Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 1 January 2026, pp: 193-204
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Machine Learning Evaluation for Hotel Cancellation Prediction with Threshold Adjustment and Cost-Based Evaluation

Evaluasi Machine Learning untuk Prediksi Pembatalan Hotel dengan Threshold Adjustment dan Cost-Based Evaluation

Efraim William Solang[1*], Franco Xander Adu[2], Agus Dharma[3], Nyoman Gunantara[4]

[1,2,3,4]Program Studi Magister Teknik Elektro, Universitas Udayana, Indonesia

E-Mail: [1]efraimwilliam@gmail.com, [2]francoxander1@gmail.com, [3]agus_dharma@unud.ac.id, [4]gunantara@unud.ac.id

Received Nov 22th 2025; Revised Dec 13th 2025; Accepted Dec 31th 2025; Available Online Jan 17th 2026
Corresponding Author: Efraim William Solang

Abstract

Hotel booking cancellations are a crucial issue that directly impacts revenue and operational planning. This study evaluates the application of threshold adjustment and cost-based evaluation to improve the quality of business decision-making. The research involves comparing several types of machine learning models using the hotel booking demand dataset. Model performance is assessed using the F0.5-Score, precision, ROC AUC, and a net revenue-based cost-based evaluation approach. The results show that Random Forest provides the best performance with an F0.5-Score of 0.8279, precision of 0.878, and ROC AUC of 0.9165. Other models such as Logistic Regression (baseline) with an F0.5-Score of 0.7816, XGBoost with an F0.5-Score of 0.8108, and ANN with an F0.5-Score of 0.8091 show relatively lower performance, indicating that this dataset is more suitable for using an ensemble learning approach. A key finding revealed that threshold adjustments based on the F0.5-Score did not always yield maximum economic benefits. The use of the threshold (0.52) was shown to produce higher net revenue values than the optimal threshold based on the F0.5-Score. This approach is expected to improve the quality of business decision-making for hotel managers in managing financial risk.

Keywords: Cost-Based Evaluation, Hotel Booking Cancellation, Hotel Booking Demand, Machine Learning

Abstrak

Pembatalan pemesanan hotel merupakan permasalahan krusial yang berdampak langsung pada pendapatan dan perencanaan operasional. Penelitian ini mengevaluasi penerapan *threshold adjustment* dan *cost-based evaluation* untuk meningkatkan kualitas pengambilan keputusan bisnis. Penelitian ini melibatkan perbandingan beberapa jenis model *machine learning* menggunakan *dataset hotel booking demand*. Kinerja model dinilai menggunakan metrik *F0.5-Score, precision,* ROC AUC, dan pendekatan *cost-based evaluation* berbasis *net revenue*. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan kinerja terbaik dengan *F0.5-Score* 0.8279, *precision* 0.878 dan ROC AUC 0.9165. Model lain seperti *Logistic Regression* (*baseline*) dengan *F0.5-Score* 0.7816, XGBoost dengan *F-.5-Score* 0.8108 dan ANN dengan *F0.5-Score* 0.8091 menunjukan performa lebih relatif lebih rendah, mengindikasikan bahwa *dataset* ini lebih cocok menggunakan pendekatan *ensamble learning*. Temuan penting mengungkapkan bahwa penyesuaian *threshold* berdasarkan *F0.5-Score* tidak selalu menghasilkan keuntungan ekonomi maksimum. Penggunaan *threshold* (0.52) terbukti menghasilkan nilai *net revenue* lebih tinggi dibandingkan *threshold* optimal berbasis F0.5-*Score*. Pendekatan ini diharapkan dapat meningkatkan kualitas pengambilan keputusan bisnis bagi manajer hotel dalam pengelolaan risiko finansial.

Kata Kunci: *Cost-Based Evaluation, Hotel Booking Demand, Machine Learning*, Pembatalan Pemesanan Hotel

1. PENDAHULUAN

Industri perhotelan merupakan bisnis yang sangat kompetitif dan dinamis, menuntut setiap pelaku usaha untuk secara konstan mengoptimalkan strategi manajemen pendapatan (*revenue management*) dan operasional. Dalam upaya meningkatkan pendapatan dan tingkat hunian (*occupancy rate*), banyak hotel

menerapkan strategi seperti *forecasting* dan *overbooking* [1]. Secara teoritis, metode *overbooking* merupakan sebuah metode yang sangat berisiko dan harus dirancang dengan matang-matang. Ketika realisasi tamu datang melebihi ekspektasi, maka akan terjadinya *denied service* (tamu tidak mendapat kamar) yang memicu biaya kompensasi, relokasi, beban operasional serta yang paling penting adalah ketidaknyamanan customer dan risiko reputasi hotel [2].

Overbooking umumnya diterapkan pada *revenue management* untuk mengantisipasi pembatalan dan *no-show* tamu. Namun, strategi ini sangat bergantung pada ketepatan prediksi pembatalan reservasi. Kesalahan prediksi, khususnya kesalahan ketika reservasi diprediksi akan dibatalkan tetapi tamu justru datang (*false positive*) dapat menyebabkan terjadinya *denied service*. Situasi ini memaksa hotel untuk memberikan kompensasi, melakukan relokasi tamu, atau bahkan menerapkan *overcompensation* guna menjaga kepuasan pelanggan, yang pada akhirnya meningkatkan beban biaya dan berisiko merusak reputasi hotel [2], [3].

Permasalahan utama yang dihadapi hotel bukan semata-mata pada penerapan *overcompensation*, melainkan pada bagaimana meminimalkan terjadinya kondisi yang memicu biaya tersebut, yaitu kesalahan prediksi pembatalan. Oleh karena itu, kemampuan dalam memprediksi pembatalan reservasi secara akurat dengan mempertimbangkan konsekuensi finansial dari setiap jenis kesalahan prediksi menjadi aspek krusial dalam pengambilan keputusan *overbooking*. Dalam konteks ini, pemanfaatan model *machine learning* tidak hanya dituntut untuk memiliki kinerja prediktif yang baik, tetapi juga mampu mendukung keputusan yang berdampak positif terhadap profitabilitas hotel secara keseluruhan.

Dalam beberapa tahun terakhir, sejumlah penelitian telah menggunakan metode *machine learning* pada *hotel booking demand dataset* untuk memprediksi pembatalan reservasi, yang umumnya berfokus pada peningkatan akurasi model dengan membandingkan algoritma model seperti *Logistic Regression*, *K-Nearest Neighbors* (KNN), *Decision Tree*, *Random Forest*, XGBoost, LightGBM, dan *Artificial Neural Network* (ANN) [4], [5], [6]. Model-model tersebut di pilih karena mewakili berbagai karakteristik algoritma yang berbeda-beda. *Logistic Regression* digunakan sebagai *baseline model linear* yang sederhana dan mudah diinterpretasikan, KNN dan *Decision Tree* digunakan untuk menangkap hubungan nonlinier sederhana, sementara model *ensemble* seperti *Random Forest*, XGBoost dan LightGBM dikenal memiliki performa yang baik pada data tabular dengan interaksi fitur yang kompleks dan heterogen [7]. Di sisi lain, ANN digunakan untuk mengevaluasi kemampuan model berbasis *neural network* dalam menangkap pola non-linier yang lebih kompleks pada data reservasi hotel. Namun demikian, penelitian-penelitian tersebut sejauh ini masih berfokus pada metrik klasifikasi konvensional seperti *accuracy*, AUC, atau *F1-Score*, tanpa mengaitkannya dengan dampak finansial nyata dari kesalahan prediksi. Selain itu, perhatian terhadap kesalahan *false positive* yang berpotensi memicu *overbooking* serta aspek explainability model masih relatif terbatas.

