

Studi Komparatif Performa Model Ensemble Learning dalam Klasifikasi Kepuasan Penumpang Maskapai Penerbangan

Fransiska Elly Renni Susanti^{*1}, Heni Ermewaningsih², Donatus Leo³, Kristian Tengker⁴

^{1,2,3,4}Rekayasa Komputer, Institut Teknologi Keling Kumang

Email: ¹fransiska_ers@itkk.ac.id, ²heni_ermewaningsih@itkk.ac.id, ³donatus_leo@itkk.ac.id,
⁴kristian_tengker@itkk.ac.id

Abstrak

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa beberapa model *ensemble learning* dalam klasifikasi kepuasan penumpang maskapai penerbangan menggunakan dataset publik dari Kaggle. Dataset terdiri dari 25.893 observasi dengan 23 variabel prediktor yang mencakup karakteristik pelanggan, kualitas pelayanan, dan pengalaman penerbangan. Tiga algoritma yang digunakan adalah *Random Forest*, *Gradient Boosting*, dan *XGBoost*. Proses pemodelan dilakukan menggunakan skema *train-test split* 80:20 serta validasi silang, dan evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-Score*, dan *ROC-AUC*. Hasil penelitian menunjukkan bahwa *XGBoost* memberikan performa terbaik pada seluruh metrik evaluasi dengan akurasi sebesar 0,963 dan *ROC-AUC* sebesar 0,995. Analisis *feature importance* menunjukkan bahwa variabel *online boarding*, tipe perjalanan bisnis, serta kualitas layanan dalam penerbangan merupakan faktor dominan yang memengaruhi kepuasan penumpang. Kontribusi utama penelitian ini adalah penyajian evaluasi komparatif model *ensemble learning* yang terintegrasi dengan analisis interpretabilitas untuk mengidentifikasi determinan utama kepuasan pelanggan pada data tabular. Temuan ini memberikan implikasi bahwa peningkatan layanan digital dan kualitas pengalaman kabin menjadi prioritas strategis dalam meningkatkan kepuasan pelanggan maskapai.

Kata kunci: *ensemble learning, kepuasan penumpang, klasifikasi, maskapai penerbangan, XGBoost*

A Comparative Study of Ensemble Learning Model Performance in Airline Passenger Satisfaction Classification

Abstract

This study aims to evaluate and compare the performance of several ensemble learning models in classifying airline passenger satisfaction using a publicly available dataset from Kaggle. The dataset consists of 25,893 observations with 23 predictor variables covering customer characteristics, service quality, and flight experience attributes. Three algorithms were employed, namely Random Forest, Gradient Boosting, and XGBoost. The modeling process was conducted using an 80:20 train-test split combined with cross-validation, and performance was evaluated using accuracy, precision, recall, F1-Score, and ROC-AUC metrics. The results indicate that XGBoost achieved the best performance across all evaluation metrics, with an accuracy of 0.963 and an ROC-AUC of 0.995. Feature importance analysis reveals that online boarding, type of travel (business), and in-flight service quality are the most influential factors affecting passenger satisfaction. The main contribution of this study lies in providing a comparative evaluation of ensemble learning models integrated with interpretability analysis to identify the key determinants of customer satisfaction in tabular data. These findings imply that enhancing digital services and improving cabin experience should be prioritized as strategic efforts to increase airline passenger satisfaction.

Keywords: *airline industry, classification, ensemble learning, passenger satisfaction, XGBoost*

1. PENDAHULUAN

Industri penerbangan menjadi salah satu aspek mobilitas dalam mobilitas global dan pertumbuhan ekonomi dunia. Menurut data dari International Air Transport Association (IATA), meredanya pandemi COVID-19 membuka kembali aktivitas masyarakat yang dahulu sempat terhenti, salah satunya dalam konteks penerbangan. Keterbukaan rute internasional membuat industri penerbangan global mengalami pertumbuhan yang signifikan. Kepuasan penumpang telah menjadi sorotan utama dalam persaingan antar maskapai dalam mempertahankan loyalitas pelanggan, memperkuat reputasi maskapai, dan meningkatkan profitabilitas jangka panjang. Penelitian

[1] menegaskan bahwa kualitas kepuasan pelanggan secara langsung mempengaruhi niat untuk kembali menggunakan jasa maskapai. Aspek layanan yang perlu ditingkatkan, seperti kenyamanan, ketepatan waktu, serta kualitas layanan dalam penerbangan dapat diukur dengan menggunakan teknik survei kepuasan penumpang yang komprehensif.

Kepuasan pelanggan sering kali diukur dengan menggunakan survei atau kuisioner yang mendeskripsikan tanggapan konsumen terhadap layanan dalam bentuk skor ordinal. Namun, hubungan antara skor rating layanan dan kepuasan sering bersifat non-linier dan kompleks menjadikan teknik statistik konvensional dianggap kurang cukup efektif dalam menangkap pola hubungan ini. Beberapa penelitian [2], [3] membuktikan bahwa pendekatan berbasis *machine learning* terbukti lebih mampu dalam menangani hubungan non-linier serta memberikan pemahaman yang lebih baik tentang kontribusi variabel input terhadap kepuasan pelanggan. Oleh karena itu, pendekatan dengan *machine learning* menjadi semakin populer karena kemampuannya dalam mengolah data besar secara efisien sekaligus mengidentifikasi pola tersembunyi.

Di sisi lain, berbagai penelitian menunjukkan bahwa pendekatan *machine learning* lebih efektif dibandingkan teknik statistik konvensional dalam menangani hubungan non-linier dan kompleksitas data keputusan pelanggan, serta mampu memberikan akurasi prediksi yang lebih tinggi dalam mengidentifikasi faktor-faktor penentu kepuasan. Beberapa studi [4], [5] empiris telah menerapkan algoritma *machine learning* untuk mengklasifikasi kepuasan penumpang maskapai, hal ini menunjukkan bahwa pendekatan *machine learning* layak digunakan untuk memprediksi kepuasan penumpang maskapai secara otomatis. Selain itu, penelitian internasional [6] menunjukkan bahwa penggunaan kombinasi analisis sentimen dengan model *machine learning* mampu memprediksi tingkat kepuasan penumpang dengan tingkat akurasi sebesar 97% dengan menggunakan gabungan data rating dan sentimen ulasan pelanggan, penelitian ini juga mendapati bahwa faktor *value for money* dan *ground service* memiliki pengaruh signifikan terhadap kepuasan pelanggan maskapai.

Selain itu, penelitian [5] yang mengeksplorasi penggunaan algoritma *supervised learning* pada kepuasan pelanggan menemukan bahwa *Random Forest* sering menunjukkan performa yang lebih unggul dibandingkan dengan algoritma lain seperti K-NN, *Decision Tree*, dan *Naive Bayes*. Ketidakseimbangan kelas merupakan masalah yang sering dialami dalam *classification tasks*, penelitian [7] menegaskan bahwa teknik integrasi seperti SMOTE dan algoritma *ensemble learning* dapat mengatasi permasalahan ketidakseimbangan kelas ini, hal ini terbukti dapat meningkatkan performa model dibandingkan dengan pendekatan tradisional. Namun, distribusi kelas pada dataset penelitian ini tergolong relatif seimbang, sehingga teknik penanganan ketidakseimbangan kelas tidak diterapkan dalam proses pemodelan.

Penelitian ini bertujuan untuk menganalisa performa beberapa model *ensemble learning* dalam klasifikasi keputusan penumpang maskapai penerbangan menggunakan dataset dari platform Kaggle. Tujuan utama penelitian ini adalah untuk mengidentifikasi dan menemukan model terbaik secara komparatif berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*, serta menemukan fitur layanan yang paling mempengaruhi kepuasan penumpang.

Beberapa penelitian lain telah menerapkan algoritma *machine learning* untuk klasifikasi kepuasan penumpang, sebagian besar studi masih berfokus pada penggunaan model tunggal atau belum melakukan evaluasi komparatif yang sistematis antar metode *ensemble learning*. Selain itu, penelitian sebelumnya cenderung menitikberatkan model untuk mengidentifikasi faktor-faktor utama yang memengaruhi kepuasan pelanggan. Oleh karena itu, penelitian ini mengisi kesenjangan tersebut dengan melakukan evaluasi komparatif terhadap model *ensemble learning* serta mengkombinasikannya dengan analisis *feature importance* untuk memberikan pemahaman yang lebih komprehensif, baik dari sisi performa maupun interpretasi model.

2. METODE PENELITIAN

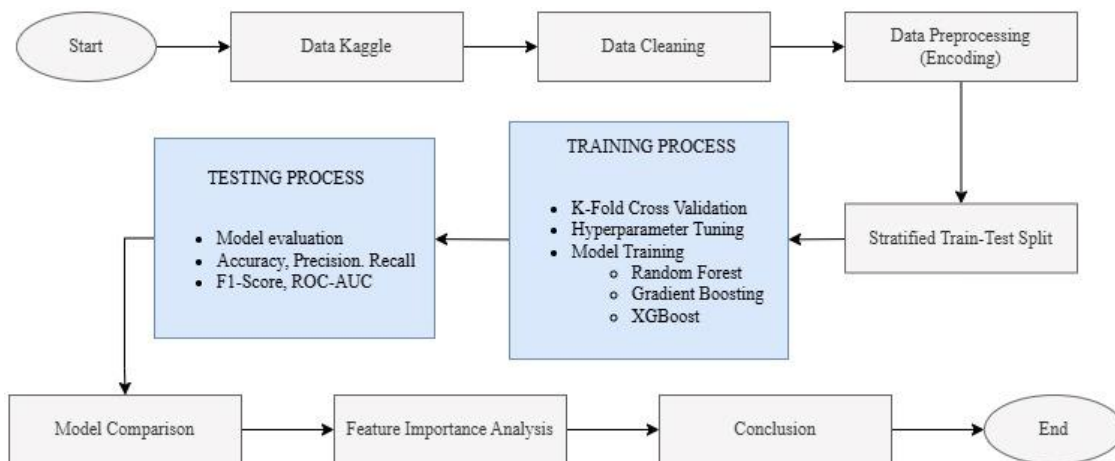
2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis *supervised machine learning* untuk menyelesaikan permasalahan klasifikasi kepuasan penumpang maskapai penerbangan. Permasalahan yang dikaji termasuk ke dalam kategori *binary classification*, karena variabel target berupa tingkat kepuasan penumpang, yang terdiri dari dua kelas, yaitu *satisfied* dan *neutral or dissatisfied*.

Pada penelitian ini, beberapa model *ensemble learning* dibandingkan untuk menganalisis performa klasifikasi. Model yang digunakan adalah *Random Forest*, *Gradient Boosting*, dan XGBoost. Ketiga model tersebut dipilih karena mewakili dua pendekatan utama dalam *ensemble learning*, yaitu *bagging* dan *boosting*, yang secara konseptual dirancang untuk meningkatkan stabilitas dan akurasi prediksi dibandingkan model tunggal.

Penelitian ini juga menggunakan skema eksperimental terkontrol dengan pemisahan data ke dalam bentuk data pelatihan (*training set*) dan data pengujian (*testing set*). Proses pelatihan menggunakan teknik *cross-validation* dan optimasi hiperparameter untuk memperoleh konfigurasi model terbaik. Evaluasi akhir dilakukan menggunakan data pengujian yang tidak terlibat dalam proses pelatihan untuk memastikan kemampuan generalisasi model.

Dengan desain eksperimental ini, penelitian bertujuan untuk membandingkan performa model *ensemble learning* dalam klasifikasi kepuasan penumpang serta mengidentifikasi variabel yang paling berkontribusi terhadap prediksi model. Selain itu, penelitian ini juga membandingkan performa model berdasarkan beberapa metrik evaluasi klasifikasi, serta mengidentifikasi variabel yang paling berpengaruh terhadap prediksi kepuasan penumpang. Dengan desain ini, penelitian ini tidak hanya menilai akurasi model tetapi juga memberikan interpretasi terhadap faktor-faktor yang berkontribusi terhadap kepuasan pelanggan.



Gambar 1. Diagram Alur Penelitian

2.2. Dataset

Dataset yang digunakan dalam penelitian ini merupakan *Airline Passenger Satisfaction Dataset* yang diperoleh dari platform Kaggle. Dataset ini berisi data survey kepuasan penumpang maskapai penerbangan yang mencakup karakteristik demografis, informasi perjalanan, serta penilaian terhadap berbagai aspek layanan penerbangan.

Setelah dilakukan proses pembersihan data dengan menghapus kolom yang tidak relevan, seperti kolom identitas (id) dan indeks tambahan, serta menghilangkan observasi yang memiliki nilai kosong, diperoleh sebanyak 25.893 observasi yang siap digunakan dalam proses pemodelan. Dataset akhir terdiri dari 23 variabel prediktor dengan satu variabel target.

Secara umum, variabel dalam dataset mempresentasikan tiga kelompok utama, yaitu karakteristik penumpang, karakteristik perjalanan, serta penilaian terhadap kualitas layanan. Karakteristik penumpang meliputi variabel seperti jenis kelamin dan usia, sedangkan karakteristik perjalanan mencakup tipe pelanggan, tipe perjalanan (*business* atau *personal*), kelas penerbangan, serta jarak penerbangan. Selain itu, sebagian besar variabel dalam dataset merupakan penilaian layanan yang diukur menggunakan skala ordinal 1 sampai 5, yang mencerminkan persepsi penumpang terhadap kualitas layanan tertentu, seperti *inflight wifi service*, *online boarding*, *seat comfort*, *inflight entertainment*, hingga *cleanliness*. Dataset juga memuat informasi terkait keterlambatan keberangkatan dan kedatangan dalam satuan menit.

Pada penelitian ini, variabel target terletak pada *satisfaction*, yang terdiri dari dua kategori, yaitu *satisfied* dan *neutral or dissatisfied*. Dengan demikian, permasalahan yang dikaji termasuk dalam kategori klasifikasi biner (*binary classification problem*).

Analisis awal terhadap distribusi kelas menunjukkan bahwa sebanyak 56,11% observasi termasuk dalam kategori *neutral or dissatisfied*, sedangkan 43,89% lainnya termasuk dalam kategori *satisfied*. Perbedaan proporsi antar kelas relatif tidak ekstrim, sehingga dataset dapat dikategorikan sebagai relatif seimbang. Berdasarkan kondisi tersebut, penelitian ini tidak menerapkan penanganan ketidakseimbangan kelas dalam tahap pemodelan.

2.3. Data Preprocessing

Tahap *preprocessing* data dilakukan untuk memastikan bahwa dataset berada dalam format yang sesuai untuk pemodelan berbasis *machine learning*. Proses ini mencakup transformasi variabel kategorikal, pemisahan data pelatihan dan pengujian, serta persiapan fitur sebelum proses pelatihan model dilakukan.

Variabel target *satisfaction* terlebih dahulu dikonversi ke dalam bentuk numerik untuk kebutuhan klasifikasi biner, di mana kategori *neutral or dissatisfied* direpresentasikan sebagai 0 dan *satisfied* direpresentasikan sebagai 1. transformasi ini dilakukan agar model klasifikasi dapat memproses target dalam format numerik.

Selanjutnya, variabel prediktor dipisahkan menjadi dua kelompok berdasarkan tipe datanya, yaitu variabel kategorikal dan variabel numerik. Variabel kategorikal seperti *Gender, Customer Type, Type of Travel, dan Class* ditransformasikan menggunakan teknik *One-Hot Encoding*. Teknik ini mengubah setiap kategori menjadi representasi biner tanpa mengasumsikan adanya hubungan ordinal antar kategori. Pendekatan ini dipilih karena *tree-based models*, seperti *Random Forest* dan *Gradient Boosting*, dapat bekerja secara efektif dengan fitur hasil encoding [8], [9].

Sementara itu, variabel numerik seperti usia, jarak penerbangan, skor layanan, serta informasi keterlambatan, tetap dipertahankan dalam bentuk aslinya tanpa dilakukan normalisasi. Hal ini dikarenakan *tree-based models* tidak sensitif terhadap skala variabel, sehingga proses standarisasi atau normalisasi tidak menjadi kebutuhan utama dalam penelitian ini.

Setelah proses transformasi fitur ditentukan, selanjutnya dataset dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) menggunakan metode *stratified train-test split* dengan proposisi sebesar 80:20 untuk memastikan bahwa proposisi masing-masing kelas target (*neutral or dissatisfied* dan *satisfied*) tetap konsisten pada data pelatihan dan pengujian. Metode *stratified* digunakan agar setiap subset mencerminkan distribusi kelas yang sama seperti pada dataset awal, sehingga evaluasi performa model tidak terdistorsi oleh perbedaan komposisi kelas. Metode ini juga membantu mengurangi bias akibat representasi kelas yang tidak proporsional di masing-masing subset, sehingga hasil evaluasi model lebih robust dan dapat digeneralisasi [10]. Proses fitting transformasi *encoding* dilakukan hanya pada data pelatihan, kemudian transformasi yang sama diterapkan pada data pengujian. Pendekatan ini bertujuan untuk mencegah terjadinya *data leakage*, yaitu kondisi di mana informasi dari data pengujian secara tidak sengaja digunakan dalam proses pelatihan model [11], [12], [13].

Seluruh tahapan *preprocessing* dibungkus dalam pipeline dan diterapkan di dalam skema *cross-validation*, sehingga setiap fold hanya menggunakan informasi subset training. Studi terkait [14] mengemukakan bukti teoritis dan empiris terhadap dampak negatif *preprocessing* yang dilakukan pada keseluruhan dataset serta rekomendasi dalam menggunakan pipeline.

2.4. Hyperparameter Tuning

Proses optimasi *hyperparameter* dilakukan menggunakan metode *Grid Search* dengan skema *Stratified K-Fold Cross Validation* sebanyak 5 fold. Pendekatan ini bertujuan untuk menemukan kombinasi parameter terbaik yang mampu memaksimalkan performa model berdasarkan metrik F1-Score.

Tabel 1. *Hyperparameter Tuning Range*

Model	Parameter	Nilai yang Diuji
Random Forest	n_estimators	100, 200
	max_depth	None, 10
Gradient Boosting	n_estimators	100, 200
	learning_rate	0.05, 0.1
	n_estimators	100, 200
XGBoost	max_depth	3, 5
	learning_rate	0.05, 0.1

Tabel 1 menunjukkan rentang *hyperparameter* yang diuji pada setiap model. Pemilihan rentang parameter didasarkan pada praktik umum dalam literatur serta karakteristik data tabular yang digunakan dalam penelitian ini.

Tabel 2. Konfigurasi *Hyperparameter* Terbaik Hasil *Grid Search*

Model	Parameter Terbaik
Random Forest	n_estimators = 200, max_depth = 10
Gradient Boosting	n_estimators = 200, learning_rate = 0.1
XGBoost	n_estimators = 200, max_depth = 5, learning_rate = 0.1

Berdasarkan hasil *Grid Search*, diperoleh kombinasi *hyperparameter* terbaik untuk masing-masing model sebagaimana ditunjukkan pada Tabel 2. Parameter optimal ini dipilih berdasarkan nilai F1-Score tertinggi pada proses *cross-validation*.

Pada model XGBoost, kombinasi *n_estimators* = 200, *max_depth* = 5, dan *learning_rate* = 0.1 menghasilkan performa terbaik. Konfigurasi ini menunjukkan bahwa model dengan kompleksitas yang lebih tinggi dan laju pembelajaran yang relatif besar mampu menangkap pola non-linier secara lebih efektif pada data kepuasan penumpang.

Sementara itu, pada *Random Forest* dan *Gradient Boosting*, peningkatan jumlah estimator juga berkontribusi terhadap skalabilitas dan akurasi model, karena semakin banyak pohon yang dibangun akan mengurangi varians prediksi dan meningkatkan generalisasi model.

2.5. Model Ensemble Learning

Model *ensemble learning* merupakan pendekatan dalam *machine learning* yang menggabungkan prediksi dari beberapa model dasar (*base learners*) untuk menghasilkan prediksi akhir yang lebih akurat dan stabil dibandingkan model tunggal. Dalam *ensemble learning*, beberapa model dilatih secara independen terhadap data yang sama atau berbeda, kemudian hasil prediksinya digabungkan melalui strategi agregasi untuk tugas klasifikasi. Secara intuitif, kombinasi antara berbagai model dapat saling meniadakan kesalahan individu dan meningkatkan kemampuan generalisasi model secara keseluruhan. Penelitian komprehensif [15] yang membahas mengenai konsep dan algoritma *ensemble learning* menyatakan bahwa teknik-teknik seperti *bagging* dan *boosting* termasuk metode yang paling banyak digunakan dalam berbagai aplikasi *machine learning* modern, khususnya pada klasifikasi data tabular.

Representasi umum dari pendekatan *ensemble learning* dapat ditulis sebagai agregasi prediksi dari *M* model dasar:

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M f_m(x) \tag{1}$$

Di mana \hat{y} adalah prediksi akhir, $f_m(x)$ adalah prediksi dari model ke-*m*, dan *M* merupakan jumlah model dalam *ensemble learning*. Pendekatan ini secara teoritis bertujuan untuk mereduksi varians dan meningkatkan stabilitas prediksi, terutama ketika model-model dasar memiliki karakteristik kesalahan yang berbeda [15].

2.5.1. Random Forest (Bagging)

Random Forest merupakan salah satu metode *bagging* yang paling umum digunakan. Cara kerja metode ini adalah dengan membangun banyak *decision tree* (pohon keputusan) pada subset berbeda dari data pelatihan dengan *bootstrap sampling*. Setiap pohon keputusan menghasilkan prediksi, dan prediksi akhir akan ditentukan melalui *majority voting* pada kelas mayoritas [15], [16]. Perbedaan antar pohon keputusan yang dihasilkan melalui *bagging* serta pemilihan fitur acak menjadikan *Random Forest* lebih stabil dan tahan terhadap *overfitting* dibandingkan dengan satu pohon tunggal. Studi empiris terbaru [16], [17] menekankan bahwa *Random Forest* tetap menjadi salah satu teknik *ensemble* yang andal untuk tugas klasifikasi dan regresi dengan tabular data dengan sering kali menunjukkan akurasi tinggi dalam evaluasi performa.

Pada *Random Forest*, rumus *majority voting* umumnya ditulis sebagai persamaan:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_M(x)\} \tag{2}$$

Di mana \hat{y} adalah prediksi akhir, $T_m(x)$ adalah prediksi dari pohon keputusan ke-*m* terhadap vektor fitur *x*, *mode*(\cdot) adalah fungsi yang memilih kelas yang paling sering muncul *majority voting* di antara seluruh prediksi pohon, dan *M* adalah jumlah pohon keputusan dalam *ensemble learning*. Pada tugas klasifikasi biner, setiap

pohon menghasilkan prediksi kelas, misalnya 0 dan 1, dan kelas dengan frekuensi tertinggi di antara seluruh pohon menjadi prediksi akhir model.

2.5.2. Gradient Boosting dan XGBoost (Boosting)

Pendekatan *boosting* merupakan strategi *ensemble learning* di mana model dasar dibangun secara berurutan, dengan setiap model mencoba memperbaiki kesalahan prediksi model sebelumnya. *Gradient Boosting* adalah varian *boosting* yang fokus pada optimasi fungsi kerugian (*loss function*) melalui penyesuaian residual dari model sebelumnya. Dalam konteks *boosting*, model akhir dibentuk dengan menjumlahkan prediksi model dasar secara bertahap sehingga kesalahan keseluruhan dapat dikurangi. XGBoost atau biasa disebut juga sebagai *Extreme Gradient Boosting* merupakan implementasi lanjutan dari *Gradient Boosting* yang menggabungkan regulasi dan optimasi lebih efisien untuk meningkatkan performa serta mengurangi *overfitting*. Pendekatan ini telah digunakan secara luas dan terbukti efektif dalam tugas klasifikasi yang kompleks [18], [19].

Secara praktis, *Gradient Boosting* membangun model aktif secara iteratif:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (3)$$

Di mana $h_m(x)$ adalah *weak learner* pada iterasi ke- m dan γ_m adalah *learning rate* yang mengontrol kontribusi model tersebut terhadap prediksi akhir. Rumus ini menunjukkan bagaimana *boosting* secara bertahap menambahkan model yang berfokus pada kesalahan yang tersisa dari model sebelumnya, sehingga model akhir mampu menangkap pola non-linier yang kompleks dari data [18].

2.5.3. Justifikasi Pemilihan Model

Pemilihan *Random Forest*, *Gradient Boosting*, dan XGBoost dalam studi ini didasarkan pada bukti literatur yang menunjukkan bahwa kedua pendekatan tersebut, baik *bagging* maupun *boosting*, secara konsisten memberikan performa superior dalam berbagai tugas klasifikasi. Evaluasi komparatif dalam studi terbaru menunjukkan bahwa model-model seperti *Random Forest* dan XGBoost sering kali menghasilkan AUC dan F1-score yang tinggi dalam skema *cross-validation*, hal ini menegaskan manfaat *ensemble learning* dibandingkan dengan model tunggal tradisional. Penggunaan berbagai metrik evaluasi dalam penelitian *ensemble learning* memastikan bahwa keputusan pemilihan model tidak hanya berdasarkan satu metrik saja, tetapi mencerminkan kemampuan model secara menyeluruh dalam menghadapi kompleksitas nyata [17].

2.6. Metode Evaluasi Model

Untuk menilai performa model klasifikasi yang diuji dalam penelitian ini, digunakan beberapa metrik evaluasi yang umum dan telah terbukti representatif dalam literatur *machine learning*. Metrik-metrik ini memberikan gambaran menyeluruh terhadap kemampuan model dalam mengklasifikasikan setiap kelas, terutama dalam tugas klasifikasi biner.

2.6.1. Confusion Matrix

Confusion matrix merupakan dasar dari sebagian besar metrik evaluasi yang menunjukkan hubungan antara prediksi model dengan label aktual. *Confusion matrix* ini selanjutnya menjadi dasar perhitungan semua metrik.

$$\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$$

Di mana: [20]

- **TP** (*True Positive*) adalah jumlah prediksi benar
- **TN** (*True Negative*) adalah jumlah prediksi negatif yang benar
- **FP** (*False Positive*) adalah jumlah prediksi positif yang salah
- **FN** (*False Negative*) adalah jumlah prediksi negatif yang salah

2.6.2. Accuracy

Akurasi merupakan proporsi total prediksi benar terhadap seluruh observasi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Accuracy memberikan gambaran umum tentang seberapa sering model menebak label dengan benar, namun dapat menyesatkan jika distribusi kelas tidak seimbang [21].

2.6.3. Precision and Recall

Precision mencerminkan proporsi dari *true positive* di antara semua prediksi positif yang dibuat model

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Sementara *Recall* yang disebut juga sebagai *Sensitivity* menunjukkan kemampuan model dalam menemukan semua *instance* positif.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Precision dan *Recall* merupakan metrik penting ketika kesalahan tertentu (misalnya *false positive* atau *false negative*) memiliki implikasi berbeda secara kontekstual, seperti dalam evaluasi diagnosis atau kepuasan pelanggan [21].

2.6.4. F1-Score

F1-Score adalah harmonisasi dari *precision* dan *recall*, memberikan satu angka yang mencerminkan keseimbangan di antara keduanya.

$$F1 - Score = \frac{2TP}{2TP+FP+FN} \quad (7)$$

F1-Score sering digunakan ketika distribusi kelas tidak sepenuhnya seimbang dan diperlukan keseimbangan antara kemampuan deteksi positif dan tingkat kesalahan. Semakin dekat nilai F1 dengan 1, menyatakan bahwa keseimbangan optimal antara *precision* dan *recall* [21].

2.6.5. ROC-AUC

Receiver Operating Characteristic (ROC) adalah kurva yang memplot *true positive rate* dengan *false positive rate* pada berbagai ambang keputusan yang berbeda. Luas area di bawah kurva tersebut (*Area Under Curve / AUC*) memberikan ukuran kapasitas model dalam membedakan antara kelas positif dan negatif tanpa bergantung pada satu threshold tertentu.

AUC memberikan ukuran seberapa baik model memisahkan kelas dalam konteks semua threshold yang mungkin. Metrik ini sangat berguna karena tidak tergantung pada threshold tunggal dan merupakan indikator konsisten performa klasifikasi biner [20], [21].

3. HASIL DAN PEMBAHASAN

3.1. Karakteristik Data dan Distribusi Kelas

Dataset yang digunakan dalam penelitian ini terdiri dari 25.976 observasi dengan 25 atribut pada tahap awal. Setelah dilakukan proses pembersihan data dengan menghapus kolom yang tidak relevan serta observasi yang mengandung nilai kosong, jumlah data yang siap digunakan dalam pemodelan menjadi 25.893 observasi dengan 23 variabel prediktor dan satu target variabel. Proses ini memastikan bahwa data yang digunakan dalam eksperimen berada dalam kondisi bersih dan layak untuk dianalisis lebih lanjut.

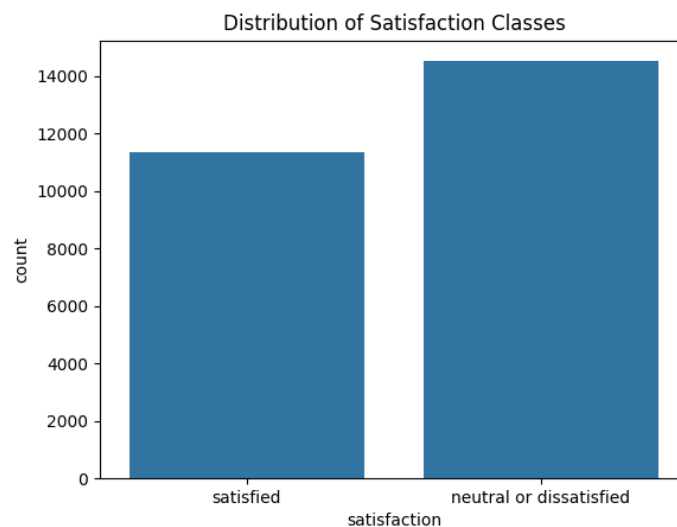
Tabel 3. Ringkasan Dataset Sebelum dan Sesudah Data Cleaning

Keterangan	Nilai
Jumlah observasi awal	25.976
Jumlah observasi setelah cleaning	25.893
Jumlah variabel prediktor	23
Jumlah variabel target	1
Total variabel akhir	24

Distribusi kelas pada variabel target pada Tabel 4 di bawah menunjukkan bahwa kategori *neutral or dissatisfied* berjumlah 14.528 observasi atau sekitar 56,11%, sedangkan *satisfied* berjumlah 11.365 observasi atau sekitar 43,89%.

Tabel 4. Distribusi Kelas Target

Kelas	Jumlah	Persentase (%)
Neutral or Dissatisfied	14.528	56,11
Satisfied	11.365	43,89
Total	25.893	100



Gambar 1. Diagram Sebaran Kategori Variabel Target

Visualisasi pada Gambar 1 memperlihatkan bahwa meskipun terdapat perbedaan jumlah antar kelas, proporsi tersebut tidak menunjukkan ketimpangan yang ekstrem. Secara keseluruhan, tahap ini menunjukkan bahwa struktur dataset dan distribusi kelas telah memenuhi prasyarat dasar untuk penerapan model klasifikasi biner berbasis *ensemble learning*.

3.2. Analisis Performa Model *Ensemble Learning*

Tabel 5. Perbandingan Performa Model *Ensemble Learning*

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.961	0.963	0.948	0.955	0.994
Gradient Boosting	0.954	0.954	0.941	0.948	0.992
XGBoost	0.963	0.966	0.949	0.957	0.995

Perbandingan performa ketiga model *ensemble learning* ditampilkan pada Tabel 5. Secara umum, seluruh model menunjukkan kinerja yang sangat tinggi dengan masing-masing nilai akurasi (*accuracy*) di atas 0,95 dengan nilai ROC-AUC yang sangat tinggi menunjukkan kemampuan model dalam membedakan kedua kelas secara konsisten. Hal ini mengindikasikan bahwa pendekatan *ensemble learning* mampu menangkap pola hubungan antara fitur layanan dan tingkat kepuasan penumpang secara efektif.

Berdasarkan Tabel 5, XGBoost menghasilkan performa terbaik dengan nilai akurasi sebesar 0,963 dengan F1-Score sebesar 0,957. Keunggulan ini juga tercermin pada nilai ROC-AUC sebesar 0,995, yang merupakan nilai tertinggi di antara ketiga model. Nilai ROC-AUC yang sangat tinggi ini menunjukkan bahwa model memiliki kemampuan diskriminatif yang sangat baik dalam membedakan kelas *satisfied* dan *neutral or dissatisfied* pada berbagai ambang keputusan. Temuan ini sejalan dengan beberapa studi sebelumnya [22], [23] yang menyatakan bahwa XGBoost unggul dalam memberikan performa pada data tabular dibandingkan dengan metode lain. Keunggulan XGBoost dibandingkan model lainnya dapat dijelaskan secara teknis melalui mekanisme regulasi (L1 dan L2) serta optimasi berbasis *gradient* yang lebih efisien. XGBoost juga

menggunakan teknik *shrinkage* dan *column subsampling* yang membantu mengurangi *overfitting* serta meningkatkan kemampuan generalisasi model. Selain itu, pendekatan *boosting* yang membangun model secara iteratif memungkinkan XGBoost untuk secara efektif memperbaiki kesalahan residual pada setiap iterasi, sehingga menghasilkan performa yang lebih optimal pada data dengan pola non-linier kompleks.

Random Forest menunjukkan performa yang sangat kompetitif dengan akurasi sebesar 0,961 dan F1-Score sebesar 0,955. Selisih performa *Random Forest* dan XGBoost relatif kecil, yang menunjukkan bahwa kedua pendekatan *ensemble learning* tersebut sama-sama efektif dalam menangani karakteristik data tabular pada penelitian ini. Stabilitas *Random Forest* dapat dikaitkan dengan mekanisme *bagging* yang mereduksi varians melalui banyak pohon keputusan. Hal ini konsisten dengan temuan peneliti sebelumnya [24], [25] yang menunjukkan bahwa *Random Forest* umumnya memberikan performa yang lebih unggul dan stabil dibandingkan algoritma tunggal seperti *Decision Tree* dalam klasifikasi, terutama dari aspek akurasi dan konsistensi prediksi.

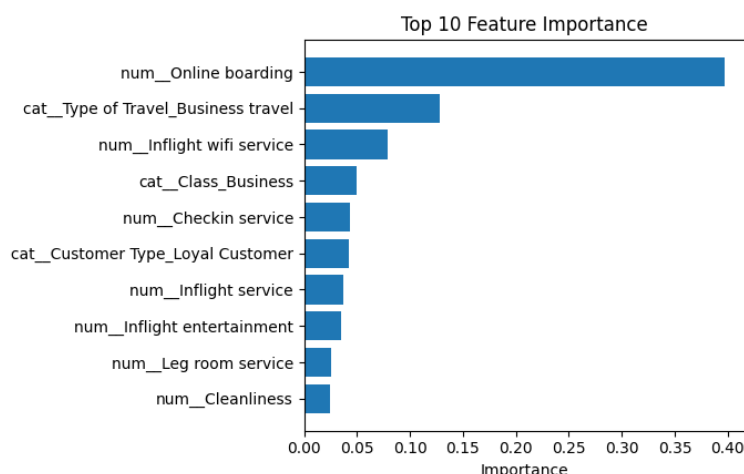
Sementara itu, *Gradient Boosting* menghasilkan performa sedikit lebih rendah dibandingkan dua model lainnya, dengan akurasi sebesar 0,954 dan F1-Score sebesar 0,948. Meskipun demikian, nilai tersebut masih berada pada kategori sangat baik. Perbedaan performa ini mengindikasikan bahwa model *boosting* dasar mungkin lebih sensitif terhadap konfigurasi parameter dibandingkan implementasi *boosting* yang telah dioptimasi seperti XGBoost. Perbedaan performa ini juga mengindikasikan bahwa *boosting* dasar sangat dipengaruhi oleh konfigurasi *hyperparameter* seperti jumlah pohon (*n_estimator*), *learning rate*, dan kedalaman maksimum (*max_depth*), yang dapat memengaruhi kemampuan model dalam menangkap hubungan non-linier pada data. Penelitian sebelumnya [26], [27] juga menyatakan bahwa sejumlah *hyperparameter* tersebut memengaruhi efektivitas *Gradient Boosting* pada tugas klasifikasi, sehingga pemilihan nilai parameter yang optimal merupakan faktor penting dalam pencapaian performa yang baik. Jika ditinjau dari keseimbangan antara *precision* dan *recall*, ketiga model menunjukkan nilai yang relatif seimbang tanpa adanya perbedaan ekstrem antara kedua metrik tersebut. Hal ini menunjukkan bahwa model tidak hanya unggul dalam memaksimalkan akurasi keseluruhan, tetapi juga konsisten dalam mengidentifikasi kedua kelas secara proporsional.

3.3. Analisis Feature Importance dan Faktor Penentu Kepuasan

Untuk memahami faktor-faktor yang paling berpengaruh dalam menentukan kepuasan penumpang, analisis dilakukan menggunakan nilai *feature importance* dari model terbaik, yaitu XGBoost. Nilai *importance* mencerminkan kontribusi relatif masing-masing variabel dalam meningkatkan kualitas pemisahan kelas selama proses pembentukan pohon keputusan dalam *boosting*. Semakin tinggi nilai *importance* suatu fitur, semakin besar peran fitur tersebut dalam proses prediksi model. Hasil perhitungan menunjukkan bahwa terdapat perbedaan kontribusi yang cukup signifikan antar variabel. Tabel 6 di bawah menyajikan sepuluh fitur dengan nilai *importance* tertinggi berdasarkan model XGBoost.

Tabel 6. Sepuluh Fitur dengan Nilai *Importance* Tertinggi

Peringkat	Fitur	Importance
1	Online boarding	0.396759
2	Type of Travel (Business travel)	0.127844
3	Inflight wifi service	0.079151
4	Class (Business)	0.049538
5	Checkin service	0.043497
6	Customer Type (Loyal Customer)	0.042313
7	Inflight service	0.036863
8	Inflight entertainment	0.034487
9	Leg room service	0.025165
10	Cleanliness	0.024007



Gambar 2. Visualisasi Sepuluh Fitur dengan Nilai *Importance* Tertinggi pada Model XGBoost.

Selain tabel, visualisasi distribusi nilai *importance* ditampilkan pada Gambar 2. Grafik tersebut memperlihatkan dominasi variabel *Online boarding* yang memiliki kontribusi jauh lebih besar dibandingkan dengan fitur lainnya. Gambar 2 menunjukkan distribusi nilai *feature importance* yang bersifat tidak merata, di mana satu fitur memiliki dominasi signifikan dibandingkan fitur lainnya. Hal ini menunjukkan adanya fitur kunci (*dominant predictor*) dalam model.

3.3.1 Analisis Teknis

Secara teknis, dominasi variabel *Online boarding* menunjukkan bahwa model secara konsisten menggunakan fitur ini sebagai pemisah utama dalam pembentukan pohon keputusan. Dalam mekanisme *boosting*, fitur yang sering digunakan pada node pemisahan awal dengan peningkatan gain tertinggi akan memperoleh nilai *importance* yang besar. Nilai *importance* sebesar 0,396 menunjukkan bahwa variabel ini memiliki kontribusi relatif paling dominan dibandingkan dengan fitur lainnya dalam proses pembentukan model, yang mengindikasikan bahwa pengalaman proses boarding digital memiliki peran signifikan dalam klasifikasi kepuasan. Fitur lain seperti *Business travel*, *Class (Business)*, serta *Customer Type (Loyal Customer)* menunjukkan bahwa segmentasi penumpang turut memengaruhi prediksi model. Hal ini berarti bahwa kepuasan tidak hanya dipengaruhi oleh kualitas layanan, tetapi juga karakteristik demografis dan perilaku pelanggan.

3.3.2 Analisis Interpretatif dan Implikasi

Dari perspektif substantif, temuan ini menunjukkan bahwa aspek digital sebelum penerbangan, khususnya proses *boarding* memiliki pengaruh yang kuat dalam pemisahan kelas kepuasan penumpang pada model klasifikasi. Dominasi *Online boarding* dan *Inflight wifi service* mengindikasikan bahwa pengalaman berbasis teknologi menjadi komponen penting dalam membentuk kepuasan pelanggan maskapai modern.

Selain itu, munculnya variabel *Business travel* dan *Class (Business)* sebagai faktor penting menunjukkan bahwa penumpang dengan tujuan bisnis atau kelas premium memiliki ekspektasi layanan yang berbeda dibandingkan penumpang lainnya. Dengan demikian, strategi peningkatan layanan tidak dapat bersifat seragam, melainkan perlu mempertimbangkan segmentasi pelanggan. Variabel seperti *Checkin service*, *Inflight service*, *Inflight entertainment*, *Leg room service*, dan *Cleanliness* menunjukkan bahwa kualitas pengalaman selama penerbangan tetap menjadi faktor signifikan, meskipun kontribusinya lebih rendah dibandingkan aspek digital dan segmentasi pelanggan.

Secara keseluruhan, hasil ini mengindikasikan bahwa kepuasan penumpang maskapai merupakan hasil kombinasi antara kualitas layanan operasional, pengalaman digital, dan karakteristik pelanggan. Model XGBoost tidak hanya memberikan performa klasifikasi terbaik, tetapi juga membantu mengidentifikasi faktor-faktor kunci yang dapat dijadikan dasar pengambilan keputusan strategis bagi maskapai.

3.4. Pembahasan

Hasil analisis *feature importance* menunjukkan bahwa variabel *Online boarding* memiliki kontribusi paling dominan dalam menentukan kepuasan penumpang. Temuan tersebut sejalan dengan penelitian terdahulu [28], yang menyatakan bahwa teknologi digital yang diadopsi oleh maskapai memiliki pengaruh signifikan terhadap kepuasan penumpang. Dalam hasil regresi yang dilaporkan [28], enam teknologi digital yang diterapkan

maskapai terbukti secara statistik memengaruhi kepuasan pelanggan. Hal ini menunjukkan bahwa transformasi digital dalam layanan penerbangan bukan sekedar inovasi operasional, tetapi juga determinan penting dalam pembentukan kepuasan pelanggan. Selain aspek digital, variabel yang berkaitan dengan pengalaman selama penerbangan seperti *Inflight service*, *Inflight entertainment*, dan *Cleanliness* juga termasuk dalam sepuluh fitur dengan nilai *importance* tertinggi. Temuan ini konsisten dengan meta-analisis yang dilakukan pada studi [29], yang menganalisis 95 studi empiris dan menemukan bahwa kualitas layanan memiliki pengaruh positif yang signifikan terhadap kepuasan penumpang. Secara khusus, dalam studi tersebut dinyatakan bahwa dimensi *inflight service* memiliki efek paling besar dan bersifat krusial terhadap kepuasan penumpang. Konsistensi ini memperkuat bahwa variabel layanan selama penerbangan yang teridentifikasi dalam model XGBoost memang memiliki dasar empiris yang kuat dalam literatur.

Variabel segmentasi seperti *Type of Travel (Business travel dan Class IBusiness)* juga menunjukkan kontribusi yang signifikan dalam model. Hal ini mengindikasikan bahwa karakteristik perjalanan dan kategori pelanggan memengaruhi persepsi terhadap kualitas layanan. Dalam konteks industri penerbangan, pelanggan dengan tujuan bisnis atau kelas premium umumnya memiliki ekspektasi layanan yang lebih tinggi terkait efisiensi, kenyamanan, dan kualitas pengalaman secara keseluruhan. Literatur mengenai kepuasan penumpang maskapai [30] secara konsisten menunjukkan bahwa persepsi terhadap kualitas layanan (*service quality*) dan nilai yang dirasakan (*perceived value*) merupakan faktor penting dalam pembentukan kepuasan pelanggan maskapai, yang selanjutnya berdampak pada perilaku lanjutan pelanggan seperti loyalitas, niat penggunaan ulang, dan komunikasi positif (*word-of-mouth*). Selaras dengan itu, studi lain [31] juga menemukan bahwa kualitas layanan memiliki pengaruh positif terhadap kepuasan penumpang, dan kepuasan tersebut memiliki kontribusi yang signifikan terhadap pembentukan perilaku pelanggan seperti *brand love* dan *word-of-mouth*. Hubungan ini memperlihatkan bahwa kepuasan bukan hanya hasil akhir dari pengalaman layanan, tetapi juga menjadi mediator penting yang menghubungkan kualitas layanan dengan loyalitas dan perilaku penggunaan ulang.

Secara metodologis, dominasi beberapa fitur tertentu dalam model XGBoost juga memperlihatkan kemampuan pendekatan *boosting* dalam mengidentifikasi variabel yang memberikan peningkatan gain terbesar dalam proses pembentukan pohon keputusan. Model tidak hanya mencapai performa klasifikasi yang tinggi, tetapi juga memberikan wawasan yang dapat diinterpretasikan mengenai faktor-faktor kunci yang memengaruhi kepuasan penumpang. Secara keseluruhan, hasil penelitian ini memperkuat temuan literatur sebelumnya yang menyatakan bahwa kepuasan penumpang maskapai dipengaruhi oleh kombinasi antara kualitas layanan dalam penerbangan, pengalaman digital sebelum dan selama perjalanan, serta karakteristik pelanggan. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada aspek komparasi model *ensemble learning*, tetapi juga memberikan bukti empiris tambahan mengenai determinan utama kepuasan pelanggan dalam konteks industri penerbangan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa model *ensemble learning*, khususnya XGBoost, memberikan performa terbaik dalam klasifikasi kepuasan penumpang maskapai penerbangan berdasarkan berbagai metrik evaluasi. Selain itu, analisis *feature importance* mengidentifikasi bahwa faktor pengalaman digital, kualitas layanan dalam penerbangan, serta segmentasi pelanggan merupakan determinan utama kepuasan penumpang. Secara praktis, hasil penelitian ini memberikan implikasi bahwa maskapai penerbangan perlu memprioritaskan peningkatan kualitas layanan digital serta pengalaman kabin untuk meningkatkan kepuasan pelanggan. Selain itu, penelitian terhadap segmen penumpang bisnis dan kelas premium dapat membantu meningkatkan persepsi kualitas layanan secara keseluruhan. Dari sisi metodologis, penelitian ini menunjukkan bahwa model *ensemble learning*, khususnya XGBoost, tidak hanya efektif dalam klasifikasi kepuasan pelanggan, tetapi juga mampu memberikan wawasan interpretatif mengenai faktor-faktor penentu kepuasan. Dengan demikian, penelitian ini berkontribusi pada pengembangan pendekatan berbasis *machine learning* dalam analisis kepuasan pelanggan di industri penerbangan. Sebagai keterbatasan, penelitian ini menggunakan dataset sekunder yang bersifat publik sehingga karakteristik populasi mungkin tidak sepenuhnya merepresentasikan kondisi maskapai tertentu. Penelitian selanjutnya dapat mempertimbangkan penggunaan data aktual dari maskapai tertentu, eksplorasi metode *explainable AI* yang lebih mendalam, serta pengujian model pada konteks industri atau wilayah berbeda. Selain itu, untuk memperkuat perbandingan performa model, penelitian selanjutnya dapat mempertimbangkan penggunaan uji statistik seperti *paired t-test* atau *Wilcoxon signed-rank test* terhadap hasil *cross-validation*.

