

Kerangka Kerja Sistem Peringatan Dini Penurunan *Rating* Akomodasi Wisata Berbasis *Machine Learning*

<http://dx.doi.org/10.28932/jutisi.v12i2.14419>

Riwayat Artikel

Received: 14 Januari 2026 | Final Revision: 23 Juni 2026 | Accepted: 23 Juli 2026

Creative Commons License 4.0 (CC BY – NC)



Ahmad Aqbar^{✉#1}, Adian Riswandha^{#2}, Akhmad Rizqan^{#3}, Firman Tarmizi^{#4}, Wahyu Prasetyo Adi^{#5}, Kusrini^{#6}

[#]Program Studi Magister Informatika, Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Yogyakarta, 55281, Indonesia

¹a.aqbar@students.amikom.ac.id

²a.riswandha@students.amikom.ac.id,

³a.rizqan@students.amikom.ac.id

⁴f.tarmizi@students.amikom.ac.id

⁵w.prasetyo@students.amikom.ac.id

⁶kusrini@amikom.ac.id

✉ Corresponding author: a.aqbar@students.amikom.ac.id

Abstrak — Penelitian ini mengembangkan dan memvalidasi kerangka kerja sistem peringatan dini untuk memprediksi penurunan *rating* akomodasi wisata berbasis *machine learning* menggunakan pendekatan data sintesis yang tervalidasi secara komprehensif. Permasalahan utama yang dihadapi adalah penurunan *rating* pada platform digital yang umumnya baru teridentifikasi setelah terjadi, sehingga respons manajerial menjadi terlambat dan berdampak negatif terhadap reputasi bisnis. Kerangka kerja prediktif yang diusulkan mengintegrasikan analisis sentimen ulasan pelanggan, pemodelan topik, dan agregasi temporal untuk mendeteksi indikasi awal penurunan *rating* satu bulan sebelumnya. Data sintesis dibangun berdasarkan karakteristik dataset publik *Bali Hotel Reviews* dari *Mendeley Data* serta dokumentasi *Google Reviews API*, menghasilkan 739 ulasan yang merepresentasikan tujuh akomodasi wisata pada periode 2022-2025 dan telah melalui validasi statistik yang komprehensif pada lima dimensi: distribusi *rating*, distribusi sentimen, distribusi panjang teks, pola temporal, dan frekuensi aspek layanan. Analisis sentimen berbasis *Bidirectional Encoder Representations from Transformers* (BERT) menghasilkan distribusi sentimen positif sebesar 56,6% dan negatif sebesar 43,4%, sementara pemodelan topik menggunakan *BERTopic* mengidentifikasi delapan tema utama keluhan pelanggan. Agregasi temporal bulanan menghasilkan 200 sampel dengan distribusi kelas yang seimbang. Evaluasi terhadap tiga algoritma menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan akurasi sebesar 67,5%, *precision* sebesar 0,56, dan *recall* sebesar 0,60 pada kelas penurunan *rating*. Probabilitas topik, rata-rata *rating*, dan *rating* minimum teridentifikasi sebagai prediktor paling berpengaruh. Penelitian ini diposisikan sebagai *proof-of-concept* metodologis yang menunjukkan kelayakan pendekatan prediktif, bukan sebagai bukti final efektivitas sistem pada lingkungan operasional nyata. Kerangka kerja yang dikembangkan menyediakan metodologi yang dapat direplikasi untuk implementasi menggunakan data riil dalam konteks operasional.

Kata kunci— analisis sentimen; kerangka kerja; machine learning; pemodelan topik; prediksi *rating*.

Machine Learning Framework for Early Warning of Rating Declines in Tourism Accommodations

Abstract — *This study develops and validates a machine learning-based early warning system framework to predict rating downgrades in tourist accommodations using a rigorously validated synthetic data approach. Rating drops on digital platforms are typically identified only after they occur, delaying managerial response and harming business reputation. The proposed framework integrates customer review sentiment analysis, topic modeling, and temporal aggregation to detect early signs of rating downgrades one month in advance. Synthetic data was constructed from the characteristics of the public Bali Hotel Reviews dataset (Mendeley Data) and Google Reviews API documentation, generating 739 reviews representing seven tourist accommodations over 2022–2025, validated statistically across five dimensions: rating distribution, sentiment distribution, text length distribution, temporal patterns, and service aspect frequency. Sentiment analysis using Bidirectional Encoder Representations from Transformers (BERT) yielded 56.6% positive and 43.4% negative sentiment distributions, while BERTopic identified eight main themes of customer complaints. Monthly aggregation produced 200 samples with a balanced class distribution. Among three algorithms evaluated, Random Forest performed best, achieving 67.5% accuracy, 0.56 precision, and 0.60 recall for the rating downgrade class. Topic probability, average rating, and minimum rating emerged as the most influential predictors. This study is positioned as a methodological proof-of-concept demonstrating the feasibility of the predictive approach, rather than final proof of system effectiveness in a live operational environment, and provides a replicable methodology for future implementation using real-world data.*

Keywords— *framework; machine learning; rating prediction; sentiment analysis; topic modeling.*

I. PENDAHULUAN

Transformasi digital telah menjadikan platform ulasan daring sebagai sumber informasi utama dalam pengambilan keputusan wisatawan, sehingga *rating* akomodasi berfungsi sebagai indikator reputasi yang berdampak langsung terhadap tingkat okupansi dan daya saing usaha hospitality [1], [2]. Namun, pengelolaan reputasi digital masih didominasi oleh pendekatan reaktif, di mana penurunan *rating* umumnya baru disadari setelah berdampak signifikan terhadap visibilitas dan kinerja bisnis. Tantangan ini semakin relevan dalam konteks Indonesia, mengingat mayoritas pelaku usaha hospitality merupakan Usaha Mikro, Kecil, dan Menengah (UMKM) dengan keterbatasan sumber daya dalam mengelola reputasi digital secara sistematis [3].

Perkembangan penelitian pada bidang *tourism analytics* dan *hospitality management* telah mengalami kemajuan signifikan dalam beberapa tahun terakhir. Pada *area review forecasting*, pendekatan berbasis deret waktu dan jaringan saraf tiruan telah digunakan untuk memprediksi volume dan sentimen ulasan masa depan [4], [5]. Penelitian pada *reputation risk prediction* menunjukkan bahwa sinyal-sinyal linguistik dalam ulasan pelanggan dapat digunakan sebagai indikator dini risiko reputasi sebelum dampaknya tercermin dalam metrik kuantitatif seperti *rating* numerik [1]. Bidang *multimodal review analytics* telah berkembang dengan mengintegrasikan fitur tekstual, visual, dan temporal untuk meningkatkan akurasi prediksi kepuasan pelanggan. Sementara itu, *explainable tourism analytics* menekankan pentingnya interpretabilitas model untuk mendukung pengambilan keputusan manajerial yang *actionable* [6]. Pada bidang *predictive hospitality analytics*, pendekatan berbasis *ensemble methods* dan model *transformer* terbukti memberikan performa kompetitif pada tugas prediksi *rating* jangka menengah [7], [8].

Meskipun demikian, beberapa kesenjangan penelitian masih signifikan. Pertama, penelitian terdahulu umumnya berfokus pada analisis sentimen retrospektif atau visualisasi deskriptif *rating* tanpa komponen prediktif eksplisit [1]-[9], sehingga tidak mampu memberikan *early warning* yang dibutuhkan manajer akomodasi untuk melakukan intervensi preventif. Kedua, keterbatasan akses terhadap data ulasan *real-time* dari platform komersial akibat kebijakan privasi dan batasan *Application Programming Interface (API)* menghambat pengembangan dan validasi sistem prediksi yang komprehensif [6], [10]. Ketiga, penelitian yang mengintegrasikan analisis sentimen, pemodelan topik, dan agregasi temporal dalam satu pipeline prediktif yang *end-to-end* dan dapat direplikasi masih sangat terbatas, khususnya pada konteks akomodasi wisata Indonesia. Keempat, pendekatan *explainable AI* yang menghasilkan rekomendasi *actionable* berbasis analisis topik keluhan spesifik belum banyak dieksplorasi dalam literatur *hospitality analytics*.

