

PERBANDINGAN KINERJA NAIVE BAYES DAN INDOBERT UNTUK ANALISIS SENTIMEN ULASAN GOOGLE MAPS BERBAHASA INDONESIA

Muas Maulana Ibrahim¹, Martanto², Fatihanursari Dikananda³, Umi Hayati⁴

Program Studi Teknik Informatika¹³⁴
Program Studi Manajemen Informatika²

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
ibrahimmaulanamalik40@gmail.com

(*) Corresponding Author : ibrahimmaulanamalik40@gmail.com
Published : 20 Juli 2026

Abstract—This research is motivated by the limited performance of sentiment analysis models on imbalanced Indonesian review data, as well as the absence of a comprehensive comparative study between traditional classification models and transformer-based models in the context of religious location reviews. The primary problem is the inability of certain models to capture semantic context while maintaining stable performance across uneven data distributions. Consequently, this study aims to compare the performance of the Naïve Bayes and IndoBERT algorithms in classifying Google Maps review sentiment. The methodology employs a quantitative approach with a comparative experimental design. The research dataset comprises 2,772 valid reviews obtained from scraping the At-Taqwa Mosque in Cirebon, processed through text cleaning, normalizing, and stemming stages within the Google Colab environment. The Multinomial Naïve Bayes (MNB) algorithm was combined with TF-IDF features and the Synthetic Minority Oversampling Technique (SMOTE) to balance classes, while the IndoBERT model was tested using a Zero-Shot Inference approach. Results demonstrate that the integration of MNB and SMOTE achieved the most optimal performance, yielding 95.38% accuracy and a Macro F1-Score of 0.95. In contrast, while IndoBERT possesses semantic advantages, it recorded a lower accuracy of 87.12% with a Macro F1-Score falling to 0.40. This low macro score confirms the presence of the Accuracy Paradox, where the model developed a bias toward the majority class (97.2% positive). The study concludes that data engineering strategies utilizing SMOTE on traditional algorithms are more effective and stable for handling imbalanced real-world review data than relying on pure deep learning models without class-balancing interventions.

Keywords: *Sentiment Analysis, Naïve Bayes, IndoBERT, Google Maps, SMOTE.*

Abstrak—Penelitian ini dilatarbelakangi oleh keterbatasan performa model analisis sentimen pada data ulasan berbahasa Indonesia yang bersifat tidak seimbang (*imbalanced*), serta belum adanya kajian komparatif yang komprehensif antara model klasifikasi tradisional dan model berbasis *transformer* pada konteks ulasan lokasi religi. Permasalahan utama terletak pada ketidakmampuan sebagian model dalam menangkap konteks semantik sekaligus menjaga stabilitas performa pada distribusi data yang tidak merata. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma *Naïve Bayes* dan *IndoBERT* dalam mengklasifikasikan sentimen ulasan Google Maps. Metode yang digunakan adalah pendekatan kuantitatif dengan desain eksperimen komparatif. *Dataset* penelitian mencakup 2.772 ulasan valid hasil *scraping* Masjid At-Taqwa Cirebon yang diproses melalui tahapan *text cleaning*, *normalizing*, dan *stemming* di lingkungan Google Colab. Algoritma *Multinomial Naïve Bayes* (MNB) dikombinasikan dengan fitur TF-IDF dan teknik *oversampling* SMOTE untuk menyeimbangkan kelas, sementara model *IndoBERT* diuji menggunakan pendekatan *Zero-Shot Inference*. Hasil penelitian menunjukkan bahwa integrasi MNB dan SMOTE menghasilkan performa paling optimal dengan akurasi 95,38% dan nilai *F1-Score Macro* mencapai 0,95. Sebaliknya, meskipun *IndoBERT* memiliki keunggulan semantik, model ini mencatatkan akurasi lebih rendah sebesar 87,12% dengan nilai *F1-Score Macro* yang jatuh pada angka 0,40. Rendahnya nilai makro tersebut mengonfirmasi terjadinya *Accuracy Paradox*, di mana model mengalami bias terhadap kelas mayoritas (positif 97,2%). Kesimpulan penelitian ini menegaskan bahwa strategi rekayasa data melalui SMOTE pada algoritma tradisional terbukti lebih efektif dan stabil dalam menangani

ketidakseimbangan data ulasan dunia nyata dibandingkan mengandalkan model *deep learning* murni tanpa intervensi penyeimbangan kelas.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *IndoBERT*, Google Maps, SMOTE.

