusion matrix dan Classification report

Gambar 6 menunjukkan bahwa NB memperoleh *accuracy* sebesar 67%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan cukup baik, namun masih terdapat beberapa kesalahan prediksi pada kelas tertentu. Berdasarkan *classification report*, kelas positif memperoleh hasil terbaik dengan *recall* 0,90 dan *F1-score* 0,72, yang berarti model cukup baik dalam mengenali komentar positif. Sementara itu, kelas negatif memiliki *precision* tinggi sebesar 0,93, tetapi *recall* masih rendah sebesar 0,32, sehingga masih banyak komentar negatif yang belum berhasil dikenali dengan tepat. Secara keseluruhan, NB cukup baik sebagai model dasar, tetapi performanya masih perlu dibandingkan dengan *Logistic Regression* untuk melihat model yang lebih optimal.masing kelas [15].

Berdasarkan Gambar 7, optimasi *Grid Search* menghasilkan model *Naïve Bayes* yang lebih stabil pada beberapa kategori sentimen. Peningkatan terutama tampak pada nilai *precision* dan *recall*, yang mengindikasikan penurunan kesalahan prediksi pada kelas tertentu [16]. Gambar 8 menunjukkan hasil penerapan *SMOTE* membantu menyeimbangkan distribusi kelas sehingga performa *Naïve Bayes* meningkat, dengan akurasi sekitar 71% serta perbaikan nilai *F1-score* pada beberapa kategori sentimen.



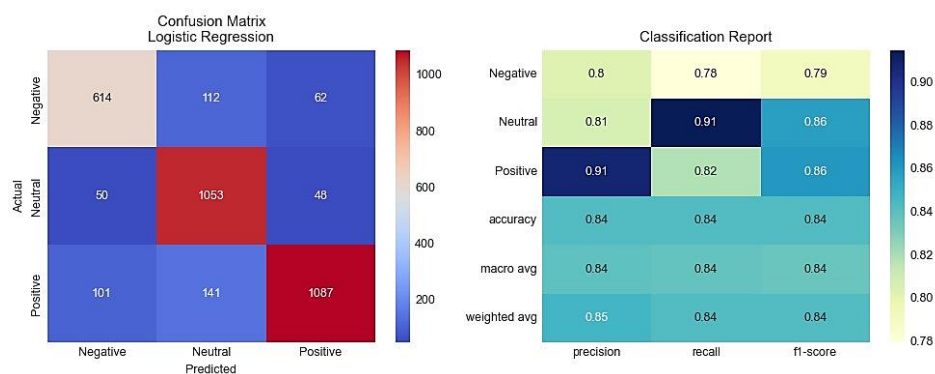
Gambar 7. CM dan CR *Naïve Bayes* dengan optimasi *Grid Search*



Gambar 8. CM dan *Classification report* dengan *SMOTE*

2.6 Evaluasi Model *Logistic Regression*

Selain *Naïve Bayes*, penelitian ini juga menggunakan algoritma *Logistic Regression* untuk melakukan klasifikasi sentimen komentar YouTube.