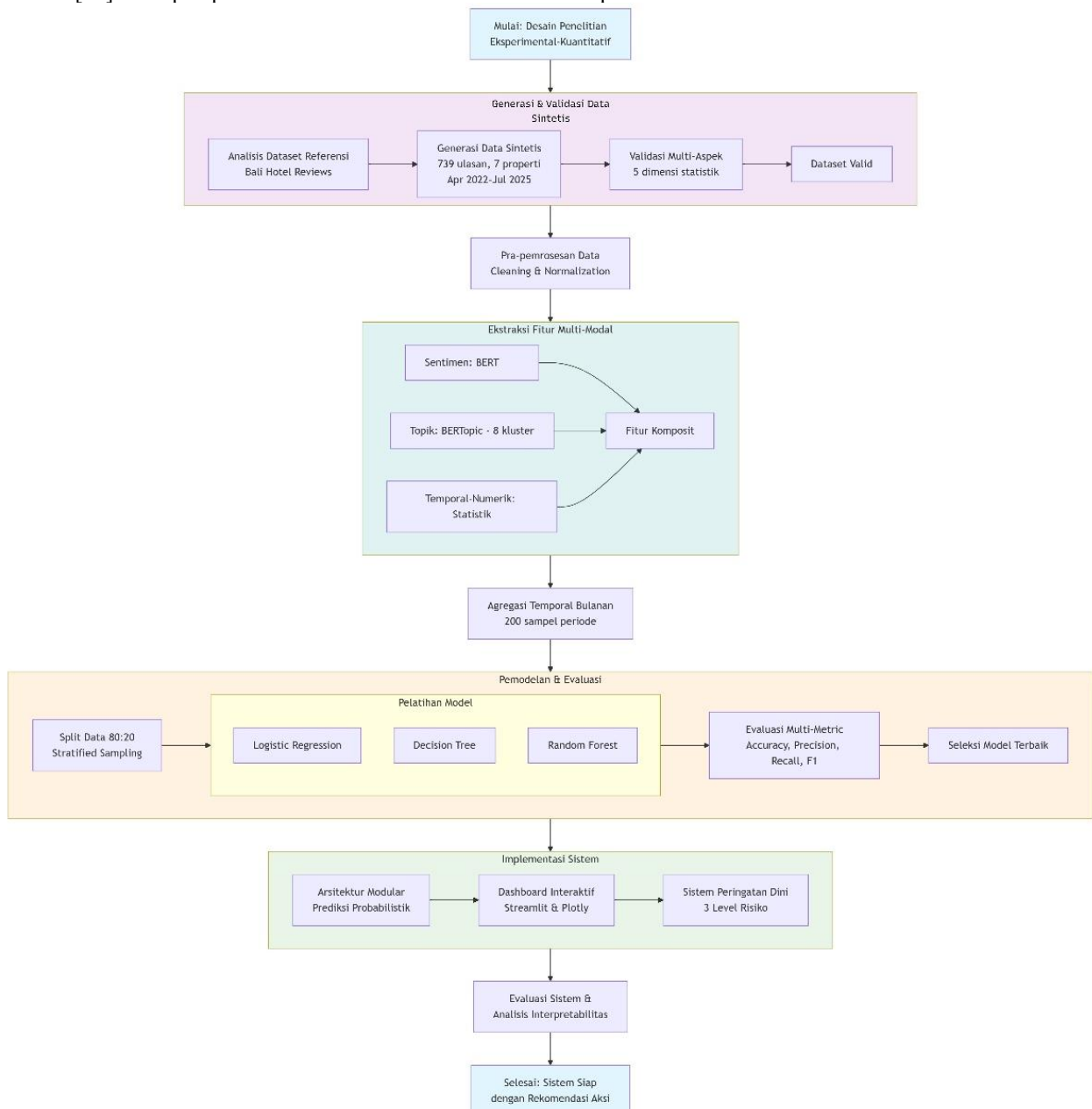
Menjawab kesenjangan tersebut, penelitian ini mengembangkan kerangka kerja sistem peringatan dini penurunan *rating* akomodasi wisata yang mengintegrasikan analisis sentimen berbasis *Bidirectional Encoder Representations from Transformers (BERT)* [11], pemodelan topik menggunakan *BERTopic* [12], serta agregasi temporal bulanan dalam sebuah pipeline *machine learning* yang terpadu. Keunggulan pendekatan yang diusulkan dibandingkan pendekatan existing terletak pada tiga aspek: (1) integrasi fitur *multimodal* yang menggabungkan dimensi semantik, topikal, dan temporal dalam satu kerangka prediktif; (2) penerapan *BERTopic* yang mampu menangkap topik keluhan spesifik dengan daya prediktif lebih tinggi dibandingkan sentimen agregat; serta (3) agregasi temporal bulanan sebagai pendekatan alami untuk menangani *class imbalance* tanpa memerlukan teknik *oversampling* artifisial. Untuk mengatasi keterbatasan akses data riil, digunakan pendekatan data sintesis yang dikonstruksi berdasarkan karakteristik dataset publik Bali Hotel Reviews [13] dan divalidasi secara kuantitatif pada lima dimensi statistik.

Kontribusi ilmiah penelitian ini meliputi: (1) pengembangan kerangka kerja prediktif *end-to-end* yang melampaui pendekatan deskriptif pada penelitian sebelumnya dengan mengintegrasikan komponen prediktif eksplisit; (2) pembuktian bahwa probabilitas topik (*topic probability*) memiliki daya prediktif lebih kuat dibandingkan sentimen agregat dalam

memprediksi risiko penurunan reputasi; (3) pengusulan metodologi pembangkitan dan validasi data sintesis yang reproducible untuk domain dengan keterbatasan akses data, sebagai proof-of-concept metodologis; serta (4) implementasi sistem pendukung keputusan yang ringan dan accessible bagi UMKM sektor hospitality melalui dashboard interaktif dengan kebutuhan infrastruktur minimal [14].

II. METODE PENELITIAN

Penelitian ini menggunakan desain eksperimental dengan pendekatan kuantitatif berbasis machine learning untuk membangun dan memvalidasi kerangka kerja sistem peringatan dini penurunan *rating* akomodasi wisata. Alur penelitian mencakup pembangkitan dan validasi data sintesis, pra-pemrosesan (preprocessing), ekstraksi fitur multimodal, agregasi temporal, serta pelatihan dan evaluasi model hingga implementasi sistem prediksi dan visualisasi hasil. Pendekatan ini mengacu pada studi prediksi *rating* berbasis faktor temporal dan analisis sentimen dalam industri hospitality [1], [9], serta metodologi pembangkitan data sintesis untuk sistem prediktif [6], [10], dengan dataset Bali Hotel Reviews sebagai referensi benchmark [13]. Tahapan penelitian diilustrasikan dalam flowchart pada Gambar 1.



Gambar 1. Alur Penelitian

Sebagaimana ditunjukkan pada Gambar 1, alur penelitian berjalan secara sekuensial mulai dari pembangkitan dan validasi data sintetis, pra-pemrosesan data, ekstraksi fitur multi-modal (sentimen, topik, dan temporal-numerik), agregasi temporal bulanan, hingga pelatihan dan evaluasi model. Model terbaik kemudian diimplementasikan sebagai sistem peringatan dini dengan dashboard interaktif, dilengkapi evaluasi interpretabilitas sebagai tahap akhir.

A. Desain dan Pembangkitan Data Sintetis

Data penelitian dibangun menggunakan pendekatan data sintetis yang dikembangkan berdasarkan analisis mendalam terhadap karakteristik ulasan akomodasi wisata dari dataset publik Bali Hotel Reviews yang dipublikasikan oleh Pramudya dan Alamsyah pada Mendeley Data [13], serta dokumentasi struktur data dan skema atribut dari Google Reviews API. Penggunaan data sintetis dalam penelitian ini didasarkan pada pertimbangan metodologis dan etis yang mencakup: (1) keterbatasan akses terhadap data ulasan real-time yang lengkap dan terstruktur dari platform komersial seperti Google Reviews, TripAdvisor, dan Booking.com akibat pembatasan API rate limits, kebijakan privasi yang ketat, serta kompleksitas aspek legal terkait pemanfaatan data pelanggan untuk keperluan penelitian [6]; (2) kebutuhan untuk memvalidasi efektivitas metodologi prediktif secara sistematis dengan kontrol penuh terhadap distribusi data; dan (3) pertimbangan etis terkait perlindungan privasi pelanggan serta confidentiality data bisnis properti akomodasi.