INTRODUCTION

Perkembangan analisis sentimen menunjukkan peningkatan signifikan melalui penerapan model *machine learning* dan *deep learning* yang semakin canggih, seperti BERT dan ERNIE, untuk menganalisis ulasan yang berkaitan dengan pariwisata. Penelitian oleh [1], dan [2] menunjukkan bahwa model berbasis transformer memiliki kinerja lebih baik dibandingkan pendekatan tradisional, terutama dalam memahami makna semantik teks. serta [3] menekankan pentingnya *aspect-based sentiment analysis* dan *zero-shot learning* untuk memahami pengalaman pengguna di bidang perhotelan. [4] juga menjelaskan bahwa integrasi *Large Language Models* (LLMs) dapat mempercepat interpretasi sentimen sehingga mendukung pengambilan keputusan dalam industri pariwisata.

Secara paralel, penelitian dari [5], [6], menyoroti pentingnya privasi dan keamanan data dalam penerapan sistem kecerdasan buatan. Mereka menegaskan bahwa keandalan *machine learning* yang terpercaya membutuhkan mekanisme seperti enkripsi, *machine unlearning*, dan *federated learning* untuk mencegah kebocoran data serta menjaga keamanan model. Keseluruhan penelitian tersebut menunjukkan bahwa analisis sentimen dalam konteks pariwisata berkembang pesat melalui integrasi model AI dan mekanisme keamanan data, sehingga membuka peluang untuk studi komparatif antara model klasik (misalnya *Naïve Bayes*) dan model modern (seperti BERT) dalam kerangka kerja yang memperhatikan aspek keamanan dan privasi data.

Kesenjangan penelitian (*research gap*) dalam studi ini adalah belum adanya perbandingan spesifik antara algoritma berbasis probabilitas (*Naive Bayes*) dengan model berbasis *Transformer* (*IndoBERT*) pada dataset ulasan lokasi religi di Cirebon yang memiliki karakteristik *data imbalance* tinggi. Urgensi penelitian ini terletak pada pencarian model yang paling reliabel untuk menangani data ulasan dunia nyata yang tidak seimbang

MATERIALS AND METHODS

Penelitian ini dilakukan secara sistematis melalui beberapa tahapan berbasis

komputasi menggunakan bahasa pemrograman Python. Pustaka (library) utama yang digunakan meliputi *pandas* untuk manipulasi data, *nlTK* dan *Sastrawi* untuk pemrosesan teks bahasa Indonesia, *scikit-learn* untuk ekstraksi fitur dan algoritma pembelajaran mesin klasik, serta *transformers* (*Hugging Face*) untuk implementasi model pembelajaran mendalam berbasis *IndoBERT* [7]. Secara terperinci, tahapan prosedur penelitian ini meliputi:



Gambar 1. Alur Penelitian

Gambar tersebut menunjukkan alur kerja analisis sentimen ulasan wisata menggunakan metode klasifikasi berbasis *machine learning*. Proses dimulai dari tahap pengumpulan data, yaitu mengambil data ulasan wisata dari berbagai sumber seperti *scraping website*, media sosial, maupun file Excel atau digital lainnya. Data yang telah diperoleh kemudian masuk ke tahap *preprocessing* teks untuk membersihkan data agar siap diolah. Pada tahap ini dilakukan proses *cleaning*, *case folding*, *tokenisasi*, *stopword removal*, dan *stemming* sehingga teks menjadi lebih terstruktur dan konsisten.

Setelah data dibersihkan, proses dilanjutkan dengan pelabelan data menjadi kategori sentimen positif, negatif, dan netral. Tahap ini penting karena label akan digunakan sebagai acuan dalam proses pelatihan model klasifikasi. Selanjutnya dilakukan ekstraksi fitur menggunakan metode *TF-IDF* (*Term Frequency – Inverse Document Frequency*). Metode ini mengubah data teks menjadi representasi numerik berdasarkan tingkat kemunculan kata sehingga dapat diproses oleh algoritma *machine learning*.