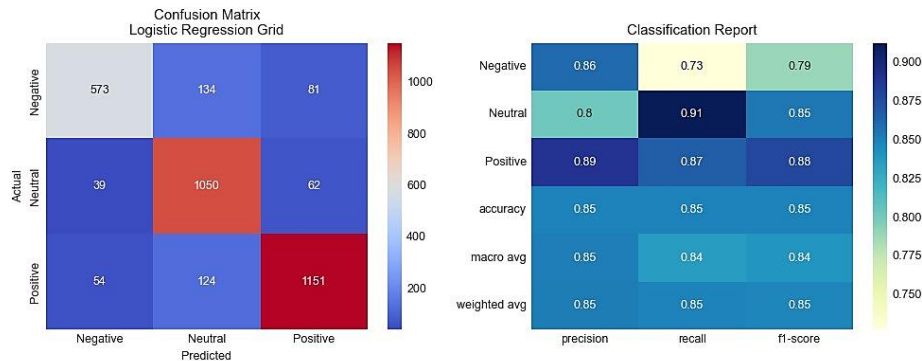


Gambar 9. *Confusion matrix* dan *Classification report* LR

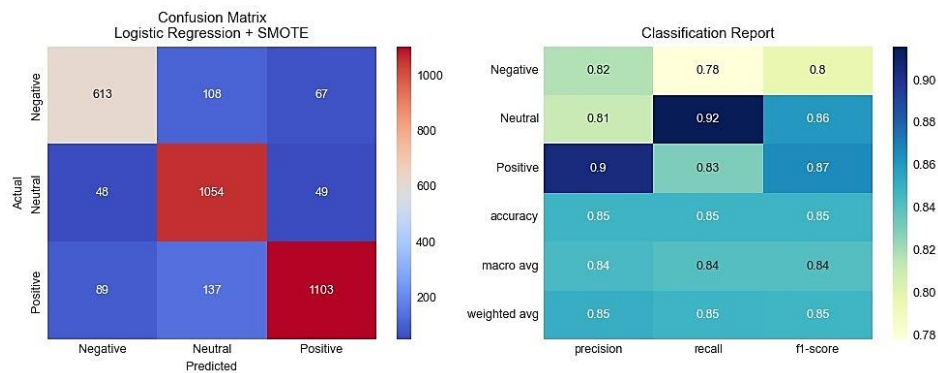
Mengacu pada Gambar 9, *Logistic Regression* menghasilkan performa yang lebih baik daripada *Naïve Bayes*, dengan akurasi sekitar 84%. Nilai *precision*, *recall*, dan *F1-score* juga relatif stabil antarkelas, menunjukkan klasifikasi yang cukup konsisten untuk sentimen positif, netral, dan negatif.

Hasil evaluasi pada Gambar 10 menunjukkan bahwa optimasi parameter mampu meningkatkan performa model *Logistic Regression* secara keseluruhan. Nilai *F1-score* pada masing-masing kategori sentimen mengalami peningkatan dibandingkan dengan model *Logistic Regression* tanpa optimasi. Untuk

meningkatkan kemampuan model dalam menangani ketidakseimbangan data, teknik SMOTE juga diterapkan pada model *Logistic Regression*.



Gambar 10. Confusion matrix dan Classification report LR Grid Search



Gambar 11. Confusion matrix dan Classification report LR SMOTE

Selanjutnya, menunjukkan hasil evaluasi model *Logistic Regression* dengan SMOTE. Model ini memperoleh *accuracy* sekitar 85%, yang menunjukkan bahwa performanya lebih baik dibandingkan dengan beberapa konfigurasi model sebelumnya. Berdasarkan *classification report*, nilai *precision*, *recall*, dan *F1-score* pada setiap kelas relatif seimbang, yaitu berada pada kisaran 0,80 hingga 0,87. Hal ini menunjukkan bahwa penerapan SMOTE membantu model dalam mengenali kelas negatif, netral, dan positif secara lebih merata. Dengan demikian, *Logistic Regression* dengan SMOTE menjadi salah satu model dengan performa terbaik dalam penelitian ini [17].

2.7 Perbandingan Performa Model Klasifikasi

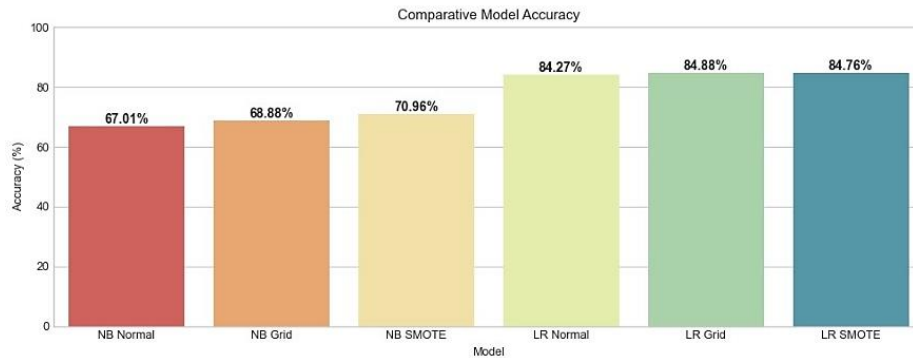
Setelah seluruh model dievaluasi, tahap selanjutnya adalah membandingkan performa algoritma NB dan LR berdasarkan nilai akurasi yang diperoleh dari setiap konfigurasi model.