Berbeda dari penelitian sebelumnya, penelitian ini berupaya mengembangkan keranga model prediksi pembatalan berbasis *machine learning* yang berorientasi pada profit hotel dan risiko *overbooking*. Pendekatan ini mengkombinasikan optimasi ambang batas kepuasan (*threshold adjustment*) yang menekankan pada peningkatan *precision* guna menimalkan kejadian *false positve*, pemodelan dampak finansial untuk menghitung potensi kerugian dan keuntungan dari hasil prediksi, serta integrasi *Explainable AI* (XAI) menggunakan *feature importance* untuk memberikan transparansi atas faktor-faktor yang memengaruhi hasil prediksi. Dengan demikian, penelitian ini tidak hanya menawarkan kontribusi akademik melalui pendekatan evaluasi berbasis profit, tetapi juga memberikan manfaat praktis sebagai dasar *decision-support* system bagi manajer hotel dalam merancang strategi *overbooking* yang lebih efektif dan berkelanjutan.

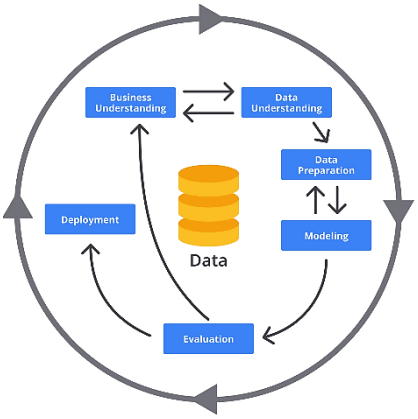
2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode *machine learning classification* untuk mengembangkan model prediksi pembatalan pemesanan hotel dengan mempertimbangkan implikasi finansial terhadap risiko *overbooking*. Pendekatan yang digunakan adalah *binary classification* dengan variabel target *is_canceled*, di mana nilai 1 menunjukkan pembatalan dan nilai 0 menunjukkan tamu hadir. Penelitian ini bertujuan untuk mengembangkan model prediktif, mengevaluasi performa model melalui *threshold adjustment* dan *cost-based evaluation*, serta meningkatkan transparansi model menggunakan pendekatan *Explainable* AI (XAI). Dengan mempertimbangkan konteks *revenue management*, evaluasi tidak hanya dilakukan menggunakan metrik statistik, tetapi juga melalui perhitungan dampak finansial, sehingga model yang dihasilkan dapat mendukung pengambilan keputusan operasional hotel.

2.1. Proses Penelitian dan Kerangka Konsepsual

Proses penelitian ini disusun menggunakan gabungan kerangka *Cross-Industry Standard Prosses for Data Mining* (CRISP-DM) dan *cost-based revenue management*. Penggunaan kerangka *hybrid* dilakukan

untuk memastikan bahwa model prediksi tidak hanya memiliki kinerja yang baik secara statistik, tetapi juga mengedepankan keuntungan yang dapat diperoleh oleh hotel [8].Gambar 1 merupakan kerangka CRISP-DM.



Gambar 1. Kerangka CRISP-DM

2.2. Pemahaman Bisnis

Dalam industri perhotelan, *overbooking* merupakan strategi yang umum diterapkan untuk meningkatkan *occupancy* dan pendapatan dengan mengantisipasi adanya pembatalan reservasi dan tamu tidak hadir. Namun, penerapan *overbooking* membawa risiko finansial dan nama baik perusahaan apabila prediksi pembatalan tidak akurat. Kesalahan prediksi dengan menerapkan *overbooking* (*false positive*) dapat menimbulkan *denied service* bagi tamu yang ingin menginap dan hotel harus memberikan kompensasi atas hal tersebut. Disisi lain, kesalahan prediksi pada tamu yang akan hadir padahal reservasi dibatalkan (*false negative*) mengakibatkan kamar kosong dan hilangnya potensial pendapatan. Oleh karena itu, tujuan bisnis dari penelitian ini adalah membangun sistem prediktif yang mampu menimalkan risiko akibat *overbooking* dan memanfaatkan model prediktif pada potensi pembatalan reservasi.

2.3. Pemahaman Data

Data yang digunakan dalam penelitian ini berasal dari *hotel booking demand dataset* yang berasal dari Negara Portugal dari tanggal 1 Juli 2015 hingga 31 Agustus 2017. Setiap satu data atau row mewakili satu pemesanan hotel [4]. Dataset ini memuat berbagai variable yang menggambarkan karakteristik pemesanan, profil tamu, serta informasi operasional hotel. Variabel-variabel tersebut meliputi jenis hotel, waktu pemesanan (*lead time*), informasi tanggal kedatangan, durasi menginap pada hari kerja dan akhir pekan, jumlah tamu berdasarkan kategori usia (dewasa, anak, dan bayi), paket makanan, serta segmen pasar dan saluran distribusi pemesanan. Selain itu, dataset juga mencakup riwayat pemesanan tamu, seperti status tamu berulang, jumlah pembatalan sebelumnya, perubahan pemesanan, tipe deposit, dan jenis kamar yang dipesan maupun yang diberikan saat *check-in*. Informasi finansial dan operasional turut disertakan melalui variabel *average daily rate* (ADR), jumlah ruang parkir yang diminta, serta jumlah permintaan khusus dari tamu.

Variabel target yang digunakan adalah *is_canceled*, dengan nilai 1 menunjukan bahwa reservasi dibatalkan, dan 0 menunjukan bahwa tamu hadir. Pemilihan dataset ini didasarkan pada relevansinya dengan permasalahan *overbooking* serta ketersediaan variabel yang memungkinkan pemodelan risiko pembatalan secara kuantitatif. Seluruh variabel akan diolah melalui proses *data cleansing*, *transformation*, dan *feature engineering*.

2.4. Data Preparation

Tahap ini dilakukan untuk memastikan datarelevan sebelum digunakan dalam proses pemodelan. Proses data diawali dengan *data cleansing*, yang mecakup penyesuaian tipe data, penganan *missing values* dan *sanity check* untuk mastikan konsistensi dan validitas data. Proses selanjutnya adalah pengananan *outlier* data pada variabel ADR yang memiliki nilai ekstrim dan berpotensi menggangu stabilitas model.

Setelah data dibersihkan, dilakukan proses *feature engineering* untuk meningkatkan daya prediktif model dengan menambahkan fitur-fitur baru. Beberapa fitur yang dikembangkan antara lain *binning* pada variabel *lead_time* untuk mengelompokkan rentang waktu pemesanan, pembuatan fitur musiman berdasarkan tanggal kedatangan untuk menggambarkan pola permintaan musiman, serta fitur kategori “musim hotel” (*high season* dan *low season*) untuk menangkap perbedaan pola permintaan antara periode puncak dan rendah. Selain itu, dilakukan *binning* pada variabel *days_in_waiting_list* untuk membedakan tingkat urgensi dan ketidakpastian pemesanan, serta pembuatan fitur regional berdasarkan variabel country dengan mengelompokkan negara ke dalam beberapa wilayah geografis guna mengurangi sparsitas data kategorikal dan mempermudah interpretasi model.