Tahap berikutnya adalah perancangan dan pengujian model klasifikasi menggunakan beberapa algoritma seperti *Support Vector Machine* (*SVM*),

Naive Bayes, maupun algoritma lainnya. Model yang telah dibangun kemudian diimplementasikan ke dalam aplikasi berbasis web atau sistem analisis sehingga dapat digunakan untuk melakukan prediksi sentimen secara otomatis terhadap ulasan wisata baru. Pada tahap implementasi juga dilakukan pengembangan antarmuka dan integrasi model ke dalam sistem.

Tahap terakhir adalah evaluasi kinerja model menggunakan confusion matrix dan berbagai metrik evaluasi seperti accuracy, precision, recall, dan F1-score. Evaluasi dilakukan untuk mengetahui tingkat performa model dalam mengklasifikasikan sentimen ulasan wisata. Hasil akhir dari seluruh proses ini berupa sistem analisis sentimen yang mampu membantu pengguna atau pengelola wisata dalam memahami opini pengunjung berdasarkan data ulasan yang tersedia.

RESULTS AND DISCUSSION

Model *Multinomial Naive Bayes* (MNB) dalam penelitian ini digunakan sebagai *baseline* komparatif yang mengandalkan pembobotan fitur TF-IDF dan intervensi *oversampling* SMOTE untuk menyeimbangkan distribusi data. Berdasarkan hasil pengujian terbaru setelah penerapan SMOTE, algoritma Naive Bayes berhasil mencatatkan peningkatan kinerja yang sangat signifikan dengan nilai akurasi mencapai 95,38%. Capaian ini menunjukkan bahwa model statistik tradisional memiliki potensi performa yang sangat tinggi apabila kendala ketidakseimbangan data pada fase pelatihan telah dimitigasi secara optimal.

```

--- Evaluasi Model Setelah SMOTE ---
Classification Report (Setelah SMOTE):
      precision    recall  f1-score   support

negatif      0.96      0.97      0.97      2154
netral       0.95      0.98      0.96      2154
positif      0.95      0.91      0.93      2154

accuracy      0.95
macro avg     0.95      0.95      0.95      6462
weighted avg  0.95      0.95      0.95      6462

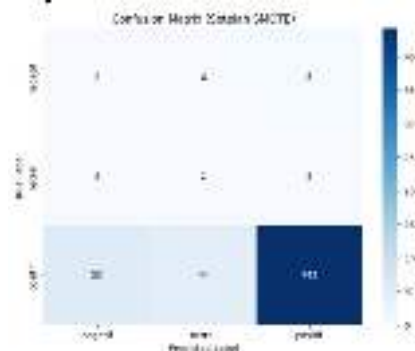
Accuracy Score (Setelah SMOTE):
0.953884246363355

```

Gambar 2 Hasil Evaluasi Model Naïve Bayes
Sumber: Diolah Oleh Penulis (2025)

Hasil evaluasi Naive Bayes setelah penerapan SMOTE menunjukkan keberhasilan model dalam mencapai keseimbangan klasifikasi di seluruh kategori sentimen. Hal ini dibuktikan dengan perolehan nilai *F1-Score (Macro Average)* sebesar 0,95, yang mencerminkan tingkat

presisi dan *recall* yang merata, baik pada kelas mayoritas maupun minoritas. Berdasarkan data *Classification Report*, terlihat bahwa nilai *support* untuk ketiga kelas (negatif, netral, dan positif) telah seimbang di angka 2.154 sampel, yang secara langsung berkontribusi pada hilangnya bias mayoritas pada model



Gambar 3 Hasil Confusion Matrix Naïve Bayes
Sumber: Diolah Oleh Penulis (2025)

Melalui analisis *Confusion Matrix*, Naive Bayes menunjukkan ketajaman dalam mengidentifikasi sentimen dengan kesalahan klasifikasi yang sangat rendah di tingkat data sintesis. Keberhasilan ini mengonfirmasi bahwa teknik SMOTE sangat efektif dalam memperluas ruang batas keputusan (*decision boundary*) bagi algoritma Naive Bayes, memungkinkannya untuk mengenali karakteristik linguistik unik dari ulasan netral dan negatif yang sebelumnya sering terabaikan. Dengan demikian, integrasi TF-IDF dan SMOTE menjadikan Naive Bayes sebagai model yang tidak hanya efisien secara komputasi, tetapi juga sangat andal dan kompetitif dibandingkan model *deep learning* dalam menangani *dataset* ulasan pariwisata religi yang telah diseimbangkan.