Tabel 1. Perbandingan Akurasi Model Klasifikasi

No	Model	Accuracy (%)
1	<i>Naïve Bayes</i> Normal	67.01
2	<i>Naïve Bayes</i> + Grid Search	68.88
3	<i>Naïve Bayes</i> + SMOTE	70.96
4	<i>Logistic Regression</i> Normal	84.27
5	<i>Logistic Regression</i> + Grid Search	84.88
6	<i>Logistic Regression</i> + SMOTE	84.76

Berdasarkan Tabel 1, model *Naïve Bayes* menunjukkan peningkatan performa setelah dilakukan optimasi parameter dan penerapan teknik SMOTE. Akurasi model *Naïve Bayes* meningkat dari 67.01% menjadi 68.88% setelah Grid Search, dan mencapai 70.96% setelah penerapan SMOTE.

Model LR menghasilkan performa yang lebih tinggi. Model *Logistic Regression* normal memperoleh akurasi 84.27%, meningkat menjadi 84.88% setelah optimasi Grid Search, dan mencapai 84.76% setelah penerapan SMOTE. Grafik perbandingan akurasi model klasifikasi dapat ditunjukkan pada Gambar 12.



Gambar 12. Grafik Perbandingan Akurasi Model Klasifikasi

Setelah seluruh model dievaluasi, hasil perbandingan menunjukkan bahwa *Logistic Regression* (LR) memiliki performa lebih baik dibandingkan dengan *Naïve Bayes* (NB) pada berbagai skenario pengujian. Pada model NB, akurasi meningkat dari 67,01% pada model normal menjadi 68,88% setelah penerapan Grid Search, dan meningkat lagi menjadi 70,96% setelah menggunakan SMOTE. Sementara itu, model LR menunjukkan hasil yang lebih tinggi, yaitu 84,27% pada model normal, 84,88% setelah penerapan Grid Search, dan 84,76% setelah menggunakan SMOTE. Hasil terbaik diperoleh pada *Logistic Regression* dengan Grid Search, yaitu sebesar 84,88%. Temuan ini menunjukkan bahwa LR lebih stabil dan lebih efektif dibandingkan dengan NB dalam mengklasifikasikan sentimen komentar YouTube berbahasa Indonesia. Selain itu, *Grid Search* dan SMOTE memberikan pengaruh positif terhadap peningkatan performa model, meskipun peningkatannya berbeda pada setiap algoritma [16].

3. KESIMPULAN

Penelitian ini melakukan evaluasi komparatif terhadap dua algoritma klasifikasi, yaitu *Naïve Bayes* (NB) dan *Logistic Regression* (LR), untuk mengklasifikasikan sentimen komentar YouTube berbahasa Indonesia pada konteks layanan digital. Data dikomentari terlebih dahulu melalui tahap *preprocessing*, kemudian direpresentasikan menggunakan TF-IDF. Selain itu, penelitian ini juga menerapkan *Grid Search* dan SMOTE untuk memperoleh konfigurasi model yang lebih optimal. Hasil evaluasi menunjukkan bahwa *Logistic Regression* memberikan performa lebih baik dibandingkan dengan *Naïve Bayes*. Model terbaik diperoleh pada *Logistic Regression* dengan *Grid Search* dengan akurasi sebesar 84,88%. Secara umum, nilai *precision*, *recall*, dan *F1-score* pada model *Logistic Regression* juga lebih stabil dibandingkan dengan *Naïve Bayes*, sehingga model ini lebih mampu mengenali kelas sentimen positif, netral, dan negatif secara seimbang. Sementara itu, *Naïve Bayes* masih menunjukkan performa yang lebih rendah, terutama pada kemampuan mengenali kelas tertentu. Dengan demikian, dapat disimpulkan bahwa *Logistic Regression* lebih sesuai digunakan dalam penelitian ini untuk klasifikasi sentimen komentar YouTube berbahasa Indonesia, karena menghasilkan performa yang lebih baik dan lebih konsisten berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