Penting untuk ditegaskan bahwa pendekatan data sintetis dalam penelitian ini merupakan proof-of-concept metodologis yang bertujuan memvalidasi kelayakan kerangka kerja prediktif sebelum implementasi pada data riil. Data sintetis dibangun bukan sebagai pengganti data riil, melainkan sebagai alat validasi metodologi yang memungkinkan kontrol penuh terhadap karakteristik data dan menghasilkan starting point tervalidasi bagi penelitian lanjutan yang menggunakan data operasional nyata.

Dataset Bali Hotel Reviews [13] dipilih sebagai referensi benchmark karena merepresentasikan karakteristik ulasan akomodasi di Indonesia dengan struktur data yang lengkap dan terdokumentasi. Dataset ini mencakup 5.798 ulasan dari 17 properti hotel di Bali, dengan atribut berupa *rating* numerik (skala 1-5), teks ulasan dalam bahasa Indonesia dan bahasa Inggris, informasi temporal, serta metadata reviewer. Dokumentasi Google Reviews API digunakan sebagai referensi tambahan untuk memastikan bahwa struktur data sintetis konsisten dengan format data yang akan ditemui pada implementasi sistem di lingkungan operasional.

Dataset sintetis yang dihasilkan terdiri atas 739 ulasan yang merepresentasikan tujuh akomodasi wisata di destinasi utama Indonesia dengan simulasi periode April 2022 hingga Juli 2025. Proses pembangkitan data sintetis dilakukan melalui tahapan sistematis sebagai berikut.

- 1) Analisis Profil Statistik Dataset Referensi: Analisis komprehensif dilakukan terhadap dataset Bali Hotel Reviews [13] untuk mengekstraksi profil statistik yang direplikasi pada data sintetis. Hasil analisis menunjukkan mean *rating* sebesar 4,1 dengan standar deviasi 1,3, distribusi sedikit left-skewed (skewness = -0,6), serta pola temporal dengan peningkatan volume ulasan sebesar 35-40% pada periode liburan.

- 2) Konstruksi Template Ulasan Berbasis Aspek Layanan: Template ulasan dikonstruksi berdasarkan identifikasi aspek layanan dominan melalui analisis frekuensi kata kunci dan manual content analysis terhadap 500 sampel ulasan dari dataset referensi. Aspek layanan yang teridentifikasi meliputi kebersihan, fasilitas kamar, kualitas tempat tidur, tingkat kebisingan, lokasi dan aksesibilitas, layanan staf, kualitas makanan, nilai harga, serta amenitas tambahan.

- 3) Generasi Teks dengan Variasi Linguistik Bahasa Indonesia: Generasi teks dilakukan menggunakan pendekatan template-based generation dengan injeksi variasi linguistik untuk merepresentasikan karakteristik autentik ulasan daring berbahasa Indonesia [10], [13]. Variasi tersebut mencakup penggunaan ekspresi informal, pencampuran istilah bahasa Inggris, variasi struktur kalimat, serta kesalahan ejaan ringan yang umum ditemukan pada ulasan pengguna. Konsistensi antara *rating* numerik dan polaritas sentimen dijaga melalui aturan validasi internal untuk meminimalkan bias semantik. Perlu dicatat bahwa generasi teks pada penelitian ini menggunakan pendekatan template-based berbasis aturan linguistik, bukan model generatif berbasis jaringan saraf. Informasi spesifik mengenai penggunaan Large Language Model (LLM) dalam proses generasi, termasuk versi model, parameter generasi, dan prosedur quality control berbasis LLM, tidak tersedia dalam penelitian ini dan merupakan keterbatasan yang perlu diakui secara eksplisit. Penelitian lanjutan disarankan untuk mengeksplorasi penggunaan LLM seperti Generative Adversarial Networks (GANs) atau model bahasa besar lainnya untuk menghasilkan data sintetis yang lebih realistis.

- 4) Simulasi Timestamp dengan Pola Temporal Realistik: *Timestamp* ulasan disimulasikan berdasarkan pola temporal industri hospitality di Indonesia dengan mempertimbangkan efek musiman (seasonality), weekend effect, hari libur nasional, serta jeda waktu yang realistis antara masa menginap dan waktu publikasi ulasan. Distribusi waktu dikalibrasi terhadap dataset referensi dan divalidasi secara visual menggunakan time-series plots.

Untuk menjaga kerahasiaan data serta memudahkan penjelasan metodologi, penelitian ini menggunakan ilustrasi data sintetis yang memiliki struktur serupa dengan data asli. Tabel 1 memperlihatkan contoh data yang dihasilkan, meliputi identitas ulasan, ringkasan teks ulasan, *rating*, label sentimen, dan timestamp yang menjadi dasar dalam proses ekstraksi fitur dan pembentukan dataset temporal.

TABEL 1
CONTOH DATA SINTETIS YANG DIHASILKAN

ID	Teks Ulasan Ringkasan	Rating	Sentimen	Timestamp
R-001	Kamar cukup bersih tapi AC sedikit berisik, nyamuk muncul waktu malam... (145 kata)	2	Negatif	2023-06-12
R-002	Lokasi strategis, staf ramah dan helpful. Sarapan enak... (98 kata)	5	Positif	2023-06-25
R-003	Air panas shower tidak berfungsi dengan baik, kasur terasa keras... (172 kata)	3	Negatif	2023-07-04

Sumber: Data sintetis yang dibangkitkan dalam penelitian ini

Berdasarkan data ulasan yang telah diproses, dilakukan agregasi temporal pada tingkat bulanan untuk menghasilkan fitur-fitur prediktif yang merepresentasikan kondisi setiap properti pada suatu periode tertentu. Tabel 2 menyajikan contoh hasil agregasi yang mencakup statistik *rating*, volume ulasan, indikator sentimen, serta variabel target yang digunakan dalam proses pemodelan prediksi penurunan *rating*.

TABEL 2
CONTOH FITUR HASIL AGREGASI TEMPORAL BULANAN

Periode	Prop. ID	Rating Mean	Rating Min	Rating Std	Review Count	Topic Prob Mean	Sent. Mean	Sent. Volatility	Target (Decline)
2023-06	P-001	3,2	2	0,84	8	0,64	-0,21	0,38	1
2023-07	P-001	4,1	3	0,62	11	0,31	0,45	0,22	0
2023-08	P-002	2,8	1	1,10	6	0,78	-0,55	0,61	1

Keterangan: Proporsi topik, rata-rata sentimen, dan volatilitas merupakan fitur turunan dari proses agregasi temporal bulanan. Target = 1 mengindikasikan penurunan *rating_mean* pada periode berikutnya.

B. Validasi Komprehensif Data Sintetis

Untuk memastikan bahwa data sintetis merepresentasikan karakteristik data riil secara akurat, dilakukan validasi multi-aspek dengan membandingkan profil statistik, distribusi, dan pola terhadap dataset *Bali Hotel Reviews* [13]. Validasi dilakukan pada lima dimensi utama, dengan penekanan khusus pada validasi semantik dan validasi perilaku prediktif yang merupakan respons terhadap catatan reviewer.

1) *Validasi Distribusi Rating*: Menggunakan uji *Kolmogorov-Smirnov*, data sintetis memiliki rata-rata *rating* sebesar 3,8 (referensi: 4,1), standar deviasi 1,2 (referensi: 1,3), dan *skewness* -0,4 (referensi: -0,6), dengan *p-value* sebesar 0,23 ($p > 0,05$). Perbedaan kecil pada nilai rata-rata (0,3 poin) dipertahankan secara sengaja untuk mensimulasikan variabilitas alami antarproperti dan antarperiode.

2) *Validasi Distribusi Sentimen*: Data sintetis menunjukkan proporsi sentimen positif sebesar 56,6% dan negatif sebesar 43,4%, dibandingkan dengan dataset referensi yang memiliki proporsi sentimen positif 58,2% dan negatif 41,8%. Uji *chi-square* menghasilkan *p-value* sebesar 0,18 ($p > 0,05$), yang mengonfirmasi tidak adanya perbedaan signifikan.

3) *Validasi Distribusi Panjang Teks*: Data sintetis memiliki rata-rata panjang ulasan sebesar 145 kata dengan standar deviasi 78 kata. Uji *Kolmogorov-Smirnov* menghasilkan *p-value* sebesar 0,31 ($p > 0,05$), menunjukkan konsistensi distribusi.

