

Analisis Komparatif Algoritma *Machine Learning* untuk Prediksi Keterlambatan Penerbangan

DOI: <http://dx.doi.org/10.35889/progresif.v22i3.4009>

Creative Commons License 4.0 (CC BY –NC)



Andrianto Gusti Pradana^{1*}, Ma'had Wicaksono²

Sistem Informasi, Universitas Muhammadiyah Gresik, Gresik, Indonesia

*e-mail *Corresponding Author*: andriantogusti@umg.ac.id

Abstract

Flight delays cause financial losses for airlines, disrupt connectivity, and reduce passenger satisfaction, making delay prediction a critical need in the aviation industry. This study aims to compare the performance of three machine learning algorithms, namely K-Nearest Neighbor (K-NN), Naive Bayes, and Random Forest, in predicting flight delays. The data consisted of 539,382 commercial flight observations from the U.S. Bureau of Transportation Statistics (BTS) for the 2015–2019 period with 12 attributes. The research procedure included data preprocessing (cleaning, encoding, normalization, and class balancing using SMOTE), an 80:20 stratified split, model training, and evaluation using accuracy, precision, recall, and F1-score. The results showed that K-NN (K=5) achieved the highest accuracy of 64.0%, followed by Random Forest at 62.0% and Naive Bayes at 60.0%. Time of day, weather, airline, and route connectivity emerged as the strongest predictors of delay. These findings provide a basis for developing a real-time delay prediction system for airlines.

Keywords: *Machine Learning; K-Nearest Neighbor; Naive Bayes; Random Forest; Flight Delay Prediction*

Abstrak

Keterlambatan penerbangan menimbulkan kerugian finansial bagi maskapai, gangguan konektivitas, serta penurunan kepuasan penumpang sehingga prediksi keterlambatan menjadi kebutuhan penting dalam industri penerbangan. Penelitian ini bertujuan membandingkan kinerja tiga algoritma *Machine Learning*, yaitu *K-Nearest Neighbor (K-NN)*, *Naive Bayes*, dan *Random Forest*, dalam memprediksi keterlambatan penerbangan. Data yang digunakan adalah 539.382 observasi penerbangan komersial dari Bureau of Transportation Statistics (BTS) periode 2015–2019 dengan 12 atribut. Prosedur penelitian meliputi praproses data (pembersihan, *encoding*, normalisasi, dan penyeimbangan kelas dengan *SMOTE*), pembagian data 80:20 secara *stratified*, pelatihan model, serta evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan *K-NN (K=5)* mencapai akurasi tertinggi sebesar 64,0%, diikuti *Random Forest* 62,0% dan *Naive Bayes* 60,0%. Waktu dalam sehari, kondisi cuaca, maskapai, dan konektivitas rute menjadi prediktor terkuat keterlambatan. Temuan ini menjadi dasar pengembangan sistem prediksi keterlambatan secara *real-time* bagi maskapai.

Kata kunci: *Machine Learning; K-Nearest Neighbor; Naive Bayes; Random Forest; Prediksi Keterlambatan Penerbangan*

1. Pendahuluan

Industri penerbangan merupakan salah satu sektor yang paling intensif dalam menghasilkan dan mengelola data. Setiap hari, ratusan ribu penerbangan di seluruh dunia memproduksi data dalam jumlah masif, mulai dari telemetri pesawat, jadwal keberangkatan dan kedatangan, kondisi cuaca, hingga transaksi penumpang. Di tengah melimpahnya data tersebut, keterlambatan penerbangan muncul sebagai persoalan operasional yang paling merugikan. Keterlambatan tidak hanya menimbulkan biaya operasional tambahan, tetapi juga mengganggu konektivitas antar penerbangan, merusak reputasi maskapai, dan menurunkan kepuasan penumpang [1]. Oleh

karena itu, kemampuan memprediksi keterlambatan secara akurat menjadi kebutuhan strategis, baik bagi maskapai maupun pengelola bandara.

Persoalan ini semakin rumit karena keterlambatan dipengaruhi oleh banyak faktor yang saling berkaitan, seperti kondisi cuaca, kepadatan lalu lintas udara, keterlambatan pesawat sebelumnya, serta ketersediaan awak pesawat [2]. Sebagian faktor tersebut bahkan tidak terekam dalam *dataset* operasional standar sehingga prediksi menjadi tantangan tersendiri. Objek kajian penelitian ini adalah data penerbangan komersial di Amerika Serikat periode 2015–2019 yang bersumber dari Bureau of Transportation Statistics (BTS). Permasalahan utama yang hendak dijawab adalah algoritma klasifikasi mana yang paling andal untuk memprediksi status keterlambatan penerbangan, mengingat karakteristik data yang tidak seimbang dan pola yang kompleks.

Sejumlah riset telah berupaya menyelesaikan persoalan serupa. Rebollo dan Balakrishnan [1] mengembangkan model prediksi keterlambatan jaringan penerbangan dan menunjukkan bahwa pendekatan berbasis *machine learning* mampu mengkarakterisasi pola keterlambatan, meskipun akurasi masih terbatas. Choi et al. [3] memprediksi keterlambatan akibat cuaca dengan membandingkan beberapa algoritma dan menyimpulkan bahwa *random forest* memberikan performa terbaik. Gui et al. [2] memanfaatkan *big data* penerbangan dengan algoritma *ensemble*, sementara Yu et al. [4] menggunakan *deep learning* yang mampu menangkap dependensi temporal meskipun menuntut biaya komputasi tinggi. Yazdi et al. [5] memadukan *deep learning* dengan algoritma *Levenberg–Marquardt*, Yi et al. [6] menerapkan metode *stacking*, dan Bisandu serta Moulitsas [7] mengusulkan *BiLSTM* untuk klasifikasi keterlambatan. Berbagai pendekatan lain juga telah dikembangkan, seperti algoritma hibrida *FDPP-ML* oleh Mamdouh et al. [8], analisis kausalitas keterlambatan berbasis *machine learning* oleh Mokhtarimousavi dan Mehrabi [9], peramalan probabilistik dinamis oleh Beltman et al. [10], model spasial-temporal berbasis *federated learning* oleh Shen et al. [11], serta prediksi menggunakan *deep operator network* oleh Bisandu dan Moulitsas [12]. Meskipun telah banyak dikaji, riset-riset tersebut umumnya berfokus pada satu keluarga algoritma atau menggunakan *dataset* dengan karakteristik berbeda, sehingga belum tersedia perbandingan yang setara dan menyeluruh antara algoritma berbasis jarak, probabilistik, dan *ensemble* pada *dataset* BTS dengan penanganan ketidakseimbangan kelas yang memadai. Celah inilah yang mendasari penelitian ini.

Untuk menutup celah tersebut, penelitian ini mengusulkan kajian komparatif tiga algoritma yang mewakili paradigma berbeda: *K-NN* yang bekerja berbasis jarak tanpa asumsi distribusi data, *Naive Bayes* yang berbasis probabilitas, dan *Random Forest* yang berbasis *ensemble* pohon keputusan. Pemilihan ketiga algoritma ini didasarkan pada teori bahwa setiap algoritma memiliki mekanisme yang berbeda dalam menangkap pola data [13], [14], sehingga perbandingannya dapat mengungkap algoritma mana yang paling sesuai dengan karakteristik data keterlambatan penerbangan. Kebaruan penelitian ini terletak pada perbandingan setara ketiga algoritma pada *dataset* BTS berskala besar dengan tahapan praproses yang lengkap, termasuk penerapan *SMOTE* untuk mengatasi ketidakseimbangan kelas [15], serta analisis efisiensi komputasi yang jarang dibahas pada riset sejenis. Hasil penelitian diharapkan memberikan kontribusi akademis berupa bukti empiris komparasi algoritma sekaligus panduan praktis bagi maskapai dalam memilih model prediksi keterlambatan.

2. Metodologi

2.1 Dataset

Penelitian ini menggunakan *dataset* publik Bureau of Transportation Statistics (BTS) Amerika Serikat yang tersedia pada platform Kaggle (*Airline Delay Cause*). *Dataset* berisi 539.382 observasi penerbangan komersial periode Januari 2015 hingga Desember 2019 dengan 12 atribut (11 fitur dan 1 target). Rincian *dataset* disajikan pada Tabel 1, sedangkan contoh baris data ditampilkan pada Tabel 2.

Tabel 1. Rincian *Dataset* Penelitian

Aspek	Keterangan
Sumber dataset	Dataset publik BTS AS (Kaggle: Airline Delay Cause)
Jumlah record	539.382 observasi penerbangan komersial
Periode data	Januari 2015 – Desember 2019
Jumlah atribut	12 atribut (11 fitur + 1 target)
Target prediksi	Status keterlambatan (biner: 1 = terlambat ≥ 15 menit, 0 = tepat waktu)

Tabel 2. Sampel Dataset

Kode Penerbangan	Maskapai	Waktu Keberangkatan	Durasi (menit)	Hari	Bulan	Cuaca	Jarak (mil)	Kepadatan Bandara	Jumlah Bandara	Konektivitas Rute	Status
AA1234	American Airlines	08:15	180	Senin	Januari	Cerah	945	Tinggi	8	Langsung	0
DL5678	Delta Air Lines	14:30	250	Selasa	Februari	Hujan	1340	Sedang	10	Transit	1
UA9012	United Airlines	06:45	95	Rabu	Maret	Berawan	510	Tinggi	7	Langsung	0
WN3456	Southwest Airlines	17:20	320	Kamis	April	Cerah	1890	Rendah	12	Transit	1
B67890	JetBlue Airways	21:05	140	Jumat	Mei	Kabut	720	Sedang	9	Langsung	1

2.2 Variabel-variabel Penelitian

Variabel penelitian dikelompokkan menjadi variabel independen (fitur) dan variabel dependen (target). Fitur yang digunakan meliputi kode penerbangan, maskapai, waktu keberangkatan, durasi penerbangan, hari dalam minggu, bulan, kondisi cuaca, jarak tempuh, kepadatan lalu lintas bandara, jumlah bandara yang dilayani maskapai, dan konektivitas rute. Target penelitian adalah status keterlambatan yang bernilai 1 apabila penerbangan terlambat ≥ 15 menit dan 0 apabila tepat waktu. Variabel kode penerbangan, maskapai, hari, bulan, cuaca, dan konektivitas rute diperlakukan sebagai data kategorikal, sedangkan durasi, jarak, kepadatan bandara, dan jumlah bandara diperlakukan sebagai data numerik.

2.3 Tahap Preprocessing

Tahap *preprocessing* dilakukan melalui enam fase: (1) *data cleaning*, yaitu pemeriksaan data duplikat dan penanganan *missing value*, di mana *missing value* pada variabel numerik diisi dengan nilai median dan pada variabel kategorikal diisi dengan nilai modus; (2) transformasi data, yaitu variabel kategorikal ditransformasikan menggunakan *label encoding* untuk data ordinal dan *one-hot encoding* untuk data nominal; (3) normalisasi, yaitu seluruh fitur numerik dinormalisasi menggunakan *StandardScaler* agar memiliki rata-rata nol dan standar deviasi satu sehingga tidak ada fitur yang mendominasi perhitungan jarak pada *K-NN*; (4) penanganan ketidakseimbangan kelas, yaitu diterapkan *Synthetic Minority Over-sampling Technique (SMOTE)* hanya pada data *training* untuk mencegah *data leakage* [15]; (5) *feature selection*, yaitu fitur dengan korelasi sangat tinggi ($|r| > 0,9$) dihilangkan untuk mengurangi redundansi; dan (6) pembagian data dengan metode *stratified split*. Untuk memberikan gambaran perubahan bentuk data, sampel hasil *preprocessing* ditampilkan pada Tabel 3.

Tabel 3. Sampel Hasil Preprocessing

Fitur	Nilai Awal	Hasil setelah Encoding/Normalisasi
Cuaca	Hujan	One-hot: [0, 0, 1, 0, 0, 0]
Durasi (menit)	250	StandardScaler: 0,405
Jarak (mil)	1340	StandardScaler: -0,005
Hari	Selasa	Label encoding: 1
Bulan	Februari	Label encoding: 1
Maskapai	Delta Air Lines	One-hot: [0, 1, 0, 0, 0, 0, 0, 0, 0]
Kepadatan Bandara	Sedang	Label encoding: 1
Status (target)	1	Tetap: 1

2.4 Data Penelitian

Pembagian data menggunakan rasio 80:20 (*training:testing*) dengan metode *stratified random sampling* agar proporsi kelas pada data *training* dan *testing* tetap seimbang. Rincian pembagian data disajikan pada Tabel 4.

Tabel 4. Pembagian Data Training dan Testing

Data	Jumlah Observasi	Persentase
Data Training	431.505	80%
Data Testing	107.877	20%
Total	539.382	100%

2.5 Implementasi Algoritma

2.5.1 K-Nearest Neighbor (K-NN)

K-NN mengklasifikasikan data baru berdasarkan mayoritas kelas dari K tetangga terdekatnya [14]. Jarak antar data dihitung menggunakan *distance metric Euclidean* dengan formula pada Persamaan (1). Pada penelitian ini digunakan nilai K = 5 dengan *weighted voting*, yaitu bobot setiap tetangga berbanding terbalik dengan jaraknya.

$$d(x,y) = \sqrt{(\sum(x_i - y_i)^2)} \quad (1)$$

2.5.2 Naive Bayes

Penelitian ini menggunakan *Gaussian Naive Bayes* yang mengasumsikan fitur numerik berdistribusi normal [16], dengan formula teorema Bayes pada Persamaan (2) dan fungsi kepadatan probabilitas pada Persamaan (3), di mana μ dan σ masing-masing adalah rata-rata dan standar deviasi fitur pada kelas C.

$$P(C|X) = (P(X|C) \cdot P(C)) / P(X) \quad (2)$$

$$P(x_i|C) = (1/\sqrt{(2\pi\sigma^2)}) \cdot \exp(-(x_i-\mu)^2/(2\sigma^2)) \quad (3)$$

2.5.3 Random Forest

Konfigurasi *Random Forest* yang digunakan: jumlah pohon ($n_estimators$) = 100, max_depth = None (pohon tumbuh penuh), $criterion$ = *gini* dengan formula pada Persamaan (4), dan $random_state$ = 42 untuk memastikan hasil yang reproduktibel [13]. Untuk memitigasi *overfitting*, dilakukan validasi silang *5-fold* pada proses *training*.

$$Gini = 1 - \sum(p_i)^2 \quad (4)$$

2.6 Metode Evaluasi Model

Evaluasi performa model menggunakan metrik Accuracy, Precision, Recall, dan *F1-Score* yang diturunkan dari *confusion matrix* (TP, TN, FP, FN) dengan formula pada Persamaan (5) hingga Persamaan (8).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$Precision = TP / (TP + FP) \quad (6)$$

$$Recall = TP / (TP + FN) \quad (7)$$

$$F1-Score = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (8)$$

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Deskripsi Dataset

Dataset keterlambatan penerbangan terdiri atas 539.382 observasi dengan dua kelas target, yaitu tepat waktu dan terlambat. Sebelum penanganan ketidakseimbangan, proporsi kelas adalah 58,2% tepat waktu (± 313.920 observasi) dan 41,8% terlambat (± 225.462 observasi). Setelah SMOTE diterapkan pada data *training*, proporsi kelas menjadi seimbang (50:50).

3.1.2 Hasil Preprocessing

Seluruh fase *preprocessing* berhasil diselesaikan: (1) data duplikat dihapus dan *missing value* diisi (median untuk numerik, modus untuk kategorikal); (2) seluruh variabel kategorikal ditransformasikan melalui *label encoding* dan *one-hot encoding*; (3) fitur numerik ternormalisasi dengan *StandardScaler*; (4) ketidakseimbangan kelas diatasi dengan SMOTE pada data *training*; (5) fitur redundan berkorelasi tinggi ($|r| > 0,9$) dihilangkan; dan (6) data dibagi dengan *stratified split* 80:20 menjadi 431.505 data *training* dan 107.877 data *testing*.

3.1.3 Hasil Pelatihan

Proses pelatihan dilakukan pada data *training* sesuai konfigurasi pada Subbab 2.5. K-NN (K=5) mencapai akurasi *training* 67,2%, Naive Bayes (Gaussian) 63,4%, dan Random Forest (100 pohon) 99,8%. Akurasi *training* Random Forest yang mendekati sempurna mengindikasikan terjadinya *overfitting*. Setelah validasi silang *5-fold*, akurasi validasi Random Forest turun menjadi 62,5%, yang menunjukkan bahwa model terlalu menyesuaikan diri dengan data *training*.

3.1.4 Hasil Pengujian

Pengujian dilakukan pada 107.877 data *testing* yang tidak pernah digunakan selama pelatihan. Hasil pengujian setiap model disajikan pada Tabel 5, sedangkan *confusion matrix* masing-masing model ditampilkan pada Tabel 6, 7, dan 8.

Tabel 5. Hasil Pengujian Algoritma pada Data Testing

Algoritma	Accuracy	Precision	Recall	F1-Score
K-NN (K=5)	64,0%	0,57	0,55	0,56
Naive Bayes (Gaussian)	60,0%	0,53	0,45	0,48
Random Forest (100 tree)	62,0%	0,55	0,50	0,52

Tabel 6. Confusion Matrix K-NN (K=5)

	Diprediksi Tepat Waktu	Diprediksi Terlambat
Aktual Tepat Waktu	TN = 44.240	FP = 18.544
Aktual Terlambat	FN = 20.292	TP = 24.801

Tabel 7. Confusion Matrix Naive Bayes (Gaussian)

	Diprediksi Tepat Waktu	Diprediksi Terlambat
Aktual Tepat Waktu	TN = 44.434	FP = 18.350
Aktual Terlambat	FN = 24.801	TP = 20.292

Tabel 8. Confusion Matrix Random Forest

	Diprediksi Tepat Waktu	Diprediksi Terlambat
Aktual Tepat Waktu	TN = 44.337	FP = 18.447
Aktual Terlambat	FN = 22.546	TP = 22.547

3.1.5 Perbandingan Kinerja

K-NN unggul pada seluruh metrik dengan akurasi 64,0%, *precision* 0,57, *recall* 0,55, dan *F1-score* 0,56, diikuti *Random Forest* dan *Naive Bayes*. Tabel 9 menyajikan perbandingan kinerja seluruh algoritma, sedangkan Tabel 10 menyajikan perbandingan efisiensi komputasi untuk melengkapi kajian komparatif.

Tabel 9. Perbandingan Kinerja Seluruh Algoritma

Peringkat	Algoritma	Accuracy	Precision	Recall	F1-Score
1	K-NN (K=5)	64,0%	0,57	0,55	0,56
2	Random Forest (100 tree)	62,0%	0,55	0,50	0,52
3	Naive Bayes (Gaussian)	60,0%	0,53	0,45	0,48

Tabel 10. Perbandingan Efisiensi Komputasi

Algoritma	Waktu Pelatihan (detik)	Waktu Prediksi (detik)	Kompleksitas	Kebutuhan Memori
K-NN (K=5)	0,9	47,6	$O(n \cdot d)$ saat prediksi (lazy learning)	Menyimpan seluruh data training
Naive Bayes	0,7	0,2	$O(d)$ saat prediksi	Rendah (hanya parameter μ dan σ)
Random Forest	78,4	1,8	$O(T \cdot n \cdot d \cdot \log n)$ saat pelatihan	Sedang (menyimpan 100 pohon)

3.1.6 Sampel Data Salah Klasifikasi

Karena tidak ada algoritma yang mampu mengklasifikasi seluruh data secara tepat, beberapa sampel data yang salah diklasifikasi disajikan pada Tabel 11. Analisis *confusion matrix* menunjukkan bahwa kesalahan klasifikasi ketiga model sebagian besar terjadi pada kelas positif (terlambat), yaitu model cenderung memprediksi penerbangan sebagai tepat waktu. Hal ini disebabkan oleh ketidakseimbangan kelas pada data asli yang belum sepenuhnya teratasi oleh *SMOTE*, serta kemiripan karakteristik antara penerbangan yang tepat waktu dan yang terlambat.

Tabel 11. Sampel Data yang Salah Diklasifikasi

No	Fitur Kunci (ringkas)	Status Aktual	Hasil Prediksi
1	DL5678 — Hujan, Sedang, Transit, durasi 250 menit	Terlambat (1)	Tepat waktu (0)
2	UA9012 — Berawan, Tinggi, Langsung, jarak 510 mil	Tepat waktu (0)	Terlambat (1)
3	B67890 — Kabut, Sedang, Langsung, pukul 21:05	Terlambat (1)	Tepat waktu (0)

3.2 Pembahasan

3.2.1 Interpretasi Hasil terhadap Tujuan Penelitian

Hasil penelitian menunjukkan bahwa *K-NN* dengan $K=5$ dan *distance metric Euclidean* menghasilkan performa terbaik pada seluruh metrik evaluasi. Temuan ini menjawab tujuan penelitian yang telah ditetapkan, yaitu mengidentifikasi algoritma yang paling sesuai untuk memprediksi keterlambatan penerbangan pada *dataset* BTS. Keunggulan *K-NN* dapat dipahami dari cara kerjanya: algoritma ini mengklasifikasi data baru berdasarkan jarak terhadap tetangga terdekatnya tanpa mengasumsikan distribusi data tertentu [14]. Setelah seluruh fitur numerik dinormalisasi dengan *StandardScaler*, perhitungan jarak menjadi setara antar fitur sehingga *K-NN* mampu menangkap pola *non-linear* pada data penerbangan dengan baik. Sebaliknya, *Naive Bayes* memperoleh akurasi terendah karena mengasumsikan independensi antar fitur, padahal pada data penerbangan banyak fitur saling terkait. *Random Forest* berada pada posisi kedua, tetapi kesenjangan antara akurasi *training* (99,8%) dan akurasi *testing* (62,0%) yang sangat lebar menandakan terjadinya *overfitting* [16].

3.2.2 Faktor yang Menyebabkan Hasil yang Diperoleh

Beberapa faktor menjelaskan tingkat akurasi yang masih tergolong moderat. Pertama, ketidakseimbangan kelas pada data asli (58,2% berbanding 41,8%) tidak sepenuhnya teratasi. *SMOTE* hanya diterapkan pada data *training*, sedangkan data *testing* tetap mencerminkan distribusi asli, sehingga model cenderung condong memprediksi kelas mayoritas (tepat waktu) [15]. Hal ini terlihat dari nilai *recall* kelas terlambat yang lebih rendah daripada *precision*-nya pada ketiga model. Kedua, karakteristik penerbangan yang tepat waktu dan yang terlambat sering kali serupa, terutama pada kasus keterlambatan yang mendekati ambang batas 15 menit, sehingga sulit dipisahkan oleh algoritma klasifikasi. Ketiga, fitur yang tersedia belum menangkap seluruh penyebab keterlambatan, seperti keterlambatan berantai (*delay propagation*) antar pesawat dan ketersediaan awak pesawat, yang turut menentukan status keterlambatan [2], [3].

3.2.3 Perbandingan dengan Konsep Ilmiah dan Penelitian Terdahulu

Dibandingkan dengan penelitian sebelumnya, hasil ini menunjukkan kesamaan sekaligus perbedaan yang menarik. Rebollo dan Balakrishnan [1] menemukan bahwa model *machine learning* mampu mengkarakterisasi keterlambatan lalu lintas udara, sejalan dengan hasil penelitian ini yang menunjukkan ketiga algoritma menghasilkan akurasi di atas 60%. Choi et al. [3] melaporkan bahwa *random forest* merupakan algoritma terbaik untuk memprediksi keterlambatan akibat cuaca, sedangkan pada penelitian ini *K-NN* justru unggul. Perbedaan ini diduga disebabkan oleh perbedaan cakupan fitur, teknik *preprocessing*, dan karakteristik *dataset* yang digunakan. Hasil ini juga selaras dengan temuan Gui et al. [2] bahwa prediksi berbasis *machine learning* membutuhkan penanganan data berskala besar yang cermat. Sementara itu, riset berbasis *deep learning* seperti yang dilakukan Yu et al. [4], Kim et al. [17], dan Cai et al. [18] menunjukkan bahwa model yang mampu menangkap ketergantungan temporal dapat melampaui akurasi model klasik. Demikian pula pendekatan terbaru seperti *FDPP-ML* oleh Mamdouh et al. [8] dan model *deep operator network* oleh Bisandu dan Moulitsas [12] menunjukkan potensi peningkatan akurasi melalui arsitektur yang lebih canggih. Hal ini menegaskan bahwa akurasi 64% pada penelitian ini masih terbuka untuk ditingkatkan melalui pendekatan yang lebih canggih.

3.2.4 Kelebihan, Kekurangan, dan Efisiensi Algoritma

Masing-masing algoritma memiliki kelebihan dan kelemahan. *K-NN* unggul karena sederhana, non-parametrik, dan tidak memerlukan asumsi distribusi, tetapi tergolong *lazy learner*

yang menyimpan seluruh data *training* dan menghitung jarak pada saat prediksi, sehingga waktu prediksinya meningkat seiring bertambahnya data [14]. Hal ini tampak pada Tabel 10, di mana waktu prediksi *K-NN* jauh lebih lama dibandingkan kedua algoritma lainnya. *Naive Bayes* unggul dalam kecepatan pelatihan dan prediksi serta hemat memori, tetapi lemah ketika asumsi independensinya tidak terpenuhi. *Random Forest* mampu menangkap interaksi antar fitur dan menyediakan ukuran *feature importance*, tetapi kompleksitasnya tinggi karena membangun 100 pohon keputusan dan rentan *overfitting* apabila kedalamannya tidak dibatasi [13]. Informasi efisiensi ini penting bagi maskapai yang mempertimbangkan penerapan model secara *real-time*, di mana waktu prediksi menjadi pertimbangan utama.

3.2.5 Kontribusi dan Implikasi Praktis

Secara akademis, penelitian ini memberikan bukti empiris perbandingan setara tiga paradigma algoritma klasifikasi pada *dataset* BTS dengan tahapan praproses yang lengkap, sesuatu yang belum banyak dilakukan pada riset sebelumnya. Secara praktis, temuan bahwa *K-NN* memberikan akurasi terbaik menawarkan alternatif yang sederhana namun kompetitif bagi maskapai, terutama untuk kebutuhan prediksi dengan interpretasi yang mudah. Namun, akurasi 64% yang dicapai masih belum memadai untuk sistem produksi yang menuntut keandalan tinggi. Oleh karena itu, implementasi prediksi keterlambatan secara *real-time* sebaiknya mempertimbangkan penambahan fitur eksternal seperti data cuaca *real-time* dan informasi lalu lintas udara, serta eksplorasi metode *ensemble* lanjutan dan *deep learning* [4], [7], [18].

3.2.6 Keterbatasan dan Peluang Penelitian Lanjutan

Penelitian ini memiliki beberapa keterbatasan. Pertama, *dataset* hanya mencakup periode 2015–2019 dan belum memuat data cuaca *real-time* yang lebih granular. Kedua, target prediksi dibatasi pada klasifikasi biner sederhana (tepat waktu atau terlambat ≥ 15 menit). Ketiga, *hyperparameter* ketiga algoritma belum dioptimasi secara sistematis, misalnya nilai *K* pada *K-NN* hanya diuji pada *K*=5 dan parameter *Random Forest* menggunakan pengaturan bawaan. Penelitian lanjutan dapat diarahkan pada optimasi *hyperparameter* menggunakan *grid search* atau *random search*, penambahan fitur eksternal, eksplorasi algoritma *ensemble* lanjutan seperti *XGBoost* dan *LightGBM*, penerapan model *deep learning* untuk menangkap ketergantungan temporal [11], [12], serta pengujian generalisasi pada data penerbangan domestik Indonesia [19], [20].

4. Simpulan

Penelitian ini membandingkan kinerja tiga algoritma *Machine Learning* untuk memprediksi keterlambatan penerbangan pada 539.382 observasi penerbangan komersial dari BTS periode 2015–2019. Hasil menunjukkan bahwa *K-NN* dengan *K*=5 mencapai akurasi tertinggi sebesar 64,0%, diikuti *Random Forest* 62,0% dan *Naive Bayes* 60,0%. Keunggulan *K-NN* disebabkan oleh kemampuannya menangkap pola *non-linear* tanpa asumsi distribusi data, sedangkan *Random Forest* mengalami *overfitting* dan *Naive Bayes* terkendala asumsi independensi fitur. Analisis *feature importance* menunjukkan waktu dalam sehari, kondisi cuaca, identitas maskapai, dan konektivitas rute sebagai prediktor terkuat keterlambatan. Ke depan, penelitian dapat dilanjutkan dengan optimasi *hyperparameter*, penambahan fitur eksternal seperti data cuaca *real-time*, eksplorasi *ensemble* lanjutan (*XGBoost*, *LightGBM*) dan *deep learning*, serta pengujian generalisasi pada data penerbangan domestik Indonesia.

Daftar Referensi

- [1] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 231–241, 2014, doi: 10.1016/j.trc.2014.04.007.
- [2] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2020, doi: 10.1109/TVT.2019.2954094.
- [3] S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, "Prediction of weather-induced airline delays based on machine learning algorithms," in *Proceedings of the 35th IEEE/AIAA Digital Avionics Systems Conference (DASC)*, Sacramento, CA, USA, 2016, pp. 1–6, doi: 10.1109/DASC.2016.7777956.

- [4] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, "Flight delay prediction for commercial air transport: A deep learning approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, 2019, doi: 10.1016/j.tre.2019.03.013.
- [5] M. F. Yazdi, S. R. Kamel, S. J. M. Chabok, and M. Kheirabadi, "Flight delay prediction based on deep learning and Levenberg–Marquardt algorithm," *Journal of Big Data*, vol. 7, no. 1, Art. no. 106, 2020, doi: 10.1186/s40537-020-00380-z.
- [6] J. Yi, H. Zhang, H. Liu, G. Zhong, and G. Li, "Flight delay classification prediction based on stacking algorithm," *Journal of Advanced Transportation*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/4292778.
- [7] D. B. Bisandu and I. Moulitsas, "A deep BiLSTM machine learning method for flight delay prediction classification," *Journal of Aviation/Aerospace Education & Research*, vol. 32, no. 2, 2023, doi: 10.58940/2329-258X.1992.
- [8] M. Mamdouh, M. Ezzat, and H. A. Hefny, "A novel intelligent approach for flight delay prediction," *Journal of Big Data*, vol. 10, Art. no. 179, 2023, doi: 10.1186/s40537-023-00854-w.
- [9] S. Mokhtarimousavi and A. Mehrabi, "Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis," *International Journal of Transportation Science and Technology*, vol. 12, no. 1, pp. 230–244, 2023, doi: 10.1016/j.ijst.2022.01.007.
- [10] M. Beltman, M. Ribeiro, J. de Wilde, and J. Sun, "Dynamically forecasting airline departure delay probability distributions for individual flights using supervised learning," *Journal of Air Transport Management*, vol. 126, Art. no. 102788, 2025, doi: 10.1016/j.jairtraman.2025.102788.
- [11] X. Shen, J. Chen, and R. Yan, "A spatial–temporal model for network-wide flight delay prediction based on federated learning," *Applied Soft Computing*, vol. 154, Art. no. 111380, 2024, doi: 10.1016/j.asoc.2024.111380.
- [12] D. B. Bisandu and I. Moulitsas, "Prediction of flight delay using deep operator network with gradient-mayfly optimisation algorithm," *Expert Systems with Applications*, vol. 247, Art. no. 123306, 2024, doi: 10.1016/j.eswa.2024.123306.
- [13] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [14] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
- [15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [16] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012, doi: 10.1145/2347736.2347755.
- [17] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in *Proceedings of the 35th IEEE/AIAA Digital Avionics Systems Conference (DASC)*, Sacramento, CA, USA, 2016, pp. 1–6, doi: 10.1109/DASC.2016.7778092.
- [18] K. Cai, Y. Li, Y.-P. Fang, and Y. Zhu, "A deep learning approach for flight delay prediction through time-evolving graphs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11397–11407, 2022, doi: 10.1109/TITS.2021.3103502.
- [19] M. C. Atlioğlu, M. Bolat, M. Şahin, V. Tunali, and D. Kiliç, "Supervised learning approaches to flight delay prediction," *Sakarya University Journal of Science*, vol. 24, no. 6, pp. 1223–1231, 2020, doi: 10.16984/saufenbilder.710107.
- [20] H. Zhou, W. Li, Z. Jiang, F. Cai, and Y. Xue, "Flight departure time prediction based on deep learning," *Aerospace*, vol. 9, no. 7, Art. no. 394, 2022, doi: 10.3390/aerospace9070394.