Tahap data *preparation* diakhiri dengan melakukan *feature selection* yang akan mengevaluasi pentingnya atau relevansi setiap karakteristik data dengan variable target. Pemilihan fitur merupakan tahap yang diperlukan untuk mencapai model klasifikasi yang kuat dan efisien pada *training model* [6].

Tahap terakhir adalah pembagian data menggunakan *sklearn.model_selection.train_test_split* sebesar 80% data *training* dan 20% data *testing* [9]. Seluruh langkah pada tahap ini memastikan bahwa data yang digunakan pada proses pemodelan berkualitas, terstruktur, dan memiliki representasi fitur yang lebih informatif terhadap pola pembatalan.

2.5. Pemodelan

Tahap pemodelan dilakukan untuk membangun dan mengevaluasi berbagai algoritma *machine learning* dalam memprediksi pembatalan pemesanan hotel. Tujuan utama tahap ini adalah untuk menentukan model terbaik dan terpercaya pada data *training* dan *test*. Proses pemodelan diawali dengan membandingkan beberapa model klasifikasi. Perbandingan model dilakukan pada algoritma *Logistic Regression* (*baseline*), KNN, *Decision Tree*, *Random Forest*, XGBoost, dan LightGBM dan ANN.

2.5.1 *Logistic Regression*

Logistic Regression digunakan sebagai *baseline model* dalam penelitian ini karena kesederhanaan, efisiensi komputasi, serta kemampuannya dalam memberikan interpretasi yang jelas terhadap hubungan antara variabel prediktor dan probabilitas pembatalan reservasi. *Logistic Regression* memodelkan probabilitas kelas menggunakan fungsi logistik (sigmoid), sehingga cocok untuk permasalahan klasifikasi biner. Secara umum, model *Logistic Regression* dapat dinyatakan pada Persamaan 1 [10].

$$p(G = k \mid \mathrm{X} = \mathrm{x}) = \frac{1}{1+(\exp-(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k))} \quad (1)$$

Meskipun memiliki keterbatasan dalam menangkap hubungan non-linier, *Logistic Regression* sering digunakan sebagai pembanding awal dalam penelitian prediksi pembatalan hotel [5], [8], [11].

2.5.2 *K-Nearest Neighbors* (KNN)

KNN merupakan algoritma berbasis instance yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur. Kedekatan antar data umumnya diukur menggunakan jarak *Euclidean*. Persamaan jarak *Euclidean* dinyatakan pada Persamaan 2 [12].

$$d(x_1 x_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (2)$$

KNN mampu menangkap pola nonlinier sederhana, namun sensitif terhadap skala fitur dan memiliki kompleksitas komputasi yang tinggi pada dataset berukuran besar.

2.5.3 *Decision Tree*

Decision Tree merupakan algoritma klasifikasi yang membagi data ke dalam struktur pohon berdasarkan aturan pemisahan (*splitting rules*) yang memaksimalkan homogenitas kelas. Algoritma ini memproses dataset berlabel untuk menghasilkan pohon keputusan yang divalidasi guna menghitung kemampuan generalisasinya [13]. Persamaan *entropy* dirumuskan pada Persamaan 3.

$$Entropy(S) = \sum_{i=1}^{n} - \pi \times log_2 \pi \quad (3)$$

Untuk menentukan atribut mana yang akan menjadi akar (*root*) atau node pemisah, algoritma ini menghitung nilai *Gain* tertinggi. Persamaann *gain* dirumuskan pada Persamaan 4.

$$Gain(S, A) = Entroyp(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} \times Entropy(Si) \quad (4)$$

Decision Tree mudah diinterpretasikan dan mampu menangkap interaksi fitur, namun cenderung mengalami *overfitting* jika tidak dilakukan pembatasan kompleksitas.

2.5.4 *Random Forest*

Random Forest merupakan metode *ensemble* berbasis *bagging* yang menggabungkan banyak *Decision Tree* untuk meningkatkan stabilitas dan akurasi model. *Random Forest* menggunakan *Gini impurity* untuk memilih pemisahan, memilih node dengan *impurity* terkecil. *Gini impuirty* dapat dihitung dengan Persamaan 5 [8].

$$I(t_{xi}) = 1 - \sum_{c=0}^{c} \left(\frac{n_{ci}}{a_i}\right)^2 \quad (5)$$

Gini index adalah rata-rata ukuran *Gini* atas berbagai nilai X, yang dapat dirumuskan dengan Persamaan 6.

$$Gini(t, X) = \sum_{i=1}^{j} \frac{a_i}{N} I(t_{xi}) \quad (6)$$

Kriteria pemisahan didasarkan pada nilai ketidakmurnian *Gini* terkecil yang dihitung di antara *m* variabel. Dalam *Random Forest*, setiap pohon menggunakan serangkaian *m* variabel yang berbeda untuk menghasilkan aturan pemisahan. Salah satu hasil dari *Random Forest* adalah pentingnya variabel. Pentingnya variabel menunjukkan tingkat korelasi antara suatu variabel dan hasil klasifikasinya. Persamaan 7 menunjukan *variabel importances* dari hasil *Random Forest*.

$$VI_j = \frac{\sum_{t=1}^{ntree} VI_m^t}{n_{tree}} \quad (7)$$

Kriteria Pemisahan didasarkan pada nilai *Gini Impurity* terkecil yang dihitung di antara variabel-variabel yang ada. Jika akurasi prediksi turun signifikan saat sebuah variabel diacak (permutasi), maka variabel tersebut memiliki korelasi kuat (penting) terhadap hasil klasifikasi. *Random Forest* dikenal sangat efektif pada data tabular dengan fitur heterogen dan interaksi kompleks, serta banyak digunakan dalam penelitian prediksi pembatalan hotel [4], [5].

2.5.5 *Extreme Gradient Boosting* (XGBoost)

XGBoost merupakan algoritma *gradient boosting* yang membangun model secara iteratif dengan meminimalkan fungsi *loss* menggunakan pendekatan gradien. Pada setiap iterasi, model menambahkan sebuah pohon keputusan baru untuk memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya dengan meminimalkan fungsi objektif berbasis gradien. Pendekatan ini memungkinkan XGBoost untuk menangkap hubungan non-linier yang kompleks pada data tabular secara efektif. Secara umum, model prediksi XGBoost pada iterasi ke-*m* dirumuskan sebagai pada Persamaan 8 [14].

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x) \quad (8)$$

dengan $h_m(x)$ sebagai pohon keputusan baru dan ρ_m sebagai *learning rate*. Fungsi objektif XGBoost menggabungkan kesalahan prediksi dan regularisasi kompleksitas model yang dapat ditunjukkan pada Persamaan 9.

$$L_{xgb} = \sum_{i=1}^{n} L\big(y_i, F(x_i)\big) + \sum_{m=1}^{m} \Omega\,(h_m) \quad (9)$$

dengan fungsi regularisasi pada persamaan 10:

$$\Omega(h) = \gamma T + \frac{1}{2}\,\lambda\,\|\omega\|^2 \quad (10)$$

Regularisasi ini berfungsi untuk mengendalikan kompleksitas pohon dan mencegah *overfitting*, sehingga XGBoost menghasilkan model yang stabil dan memiliki kemampuan generalisasi yang baik.