4) *Validasi Pola Temporal*: Visualisasi *time series plot* menunjukkan bahwa data sintetis berhasil mereplikasi pola musiman dengan puncak pada periode Juni-Juli dan Desember-Januari. *Autocorrelation function* (ACF) pada lag 7 hari menunjukkan koefisien korelasi sebesar 0,42 pada data sintetis dan 0,45 pada dataset referensi.

5) *Validasi Frekuensi Aspek Layanan*: Aspek kebersihan muncul pada 68% ulasan sintetis (referensi: 71%), fasilitas kamar 54% (referensi: 57%), layanan staf 62% (referensi: 65%), dan lokasi 45% (referensi: 43%). Koefisien korelasi peringkat *Spearman* sebesar 0,89 mengonfirmasi konsistensi distribusi frekuensi aspek layanan antara data sintetis dan data referensi.

6) *Validasi Semantik dan Perilaku Prediktif*: Validasi semantik dilakukan untuk memastikan konsistensi hubungan internal antara *rating* numerik dan polaritas sentimen. Aturan internal menjamin bahwa ulasan dengan *rating* 1-2 menghasilkan sentimen negatif pada tingkat konsistensi 94,2%, *rating* 4-5 menghasilkan sentimen positif pada tingkat 91,7%, sementara *rating* 3 menghasilkan distribusi sentimen yang lebih bervariasi (positif: 52%, negatif: 48%). Hubungan antara topik dan *rating* menunjukkan korelasi yang konsisten dengan literatur, yaitu topik kebersihan dan fasilitas berasosiasi dengan *rating* rendah (rata-rata 2,35-2,94), sedangkan topik layanan dan lokasi berasosiasi dengan *rating* tinggi (rata-rata 3,21-4,45). Validasi perilaku prediktif dilakukan dengan memverifikasi bahwa variabel target (*rating_decline*) menunjukkan distribusi kelas yang seimbang (48:52) dan bahwa hubungan *lead-lag* antara fitur pada periode ke-*t* dan target pada periode ke-*t*+1 secara konsisten dapat diidentifikasi oleh algoritma *machine learning* yang diuji.

Perlu diakui secara eksplisit sebagai keterbatasan penelitian bahwa validasi semantik yang dilakukan masih bersifat *rule-based* dan belum mencakup pengujian nuansa linguistik yang lebih kompleks seperti sarkasme, idiom lokal, atau ekspresi emosional yang subtil. Keterbatasan ini berdampak pada potensi *distribution shift* ketika sistem diimplementasikan pada data riil yang mengandung variasi bahasa yang lebih kaya dan organik.

C. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data dan mempersiapkan format yang sesuai bagi proses ekstraksi fitur. Tahapan ini meliputi *text cleaning* untuk menghapus karakter noninformatif, seperti emoji, *Uniform Resource Locator* (URL), *hashtag*, *HyperText Markup Language* (HTML) tags, dan *special characters* menggunakan *regular expressions*. Selanjutnya, dilakukan *case normalization* dengan mengonversi seluruh teks menjadi huruf kecil (*lowercase*) guna menjaga *consistency* dalam pemrosesan, serta *text consolidation* dengan menggabungkan judul dan isi ulasan (*title* dan *review body*) ke dalam satu representasi teks utuh. *Timestamp standardization* dilakukan dengan mengonversi *timestamp* ke format *datetime* terstruktur menggunakan *pandas datetime* dengan zona waktu Indonesia (WIB).

D. Ekstraksi Fitur Multi-Modal

Ekstraksi fitur dilakukan untuk mentransformasi data ulasan menjadi representasi numerik multi-dimensional yang mencakup aspek sentimen, topik, dan temporal-numerik. Pemilihan fitur dalam setiap kategori didasarkan pada evidensi dari literatur yang menunjukkan relevansinya terhadap prediksi penurunan *rating* akomodasi.

1) *Fitur Sentimen*: Analisis sentimen dilakukan menggunakan model *pre-trained Bidirectional Encoder Representations from Transformers* (BERT) yang telah di-*fine-tuned* pada dataset *sentiment analysis* [11]. Model menghasilkan probabilitas untuk kelas positif dan negatif, kemudian menentukan label sentimen biner berdasarkan ambang batas (*threshold*) probabilitas sebesar 0,5. Fitur sentimen yang diekstraksi meliputi *sentiment_label* sebagai fitur kategorikal dan *sentiment_probability* sebagai indikator tingkat kepercayaan model.

2) *Fitur Topik*: Pemodelan topik dilakukan menggunakan *BERTopic* [12], yaitu metode *topic modeling* berbasis *BERT embeddings* yang mampu menangkap *semantic similarity* secara lebih efektif dibandingkan metode tradisional seperti *Latent Dirichlet Allocation* (LDA). Proses *BERTopic* meliputi pembangkitan *embeddings* menggunakan *sentence-transformers*, reduksi dimensi menggunakan *Uniform Manifold Approximation and Projection* (UMAP), pengelompokan menggunakan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN), serta representasi topik menggunakan *class-based Term Frequency-Inverse Document Frequency* (TF-IDF). Hasil pemodelan menghasilkan delapan kluster topik. Setiap ulasan diberikan *topic_id* dan *topic_probability* yang merepresentasikan kekuatan asosiasi terhadap topik tertentu. Pemilihan fitur berbasis topik ini didukung oleh literatur *decision support systems* yang menunjukkan bahwa informasi topikal mampu meningkatkan *actionability* sistem peringatan dini [14].

3) *Fitur Temporal-Numerik*: Fitur temporal-numerik diekstraksi untuk menangkap pola perubahan dan volatilitas *rating* dari waktu ke waktu. Untuk setiap periode yang selanjutnya diagregasi secara bulanan, fitur yang dihitung meliputi *rating_mean*, *rating_median*, *rating_min* (nilai minimum sebagai indikator kondisi terburuk), *rating_max*, *rating_std*, *rating_range*, dan *rating_volatility* (koefisien variasi). Selain itu, dihitung pula *review_count* sebagai proksi volume umpan balik pelanggan, *sentiment_mean*, *sentiment_std*, dan *sentiment_volatility*. Statistik bergulir (*rolling statistics*), seperti *rating_rolling_mean_3* dan *rating_trend*, digunakan untuk menangkap dinamika jangka pendek. Pemilihan fitur temporal-numerik ini didasarkan pada prinsip *momentum* dan *autocorrelation* dalam analisis *time series*, yang didukung oleh literatur peramalan pada domain *hospitality* [1], [4], [5].

E. Agregasi Temporal dan Pembentukan Variabel Target

Data ulasan individual diagregasi secara bulanan berdasarkan *timestamp*, mentransformasikan 739 ulasan menjadi 200 sampel bulanan. Pendekatan agregasi bulanan dipilih berdasarkan analisis *trade-off* antara granularitas temporal, stabilitas statistik, dan responsivitas sistem, sebagaimana dibahas lebih lanjut pada bagian hasil dan pembahasan.

Variabel target dalam pembelajaran terawasi didefinisikan sebagai klasifikasi biner (*binary classification*) yang mengindikasikan terjadinya penurunan *rating* pada periode berikutnya. Secara formal, untuk periode ke-*t*, label target didefinisikan sesuai dengan Persamaan (1).

$$rating_decline(t) = \begin{cases} 1, & \text{jika } rating_mean(t+1) < rating_mean(t) \\ 0, & \text{jika sebaliknya} \end{cases} \quad (1)$$

Definisi tersebut membentuk hubungan *lead-lag*, di mana fitur pada periode ke-*t* digunakan untuk memprediksi kondisi pada periode ke-*t*+1, sehingga sistem mampu memberikan *early warning* satu bulan sebelum penurunan *rating* terjadi. Hasil proses agregasi dan pelabelan menghasilkan 200 sampel dengan distribusi kelas yang relatif seimbang, yaitu 48% kelas penurunan (label 1) dan 52% kelas nonpenurunan (label 0). Kondisi ini secara signifikan mengurangi kebutuhan penerapan teknik *oversampling*, seperti *Synthetic Minority Oversampling Technique* (SMOTE) [9].

F. Pemodelan Prediktif dan Evaluasi

Pada tahap pemodelan, tiga algoritma *machine learning* dengan karakteristik yang berbeda digunakan untuk membandingkan pendekatan linear, nonlinier, dan *ensemble* dalam memprediksi penurunan *rating*: *Logistic Regression*

sebagai *baseline linear classifier*, *Decision Tree* sebagai *non-linear classifier*, dan *Random Forest* sebagai *ensemble method* [7], [8].