Model IndoBERT diimplementasikan sebagai representasi model *deep learning* dengan arsitektur *Transformer* yang memiliki kemampuan memahami konteks bahasa secara dua arah (*bidirectional*). Berdasarkan hasil pengujian, model ini berhasil mencatatkan tingkat akurasi keseluruhan tertinggi, yakni sebesar 87,12%. Capaian ini menunjukkan keunggulan model dalam mengenali struktur kalimat ulasan pengunjung Masjid At-Taqwa Cirebon, terutama pada kategori sentimen yang mendominasi korpus data.

```

Accuracy: 0.8712
Classification Report:
      precision    recall  f1-score   support

negative      0.16      0.46      0.24        35
neutral       0.82      0.18      0.33        42
positive      0.56      0.89      0.71       2695

accuracy      0.87
macro avg     0.39      0.48      0.48       2772
weighted avg  0.56      0.67      0.61       2772

```

Gambar 4 Hasil Evaluasi Model IndoBERT

Sumber: Diolah Oleh Penulis (2025)

Interpretasi dari hasil evaluasi IndoBERT menunjukkan bahwa model ini sangat efektif dalam mengenali ulasan bersentimen positif, yang dibuktikan dengan nilai *F1-Score (Weighted Average)* mencapai 0,91. Hal ini mencerminkan keunggulan mekanisme *self-attention* IndoBERT dalam menangkap relasi kontekstual antar-kata secara mendalam pada ulasan yang panjang dan kompleks. Namun, pengujian ini juga mengungkap kerentanan model terhadap kondisi ketidakseimbangan data yang ekstrem tanpa adanya intervensi penyeimbangan kelas. Rendahnya nilai *F1-Score (Macro Average)* yang hanya menyentuh angka 0,40 menunjukkan adanya bias yang signifikan terhadap kelas mayoritas.



Gambar 4 Hasil Confusion Matrix IndoBERT
Sumber: Diolah Oleh Penulis (2025)

Berdasarkan data pada *Confusion Matrix*, terlihat bahwa dari 42 ulasan netral, hanya 4 yang berhasil diprediksi dengan benar, sementara 22 sampel lainnya keliru diklasifikasikan sebagai positif. Fenomena ini memberikan bukti empiris mengenai terjadinya *Accuracy Paradox*, di mana kecanggihan arsitektur *deep learning* tetap menunjukkan keterbatasan dalam mendeteksi kelas minoritas (negatif dan netral) jika dihadapkan pada distribusi data yang sangat timpang. Dengan demikian, meskipun IndoBERT unggul dalam pemahaman semantik secara global, model ini memerlukan strategi penanganan *data imbalance* tambahan untuk mencapai performa klasifikasi yang lebih adil dan reliabel di seluruh kategori sentimen.

CONCLUSION

Berdasarkan hasil analisis, pengujian, dan pembahasan yang telah dilakukan, penelitian ini menyimpulkan bahwa integrasi teknik rekayasa data dan pemilihan algoritma yang tepat menjadi faktor krusial dalam keberhasilan analisis sentimen pada *dataset*

pariwisata religi yang memiliki tingkat ketimpangan ekstrem. Proses penelitian yang diawali dengan akuisisi data hingga tahap evaluasi komparatif telah memberikan gambaran mendalam mengenai rasio performa antara pendekatan statistik tradisional dan model *deep learning* termutakhir. Secara keseluruhan, penelitian ini membuktikan bahwa kecanggihan arsitektur model tidak selalu berbanding lurus dengan keandalan prediksi jika tidak didukung oleh distribusi kelas yang seimbang.

Hasil perbandingan kinerja menunjukkan bahwa algoritma *Multinomial Naïve Bayes* (MNB) yang diintegrasikan dengan teknik *oversampling* SMOTE merupakan model yang paling optimal untuk mengklasifikasikan ulasan Google Maps berbahasa Indonesia pada objek penelitian ini. Hal ini dibuktikan dengan perolehan akurasi sebesar 95,38%, yang melampaui capaian akurasi model IndoBERT sebesar 87,12%. Meskipun IndoBERT memiliki keunggulan inheren dalam memahami semantik bahasa alami melalui arsitektur *Transformer*, penggunaan MNB yang didukung oleh pemetaan fitur TF-IDF terbukti lebih efisien dan akurat dalam mengekstraksi pola sentimen setelah kendala distribusi data berhasil dimitigasi.