2.5.6 LightGBM

LightGBM merupakan pengembangan dari *gradient boosting* yang dirancang untuk efisiensi komputasi pada dataset besar dan berdimensi tinggi. LightGBM menggunakan pendekatan *leaf-wise growth* yang memungkinkan konvergensi lebih cepat dibandingkan metode *level-wise*. Teknik lain seperti *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB) dirancang untuk mempercepat proses pelatihan dan mengurangi penggunaan memori dengan cara memprioritaskan sampel data yang memiliki gradien besar serta menggabungkan fitur-fitur yang jarang (*sparse*) secara efisien tanpa mengorbankan akurasi model [15]. Karena kemampuannya dalam menangani data tabular secara efisien, LightGBM sering digunakan dalam sistem prediksi berbasis bisnis dan kompetisi *data mining*.

2.5.7 *Artificial Neural Network*

ANN merupakan model berbasis jaringan saraf tiruan yang mampu memodelkan hubungan non-linier kompleks melalui beberapa lapisan neuron. Setiap neuron melakukan transformasi linear yang diikuti fungsi aktivasi non-linier. Persamaan dasar neuron dinyatakan pada Persamaan 11 [16].

$$z = f(\sum_{i=1}^{n} w_i x_i + b) \quad (11)$$

Dimana x_i merupakan *input*, w_i adalah bobot koneksi, b adalah *bias*, dan f merupakan fungsi aktivasi non-linier, seperti *sigmoid* atau *ReLU*.

Pada jaringan dengan satu *hidden layer*, proses ini berlangsung secara berlapis, di mana *output* dari neuron pada *hidden layer* menjadi input bagi neuron pada *output layer*. Fungsi aktivasi *sigmoid* digunakan untuk memetakan nilai keluaran ke rentang probabilitas, sehingga sesuai untuk tugas klasifikasi pembatalan reservasi.

Dalam penelitian ini, ANN digunakan sebagai model pembanding untuk mengevaluasi kemampuan pendekatan neural network dalam menangkap pola non-linier pada data reservasi hotel. Namun, sejalan dengan temuan pada penelitian sebelumnya, performa ANN pada data tabular cenderung tidak selalu unggul dibandingkan model *ensemble* berbasis pohon keputusan [7].

2.5.8 Skema Evaluasi dan Pemilihan Model

Seluruh algoritma model dievaluasi menggunakan skema *Stratified K-Fold Cross-Validation* untuk menjaga proporsi kelas tetap seimbang. Tahap perbandingan ini dilakukan untuk mengidentifikasi model mana yang menunjukan keseimbangan terbaik antara metrik evaluasi utama, yaitu *F0.5-Score*, *precision*, dan ROC AUC.

Hasil proses digunakan sebagai dasar dalam tahap pemilihan model (*model selection*), di mana model dengan performa terbaik dipilih sebagai kandidat utama untuk tahap selanjutnya. Model selection dilakukan berdasarkan evaluasi metrik yang relevan dengan tujuan bisnis, yaitu *F0.5-Score*, *precision*, dan *ROC AUC*, menggunakan skema *cross-validation* untuk memastikan stabilitas performa sebelum dilanjutkan ke tahap optimasi parameter.

Proses modeling dilanjukan dengan melakukan perbandingan antara data asli dan data yang diolah menggunakan metode *oversampling Synthetic Minority Oversampling Technique* (SMOTE). Tujuannya adalah untuk menilai pengaruh ketidakseimbangan kelas terhadap performa model. SMOTE diterapkan hanya pada data pelatihan agar tidak menimbulkan bias pada data uji [17].

Tahap terakhir adalah *hyperparameter tunnng* yang dilakukan menggunakan pendekatan *RandomizedSearch Cross-Validation* untuk menentukan kombinasi parameter optimal bagi model terpilih. Proses *tuning* dilakukan pada *model tree-based ensemble* dengan *parameter* yang disesuaikan untuk masing-masing algoritma, seperti jumlah *estimator*, kedalaman pohon, tingkat pembelajaran (*learning rate*), serta jumlah daun (*num_leaves*) dan *regularization term*. Model terbaik hasil *tuning* kemudian disimpan sebagai *model final* untuk tahap evaluasi berbasis biaya (*cost-based evaluation*) dan optimasi ambang batas (*threshold optimization*).

2.6. *Cost-Based Evaluation* dan *Threshold Optimization*

Tahap *cost-based evaluation* dan *threshold optimization* bertujuan untuk menghubungkan hasil prediksi model dengan dampak finansial nyata yang ditimbulkan pada operasional hotel. Pendekatan ini digunakan untuk menilai sejauh mana model prediksi dapat meminimalkan risiko kerugian akibat pembatalan, serta menentukan ambang keputusan (*decision threshold*) yang menghasilkan pendapatan bersih (*net revenue*) paling optimal.

Secara umum, model klasifikasi menghasilkan keluaran berupa probabilitas, yang kemudian dikonversi menjadi keputusan biner berdasarkan nilai ambang batas (*threshold*). Nilai *threshold default* sebesar 0.5 sering kali tidak optimal karena tidak mempertimbangkan ketidakseimbangan kelas maupun perbedaan biaya kesalahan prediksi. Oleh karena itu, dalam penelitian ini percobaan nilai ambang pada rentang 0.1 hingga 0.9 dengan interval 0.02 untuk mengevaluasi perubahan performa model terhadap metrik statistik dan finansial.

Pendekatan ini menggabungkan dua aspek evaluasi utama, yaitu *statistical-based evaluation* dan *cost-based evaluation*. Pada aspek pertama, performa model diukur menggunakan metrik *precision*, *recall*, dan *F0.5-Score*, di mana *F0.5-Score* dipilih sebagai evaluasi metrik utama karena menekankan pentingnya *precision* guna meminimalkan *false positive*. Namun, metrik statistik tersebut tidak sepenuhnya mencerminkan dampak ekonomi dari kesalahan prediksi. Oleh karena itu, dilakukan perluasan analisis melalui *cost-based evaluation* dengan mengonversi hasil *confusion matrix* menjadi nilai pendapatan bersih (*net revenue*) berdasarkan asumsi biaya aktual [18]. Berdasarkan pendekatan *cost-sensitive learning*, nilai *net revenue* (NR) dihitung menggunakan Persamaan 12.

$$NR=(RTN+RTP)-(LFP+LFN) \quad (12)$$

Komponen evaluasi berbasis biaya dalam penelitian ini terdiri dari RTN yang merepresentasikan pendapatan dari pemesanan yang terealisasi (*true negatives*), dan RTP sebagai nilai penghematan dari pembatalan yang berhasil diprediksi secara akurat (*true positives*). Di sisi lain, analisis ini juga

mempertimbangkan potensi kerugian, yaitu LFP sebagai biaya kompensasi akibat kesalahan prediksi pembatalan pada tamu yang sebenarnya datang (FP), serta LFN yang merupakan kehilangan pendapatan akibat pembatalan yang tidak terdeteksi oleh sistem (FN).

Dalam penelitian ini, biaya kesalahan klasifikasi LFP dimodelkan sebesar 1,5 kali nilai ADR. Penilaian ini dipilih sebagai skenario konservatif yang merepresentasikan kombinasi biaya kehilangan kamar, relokasi tamu, dan kompensasi layanan sebagaimana umum dilaporkan dalam literatur *overbooking* [19]. Pendekatan ini mengadaptasi konsep *profit-sensitive confusion matrix* sebagaimana diperkenalkan oleh Ariza-Garzón et al. (2024) [20], di mana setiap komponen pada *confusion matrix* (TP, TN, FP, FN) memiliki nilai finansial berbeda sesuai dampaknya terhadap pendapatan bersih. Dengan demikian, fungsi *net revenue* didefinisikan sebagai gabungan antara pendapatan aktual, penghematan dari prediksi benar, serta kerugian akibat kesalahan prediksi.