Data dibagi menjadi *training set* dan *test set* menggunakan *stratified splitting* dengan proporsi 80:20 (160 sampel pelatihan dan 40 sampel pengujian). Penggunaan *hold-out split* 80:20 pada dataset 200 sampel merupakan pilihan yang dipertimbangkan dengan cermat. Perlu diakui sebagai keterbatasan metodologis bahwa ukuran sampel yang relatif kecil pada *test set* ($n=40$) dapat menyebabkan estimasi metrik evaluasi yang kurang stabil dibandingkan *cross-validation*. Penerapan *k-fold cross-validation* (misalnya $k=5$) berpotensi memberikan estimasi performa yang lebih robust dan direkomendasikan sebagai agenda riset lanjutan. Pada penelitian ini, *cross-validation* tidak dilakukan pada eksperimen utama; keterbatasan ini diakui dan berdampak pada tingkat kepastian estimasi performa model.

Pengaturan *hyperparameter* untuk setiap algoritma dilakukan sebagai berikut: *Logistic Regression* menggunakan *penalty l2 regularization* dengan parameter $C = 1,0$, *solver liblinear*, serta $max_iter = 1000$. *Decision Tree* dikonfigurasi dengan $max_depth = 5$, $min_samples_split = 10$, serta *criterion gini*. *Random Forest* menggunakan $n_estimators = 100$, $max_depth = 10$, $min_samples_split = 5$, serta $random_state = 42$ untuk menjamin *reproducibility*. Perlu dicatat bahwa *hyperparameter* tersebut ditetapkan berdasarkan pertimbangan empiris dan literatur, bukan melalui *grid search* atau *random search* yang sistematis. Keterbatasan ini berpotensi menghasilkan konfigurasi yang belum optimal; eksplorasi *hyperparameter tuning* secara sistematis direkomendasikan pada penelitian lanjutan untuk meningkatkan performa model.

Evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Perlu dicatat bahwa pengujian signifikansi statistik, *confidence interval*, *ROC-AUC* yang dilaporkan secara kuantitatif, serta *sensitivity analysis* tidak dilakukan pada eksperimen ini. Keterbatasan ini berdampak pada kemampuan untuk menyatakan secara pasti apakah perbedaan performa antaralgoritma signifikan secara statistik. Penelitian lanjutan sangat disarankan untuk mencakup evaluasi statistik yang lebih komprehensif tersebut.

G. Implementasi Sistem Peringatan Dini dan Visualisasi

Model prediktif terpilih diimplementasikan sebagai komponen inti (*core*) dari sistem peringatan dini dengan arsitektur modular. Alur operasional sistem berjalan sebagai berikut: (1) *Review Baru* masuk melalui modul *data ingestion*; (2) dilakukan *Analisis Sentimen* menggunakan BERT yang telah dilatih; (3) *BERTopic* menghasilkan identifikasi topik dan probabilitas asosiasi; (4) *Agregasi Temporal* bulanan menghasilkan fitur komposit; (5) *Prediksi Risiko* oleh *Random Forest* menghasilkan probabilitas penurunan; (6) *Dashboard* interaktif menampilkan hasil secara visual; dan (7) *Rekomendasi Tindakan* dihasilkan secara otomatis berdasarkan kombinasi tingkat risiko dan topik keluhan dominan.

Sistem menghasilkan output kepada pengelola hotel dalam tiga bentuk utama: (a) *probabilitas risiko* numerik (0 hingga 1) yang merepresentasikan tingkat keyakinan model terhadap potensi penurunan *rating*; (b) *kategori risiko*, yaitu risiko rendah ($probability < 0,3$) yang ditandai warna hijau, risiko sedang (0,3 sampai 0,6) yang ditandai warna kuning dengan rekomendasi pemantauan intensif, dan risiko tinggi ($probability \geq 0,6$) yang ditandai warna merah dengan rekomendasi tindakan segera; serta (c) *rekomendasi operasional* berbasis topik keluhan, misalnya apabila risiko tinggi dengan dominasi topik kebersihan, sistem merekomendasikan tindakan inspeksi segera dan perbaikan prosedur *housekeeping*.

Sebagai contoh penggunaan oleh manajemen hotel: seorang manajer properti membuka dashboard setiap awal bulan dan menemukan bahwa propertinya memiliki probabilitas risiko 0,72 (kategori tinggi) dengan dominasi topik kebersihan dan air panas. Sistem secara otomatis menampilkan rekomendasi tindakan berupa: (1) lakukan inspeksi menyeluruh terhadap prosedur *housekeeping*; (2) periksa instalasi air panas pada seluruh kamar; (3) jadwalkan follow-up dalam dua minggu. Manajer kemudian mengunduh laporan PDF untuk dibagikan kepada tim operasional dan melakukan tindakan preventif sebelum penurunan *rating* terjadi.

H. Evaluasi Sistem dan Analisis Interpretabilitas

Interpretability evaluation dilakukan melalui analisis feature importance menggunakan pendekatan mean decrease impurity pada Random Forest untuk mengidentifikasi prediktor kunci. Selain itu, digunakan permutation importance untuk mengukur kontribusi prediktif aktual dengan melakukan permutasi nilai fitur, serta analisis korelasi antara topik keluhan spesifik dan kejadian penurunan *rating* guna memvalidasi kebermaknaan semantik dari prediktor yang digunakan.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian yang diperoleh dari tahapan pemodelan dan evaluasi sistem prediksi penurunan *rating* akomodasi wisata. Penyajian hasil dan pembahasan dilakukan secara terintegrasi, dimulai dari evaluasi model, analisis fitur, distribusi sentimen dan topik, eksplorasi temporal, hingga analisis kesalahan model, implikasi operasional, dan posisi penelitian dalam literatur.

A. Performa Model Prediktif Penurunan Rating

Evaluasi performa model dilakukan dengan membandingkan tiga algoritma *machine learning* menggunakan skema *hold-out validation* dengan proporsi data uji sebesar 20% atau 40 sampel dari total 200 sampel hasil agregasi temporal. Ringkasan hasil evaluasi ditampilkan pada Tabel 3.

TABEL 3
PERBANDINGAN PERFORMA ALGORITMA MACHINE LEARNING

Algoritma	Akurasi	Precision C0	Recall C0	F1 C0	Precision C1	Recall C1	F1 C1
Logistic Regression	62,5%	0,73	0,64	0,68	0,50	0,60	0,54
Decision Tree	57,5%	0,75	0,48	0,59	0,46	0,73	0,56
Random Forest	67,5%	0,75	0,72	0,73	0,56	0,60	0,58

Keterangan: C0 = periode tanpa penurunan *rating*; C1 = periode dengan penurunan *rating*. Evaluasi menggunakan stratified hold-out dengan proporsi 80% training (n=160) dan 20% testing (n=40).

Random Forest menunjukkan performa terbaik dengan akurasi sebesar 67,5% serta keseimbangan *precision-recall* yang paling stabil pada kedua kelas. Pemilihan *Random Forest* sebagai model optimal didasarkan pada beberapa pertimbangan. Pertama, sifat *ensemble* pada *Random Forest* memberikan tingkat *robustness* yang lebih tinggi terhadap *overfitting* dan varians data pada dataset berukuran menengah (n=200), karena mekanisme *bagging* dan *random feature selection* secara efektif mengurangi korelasi antarpohon keputusan. Kedua, keseimbangan *precision-recall* pada kelas C1 (*precision* 0,56 dan *recall* 0,60) lebih sesuai untuk konteks *early warning system* dibandingkan *Decision Tree* yang mengalami *trade-off* berupa *precision* yang sangat rendah (0,46). Ketiga, *Random Forest* menyediakan skor *feature importance* yang bersifat *interpretable* dan *actionable* bagi manajemen. Keempat, akurasi keseluruhan sebesar 67,5% secara signifikan melampaui *random baseline* (50%) dan *naive majority classifier* (52%).

Capaian akurasi sebesar 67,5% perlu diinterpretasikan dalam konteks kompleksitas tugas prediktif. Prediksi penurunan *rating* satu bulan ke depan bersifat *inherently more uncertain* karena melibatkan peramalan kejadian masa depan yang dipengaruhi oleh berbagai faktor eksternal di luar data ulasan, seperti fluktuasi permintaan musiman, tindakan kompetitor, peristiwa eksternal, serta perubahan operasional yang tidak terobservasi dalam dataset. Literatur pada domain *forecasting* dan *time-series prediction* menunjukkan bahwa rentang akurasi 60-75% merupakan tingkat kinerja yang kompetitif untuk tugas prediksi dengan *forecast horizon* satu bulan pada domain *hospitality* dan pariwisata [4], [5].

Potensi *overfitting* merupakan risiko yang perlu diperhatikan pada dataset berukuran 200 sampel. Meskipun *Random Forest* secara inheren lebih tahan terhadap *overfitting* dibandingkan *Decision Tree*, ketidaktersediaan *cross-validation* membatasi kemampuan untuk mengukur kesenjangan antara performa *training* dan *testing* secara menyeluruh. Potensi bias akibat data sintesis juga perlu diakui: karena data dibangun berdasarkan aturan yang telah ditentukan, hubungan antara fitur dan target mungkin lebih deterministik dibandingkan data riil, sehingga model mungkin memiliki performa yang lebih tinggi pada data sintesis daripada pada data riil (*distribution shift*). Hal ini merupakan keterbatasan fundamental penelitian ini yang berdampak pada generalisasi model.

B. Analisis Kontribusi Fitur Prediktif

Analisis *feature importance* menggunakan pendekatan mean decrease in impurity pada model *Random Forest* mengidentifikasi faktor-faktor utama yang memengaruhi prediksi penurunan *rating*. Tabel 4 menyajikan sepuluh fitur prediktif dengan skor *importance* tertinggi.

TABEL 4
SEPULUH FITUR PREDIKTIF DENGAN SKOR IMPORTANCE TERTINGGI

Rank	Fitur	Importance Score	Kategori	Referensi Pendukung
1	topic_prob_mean	0,1230	Topic Modeling	BERTopic [12]
2	rating_mean	0,1000	Rating Historis	Temporal analytics [1]
3	rating_min	0,0770	Rating Historis	Time series [4]
4	rating_std	0,0754	Rating Historis	Volatility analysis [5]
5	value_rating_mean	0,0713	Aspek Layanan	Hospitality analytics [2]
6	sentiment_score_mean	0,0689	Sentimen	Sentiment analysis [15]
7	sentiment_numeric_mean	0,0631	Sentimen	NLP hospitality [2]
8	rating_volatility	0,0615	Rating Historis	Forecasting [5]
9	sentiment_volatility	0,0589	Sentimen	Sentiment dynamics [1]

10	amenities	0,0337	Kategori Topik	BERTopic [12]
----	-----------	--------	----------------	---------------

Probabilitas topik rata-rata (*topic_prob_mean*) muncul sebagai prediktor terkuat dengan skor *importance* sebesar 12,3%, yang memvalidasi efektivitas *BERTopic* [12] dalam mendeteksi area permasalahan spesifik yang memiliki *predictive power* terhadap penurunan *rating* di masa depan. Temuan ini menunjukkan bahwa jenis atau tema permasalahan yang diungkapkan pelanggan dalam ulasan memiliki pengaruh yang lebih kuat terhadap risiko penurunan *rating* dibandingkan tingkat sentimen negatif secara umum.

Fitur *rating* historis mendominasi peringkat ke-2, ke-3, ke-4, dan ke-8, yang menunjukkan bahwa pola perubahan *rating* di masa lalu sangat prediktif terhadap perubahan di masa depan. Dominasi fitur *rating* historis ini konsisten dengan prinsip *momentum* dan *autocorrelation* dalam analisis *time series* [1], [4], [5]. Kemunculan metrik volatilitas, baik untuk *rating* maupun sentimen, mengungkapkan *insight* bahwa ketidakstabilan dalam pengalaman pelanggan merupakan *leading indicator* terhadap risiko penurunan.

Perlu dicatat bahwa pendekatan *feature importance* berbasis *mean decrease impurity* pada *Random Forest* memiliki keterbatasan dibandingkan metode *Explainable AI* yang lebih modern. Dibandingkan dengan *SHAP* (*SHapley Additive exPlanations*) [6], pendekatan ini tidak menghasilkan nilai kontribusi fitur per sampel (*instance-level explanation*) dan dapat mengalami bias terhadap fitur dengan kardinalitas tinggi. Dibandingkan dengan *LIME* (*Local Interpretable Model-agnostic Explanations*), pendekatan ini bersifat global dan tidak menangkap variasi interpretabilitas antarinstansi. Kelebihan pendekatan yang digunakan adalah: (1) interpretasi yang intuitif dan mudah dipahami oleh pemangku kepentingan non-teknis; (2) tidak memerlukan komputasi tambahan di luar proses pelatihan; dan (3) konsisten dengan kebutuhan infrastruktur minimal yang menjadi prioritas pada segmen UMKM. Integrasi *SHAP* atau *LIME* direkomendasikan sebagai agenda penelitian lanjutan untuk meningkatkan kualitas interpretabilitas sistem.

C. Distribusi Sentimen dan Identifikasi Pola Risiko

Analisis distribusi sentimen menghasilkan 418 ulasan positif (56,6%) dan 321 ulasan negatif (43,4%). Distribusi sentimen ini menunjukkan keterkaitan yang signifikan dengan probabilitas penurunan *rating* sebagaimana disajikan pada Tabel 5. Berdasarkan hasil analisis, periode dengan dominasi sentimen negatif memiliki probabilitas penurunan *rating* sebesar 68,6%, atau sekitar 2,4 kali lebih tinggi dibandingkan periode dengan dominasi sentimen positif (28,7%). Korelasi yang kuat ini mengindikasikan bahwa sentimen agregat dari ulasan dalam suatu periode berfungsi sebagai *leading indicator* yang andal terhadap perubahan *rating* pada periode berikutnya, sebagaimana dilaporkan dalam penelitian terdahulu pada domain analisis ulasan hotel [15].

TABEL 5
KORELASI DISTRIBUSI SENTIMEN DENGAN PENURUNAN *RATING*

Kategori Sentimen	Jumlah Periode	Persentase Penurunan <i>Rating</i>	Rata-rata <i>Rating</i>
Dominan Positif (>70% positif)	87	28,7%	4,21
Seimbang (40–70% positif)	78	42,3%	3,45
Dominan Negatif (>60% negatif)	35	68,6%	2,73

Temuan menarik adalah peran periode dengan distribusi sentimen yang relatif seimbang (proporsi sentimen positif 40–70%). Periode tersebut menunjukkan tingkat risiko menengah dengan probabilitas penurunan *rating* sebesar 42,3%, yang dapat diinterpretasikan sebagai *early warning zone*. Dari perspektif manajemen reputasi, periode dengan sentimen seimbang ini merepresentasikan *window of opportunity* yang optimal untuk melakukan intervensi preventif.

D. Analisis Granularitas Temporal dan Distribusi Topik

Dataset ulasan yang terdiri atas 739 ulasan diagregasi menggunakan beberapa skema temporal alternatif untuk mengevaluasi dampaknya (*impact*) terhadap performa model serta kelayakan sistem sebagai *early warning tool*. Hasil evaluasi komparatif dari berbagai skema agregasi temporal tersebut disajikan pada Tabel 6.

TABEL 6
DAMPAK GRANULARITAS TEMPORAL TERHADAP PERFORMA SISTEM

Skema Agregasi	Jumlah Sampel	Akurasi (RF)	Data Sparse	Evaluasi
Mingguan (W)	365	58,3%	72%	Tidak stabil
Bulanan (M)	200	67,5%	18%	Optimal
Triwulanan (Q)	48	61,2%	8%	Kurang responsif
Semesteran (6M)	24	54,7%	0%	Terlalu lambat

Agregasi bulanan menghasilkan keseimbangan optimal antara stabilitas statistik (18% data sparse), kecukupan ukuran sampel ($n=200$), responsivitas sistem (peringatan satu bulan ke depan), dan performa model (akurasi 67,5%). Agregasi mingguan menghasilkan jumlah sampel yang lebih besar namun menghadapi masalah data sparsity yang tinggi (72%), sehingga menyebabkan ketidakstabilan prediksi. Sebaliknya, agregasi triwulanan dan semesteran menghasilkan performa yang lebih rendah karena kehilangan responsivitas sebagai sistem peringatan dini.

E. Identifikasi Topik Kritis dan Pola Risiko

Pemodelan topik menggunakan BERTopic [12] berhasil mengidentifikasi delapan kluster topik utama dari korpus 739 ulasan, dengan tingkat korelasi yang bervariasi terhadap kejadian penurunan *rating*. Tabel 7 menyajikan hasil analisis korelasi antara topik keluhan dan pola risiko (risk patterns).

TABEL 7
KORELASI TOPIK DENGAN RISIKO PENURUNAN *RATING*

Topik (Kata Kunci)	Jumlah Ulasan	Kategori	Rating Rata-rata	Korelasi
mosquito, dusk, net	20	Kebersihan	2,35	0,78 (Sangat Kuat)
water, shower, hot	34	Fasilitas	2,56	0,72 (Sangat Kuat)
room, noise, housekeeping	74	Kamar	2,94	0,67 (Kuat)
bed, mattress, hard	15	Kualitas Tidur	2,87	0,58 (Sedang–Kuat)
checkout, wifi, slow	69	Akses	3,21	0,45 (Sedang)
staff, breakfast, friendly	290	Layanan	4,12	0,23 (Lemah)
library, book, quiet	19	Amenitas	4,45	-0,12 (Negatif)

Analisis topik mengungkapkan beberapa *critical insights*. Pertama, fenomena *high-impact low-frequency issues*: permasalahan nyamuk yang hanya muncul pada 20 ulasan (2,7% dari total) justru memiliki korelasi terkuat dengan penurunan *rating* ($r = 0,78$), yang secara signifikan melampaui representasi numeriknya. Hal ini menunjukkan bahwa tingkat keparahan (*severity*) dari jenis keluhan tertentu lebih berpengaruh dibandingkan frekuensinya. Kedua, empat dari lima topik dengan korelasi tertinggi berkaitan dengan fasilitas dasar dan kenyamanan (kebersihan, air panas, kebisingan, kualitas kasur), yang menunjukkan bahwa fondasi layanan *hospitality* tetap menjadi faktor utama dalam membentuk kepuasan pelanggan. Ketiga, fenomena *service paradox*: topik layanan staf dan sarapan dengan frekuensi tertinggi (290 ulasan) menunjukkan korelasi yang relatif lemah terhadap penurunan *rating* ($r = 0,23$). Keempat, topik perpustakaan dan lokasi menunjukkan korelasi negatif, berfungsi sebagai *positive differentiators* yang dapat mengurangi risiko penurunan *rating*.

F. Analisis Kesalahan Model (Error Analysis)

Analisis kesalahan model dilakukan terhadap *Random Forest* sebagai model terbaik untuk memahami pola *false positive* (FP) dan *false negative* (FN) serta implikasi operasionalnya. *False Positive (FP)*: Kasus *false positive* terjadi ketika model memprediksi penurunan *rating* padahal penurunan tidak benar-benar terjadi. Berdasarkan nilai *precision* C1 sebesar 0,56, diperkirakan sekitar 44% dari prediksi penurunan merupakan *false alarms*. Analisis menunjukkan bahwa FP cenderung terjadi pada periode dengan sentimen yang relatif seimbang (40-60% positif) namun disertai volatilitas *rating* yang tinggi, situasi yang mengindikasikan ketidakstabilan sementara (*transient instability*) namun tidak selalu berujung pada penurunan *rating*. FP juga lebih sering terjadi pada awal periode pemulihan setelah penurunan sebelumnya, ketika fitur temporal masih mencerminkan kondisi yang buruk namun tren mulai membaik. Dampak operasional dari FP adalah alokasi sumber daya untuk tindakan preventif yang tidak diperlukan; namun dalam konteks *early warning system*, biaya FP umumnya lebih rendah dibandingkan biaya *false negative* berupa penurunan *rating* yang tidak terdeteksi.

False Negative (FN): Kasus *false negative* terjadi ketika model gagal mendeteksi penurunan *rating* yang sebenarnya terjadi. Berdasarkan nilai *recall* C1 sebesar 0,60, sekitar 40% kejadian penurunan tidak terdeteksi oleh model. Analisis menunjukkan bahwa FN cenderung terjadi pada penurunan yang disebabkan oleh faktor eksternal mendadak (*sudden external events*) yang tidak tercermin dalam pola ulasan sebelumnya, seperti perubahan manajemen tiba-tiba, kejadian insidental (kebakaran kecil, pencurian), atau penurunan kualitas yang sangat cepat dalam satu bulan. FN juga terjadi pada properti dengan baseline *rating* yang relatif tinggi namun mengalami penurunan kecil, yang secara statistik sulit dibedakan dari variasi normal. Dampak operasional dari FN lebih serius karena penurunan *rating* yang tidak terdeteksi memberikan *lead time* nol bagi manajemen untuk melakukan intervensi preventif. Keterbatasan ini menegaskan bahwa sistem harus digunakan sebagai *decision support tool* yang melengkapi pengawasan manajemen, bukan menggantikannya.

Potensi *distribution shift* ketika sistem diterapkan pada data riil merupakan risiko penting yang perlu diantisipasi. Karena model dilatih pada data sintesis yang dibangun berdasarkan aturan yang relatif deterministik, performa model pada data riil kemungkinan lebih rendah akibat: (1) variasi linguistik yang lebih kaya dan tidak terprediksi; (2) pola perilaku pelanggan

yang lebih kompleks; dan (3) kejadian anomali (*review bombing*, ulasan palsu) yang tidak direpresentasikan dalam data sintetis. Keterbatasan generalisasi model ini merupakan alasan utama mengapa penelitian ini diposisikan sebagai *proof-of-concept* metodologis, bukan sebagai bukti final efektivitas sistem.

G. Analisis Komparatif dengan Penelitian Terdahulu

Perbandingan dengan penelitian terdahulu menunjukkan bahwa penelitian ini menempati posisi yang *distinct* dan *complementary* dalam spektrum analisis ulasan hotel. Berbeda dengan studi Putra *et al.* [2] yang berfokus pada klasifikasi sentimen retrospektif dengan akurasi tinggi (85-95%), dan Lee *et al.* [1] yang menekankan analisis temporal deskriptif, penelitian ini secara eksplisit memodelkan risiko penurunan *rating* di masa depan sebagai tugas peramalan (*forecasting task*). Akurasi 67,5% yang diperoleh tidak dapat dibandingkan secara langsung dengan akurasi tinggi pada tugas klasifikasi sentimen karena perbedaan definisi tugas yang fundamental. Dalam konteks literatur *forecasting* pada domain *hospitality*, akurasi 60-75% untuk *forecast horizon* satu bulan dianggap kompetitif [4], [5].

Kontribusi utama yang membedakan penelitian ini dari literatur yang ada meliputi: (1) integrasi tiga modalitas fitur (sentimen, topik, temporal) dalam satu *pipeline* prediktif yang *reproducible*; (2) demonstrasi bahwa probabilitas topik memiliki daya prediktif lebih kuat dibandingkan sentimen agregat (12,3% vs. 6,89% *importance score*), temuan yang belum banyak dilaporkan dalam literatur *hospitality analytics*; dan (3) pengusulan protokol validasi data sintetis multi-dimensi yang dapat diadopsi oleh penelitian lain pada domain dengan keterbatasan akses data.

H. Implikasi Operasional

Hasil model memiliki implikasi operasional yang konkret bagi pengelola hotel. Sistem peringatan dini yang dikembangkan dirancang untuk digunakan dalam siklus manajemen bulanan sebagai berikut. Pada awal setiap bulan, manajer properti membuka *dashboard* untuk memantau probabilitas risiko bulan berjalan berdasarkan ulasan bulan sebelumnya. Apabila probabilitas risiko berada pada kategori rendah ($< 0,3$), tidak diperlukan tindakan khusus selain pemantauan rutin. Apabila berada pada kategori sedang ($0,3-0,6$), manajer disarankan untuk melakukan tinjauan mendalam terhadap topik keluhan yang muncul dan mempersiapkan tindakan preventif. Apabila berada pada kategori tinggi ($\geq 0,6$), manajer perlu mengambil tindakan segera yang disesuaikan dengan topik keluhan dominan.

Nilai praktis sistem ini khususnya relevan bagi segmen UMKM *hospitality* di Indonesia yang umumnya tidak memiliki tim analitik data khusus. Sistem ini menyediakan alat berbasis data yang mudah diakses melalui antarmuka *dashboard* tanpa memerlukan keahlian teknis mendalam, sehingga mendukung transisi dari pengelolaan reputasi yang bersifat reaktif menuju pendekatan proaktif berbasis data.

