

PEMODELAN TINGKAT PREDIKSI KETERLAMBATAN PENERBANGAN, TERHADAP DATASET PENERBANGAN KOMERSIAL

Redrian Alfa Maulana, Muhammad Haykal Islam Baskoro, Wicaksana Surya Dewantara,
Muhammad Gifary Madyan, Alif Akmal Winasis, Nurul Afifah Ariffudin

Informatika, Universitas Pembangunan Nasional “Veteran” Jakarta

Jalan RS Fatmawati Jakarta, Indonesia

2210511133@mahasiswa.upnvj.ac.id

ABSTRAK

Dinamika industri penerbangan sangat bergantung pada ketepatan waktu operasional. Namun, keterlambatan penerbangan (*flight delay*) menjadi permasalahan krusial yang terus berulang, merugikan maskapai secara finansial, dan menurunkan tingkat kepuasan penumpang. Penelitian ini bertujuan untuk membangun model komputasi cerdas yang mampu memprediksi tingkat keterlambatan penerbangan sebagai sistem peringatan dini. Metode penelitian yang diusulkan menggunakan pendekatan *Ensemble Learning* dengan menggabungkan algoritma *Random Forest* dan *Linear Regression*. Pola penyebab keterlambatan diekstraksi dari data historis penerbangan berdasarkan fitur operasional seperti maskapai, jadwal, dan jarak tempuh. Penggabungan ini dilakukan agar kehandalan *Random Forest* pada data non-linear dapat diimbangi oleh efisiensi komputasi *Linear Regression*. Hasil pengujian menunjukkan bahwa model prediksi berhasil memberikan performa yang sangat baik dengan perolehan nilai *Mean Squared Error* (MSE) sebesar 178.71 dan skor *R-Squared* mencapai 0.906. Kinerja ini membuktikan model mampu memprediksi 90,6% variabilitas keterlambatan secara presisi, sehingga sangat layak direkomendasikan sebagai pendukung keputusan manajemen maskapai.

Kata kunci : *Prediksi Keterlambatan Penerbangan, Data Mining, Ensemble Learning, Random Forest, Linear Regression.*

1. PENDAHULUAN

Dalam konteks dinamika industri penerbangan, keterlambatan penerbangan bukan hanya menjadi isu operasional semata, namun juga menimbulkan dampak yang signifikan dalam berbagai aspek. Keterlambatan penerbangan berkaitan langsung dengan penurunan kepuasan pelanggan, inefisiensi operasional maskapai, dan kerugian ekonomi yang luas. Fenomena ini terkait erat dengan aspek waktu yang menjadi inti operasional: perbedaan antara jadwal keberangkatan atau kedatangan yang direncanakan dengan waktu aktual penerbangan. Kondisi tersebut menjadikan manajemen ketepatan waktu sebagai fokus perhatian yang tak terelakkan dalam industri penerbangan modern.

Di balik urgensi tersebut, permasalahan utama yang dihadapi oleh pihak maskapai saat ini adalah sulitnya memproyeksikan potensi keterlambatan secara proaktif akibat tingginya kompleksitas variabel operasional, seperti jadwal, rute, dan rotasi pesawat. Ketidakmampuan untuk memprediksi *delay* secara dini menyebabkan pihak manajemen sering kali hanya dapat melakukan tindakan mitigasi yang bersifat reaktif setelah keterlambatan terjadi, sehingga memperbesar nilai kerugian.

Berangkat dari permasalahan tersebut, tujuan utama penelitian ini diinisiasi sebagai langkah terobosan untuk mengembangkan sebuah model komputasi prediktif yang mampu mengidentifikasi dan memperkirakan kemungkinan keterlambatan penerbangan secara presisi. Model ini berfokus untuk menghasilkan metrik prediksi yang akurat guna

menganalisis akar masalah keterlambatan berdasarkan data historis.