Perlu dicatat bahwa nilai biaya ini bersifat fleksibel dan dapat disesuaikan dengan kebijakan serta struktur biaya masing-masing hotel. Namun demikian, pendekatan *cost-based* yang digunakan dalam penelitian ini tetap relevan secara metodologis karena fokus utama penelitian adalah pada mekanisme optimasi *threshold* berbasis dampak ekonomi, bukan pada nilai biaya absolut semata.

2.7. Evaluasi Model

Tahap evaluasi model dilakukan untuk menilai kinerja model dan menentukan konfigurasi terbaik yang tidak hanya unggul secara statistik, tetapi juga memberikan dampak positif terhadap profitabilitas hotel. Evaluasi dilakukan melalui dua pendekatan utama, yaitu *statistical-based evaluation* dan *cost-based evaluation*. Pendekatan pertama berfokus pada performa prediktif model dalam konteks klasifikasi, sedangkan pendekatan kedua mengukur implikasi finansial dari keputusan model terhadap pendapatan bersih (*net revenue*) [21].

Pada *statistical-based evaluation*, model diuji menggunakan berbagai metrik umum untuk klasifikasi biner, yaitu *precision*, *recall*, *F0.5-Score*, dan ROC AUC. Metrik *precision* digunakan untuk menilai seberapa akurat model dalam mengidentifikasi pembatalan yang benar, sementara *recall* mengukur kemampuan model mendeteksi seluruh kasus pembatalan actual [22]. Metrik *F0.5-Score* digunakan sebagai indikator utama karena menekankan bobot yang lebih besar pada *precision* [23], mengingat *false positive* (prediksi pembatalan padahal tamu hadir) menimbulkan kerugian yang lebih besar bagi hotel. Selain itu, nilai ROC AUC digunakan untuk mengukur kemampuan diskriminatif model secara keseluruhan terhadap dua kelas target [24].

Sementara itu, *cost-based evaluation* digunakan untuk menghubungkan hasil klasifikasi dengan dampak ekonomi yang timbul dari kesalahan prediksi. Dalam konteks ini, *true negative* dianggap sebagai sumber pendapatan aktual, *true positive* sebagai bentuk penghematan dari pembatalan yang berhasil diantisipasi, sedangkan *false positive* dan *false negative* dikonversi menjadi kerugian finansial. Pendekatan evaluasi berbasis biaya ini sejalan dengan penelitian terkini yang menekankan pentingnya *application-centric evaluation*, di mana pemilihan model harus mempertimbangkan konteks operasional dan dampak bisnis dari kesalahan klasifikasi [25]. Perhitungan nilai *net revenue* dilakukan berdasarkan formula yang telah dijelaskan pada subbab sebelumnya, dengan parameter kompensasi $\gamma = 1.5$ dan faktor penghindaran kerugian $\rho = 1.0$. Pendekatan ini memungkinkan evaluasi model dilakukan secara lebih realistis terhadap kondisi bisnis perhotelan yang sebenarnya.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Eksperimen dalam penelitian ini dilakukan secara bertahap dan sistematis menggunakan *dataset hotel booking demand* . Fitur yang digunakan mencakup fitur numerik, kategorikal, ordinal, dan biner, yang diproses melalui tahapan preprocessing seperti normalisasi, *encoding*, dan transformasi data.

3.2. Model *Machine Learning* dan Perbandingan Algoritma

Evaluasi kinerja model dilakukan menggunakan pendekatan *Stratified K-Fold Cross-Validation* guna memastikan distribusi kelas yang seimbang pada setiap lipatan data [26]. Metrik evaluasi yang digunakan meliputi *F0.5-Score*, *precision*, dan ROC AUC. Pemilihan *F0.5-Score* didasarkan pada kebutuhan bisnis industri perhotelan yang lebih menekankan pada ketepatan prediksi pembatalan dibandingkan kemampuan mendeteksi seluruh kasus pembatalan. Tabel 1 merupakan perbandingan kinerja model klasifikasi.

Table 1. Kinerja Model Klasifiaksi

Model	*F0.5* Mean	*F0.5* Std	*Precision* Mean	ROC_AUC Mean
Random Forest	0.8179	0.0016	0.8362	0.9177
XGBoost	0.8108	0.0036	0.0839	0.9096
LigtGBM	0.8091	0.0018	0.8505	0.9036
ANN	0.8091	0.0045	0.8389	0.911

Model	*F0.5* Mean	*F0.5* Std	*Precision* Mean	ROC_AUC Mean
Logistic Regression	0.7816	0.0033	0.8359	0.8613
KNN	0.7586	0.0044	0.7696	0.8721
Decision Tree	0.7505	0.0039	0.7492	0.8096

Berdasarkan perbandingan model klasifikasi untuk prediksi pembatalan pemesanan hotel, *Random Forest* menunjukan model unggul, khususnya pada metrik *F0.5-Score*. Model *ensemble* berbasis *gradien* seperti XGBoost dan LightGBM bersifat kompetitif pada metrik ROC AUC dan *precision*, namun tetap berada di bawah *Random Forest* dalam nilai *F0.5-Score*. Sementara itu, model klasik dan ANN menunjukkan performa yang relatif lebih rendah karena keterbatasan dalam menangani hubungan non-linear dan fitur kategorikal yang kompleks.

3.3. Perbandingan Data Asli dan *Oversampling*

Perbandingan dilakukan bertujuan untuk mengevaluasi apakah teknik *oversampling* mampu meningkatkan kinerja model dalam menangani ketidakseimbangan kelas, khususnya pada kasus pembatalan pemesanan hotel yang memiliki proporsi kelas minoritas lebih kecil. Tabel 2 merupakan hasil perbandingan kinerja *Random Forest* antara menggunakan data asli dan data *oversampling*

Table 2. Perbandingan Kinerja Random Forest Menggunakan Data Asli dan Data *Oversampling*

Model	Skenario	*F0.5*	*Precision 1*	ROC_AUC
Random Forest	*Data Asli*	0.8202	0.84	0.9175
	Oversampling	0.8096	0.82	0.9177

Berdasarkan hasil yang ditunjukkan pada Tabel 2, penggunaan data asli menghasilkan kinerja model yang lebih baik dibandingkan data *oversampling*. Hal ini menunjukkan bahwa meskipun *oversampling* dapat meningkatkan keseimbangan distribusi kelas, teknik tersebut tidak selalu memberikan peningkatan performa yang relevan terhadap tujuan bisnis yang dihadapi. Oleh karena itu, data asli dipilih untuk digunakan pada tahapan analisis selanjutnya.

3.4. *Hyperparameter Tuning Random Forest*

Hyperparameter tuning Random Forest dilakukan untuk mengoptimalkan konfigurasi model sehingga diperoleh performa yang lebih baik dan stabil. Pemilihan nilai parameter yang tidak optimal berpotensi menyebabkan *overfitting* atau *underfitting*, sehingga diperlukan proses pencarian parameter yang sistematis [27].

Hyperparameter tuning dilakukan menggunakan pendekatan *cross-validation* dengan skema *Stratified K-Fold* untuk menjaga proporsi kelas pada setiap lipatan data. Evaluasi performa difokuskan pada metrik *F0.5-Score* dan *precision*, karena kedua metrik tersebut memberikan penalti lebih besar terhadap kesalahan *false positive*, yang berdampak langsung terhadap potensi pendapatan hotel. Tabel 3 menunjukan perbandingan kinerja model klasifikasi sebelum dan sesudah *tuning*.

Table 3. Perbandingan Kinerja Model Klasifikasi

Metode	*F0.5*	*Precision 1*	*ROC_AUC*
Hyperparameter Tuning	0.8253	0.87	0.9165
Default	0.8202	0.84	0.9165

Tabel 3 menunjukan bahwa hasil *hyperparameter tuning* meningkatkan kinerja model *Random Forest*. Peningkatan ini terlihat pada nilai *F0.5-score* dan *precision*, yang menunjukkan bahwa model hasil *tuning* lebih efektif dalam mengurangi kesalahan klasifikasi yang berpotensi merugikan secara bisnis. Model *Random Forest* dengan parameter terbaik akan digunakan sebagai dasar untuk tahapan analisis selanjutnya.

3.5. *Threshold Adjustment* dan *Cost-Based Evaluation*

Meskipun model *Random Forest* hasil *hyperparameter tuning* menunjukkan kinerja yang baik berdasarkan metrik klasifikasi konvensional, penggunaan *default threshold* belum tentu menghasilkan keputusan yang optimal dari sudut pandang bisnis. Dalam pendekatan *cost-sensitive learning*, keputusan *threshold* tidak lagi diasumsikan konstan 0.5 (*default*), melainkan dapat diubah untuk meminimalkan biaya *misclassification* sesuai rasio biaya FP/FN. Pendekatan ini telah banyak dibahas dalam literatur *cost-sensitive learning* yang menekankan penyesuaian *threshold* berdasarkan *cost proportion* dan metrik *cost-dependent* [28]. Oleh karena itu, penelitian ini menerapkan *threshold adjustment* untuk mengevaluasi bagaimana perubahan nilai ambang keputusan memengaruhi kinerja model, baik dari sisi metrik klasifikasi maupun dari sisi *cost-based evaluation*. Pendekatan ini bertujuan untuk mencari nilai *threshold* yang paling sesuai dengan tujuan bisnis, bukan semata-mata berdasarkan performa statistik [18].

3.5.1 *Threshold Adjustment* dan *Cost-Based Evaluation*

Threshold pertama yang dievaluasi adalah *threshold* yang menghasilkan nilai *F0.5-Score* terbaik. Pemilihan *F0.5-Score* sebagai acuan didasarkan pada karakteristik permasalahan, di mana kesalahan f*alse positive* memiliki dampak yang lebih besar dibandingkan *false negative*. Dengan memberikan bobot lebih besar pada *precision*, *F0.5-Score* diharapkan mampu menghasilkan model yang lebih konservatif dalam memprediksi pembatalan. Tabel 4 merupakan perbandingan kinerja model berdasarkan *threshold default* dan *threshold* terbaik *F0.5-Score*.

Table 4. Perbandingan Kinerja Model Berdasarkan *Threshold*

Threshold	*F0.5*	*Precision 1*	ROC_AUC	*Net revenue* (£)
Best Threshold (0.62)	0.8297	0.8297	0.9174	4,992,424
Default (0.5)	0.8254	0.8253	0.9165	5,213,242

Hasil evaluasi menunjukkan bahwa penyesuaian threshold yang memaksimalkan *F0.5-Score* mampu meningkatkan *precision* dibandingkan penggunaan *threshold default*. Namun, peningkatan ini diikuti dengan penurunan jumlah prediksi positif, yang berdampak pada berkurangnya jumlah intervensi terhadap potensi pembatalan. Meskipun ada peningkatan skor metrik dari nilai *default threshold*, peningkatan tersebut tidak serta merta meningkatkan nilai *net revenue*. Hal ini menunjukan bahwa skor metrik terbaik belum tentu merupakan model yang harus digunakan dalam konteks bisnis. Hasil ini sejalan dengan penerapan *cost-sensitive learning* dalam berbagai domain praktis di mana keputusan klasifikasi berbeda memiliki implikasi ekonomi yang nyata [29].

Hasil mengindikasikan bahwa perubahan kecil pada nilai *threshold* dapat menghasilkan variasi yang signifikan terhadap *net revenue*, terutama pada kondisi permintaan tinggi. Hal ini menunjukkan bahwa pemilihan *threshold* bersifat kontekstual dan perlu disesuaikan dengan strategi manajemen risiko hotel

3.5.2 *Cost-Based Evaluation* Berdasarkan *Net Revenue*

Perhitungan *net revenue* dilakukan dengan mempertimbangkan manfaat dan kerugian yang timbul akibat keputusan klasifikasi model, seperti biaya intervensi terhadap pelanggan dan potensi pendapatan yang berhasil diselamatkan. Tabel 5 merupakan perbandingan *net revenue* pada berbagai nilai *threshold*.

Table 5. Perbandingan *Net revenue* Pada Berbagai Nilai *Threshold*

Threshold	*F0.5*	*Precision 1*	*Net revenue* (£)
0.52	0.8298	0.8755	5,214,886
0.5	0.8254	0.8681	5,213,242
0.54	0.8284	0.8861	5,208,860
0.56	0.8288	0.8924	5,184,555
0.62	0.8297	0.9207	4,992,424

Hasil perhitungan menunjukkan bahwa penggunaan *threshold* sebesar 0.52 menghasilkan *nilai net revenue* yang lebih tinggi dibandingkan *threshold* yang dioptimalkan berdasarkan *F0.5-Score* (0.62). Secara kuantitatif, *threshold* 0.62 menghasilkan *net revenue* sebesar 4,992,424 Poundsterling, sedangkan *threshold* berdasarkan *cost-based evaluation* menghasilkan *net revenue* sebesar 5,182,444 Poundsterling.

3.5.3 Analisis Perbandingan Threshold Berbasis Metrik dan Bisnis

Perbedaan hasil antara *threshold* terbaik berdasarkan *F0.5-Score* dan *threshold* terbaik berdasarkan *net revenue* menunjukkan bahwa optimasi berbasis metrik klasifikasi tidak selalu sejalan dengan optimasi berbasis tujuan bisnis. *Threshold* yang menghasilkan nilai *F0.5-Score* tertinggi cenderung lebih ketat dalam memprediksi pembatalan, sehingga mengurangi risiko *false positive*, tetapi juga membatasi peluang penyelamatan pendapatan.

Sebaliknya, *threshold* 0.52 memberikan keseimbangan yang lebih baik antara jumlah prediksi pembatalan dan biaya intervensi, sehingga menghasilkan *net revenue* yang lebih tinggi. Temuan ini mengindikasikan bahwa dalam konteks pengambilan keputusan operasional, threshold yang optimal sebaiknya ditentukan berdasarkan evaluasi berbasis biaya, bukan hanya berdasarkan metrik klasifikasi konvensional.

Dengan demikian, hasil penelitian ini menegaskan pentingnya integrasi antara *machine learning*, *threshold adjustment*, dan *cost-based evaluation* dalam membangun sistem pendukung keputusan yang tidak hanya akurat secara statistik, tetapi juga memberikan nilai tambah secara ekonomi.