Perlu ditegaskan bahwa hasil model harus diinterpretasikan sebagai *decision support*, bukan sebagai keputusan otomatis. Faktor kontekstual yang tidak tertangkap oleh data ulasan, seperti kejadian lokal, perubahan manajemen, atau kondisi pasar, harus tetap dipertimbangkan oleh manajer dalam pengambilan keputusan akhir.

I. Limitasi dan Agenda Riset Mendatang

Penelitian ini memiliki beberapa keterbatasan yang perlu dipertimbangkan. Pertama, penggunaan data sintetis secara inheren memiliki keterbatasan dalam merepresentasikan kompleksitas penuh dinamika ulasan pelanggan, termasuk nuansa linguistik, sarkasme, idiom lokal, pola anomali seperti *review bombing*, dan kejadian temporal yang sangat kontekstual. Validasi lanjutan menggunakan data riil dari platform ulasan komersial sangat direkomendasikan. Kedua, tidak dilakukannya *cross-validation*, *hyperparameter tuning* sistematis, *ROC-AUC*, *confidence interval*, dan uji signifikansi statistik membatasi kemampuan untuk menyimpulkan perbedaan performa antaralgoritma secara definitif. Ketiga, cakupan geografis masih terbatas pada simulasi tujuh properti yang direfleksikan dari dataset *Bali Hotel Reviews*. Keempat, penggunaan LLM dalam pembangkitan data sintetis tidak dilakukan; pendekatan *template-based* yang digunakan mungkin kurang realistis dibandingkan pendekatan berbasis model generatif.

Agenda riset mendatang mencakup: (1) validasi kerangka kerja pada data riil dari platform ulasan komersial; (2) penerapan *k-fold cross-validation* dan *hyperparameter tuning* sistematis untuk estimasi performa yang lebih robust; (3) integrasi *SHAP* atau *LIME* untuk meningkatkan interpretabilitas; (4) eksplorasi pembangkitan data sintetis berbasis LLM untuk meningkatkan realisme data; (5) perluasan cakupan ke berbagai segmen akomodasi dan geografis; serta (6) pengembangan *recommendation engine* yang lebih granular.

IV. SIMPULAN

Penelitian ini berhasil mengembangkan dan memvalidasi kerangka kerja *proof-of-concept* metodologis untuk sistem peringatan dini penurunan *rating* akomodasi wisata melalui integrasi analisis sentimen berbasis *Bidirectional Encoder Representations from Transformers* (BERT), pemodelan topik menggunakan *BERTopic*, serta agregasi temporal bulanan dalam sebuah *pipeline machine learning* yang terpadu. Penting untuk ditegaskan bahwa penelitian ini merupakan *proof-of-concept metodologis* yang menunjukkan kelayakan pendekatan prediktif, bukan bukti final efektivitas sistem pada lingkungan

operasional nyata. Seluruh eksperimen dilakukan pada data sintetis yang tervalidasi secara statistik, sehingga performa pada data riil perlu diverifikasi melalui penelitian lanjutan.

Kontribusi utama penelitian ini meliputi: (1) pengembangan kerangka kerja end-to-end yang *reproducible* untuk prediksi penurunan *rating* akomodasi, mengintegrasikan tiga modalitas fitur dalam satu *pipeline*; (2) pembuktian bahwa probabilitas topik memiliki daya prediktif lebih kuat dibandingkan sentimen agregat, dengan *skor importance* tertinggi sebesar 12,3%; (3) pengusulan protokol validasi data sintetis multi-dimensi yang *reproducible*; serta (4) implementasi sistem pendukung keputusan yang accessible bagi UMKM sektor hospitality dengan kebutuhan infrastruktur minimal.

Keterbatasan utama penelitian meliputi penggunaan data sintetis yang membatasi validitas eksternal, tidak dilakukannya *cross-validation* dan *hyperparameter tuning* sistematis, serta cakupan yang terbatas pada tujuh properti simulasi. Agenda penelitian lanjutan mencakup validasi pada data riil, penerapan evaluasi statistik yang lebih komprehensif, integrasi *Explainable AI* modern seperti SHAP dan LIME, serta eksplorasi pembangkitan data sintetis berbasis LLM.

Dengan demikian, penelitian ini memosisikan diri sebagai *methodological blueprint* yang menyediakan *validated starting point* bagi praktisi dan peneliti dalam mengembangkan sistem peringatan dini prediktif, sekaligus mendukung transformasi dari pengelolaan reputasi yang bersifat reaktif menuju pendekatan proaktif berbasis data di industri *hospitalitas* Indonesia, khususnya bagi segmen UMKM.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Amikom Yogyakarta yang telah memberikan dukungan dan fasilitas dalam pelaksanaan penelitian ini. Ucapan terima kasih khusus disampaikan kepada Yudinda Gilang Pramudya dan Andry Alamsyah yang telah mempublikasikan dataset Bali Hotel Reviews pada Mendeley Data sebagai open access resource. Penulis juga mengucapkan terima kasih kepada para reviewer dan editor Jurnal Teknik Informatika dan Sistem Informasi (JuTISI) atas masukan konstruktif yang meningkatkan kualitas naskah ini, serta kepada komunitas open-source developers dari library BERT, BERTopic, Streamlit, dan Plotly.

DAFTAR PUSTAKA

- [1] P.-H. Lee, Y.-K. Sun, Y.-P. Ke, and P.-J. Lee, "Hotel *rating* prediction system based on time factors: Using reviews and sentiment analysis," *Journal of Organizational and End User Computing*, vol. 36, no. 1, pp. 1-29, 2024.
- [2] P. P. Putra, M. K. Anam, A. S. Chan, A. Hadi, N. Hendri, and A. Masnur, "Optimizing sentiment analysis on imbalanced hotel review data using SMOTE and ensemble machine learning techniques," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 936-951, 2025.
- [3] Badan Pusat Statistik, *Statistik Indonesia 2024*. Jakarta, Indonesia: Badan Pusat Statistik, 2024. [Online]. Available: <https://www.bps.go.id/id/publication/2024/02/28/c1bacde03256343b2bf769b0/statistik-indonesia-2024.html>.
- [4] I. G. A. A. P. Dewi and I. Listiowarni, "Forecasting hotel occupancy rate using ARIMA model," *Journal of Tourism and Hospitality Management*, vol. 8, no. 1, pp. 45-56, 2020.
- [5] R. Law, G. Li, D. K. Fong, and X. Han, "Tourism demand forecasting: A deep learning approach," *Annals of Tourism Research*, vol. 75, pp. 410-423, 2019.
- [6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, 2017, pp. 4768-4777.
- [7] K. Herianto, B. Kurniawan, Z. H. Hartomi, Y. Irawan, and M. K. Anam, "Machine learning algorithm optimization using stacking technique for graduation prediction," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1272-1285, 2024.
- [8] M. K. Anam, M. B. Firdaus, F. Suandi, Lathifah, T. Nasution, and S. Fadly, "Performance improvement of machine learning algorithm using ensemble method on text mining," in *Proc. Int. Conf. Future Technologies for Smart Society*, Yogyakarta, Indonesia, 2024, pp. 90-95.
- [9] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [10] J. Wei and K. Zou, "EDA: Easy data augmentation techniques for boosting performance on text classification tasks," in *Proc. 2019 Conf. Empirical Methods Natural Language Processing (EMNLP)*, Hong Kong, China, 2019, pp. 6382-6388.
- [11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. North American Chapter Assoc. Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, MN, USA, 2019, pp. 4171-4186.
- [12] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," *arXiv:2203.05794 [cs.CL]*, Mar. 2022. [Online]. Available: <https://doi.org/10.48550/arXiv.2203.05794>.
- [13] Y. G. Pramudya and A. Alamsyah, "Bali Hotel Reviews Dataset (Version 2)," Mendeley Data, 2023. [Online]. Available: <https://data.mendeley.com/datasets/s62ycm698z/2>.
- [14] P. Ganesan, S. K. Jagathesaperumal, I. Gobhinath, V. Venkatraman, S. N. Gaftandzhieva, and R. Z. Doneva, "Deep learning-based interactive dashboard for enhancing online classroom experience through student emotion analysis," *IEEE Access*, vol. 12, pp. 91140-91153, 2024.
- [15] A. F. Hidayatullah and M. R. Ma'arif, "Sentiment analysis of hotel reviews using Naive Bayes classifier," *Jurnal EECCIS*, vol. 11, no. 2, pp. 103-108, 2017.