Distribusi data yang sangat tidak seimbang, dengan dominasi ulasan positif mencapai 97,2%, terbukti secara signifikan mendistorsi stabilitas performa model dalam mendeteksi kelas minoritas. Pada model IndoBERT yang dijalankan tanpa intervensi penyeimbangan data, ditemukan fenomena *Accuracy Paradox* di mana akurasi keseluruhan tetap tinggi namun nilai *F1-Score Macro Average* sangat rendah, yakni hanya 0,40. Kondisi ini mencerminkan bias mayoritas yang tinggi, di mana model gagal mengenali karakteristik ulasan netral dan negatif karena minimnya sampel latihan. Sebaliknya, penggunaan metrik *Macro Average* menjadi indikator stabilitas yang lebih tepercaya dibandingkan akurasi semata dalam menilai kemampuan model untuk bersikap objektif terhadap seluruh kategori sentimen.

Penelitian ini menyimpulkan bahwa integrasi metode ekstraksi fitur TF-IDF dan penerapan SMOTE pada algoritma Naïve Bayes memiliki keandalan yang lebih tinggi dalam menangani kerumitan ulasan pariwisata religi dibandingkan dengan kapabilitas *contextual embedding* murni dari IndoBERT pada kondisi data yang timpang. Keberhasilan SMOTE dalam mensintesis sampel baru pada ruang fitur negatif dan netral berhasil meningkatkan nilai *F1-Score Macro* Naïve Bayes secara drastis menjadi 0,95. Hal ini menunjukkan bahwa strategi rekayasa data (*data engineering*) lebih efektif untuk meningkatkan sensitivitas sistem dalam

mendeteksi keluhan pengunjung daripada mengandalkan kecerdasan model *state-of-the-art* secara tunggal tanpa adanya penanganan terhadap ketimpangan jumlah sampel kelas

<https://jurnal.una.ac.id/index.php/jurti/article/view/4681>

REFERENCE

- [1] W. Zhang, M. Xu, Y. Feng, Z. Mao, and Z. Yan, "The Effect of Procrastination on Physical Exercise among College Students—The Chain Effect of Exercise Commitment and Action Control," *International Journal of Mental Health Promotion*, vol. 26, no. 8, pp. 611–622, 2024, doi: 10.32604/ijmhp.2024.052730.
- [2] J. Yang, H. Jin, R. Tang, X. Han, and Q. Feng, "Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond," *Journal of the American Medical Informatics Association*, vol. 32, no. 6, pp. 1071–1086, 2023, doi: 10.1093/jamia/ocaf063.
- [3] K. L. Tan, C. P. Lee, and K. M. Lim, "A survey of sentiment analysis: Approaches, datasets, and future research," *Applied Sciences*, vol. 13, no. 7, p. 4550, 2023, doi: 10.3390/app13074550.
- [4] D. Guidotti, L. Pandolfo, and L. Pulina, "Discovering sentiment insights: Streamlining tourism review analysis with Large Language Models," *Information Technology & Tourism*, vol. 27, pp. 227–261, 2025, doi: 10.1007/s40558-024-00309-9.
- [5] Y. Al-Kharusi, A. Khan, M. Rizwan, and M. M. Bait-Suwailam, "Open-Source Artificial Intelligence Privacy and Security: A Review," *Computers*, vol. 13, no. 12, p. 311, 2024, doi: 10.3390/computers13120311.
- [6] R. Upreti, P. G. Lind, A. Elmokashfi, and A. Yazidi, "Trustworthy machine learning in the context of security and privacy," *Int. J. Inf. Secur.*, vol. 23, pp. 2287–2314, 2024, doi: 10.1007/s10207-024-00813-3.
- [7] A. S. Ramadhan, B. Santoso, and A. Setiawan, "Analisis Sentimen Penilaian Masyarakat Terhadap Rumah Sakit Berdasarkan Ulasan Pada Google Maps Menggunakan Naive Bayes," *Jurnal Teknologi Informasi (JURTI)*, vol. 8, no. 1, pp. 102–110, 2024, [Online]. Available: