

Effectiveness And Efficiency Of LLM Models Vs Traditional Machine Learning In Sentiment Analysis Of Indonesian Language Product Reviews

Galih Setiawan Nurohim ¹⁾; Budi Al Amin ²⁾; Heribertus Ary Setyadi ³⁾
^{1,2)} Universitas Bina Sarana Informatika

Email: ¹⁾ galih.glt@bsi.ac.id ;²⁾ budi.bdm@bsi.ac.id ;³⁾ heribertus.hbs@bsi.ac.id

How to Cite :

Nurohim, G, S., Amin, B, A., Setyadi, H, A. (2025). Effectiveness And Efficiency Of LLM Models Vs Traditional Machine Learning In Sentiment Analysis Of Indonesian Language Product Reviews. Jurnal Media Computer Science, 4(2)

ARTICLE HISTORY

Received [12 Juni 2025]

Revised [03 Juli 2025]

Accepted [05 Juli 2025]

KEYWORDS

Sentiment Analysis, Bisnis Digital, Machine Learning, Naive Bayes, Transformer, Return on Investment (ROI).

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap kinerja dan efisiensi beberapa model machine learning dalam tugas analisis sentimen pada ulasan pelanggan berbahasa Indonesia. Di era bisnis digital, pemahaman cepat dan akurat terhadap opini pelanggan merupakan aset strategis untuk pengambilan keputusan, mulai dari pengembangan produk hingga strategi pemasaran. Empat model dievaluasi: dua model berbasis Transformer (*agufsamudra/indo-sentiment-analysis* dan *ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa*), Naive Bayes, dan K-Nearest Neighbors (KNN) pada dataset berisi 5.400 ulasan produk. Metrik evaluasi yang digunakan adalah Akurasi, Presisi, Recall, dan F1-Score. Hasil menunjukkan bahwa model Naive Bayes dan model Transformer '*agufsamudra/indo-sentiment-analysis*' mencapai kinerja tertinggi dengan F1-Score dan akurasi sekitar 95%, secara signifikan mengungguli model Transformer lainnya (90%) dan KNN (47%). Temuan krusial dari penelitian ini adalah kinerja model klasik Naive Bayes yang setara dengan model Transformer canggih. Dari perspektif akuntansi dan bisnis, ini mengimplikasikan bahwa solusi dengan efisiensi komputasi yang jauh lebih tinggi (Naive Bayes) dapat memberikan Return on Investment (ROI) yang lebih optimal untuk implementasi sistem pemantauan sentimen pelanggan skala besar.

ABSTRACT

*This research aims to conduct a comparative analysis of the performance and efficiency of several machine learning models in the task of sentiment analysis on Indonesian language customer reviews. In the digital business era, a quick and accurate understanding of customer opinions is a strategic asset for making decisions, from product development to marketing strategy. Four models were evaluated: two Transformer-based models (*agufsamudra/indo-sentiment-analysis* and *ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa*), Naive Bayes, and K-Nearest Neighbors (KNN) on a dataset of 5,400 product reviews. The evaluation metrics used are Accuracy, Precision, Recall, and F1-Score. The results show that the Naive Bayes model and the Transformer model '*agufsamudra/indo-sentiment-analysis*' achieve the highest performance with an F1-Score and accuracy of around 95%, significantly outperforming other Transformer models (90%) and KNN (47%). The crucial finding of this*

research is that the performance of the classical Naive Bayes model is equivalent to the state-of-the-art Transformer model. From an accounting and business perspective, this implies that solutions with much higher computational efficiency (Naive Bayes) can provide a more optimal Return on Investment (ROI) for large-scale implementation of customer sentiment monitoring systems.

PENDAHULUAN

Dalam era digital saat ini, aplikasi mobile menjadi bagian penting dalam kegiatan sehari-hari, mulai dari layanan hiburan, belanja online maupun transaksi keuangan. Seiring meningkatnya penggunaan aplikasi, ulasan pengguna atau review dari user di Google play store menjadi sumber informasi yang berharga baik bagi pengembang aplikasi maupun bisnis untuk memahami pengalaman dan persepsi user terhadap layanan bisnis tersebut serta, aset strategis bagi pengembang aplikasi, pelaku bisnis digital, dan pengambil keputusan dalam merancang kebijakan berbasis data. Ulasan ini mencerminkan sentimen pelanggan terhadap produk atau layanan dan dapat dimanfaatkan untuk mendeteksi kepuasan, keluhan, serta kebutuhan konsumen secara real-time.

Analisis Sentimen adalah teknik dalam bidang pemrosesan bahasa alami (NL) yang memiliki tujuan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks menjadi sentimen positif dan negatif. Dalam konteks ini, berbagai algoritma machine learning telah digunakan, mulai dari metode klasik seperti Naive Bayes dan K-Nearest Neighbor (KNN) hingga pendekatan berbasis deep learning seperti BERT dan LLM (Large Language Models).

Naive Bayes menjadi pilihan yang menarik karena kemudahannya dan efisiensinya dalam mengklasifikasikan teks, bahkan ketika dataset berukuran besar. Misalnya penelitian oleh Adelia Irawan et al. (2024) menunjukkan bahwa Naive Bayes mampu memberikan hasil akurat dalam klasifikasi sentimen aplikasi bank digital. Sementara itu, KNN mengandalkan kedekatan antar data dan dapat menangkap pola lokal yang tidak ditangkap oleh model probabilistik, sebagaimana dijelaskan oleh Pratomanto et al. (2024) dalam sentimen pada ulasan aplikasi Vidio.

Namun, meskipun kedua algoritma ini banyak digunakan, studi yang membandingkan efektivitas dan efisiensi Naive Bayes dan KNN dalam domain ulasan aplikasi mobile berbahasa Indonesia masih terbatas. Hal ini menjadi celah penelitian yang penting, mengingat karakteristik bahasa Indonesia dan konteks lokal pengguna yang berbeda dengan studi-studi internasional.

Di sisi lain, kemajuan teknologi kecerdasan buatan membawa pendekatan baru dalam analisis sentimen, seperti transformer-based models dan Large Language Models (LLMs). Penelitian oleh al Hafidh & Al-Karawi (2025) menunjukkan bahwa BERT mampu meningkatkan akurasi analisis sentimen pada ulasan elektronik Amazon. Bahkan, LLM kini digunakan dalam konteks yang lebih kompleks seperti keuangan, sebagaimana ditunjukkan oleh Kirtac & Germano (2024) dalam pemanfaatannya untuk sentiment trading.

Meskipun model-model besar menawarkan performa yang sangat baik, mereka juga menuntut sumber daya komputasi tinggi dan waktu prediksi yang lama, yang tidak selalu sesuai untuk implementasi di lingkungan terbatas sumber daya. Oleh karena itu, penting untuk mengevaluasi kembali kelebihan metode klasik seperti Naive Bayes dan KNN dalam konteks aplikasi nyata dengan bahasa Indonesia, serta membandingkannya secara menyeluruh baik dari aspek akurasi, waktu prediksi, maupun stabilitas prediksi.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan melakukan analisis komparatif antara algoritma Naive Bayes, K-Nearest Neighbor (KNN), dan model Large Language Model (LLM) lokal dalam mengklasifikasikan sentimen ulasan pengguna terhadap produk berbahasa Indonesia. Fokus utama diarahkan pada aspek akurasi klasifikasi, efisiensi waktu prediksi, serta kelayakan implementasi dalam konteks bisnis digital di Indonesia.

Berbeda dengan penelitian sebelumnya seperti oleh Pratmanto et al. (2024) yang menggunakan data ulasan aplikasi Vidio, studi ini memilih untuk menggunakan dataset PRDECT-ID (Sutoyo et al., 2022), yaitu kumpulan ulasan produk dari marketplace lokal yang lebih beragam secara tematik dan kontekstual. Dataset ini terdiri dari 5.400 ulasan produk berbahasa Indonesia, dikurasi dari 29 kategori produk dan telah diberi anotasi emosi oleh pakar psikologi klinis. Pendekatan ini dipilih untuk memastikan generalisasi model terhadap berbagai jenis ulasan produk yang umum dijumpai dalam ekosistem bisnis digital Indonesia.

Dengan demikian, penelitian ini tidak hanya memperluas cakupan data dari sisi domain (tidak terbatas pada aplikasi hiburan), tetapi juga menyediakan evaluasi menyeluruh terhadap performa model dari sisi akurasi, kecepatan prediksi, dan skalabilitas implementasi, yang dapat menjadi acuan bagi pengambil keputusan dalam mengadopsi teknologi analitik berbasis kecerdasan buatan di era ekonomi digital.

LANDASAN TEORI

Analisa Sentimen

Analisis sentimen merupakan cabang dari Natural Language Processing (NLP) yang bertujuan untuk mengidentifikasi opini atau perasaan dalam suatu teks. Dalam konteks bisnis digital, analisis ini berguna untuk memahami persepsi pelanggan terhadap produk atau layanan melalui ulasan yang ditulis secara bebas di platform digital seperti marketplace atau media sosial. Dengan analisis sentimen, pelaku bisnis dapat mengevaluasi kualitas layanan dan mengambil keputusan berbasis data pelanggan secara cepat dan efisien (Adelia Irawan et al., 2024).

Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang sering digunakan dalam analisis teks. Algoritma ini mengasumsikan independensi antar fitur dan menggunakan Teorema Bayes untuk menghitung probabilitas sebuah dokumen termasuk ke dalam kelas tertentu. Kelebihan Naive Bayes terletak pada efisiensi waktu pelatihan dan kemampuannya menangani dataset besar. Dalam konteks analisis ulasan aplikasi, model ini telah terbukti memberikan hasil akurat dengan biaya komputasi yang rendah (Pratmanto et al., 2024; Adelia Irawan et al., 2024).

Algoritma K-Nearest Neighbor (KNN)

KNN adalah algoritma berbasis instance-based learning yang mengklasifikasikan data berdasarkan kedekatan jarak dengan data latih. Pada analisis sentimen, KNN bekerja dengan mencari k tetangga terdekat dari sebuah teks baru, lalu menentukan kelas berdasarkan mayoritas dari tetangga tersebut. Keunggulan utama KNN adalah kemampuannya menangkap pola lokal tanpa pelatihan model eksplisit, namun waktu prediksi bisa lambat karena pencarian dilakukan terhadap seluruh data latih (Pratmanto et al., 2024).

Large Language Models (LLM)

LLM seperti BERT, RoBERTa, dan GPT merupakan model berbasis Transformer yang dilatih pada skala data besar dan mampu memahami konteks bahasa secara mendalam. Dalam konteks analisis sentimen, LLM dapat menangkap nuansa makna yang sulit diidentifikasi oleh model klasik. Penelitian oleh al Hafidh & Al-Karawi (2025) menunjukkan bahwa BERT dapat secara signifikan meningkatkan akurasi dalam klasifikasi sentimen pada ulasan elektronik. Namun demikian, LLM menuntut sumber daya komputasi yang besar, dan waktu inferensi yang lebih lama (Kirtac & Germano, 2024).

Dataset PRDECT-ID

PRDECT-ID adalah dataset yang berisi 5.400 ulasan produk dalam Bahasa Indonesia yang dikurasi dari 29 kategori produk di marketplace lokal. Dataset ini telah dianotasi oleh ahli psikologi

dengan label emosi dan sentimen, sehingga menjadi aset penting dalam pengembangan model NLP berbahasa Indonesia. Ketersediaan data seperti ini memungkinkan studi komparatif antara berbagai algoritma dalam konteks lokal, yang sebelumnya masih jarang dilakukan (Sutoyo et al., 2022).

METODE PENELITIAN

Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimental. Tujuan utamanya adalah untuk membandingkan secara objektif kinerja dari empat model klasifikasi dengan arsitektur yang berbeda—dua berbasis Large Language Model (LLM) dan dua berbasis machine learning tradisional—dalam tugas analisis sentimen pada teks berbahasa Indonesia.

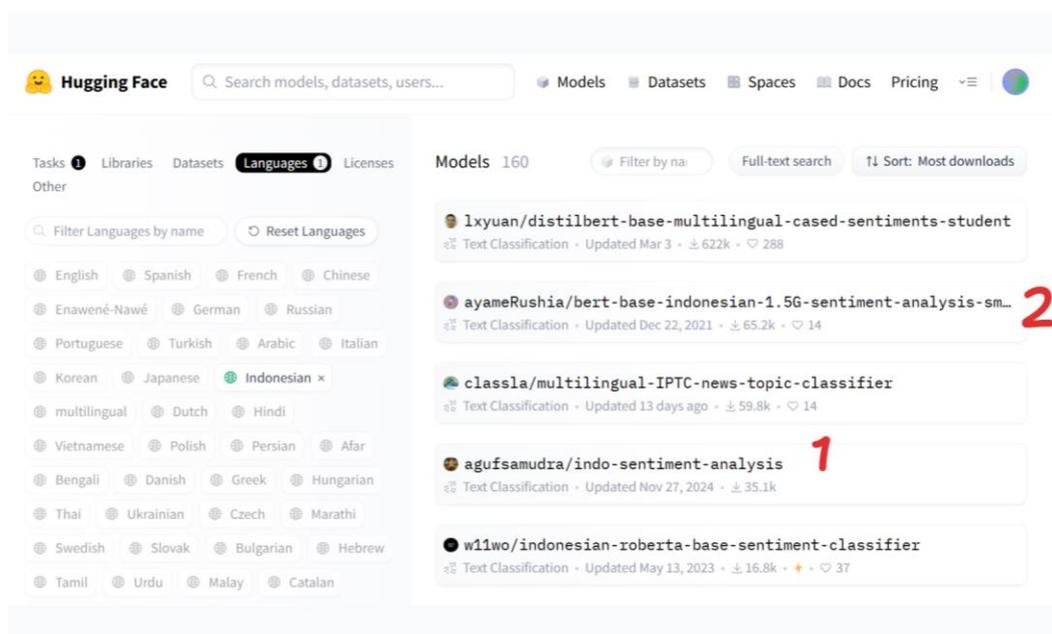
Sumber Data dan Pra-pemrosesan

Data yang digunakan dalam penelitian ini berasal dari dataset publik PRDECT-ID yang berisi ulasan pelanggan untuk berbagai produk. Proses pra-pemrosesan awal meliputi eliminasi entri dengan nilai kosong (null values) untuk memastikan kualitas data, menghasilkan korpus data final sebanyak 5.400 ulasan. Distribusi kelas sentimen pada dataset ini relatif seimbang, terdiri dari 2.821 ulasan negatif (52.2%) dan 2.579 ulasan positif (47.8%), yang ideal untuk tugas klasifikasi biner.

Model yang Diuji dan Justifikasi Pemilihan

Empat model dipilih untuk dievaluasi, merepresentasikan pendekatan state-of-the-art dan klasik dalam NLP.

1. Model Berbasis LLM: Dua model dipilih berdasarkan popularitas, adopsi oleh komunitas, dan tingkat penggunaan tertinggi (diukur dari jumlah unduhan) pada platform Hugging Face untuk tugas analisis sentimen Bahasa Indonesia: `agufsamudra/indo-sentiment-analysis` dan `ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa`. Pemilihan berdasarkan kriteria ini bertujuan untuk menguji model yang secara de-facto dianggap efektif oleh komunitas praktisi.



Gambar 1. Pemilihan model LLM

2. Model Machine Learning Tradisional: Dua model klasik disertakan sebagai titik referensi (baseline) untuk mengevaluasi apakah peningkatan kompleksitas komputasi pada LLM memberikan peningkatan kinerja yang sepadan. Multinomial Naive Bayes: Dipilih karena efisiensinya yang tinggi dan kinerjanya yang kuat secara historis pada tugas klasifikasi teks dan K-Nearest Neighbors (KNN): Dipilih untuk merepresentasikan pendekatan berbasis instan (instance-based learning).

HASIL DAN PEMBAHASAN

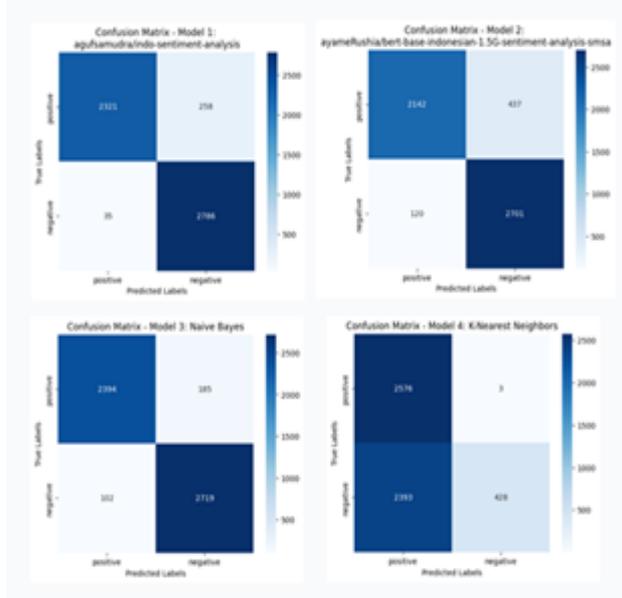
Hasil Evaluasi Empiris Model Sentiment Analysis

Penelitian ini membandingkan performa dua model Large Language Models (LLM) dan dua algoritma Machine Learning (ML) tradisional dalam mengklasifikasikan ulasan pelanggan berbahasa Indonesia dari dataset PRDECT-ID. Dataset terdiri atas 5.400 entri, dengan distribusi sentimen negatif sebanyak 2.821 dan positif sebanyak 2.579. Keempat model diuji menggunakan metrik umum dalam klasifikasi: precision, recall, f1-score, dan accuracy. Hasil evaluasi ringkas disajikan pada Tabel 1, diikuti oleh rincian Confusion Matrix pada Tabel 2.

Tabel 1. Evaluasi Model

Model	Accuracy	F1-Score Pos	F1-Score Neg	Distribusi Prediksi (Pos/Neg)
agufsamudra/indo-sentiment-analysis (LLM)	0.95	0.94	0.95	2.356 / 3.044
ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa (LLM)	0.90	0.88	0.91	2.262 / 3.138
Naive Bayes (Tradisional)	0.95	0.94	0.95	2.496 / 2.904
K-Nearest Neighbors (KNN)	0.56	0.68	0.26	4.969 / 431

Dari hasil di atas, terlihat jelas bahwa model LLM agufsamudra dan model tradisional Naive Bayes menunjukkan kinerja superior dan secara statistik identik pada semua metrik utama (Akurasi dan F1-Score 95%). Model LLM ayameRushia berada di tingkat kedua dengan kinerja yang solid, sementara K-Nearest Neighbors (KNN) menunjukkan kegagalan total dalam tugas ini.



Gambar 1 Confusion Matrix

Tabel 2. Tabel Hasil Evaluasi

Model	Arsitektur	Waktu Inferensi (detik)	F1-Score Neg
agufsamudra/indo-sentiment-analysis (LLM)	LLM (BERT)	49.1217	GPU (Direkomendasikan)
ayameRushia/bert-base-indonesian-1.5G-sentiment-analysis-smsa (LLM)	LLM (BERT)	49.1891	GPU (Direkomendasikan)
Naive Bayes (Tradisional)	Tradisional	0.0018	CPU
K-Nearest Neighbors (KNN)	Tradisional	0.0865	CPU

Analisis dan Pembahasan Kinerja Model

Analisis mendalam pada pola kesalahan setiap model memberikan wawasan yang lebih dari sekadar angka akurasi. Dimulai dari model berkinerja terendah, kegagalan prediktif K-Nearest Neighbors (KNN) terkonfirmasi secara visual. Dari 2.821 ulasan negatif, Confusion Matrix menunjukkan model ini secara keliru mengklasifikasikannya sebagai positif sebanyak 2.393 kali (False Positive). Ini membuktikan bahwa model KNN dalam kasus ini tidak belajar membedakan sentimen, melainkan hanya cenderung menebak kelas mayoritas, membuatnya tidak berguna untuk aplikasi bisnis.

Berpindah ke model-model dengan kinerja teratas, temuan yang paling menarik justru muncul dari perbandingan antara LLM agufsamudra dan Naive Bayes. Meskipun Tabel 1 menunjukkan efektivitas mereka nyaris identik, Tabel 2 dan 3 mengungkapkan bahwa profil efisiensi dan profil risiko mereka sangatlah berbeda.

Dari sisi efisiensi, perbedaannya sangat dramatis. Data pada Tabel 3 mengkuantifikasi hal ini: Naive Bayes mampu menyelesaikan prediksi dalam 0.0018 detik, sementara LLM membutuhkan ~49 detik. Artinya, Naive Bayes beroperasi lebih dari 27.000 kali lebih cepat pada studi kasus ini.

Di satu sisi, model LLM agufsamudra dapat dikarakterisasi sebagai model 'Teliti' yang berisiko rendah. Kekuatan utamanya adalah kemampuannya menekan jumlah False Positive (FP) hingga ke angka yang sangat minim, yaitu hanya 35 kasus. Ini berarti, risiko sistem mengabaikan keluhan riil dari pelanggan sangatlah kecil. Namun, "ketelitian" ini dibayar dengan jumlah False Negative (FN) yang lebih tinggi, yaitu 258 kasus, di mana ulasan positif terlewat dan keliru dilabeli sebagai negatif.

Di sisi lain, Naive Bayes berperan sebagai model 'Generalis' yang lebih seimbang dan efisien. Keunggulannya adalah kemampuannya menangkap lebih banyak ulasan positif, dengan tingkat False Negative (FN) yang jauh lebih rendah, yaitu hanya 185 kasus. Namun, keunggulan ini datang dengan konsekuensi tingkat False Positive (FP) yang hampir tiga kali lipat lebih tinggi dari LLM, yaitu sebanyak 102 kasus. Sementara itu, model LLM sekunder, ayameRushia, tidak menunjukkan profil yang menarik karena menggabungkan kelemahan keduanya: tingkat False Positive yang cukup tinggi (120) sekaligus False Negative yang sangat tinggi (437).

Perbedaan profil risiko ini memiliki implikasi strategis yang langsung bagi para pengambil keputusan di dunia bisnis dan akuntansi. Keputusan pemilihan model tidak lagi sekadar "mana yang paling akurat?", melainkan "profil risiko mana yang paling sesuai dengan strategi dan toleransi risiko perusahaan?"

Untuk organisasi yang memprioritaskan efisiensi biaya dan optimalisasi peluang—misalnya UMKM atau bisnis retail yang ingin memaksimalkan pengumpulan testimoni positif untuk pemasaran—Naive Bayes menjadi pilihan yang paling rasional. Model ini paling hemat secara komputasi dan paling efektif dalam menangkap sentimen positif, meskipun perusahaan harus menerima adanya risiko yang sedikit lebih tinggi terkait keluhan pelanggan yang mungkin terlewat.

Sebaliknya, bagi organisasi yang beroperasi di sektor sensitif di mana mitigasi risiko adalah prioritas utama—seperti jasa keuangan, travel premium, atau kesehatan, di mana satu keluhan terlewat dapat menimbulkan kerugian reputasi atau finansial yang besar—model LLM agufsamudra

menawarkan proposisi nilai yang lebih kuat. Biaya komputasi yang lebih tinggi dapat dipandang sebagai "premi asuransi" yang dibayarkan untuk meminimalkan risiko salah mengklasifikasikan keluhan pelanggan yang krusial.

Dengan demikian, keputusan untuk mengadopsi model tertentu harus didasarkan pada analisis biaya-manfaat dan risiko yang cermat, bukan sekadar mengejar teknologi yang paling canggih. Penelitian ini membuktikan bahwa solusi yang lebih sederhana dan efisien seringkali merupakan pilihan yang paling cerdas secara bisnis, tergantung pada tujuan strategis yang ingin dicapai.

KESIMPULAN DAN SARAN

Kesimpulan

1. Paritas Efektivitas, Disparitas Efisiensi yang Ekstrem: Tidak ditemukan bahwa model LLM secara inheren lebih unggul dalam hal akurasi. Kinerja efektivitas model machine learning tradisional (Naive Bayes) terbukti setara dengan model LLM terancang (agufsamudra), keduanya mencapai akurasi ~95%. Namun, kesetaraan ini diiringi oleh perbedaan efisiensi yang masif, di mana Naive Bayes terbukti beroperasi lebih dari 27.000 kali lebih cepat dan tidak memerlukan perangkat keras khusus (GPU).
2. Profil Risiko yang Berbeda untuk Model Terbaik: Meskipun akurasinya identik, model-model terbaik menunjukkan pola kesalahan yang berbeda secara strategis. Model LLM (agufsamudra) unggul dalam meminimalkan False Positive, menjadikannya pilihan berisiko rendah untuk tidak melewatkan keluhan pelanggan. Sebaliknya, Naive Bayes lebih unggul dalam meminimalkan False Negative, membuatnya lebih andal dalam menangkap semua testimoni positif.
3. Ketidakcocokan Model Berbasis Jarak (KNN): Pemilihan arsitektur model sangat krusial. Model berbasis jarak seperti K-Nearest Neighbors terbukti sama sekali tidak cocok untuk tugas klasifikasi teks pada data ini, menghasilkan prediksi yang sangat bias dan kinerja yang tidak dapat diandalkan.
4. Indikasi Bias pada Deteksi Sentimen: Terdapat indikasi bahwa ekspresi sentimen negatif dalam dataset ulasan produk ini lebih mudah untuk dideteksi secara konsisten oleh berbagai model dibandingkan ekspresi positif, yang mungkin mengindikasikan penggunaan kosakata yang lebih eksplisit dalam keluhan.

Saran

Berdasarkan kesimpulan di atas, rekomendasi yang diberikan tidak bersifat satu ukuran untuk semua, melainkan terbagi menjadi saran praktis untuk pelaku bisnis dan saran akademis untuk penelitian selanjutnya.

Saran Praktis (untuk Pelaku Bisnis)

Rekomendasi implementasi harus didasarkan pada strategi, skala, dan toleransi risiko organisasi:

1. Untuk UMKM, Startup, dan Efisiensi Skala Besar: Naive Bayes adalah pilihan yang paling direkomendasikan. Kombinasi antara biaya implementasi yang sangat rendah (berjalan di CPU), kecepatan inferensi yang superior (cocok untuk aplikasi *real-time*), dan tingkat akurasi yang sangat tinggi menjadikannya solusi dengan *Return on Investment* (ROI) tertinggi untuk sebagian besar kasus penggunaan umum.
2. Untuk Perusahaan dengan Prioritas Mitigasi Risiko: Bagi perusahaan skala besar atau yang beroperasi di sektor sensitif (misal: jasa keuangan, kesehatan) di mana satu keluhan terlewat dapat menimbulkan kerugian besar, model LLM agufsamudra menjadi pilihan yang lebih dapat dipertanggungjawabkan. Biaya komputasi yang lebih tinggi dapat dianggap sebagai "premi asuransi" untuk meminimalkan risiko reputasi.
3. Implementasi Aplikatif: Terlepas dari model yang dipilih, nilainya akan meningkat secara eksponensial jika diintegrasikan ke dalam dasbor analitik visual atau sistem peringatan otomatis

(*alert system*). Hal ini memungkinkan pemilik bisnis untuk memonitor sentimen pasar dan merespons keluhan pelanggan secara proaktif.

Saran Akademis (untuk Penelitian Lanjutan)

1. Eksplorasi Domain Spesifik: Melakukan studi komparatif serupa pada dataset dari domain yang lebih spesifik (misal: ulasan aplikasi finansial, komentar politik) untuk menguji apakah dominasi efisiensi Naive Bayes tetap bertahan.
2. Analisis Biaya Kuantitatif: Melakukan analisis Total Cost of Ownership (TCO) yang formal, dengan memperhitungkan tidak hanya biaya komputasi tetapi juga biaya pengembangan, pemeliharaan, dan sumber daya manusia untuk setiap pendekatan.
3. Pengembangan Model Hibrida: Meneliti kemungkinan model hibrida yang menggabungkan kecepatan representasi TF-IDF dari Naive Bayes dengan kemampuan pemahaman konteks dari lapisan atensi pada model Transformer untuk mencari titik optimal baru antara efektivitas dan efisiensi.

DAFTAR PUSTAKA

- Adelia Irawan, F., Rialdy Atmadja, A., Wahana, A., Informatika, T., & Sunan Gunung Djati Bandung, U. (2024). Analisis Sentimen Ulasan Aplikasi Bank Digital Menggunakan Algoritma Naive Bayes. In *Journal of Computer Science and Information Technology E-ISSN* (Vol. 4, Issue 2).
- al Hafidh, N., & Al-Karawi, A. (2025). Advanced Sentiment Analysis of Amazon Electronics Reviews Leveraging BERT: Model Optimization and Evaluation. *Procedia Computer Science*, 258, 3608–3618. <https://doi.org/10.1016/J.PROCS.2025.04.616>
- Alamsyah, A., & Saviera, F. (n.d.). A Comparison of Indonesia's E-Commerce Sentiment Analysis for Marketing Intelligence Effort (case study of Bukalapak, Tokopedia and Elevenia).
- Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, 14(3), 348-354.
- Kirtac, K., & Germano, G. (2024). Sentiment trading with large language models. *Finance Research Letters*, 62, 105227. <https://doi.org/10.1016/J.FRL.2024.105227>
- Pratmanto, D., Widayanto, A., Meisella Kristania, Y., & Wijianto, R. (2024). Analisis Perbandingan Algoritma Naive Bayes Dan KNN Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Vidio Di Google Play Store. *CONTEN: Computer and Network Technology*, 4(2), 119–124. <http://jurnal.bsi.ac.id/index.php/conten>
- Prayoga Siswono, A., Fauzi, S., Zalva Surya Hermawan, A., Riyandi, A., Panjaitan No, J. di, Kidul, P., Purwokerto Selatan, K., & Banyumas, K. (2024). Analisis Sentimen Pelantikan Presiden Indonesia 2024 Menggunakan Model Klasifikasi dan Algoritma Naive Bayes (Vol. 4, Issue 1).
- Sutoyo, R., Achmad, S., Chowanda, A., Andangsari, E. W., & Isa, S. M. (2022). PRDECT-ID: Indonesian product reviews dataset for emotions classification tasks. *Data in Brief*, 44, 108554. <https://doi.org/10.1016/J.DIB.2022.108554>