DAFTAR PUSTAKA

- [1] S. Chung, J. W. Park, and S. Lee, "The Influence of CSR on Airline Loyalty through the Mediations of Passenger Satisfaction, Airline Brand, and Airline Trust: Korean Market Focused," *Sustainability (Switzerland)*, vol. 14, no. 8, Apr. 2022, doi: 10.3390/su14084548.
- [2] J. Wang, J. Wu, S. Sun, and S. Wang, "The relationship between attribute performance and customer satisfaction: An interpretable machine learning approach," *Data Science and Management*, vol. 7, no. 3, pp. 164–180, Sep. 2024, doi: 10.1016/j.dsm.2024.01.003.
- [3] S. Li, B. Zhu, Y. Zhang, F. Liu, and Z. Yu, "A Two-Stage Nonlinear User Satisfaction Decision Model Based on Online Review Mining: Considering Non-Compensatory and Compensatory Stages," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, no. 1, pp. 272–296, Mar. 2024, doi: 10.3390/jtaer19010015.
- [4] A. I. Fadri, A. Zahfran, T. Irak, N. H. Firjatullah, and J. E. Herianto, "Comparison of Supervised Learning Algorithms for Predicting Airline Passenger Satisfaction," *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, vol. 2, no. 1, pp. 42–52, Mar. 2025, doi: 10.57152/ijatis.v2i1.1868.
- [5] J. Suwanto *et al.*, "Comparison of Classification Algorithm in Classifying Airline Passenger Satisfaction," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 6, no. 1, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v6i1.2848.
- [6] R. Murugesan, R. A P, N. N, and R. Balanathan, "Forecasting airline passengers' satisfaction based on sentiments and ratings: An application of VADER and machine learning techniques," *J. Air Transp. Manag.*, vol. 120, Sep. 2024, doi: 10.1016/j.jairtraman.2024.102668.
- [7] B. Laksono, I. Kurniawati, A. B. Sriwiyanta, Z. Zen Zaenudin, J. A. Ramadha, and D. Alfian, "Integration of SMOTE and Ensemble Models for Predicting Airline Passenger Satisfaction," *Innovation in Research of Informatics (INNOVATICS)*, vol. 7, no. 1, pp. 77–85, 2025, doi: 10.37058/innovatics.v7i1.14001.
- [8] W. Zhu, R. Qiu, and Y. Fu, "Comparative Study on the Performance of Categorical Variable Encoders in Classification and Regression Tasks," *ArXiv*, Jan. 2024, doi: 10.48550/arXiv.2401.09682.
- [9] F. Bolikulov, R. Nasimov, A. Rashidov, F. Akhmedov, and Y. I. Cho, "Effective Methods of Categorical Data Encoding for Artificial Intelligence Algorithms," *Mathematics*, vol. 12, no. 16, Aug. 2024, doi: 10.3390/math12162553.
- [10] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off between training and testing ratio in machine learning for medical image processing," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2245.
- [11] Q. Dong, "Leakage Prediction in Machine Learning Models When Using Data from Sports Wearable Sensors," *Comput. Intell. Neurosci.*, vol. 2022, p. 9, 2022, doi: 10.1155/2022/5314671.
- [12] C. Yang, R. A. Brower-Sinning, G. Lewis, and C. Kästner, "Data Leakage in Notebooks: Static Detection and Better Processes," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Sep. 2022. doi: 10.1145/3551349.3556918.
- [13] A. Apicella, F. Isgrò, and R. Prevede, "Don't push the button! Exploring data leakage risks in machine learning and transfer learning," *Artif. Intell. Rev.*, vol. 58, no. 11, Nov. 2025, doi: 10.1007/s10462-025-11326-3.
- [14] A. Moscovich and S. Rosset, "On the cross-validation bias due to unsupervised preprocessing," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 84, no. 4, pp. 1474–1502, Sep. 2022, doi: 10.1111/rssb.12537.
- [15] I. Domor Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 1–1, Sep. 2022, doi: 10.1109/ACCESS.2022.3207287.
- [16] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Syst. Appl.*, vol. 237, Mar. 2024, doi: 10.1016/j.eswa.2023.121549.
- [17] F. E. Arévalo-Cordovilla and M. Peña, "Evaluating ensemble models for fair and interpretable prediction in higher education using multimodal data," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-15388-9.
- [18] L. W. Rizkallah, "Enhancing the performance of gradient boosting trees on regression problems," *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01071-3.

-
- [19] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Syst. Appl.*, vol. 244, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.
- [20] O. Rainio, J. Teuvo, 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.
- [21] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," *Lecture Notes in Networks and Systems*, vol. 724 LNNS, pp. 15–25, 2023, doi: 10.1007/978-3-031-35314-7_2.
- [22] I. M. Rajagukguk, R. Hartanto, Julian, and R. Halim, "Comparative Analysis of XGBoost, Random Forest, and Logistic Regression for Classifying Jakarta's Air Pollution Index (ISPU)," *Procedia Comput. Sci.*, vol. 269, pp. 108–120, 2025, doi: 10.1016/j.procs.2025.08.264.
- [23] L. R. Sitompul, A. A. Nababan, M. L. Manihuruk, W. A. Ponsen, and S. Supriyandi, "Comparison of Xgboost, Random Forest and Logistic Regression Algorithms in Stroke Disease Classification," *Sinkron*, vol. 9, no. 2, pp. 957–968, Jun. 2025, doi: 10.33395/sinkron.v9i2.14794.
- [24] R. Septiawan Putra, H. Fachri Satia Simbolon, A. Linhar, F. Izhari, and U. Syekh Ali Hasan Ahmad Addary, "Perbandingan Algoritma Decision Tree dan Random Forest dalam Klasifikasi Kepuasan Pengguna Sistem Informasi Akademik," *TECHSI - Jurnal Teknik Informatika*, vol. 16, p. 2025, doi: 10.29103/techsi.v16i2.25799.
- [25] A. N. S. Kinasih, A. N. Handayani, J. T. Ardiansah, and N. S. Damanhuri, "Comparative analysis of decision tree and random forest classifiers for structured data classification in machine learning," *Science in Information Technology Letters*, vol. 5, no. 2, pp. 13–24, Nov. 2024, doi: 10.31763/sitech.v5i2.1746.
- [26] Harminto Mulyo and Akhmad Khanif Zyen, "Pengaruh Hyperparameter Tuning Gradient Boosting Terhadap Prediksi Pemilihan Program Studi Mahasiswa Baru," *Bulletin of Computer Science Research*, vol. 5, no. 2, pp. 131–137, Feb. 2025, doi: 10.47065/bulletincsr.v5i2.454.
- [27] R. Arya Andika and C. Dewi, "Importance of Feature Selection for Multiple Disease Classification," *Jurnal Buana Informatika*, vol. 16, no. 1, pp. 34–45, 2025, [Online]. Available: <https://ojs.uajy.ac.id/index.php/jbi/article/view/11354>
- [28] N. Shiwakoti, Q. Hu, M. K. Pang, T. M. Cheung, Z. Xu, and H. Jiang, "Passengers' Perceptions and Satisfaction with Digital Technology Adopted by Airlines during COVID-19 Pandemic," *Future Transportation*, vol. 2, no. 4, pp. 988–1009, Dec. 2022, doi: 10.3390/futuretransp2040055.
- [29] M. S. Eshaghi, M. Afshardoost, G. Lohmann, and B. D. Moyle, "Drivers and outcomes of airline passenger satisfaction: A Meta-analysis," *Journal of the Air Transport Research Society*, vol. 3, Dec. 2024, doi: 10.1016/j.jatrs.2024.100034.
- [30] H. Ragab, A. I. Polo-Peña, and A. A. Mahrous, "The effect of airline service quality, perceived value, emotional attachment, and brand loyalty on passengers' willingness to pay: The moderating role of airline origin," *Case Stud. Transp. Policy*, vol. 18, Dec. 2024, doi: 10.1016/j.cstp.2024.101313.
- [31] Anita Dyah Nur'aini, Mariani, and Fauzan Qamaar, "The influence of service quality, passenger satisfaction, perceived value, and airline brand love on word-of-mouth behavior among domestic airline users in Indonesia," *Priviet Social Sciences Journal*, vol. 4, no. 6, Jun. 2024, doi: 10.55942/pssj.v4i6.237.