REFERENSI

- [1] 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.
- [2] A. R. Abdillah and F. N. Hasan, "Analisis Sentimen Terhadap Kandidat Calon Presiden Berdasarkan Tweets Di Sosial Media Menggunakan Naive Bayes Classifier," *SMATIKA JURNAL*, vol. 13, no. 01, pp. 117–130, Jul. 2023, doi: 10.32664/smatika.v13i01.750.
- [3] A. Asro, N. Azizah, and Sudaryono, "Public Sentiment Analysis on the 2024 Presidential Election Using Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) On Social Media Data," vol. 1, p. 0226, Jun. 2024.
- [4] J. Zoti, M. Rahman, S. Ahmed, A.-A. Akib, and S. Khan, *Sentiment Analysis of YouTube Comments: A Comprehensive Study of Machine Learning Models*. 2025.
- [5] 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," in *2025 13th International Conference on Awareness Science and Technology (iCAST)*, IEEE, Nov. 2025, pp. 1–6. doi: 10.1109/iCAST68191.2025.11300032.
- [6] N. Sulistianingsih and I. N. Switrayana, "Enhancing Sentiment Analysis for the 2024 Indonesia Election Using SMOTE-Tomek Links and Binary *Logistic Regression*," *International Journal of*

- Education and Management Engineering*, vol. 14, no. 3, pp. 22–32, Jun. 2024, doi: 10.5815/ijeme.2024.03.03.
- [7] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, “Transformer-based models for sentiment analysis of YouTube video comments,” *Sci. Afr.*, vol. 29, p. e02836, Sep. 2025, doi: 10.1016/j.sciaf.2025.e02836.
- [8] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, “The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm,” *Procedia Comput. Sci.*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.
- [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] H. Henderi, “Utilizing Sentiment Analysis for Reflect and Improve Education in Indonesia,” *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 189–200, Jan. 2024, doi: 10.47738/jads.v6i1.527.
- [11] S. W. Iriananda, R. W. Budiawan, A. Y. Rahman, and I. Istiadi, “Optimasi Klasifikasi Sentimen Komentar Pengguna Game Bergerak Menggunakan Svm, Grid Search Dan Kombinasi N-Gram,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 743–752, Aug. 2024, doi: 10.25126/jtiik.1148244.
- [12] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, “Text Preprocessing for Text mining in Organizational Research: Review and Recommendations,” *Organ. Res. Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.
- [13] 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.
- [14] X. Jiang and C. Xu, “Deep Learning and Machine learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data,” *J. Clin. Med.*, vol. 11, no. 19, p. 5772, Sep. 2022, doi: 10.3390/jcm11195772.
- [15] A. S. Aribowo, N. H. Cahyana, and Y. Fauziah, “Enhancing Semi-Supervised Sentiment Analysis Through Hyperparameter Tuning Within Iterations: A Comparative Study Using Grid Search and Random Search,” 2024, pp. 248–260. doi: 10.2991/978-94-6463-366-5_23.
- [16] A. D. Rachmatsyah, T. Sugihartono, and K. Irfan, “Perbandingan Teknik Optimasi Grid Search Dan Randomized Search Dalam Meningkatkan Akurasi Metode Klasifikasi Svm Pada Sentimen Ulasan Pengguna Aplikasi Jkn Mobile,” *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 8, no. 1, pp. 13–22, Dec. 2024, doi: 10.36080/skanika.v8i1.3328.