3.6. Evaluasi Model

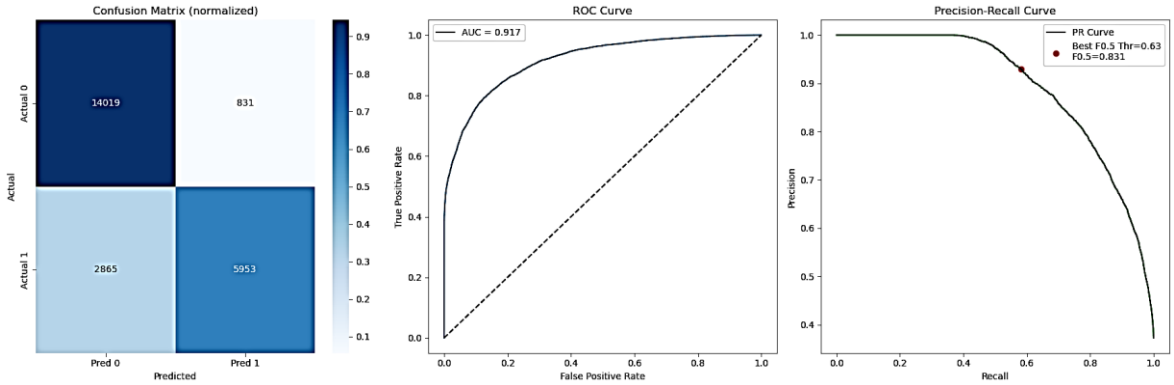
Berdasarkan seluruh tahapan eksperimen yang telah dilakukan, *Random Forest* terpilih sebagai model terbaik. Model ini menunjukkan performa yang konsisten dan unggul dibandingkan model lain pada berbagai

skenario pengujian, baik menggunakan data asli maupun data hasil *oversampling*, serta sebelum dan sesudah proses *hyperparameter tuning*. Hasil akhir metriks model *Random Forest* dapat dilihat pada Tabel 6.

Table 6. Hasil Akhir Performa Model *Random Forest*

Model	*F0.5*	*Precision 1*	*Recall 1*	ROC AUC
Random Forest	0.8279	0.878	0.6751	0.9165

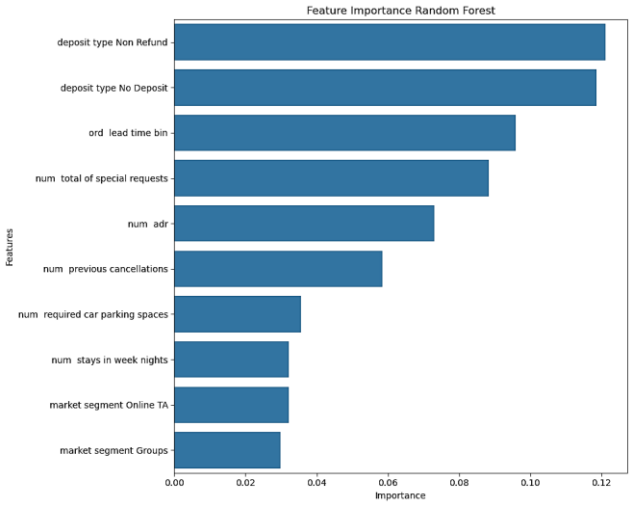
Tabel 6 menunjukkan hasil akhir performa model Random Forest yang mencapai nilai *F0.5-Score* sebesar 0,8279 dan skor ROC AUC mencapai 0,9165, yang mengindikasikan kemampuan klasifikasi yang sangat kuat. Untuk memberikan gambaran yang lebih komprehensif mengenai kinerja model, evaluasi selanjutnya dilakukan menggunakan *confusion matrix*, *ROC curve*, dan *precision–recall curve* ditunjukkan pada Gambar 2.



Gambar 2. Evaluasi Kinerja Model Random Forest

Confusion matrix menunjukkan bahwa model mampu mengidentifikasi sebagian besar kasus pembatalan dengan baik melalui jumlah *true positive* yang relatif tinggi, sementara kesalahan prediksi berupa *false positive* masih berada pada tingkat yang dapat diterima. Kurva ROC dengan nilai AUC yang tinggi menegaskan kemampuan diskriminasi model *Random Forest* yang sangat baik. Kurva *precision-recall curve* menunjukan bahwa *precision* tetap terjaga meskipun *recall* meningkat. Hasil ini konsisten dengan penggunaan F0.5-Score sebagai metrik utama, yang menekankan ketepatan prediksi pembatalan dan selaras dengan tujuan optimasi berbasis biaya.

Selain memiliki performa prediktif yang tinggi, model ini juga memungkinkan dilakukannya analisis mendalam terhadap faktor-faktor yang memengaruhi keputusan pembatalan. *Explainable Artificial Intelligence* (XAI) diterapkan pada tingkat global untuk mengidentifikasi faktor-faktor utama yang memengaruhi prediksi pembatalan reservasi. Mengingat model terbaik yang diperoleh adalah *Random Forest*, analisis XAI dilakukan menggunakan *feature importance* untuk mengevaluasi kontribusi relatif masing-masing fitur terhadap keputusan model secara keseluruhan. Gambar 3 merupakan hasil analisis *feature importance* pada *dataset hotel booking demand*.



Gambar 3. *Feature Importance* Model *Random Forest*

Hasil analisis menunjukkan bahwa fitur-fitur yang berkaitan dengan karakteristik pemesanan dan perilaku pelanggan, seperti *deposit type*, *lead time*, *total special reqest*, *adr*, dan sebagainya memiliki kontribusi paling besar terhadap keputusan model. Temuan ini mengindikasikan bahwa faktor waktu pemesanan dan komitmen finansial pelanggan memainkan peran penting dalam menentukan risiko pembatalan. Informasi ini tidak hanya meningkatkan transparansi model, tetapi juga memberikan wawasan yang bernilai bagi manajemen hotel dalam memahami pola perilaku pelanggan. Penelitian ini menggunakan *feature importance* sebagai bentuk *explainability global* yang sederhana dan mudah diinterpretasikan oleh praktisi bisnis. Pendekatan ini dipilih untuk menjaga keseimbangan antara kompleksitas teknis dan kebutuhan interpretasi manajerial, sementara metode XAI lanjutan seperti SHAP dapat menjadi arah penelitian selanjutnya.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi penerapan *machine learning* dalam prediksi pembatalan pemesanan hotel dengan menekankan pendekatan *threshold adjustment* dan *cost-based evaluation*. Berdasarkan hasil *benchmarking* terhadap beberapa model klasifikasi, *Random Forest* terbukti memberikan kinerja paling stabil dan unggul dibandingkan model lain, termasuk ANN, baik dari sisi *F0.5-Score*, *Precision*, maupun ROC AUC. Hasil ini menunjukkan bahwa model berbasis *ensemble tree* lebih efektif dalam menangkap pola kompleks pada *dataset hotel booking demand*.

Eksperimen lanjutan menunjukkan bahwa penggunaan data asli menghasilkan performa yang lebih baik dibandingkan pendekatan *oversampling*, serta *hyperparameter tuning* mampu meningkatkan kinerja model secara moderat namun konsisten. Penyesuaian *threshold* klasifikasi berdasarkan *F0.5-Score* terbukti meningkatkan keseimbangan antara precision dan recall kelas positif, namun tidak selalu menghasilkan keuntungan ekonomi maksimum.

Hasil evaluasi berbasis biaya menunjukkan bahwa *threshold* 0.52 menghasilkan nilai *net revenue* sejumlah 5,182,444 Poundsterling yang lebih tinggi dibandingkan *threshold* optimal berbasis *F0.5-Score*. Temuan ini mengindikasikan bahwa optimalitas metrik klasifikasi tidak selalu sejalan dengan optimalitas keputusan bisnis. Oleh karena itu, penggunaan *cost-based evaluation* menjadi pendekatan yang lebih relevan dalam konteks aplikasi nyata, khususnya untuk sistem pendukung keputusan di industri perhotelan.

Kebaruan dari penelitian ini terletak pada integrasi komprehensif antara proses pemilihan model, penyesuaian *threshold*, dan evaluasi berbasis nilai ekonomi dalam satu kerangka evaluasi yang utuh. Penelitian ini tidak hanya menilai performa model dari sisi statistik, tetapi juga dari dampak finansial yang dihasilkan, sehingga memberikan kontribusi praktis dalam pengembangan sistem prediksi pembatalan hotel yang berorientasi pada optimalisasi pendapatan. Penelitian selanjutnya dapat memperluas validasi model pada data lintas hotel atau periode waktu yang berbeda untuk menguji generalisasi pendekatan yang diusulkan pada penelitian ini.

REFERENSI

[1] S. Ivanov and V. Zhechev, "TOURISM Review Hotel revenue Management - A Critical Literature Review," Jan. 2012.

[2] B. M. Noone and C. H. Lee, "Hotel overbooking: The effect of overcompensation on customers' reactions to denied service," *Journal of Hospitality and Tourism Research*, vol. 35, no. 3, pp. 334–357, Aug. 2011, doi: 10.1177/1096348010382238.

[3] Z. Kenesei and Z. Bali, "Overcompensation as a service recovery strategy: the financial aspect of customers' extra effort," *Service Business*, vol. 14, no. 2, pp. 187–216, Jun. 2020, doi: 10.1007/s11628-020-00413-w.

[4] N. Antonio, A. de Almeida, and L. Nunes, "Hotel booking demand datasets," *Data Brief*, vol. 22, pp. 41–49, Feb. 2019, doi: 10.1016/j.dib.2018.11.126.

[5] T. Y. Alkan, "A Comparative Study of Machine Learning and Deep Learning Approaches For Hotel Booking Cancellation Prediction," *International Journal Of Scientific Research In Engineering & Technology*, pp. 11–18, May 2025, doi: 10.59256/ijsreat.20250503002.

[6] A. Onan and S. KorukoGlu, "A feature selection model based on genetic rank aggregation for text sentiment classification," *J Inf Sci*, vol. 43, no. 1, pp. 25–38, Feb. 2017, doi: 10.1177/0165551515613226.

[7] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on tabular data?," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.08815

[8] Z. A. Andriawan *et al.*, "Prediction of Hotel Booking Cancellation using CRISP-DM," in *ICICoS 2020 - Proceeding: 4th International Conference on Informatics and Computational Sciences*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICICoS51170.2020.9299011.

[9] A. Onan, "Consensus Clustering-Based Undersampling Approach to Imbalanced Learning," *Sci Program*, vol. 2019, 2019, doi: 10.1155/2019/5901087.

[10] S. Domínguez-Almendros, N. Benítez-Parejo, and A. R. Gonzalez-Ramirez, "Logistic regression models," *Allergol Immunopathol (Madr)*, vol. 39, no. 5, pp. 295–305, Sep. 2011, doi: 10.1016/j.aller.2011.05.002.

[11] N. Antonio, A. De Almeida, and L. Nunes, "An automated machine learning based decision support system to predict hotel booking cancellations," *Data Sci J*, vol. 18, no. 1, 2019, doi: 10.5334/dsj-2019-032.

[12] M. Sagala, "Algorithm Modified K-Nearest Neighbor (M-KNN) for Classification of Attention Deficit Hyperactive Disorder (ADHD) in Children," 2019. [Online]. Available: http://login.seaninstitute.org/index.php/Loginρ11Journalhomepage:http://login.seaninstitute.org/index.php/Login

[13] F. Akbar, H. Wira Saputra, A. Karel Maulaya, and M. Fikri Hidayat, "Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," vol. 2, pp. 61–67, Oct. 2022.

[14] C. Bentéjac, A. Csörgö, and G. Martínez-Muñoz, "A Comparative Analysis of XGBoost," Nov. 2019. doi: 10.48550/arXiv.1911.01914.

[15] P. Florek and A. Zagdański, "Benchmarking state-of-the-art gradient boosting algorithms for classification," May 2023, [Online]. Available: http://arxiv.org/abs/2305.17094

[16] J. G. Cabello, "Mathematical Neural Networks," *Axioms*, vol. 11, no. 2, Feb. 2022, doi: 10.3390/axioms11020080.

[17] M. Mujahid *et al.*, "Data Oversampling and Imbalanced Datasets: An Investigation of Performance For Machine Learning and Feature Engineering," *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00943-4.

[18] P. Peykani, M. Peymany Foroushany, C. Tanasescu, M. Sargolzaei, and H. Kamyabfar, "Evaluation of Cost-Sensitive Learning Models in Forecasting Business Failure of Capital Market Firms," *Mathematics*, vol. 13, no. 3, Feb. 2025, doi: 10.3390/math13030368.

[19] N. Phumchusri and P. Maneesophon, "Optimal overbooking decision for hotel rooms revenue management," *Journal of Hospitality and Tourism Technology*, vol. 5, no. 3, pp. 261–277, Oct. 2014, doi: 10.1108/JHTT-03-2014-0006.

[20] M. J. Ariza-Garzón, J. Arroyo, M. J. Segovia-Vargas, and A. Caparrini, "Profit-sensitive machine learning classification with explanations in credit risk: The case of small businesses in peer-to-peer lending," *Electron Commer Res Appl*, vol. 67, Sep. 2024, doi: 10.1016/j.elerap.2024.101428.

[21] C. Yaiprasert and A. N. Hidayanto, "AI-powered ensemble machine learning to optimize cost strategies in logistics business," *International Journal of Information Management Data Insights*, vol. 4, no. 1, Apr. 2024, doi: 10.1016/j.jjimei.2023.100209.

[22] L. Cohen, Y. Mansour, S. Moran, and H. Shao, "Probably Approximately Precision and Recall Learning," Oct. 2025, [Online]. Available: http://arxiv.org/abs/2411.13029

[23] S. Beddar-Wiesing, A. Moallemy-Oureh, M. Kempkes, and J. M. Thomas, "Absolute Evaluation Measures for Machine Learning: A Survey," Jul. 2025, [Online]. Available: http://arxiv.org/abs/2507.03392

[24] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

[25] B. Hutchinson, N. Rostamzadeh, C. Greer, K. Heller, and V. Prabhakaran, "Evaluation Gaps in Machine Learning Practice," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Jun. 2022, pp. 1859–1876. doi: 10.1145/3531146.3533233.

[26] M. Bhagat and B. Bakariya, "A Comprehensive Review of Cross-Validation Techniques in Machine Learning," 2025.

[27] P. Probst, M. Wright, and A.-L. Boulesteix, "Hyperparameters and Tuning Strategies for Random Forest," Feb. 2019, doi: 10.1002/widm.1301.

[28] V. Komisarenko and M. Kull, "Cost-sensitive classification with cost uncertainty: do we need surrogate losses?," *Mach Learn*, vol. 114, no. 6, Jun. 2025, doi: 10.1007/s10994-024-06634-8.

[29] N. Lawrance, M.-A. Guerry, and G. Petrides, "Cost-Sensitive Stacking: an Empirical Evaluation," Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.01748