Sebagai rencana penyelesaian masalah, penelitian ini mengusulkan integrasi teknologi komputasional yang sistematis. Pada tahap awal, teknik *data mining* dimanfaatkan untuk mengekstraksi dan mengeksplorasi pola-pola tersembunyi serta tren dari himpunan data (*dataset*) historis penerbangan. Selanjutnya, algoritma *machine learning* diterapkan menggunakan metode *Ensemble Learning* yang menggabungkan pendekatan *Random Forest* dan *Linear Regression*. Pengintegrasian ini dipilih untuk menutupi kelemahan model tunggal; di mana *Random Forest* sangat handal dalam menangani kompleksitas fitur non-linear, sementara *Linear Regression* memetakan tren linier secara efisien. Hasil analitis dan pemodelan prediktif ini pada akhirnya dirancang untuk menjadi landasan algoritma dalam pengembangan Sistem Pendukung Keputusan (*Decision Support System*) cerdas, yang siap dimanfaatkan oleh manajemen operasional maskapai.

Untuk memberikan gambaran yang komprehensif, sistematika penulisan makalah ini disusun sebagai berikut: Bagian 2 menguraikan tinjauan pustaka dari penelitian-penelitian terdahulu yang relevan. Bagian 3 menjelaskan metodologi penelitian, termasuk tahapan pengumpulan data, pra-pemrosesan, dan skema evaluasi *machine learning*. Bagian 4 memaparkan hasil eksperimen, visualisasi pola data, dan diskusi mengenai implikasi temuan. Terakhir, Bagian 5 merangkum kesimpulan utama dari penelitian ini beserta saran untuk arah pengembangan riset di masa mendatang.

2. TINJAUAN PUSTAKA

Penelitian mengenai prediksi keterlambatan penerbangan telah berevolusi dari model statistik konvensional menuju pendekatan *machine learning* yang kompleks. Raharjo dkk. [1] dalam studinya mengenai prediksi durasi *delay* menggunakan algoritma *Random Forest* mencatatkan tingkat keakuratan prediksi yang sangat baik, membuktikan bahwa model berbasis pohon keputusan (*decision tree*) mampu memetakan variabel non-linear terkait cuaca dan jadwal penerbangan. Namun, penelitian tersebut masih terbatas pada observasi maskapai tunggal sehingga model rentan mengalami *overfitting* bila dihadapkan pada data dari maskapai lain yang memiliki kebijakan operasional berbeda.

Untuk mengatasi limitasi tersebut, penelitian lain oleh Adinarta [6] mencoba mengekstraksi data historis dari berbagai maskapai menggunakan *Stacking Ensemble Method*, yang menggabungkan *Random Forest* dengan *XGBoost*. Hasilnya menunjukkan peningkatan akurasi, namun beban komputasinya menjadi sangat tinggi. Di sisi lain, Winata [2] melakukan komparasi performa klasifikasi antara algoritma *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan *Random Forest*, yang pada akhirnya mengonfirmasi bahwa *Random Forest* memiliki tingkat ketahanan (*robustness*) terbaik terhadap variasi data penerbangan yang berisik (*noisy*).

Selain algoritma berbasis pohon keputusan, berbagai riset juga secara masif mengeksplorasi metode dan faktor determinan lainnya. Dari segi faktor operasional, Putri [4] menegaskan bahwa anomali cuaca merupakan variabel yang sangat krusial dan berdampak langsung pada perencanaan penerbangan, sejalan dengan temuan Kautsarina [5] yang secara khusus mengimplementasikan model *Random Survival Forest* untuk memprediksi probabilitas *delay* akibat cuaca. Lebih lanjut, komparasi berbagai model *machine learning* juga terus dievaluasi untuk mencari arsitektur terbaik. Kothari dkk. [7] dan Singh [8] melakukan uji komparasi secara komprehensif untuk menyeleksi algoritma dasar dengan performa paling optimal pada *dataset* maskapai penumpang. Sementara itu, dalam pengembangan metode yang lebih mutakhir, Ardhana dkk. [3] memanfaatkan *Neural Network*, dan Ishaya [9] menggunakan *Hybrid Shallow Machine Learning* untuk mempercepat proses analisis prediksi *delay*. Berbagai tren pendekatan algoritma prediktif ini secara luas telah dievaluasi relevansinya oleh Ahmed [10] guna melihat metode mana yang paling signifikan dalam meningkatkan akurasi.

Melihat celah dari literatur-literatur terdahulu, terutama tingginya beban komputasi pada *ensemble* algoritma berat, penelitian ini diusulkan untuk menawarkan jalan tengah. Model yang diusulkan memadukan kehandalan *Random Forest* dalam menangani data kompleks dengan kesederhanaan dan efisiensi algoritma *Linear Regression*. Penggabungan ini diharapkan dapat menghasilkan model prediksi

keterlambatan pesawat komersial yang tidak hanya memiliki akurasi tinggi, tetapi juga efisien secara komputasional untuk diterapkan pada *dataset* berskala besar.

3. METODE PENELITIAN

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari platform repositori publik, Kaggle. Pengumpulan data ini merupakan tahap awal dari siklus *Knowledge Discovery in Databases* (KDD) dalam proses *data mining*, di mana himpunan data berskala masif dipersiapkan sebelum diekstraksi polanya. Penggunaan *dataset* sekunder dengan nama file *flights.csv* ini dipilih karena menyediakan rekaman historis penerbangan yang sangat komprehensif. Penggunaan repositori ini dinilai jauh lebih rasional dan efisien dibandingkan melakukan ekstraksi mandiri (*crawling/scraping*) melalui layanan *Application Programming Interface* (API) pelacakan penerbangan yang umumnya memiliki limitasi akses kuota historis. *Dataset* asli dari populasi ini terdiri dari jutaan entri, namun untuk keperluan efisiensi beban komputasi pemodelan, penelitian ini mengekstrak sampel acak sebanyak 50.000 entri penerbangan. Dari total 30 atribut yang tersedia, dipilih 11 atribut yang paling relevan dengan target prediksi, di antaranya meliputi waktu keberangkatan, rute bandara asal dan tujuan, jarak tempuh, hingga riwayat keterlambatan keberangkatan dan kedatangan.

3.2. Pra-pemrosesan Data dan Feature Engineering

Tahap pra-pemrosesan (*preprocessing*) dilakukan menggunakan pustaka *Pandas* pada bahasa pemrograman *Python* untuk memastikan kualitas dan integritas data sebelum dimasukkan ke dalam algoritma pemodelan. Langkah pertama yang krusial adalah penanganan nilai yang hilang (*missing values*) menggunakan fungsi *dropna()*. Pembersihan ini dipilih dibandingkan metode imputasi (pengisian nilai rata-rata) untuk menghindari manipulasi buatan yang dapat mendistorsi pola asli penyebab keterlambatan penerbangan.

Langkah kedua adalah tahap *Feature Engineering* melalui transformasi tipe data. Algoritma *Linear Regression* dan *Random Forest* bekerja berbasis kalkulasi matematis spasial, sehingga tidak dapat memproses fitur kategorikal berskala teks seperti nama maskapai (AIRLINE), kode bandara asal (ORIGIN_AIRPORT), dan bandara tujuan (DESTINATION_AIRPORT). Oleh karena itu, dilakukan teknik *Label Encoding* menggunakan metode *cat.codes*. Teknik ini mengonversi setiap kategori teks unik menjadi representasi bilangan bulat (*integer*). Data yang telah sepenuhnya berwujud numerik ini kemudian dipartisi menjadi dua dimensi: variabel matriks fitur (X) yang berisi indikator penerbangan, dan variabel vektor target (y) yang berisi

nilai durasi keterlambatan kedatangan (*ARRIVAL_DELAY*).

3.3. Pengembangan dan Pelatihan Model

Penelitian ini menggunakan pendekatan *supervised learning* untuk melakukan prediksi. Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) menggunakan fungsi *train_test_split*. Untuk mengekstraksi pola dari data latih, terdapat dua algoritma utama yang dilatih secara independen, yaitu *Linear Regression* dan *Random Forest Regressor*.

Algoritma pertama yang diterapkan adalah *Linear Regression*, yang merupakan algoritma fundamental dalam *machine learning* untuk memodelkan hubungan antara variabel dependen (target) kontinu dengan satu atau lebih variabel independen (fitur). Dalam konteks penelitian ini, variabel targetnya adalah durasi keterlambatan pesawat. Model regresi linier ini bekerja dengan mencari garis linier terbaik (garis regresi) yang meminimalkan jumlah kuadrat galat (*Sum of Squared Errors*) antara nilai yang diprediksi dengan nilai aktual. Pendekatan ini sangat efisien secara komputasi dan memberikan interpretasi yang sangat transparan terkait seberapa besar bobot setiap fitur dalam memicu terjadinya keterlambatan.

Sebagai algoritma kedua yang difungsikan untuk memodelkan pola yang lebih rumit, penelitian ini menggunakan *Random Forest Regressor*. Berbeda dengan karakteristik regresi linier dasar, *Random Forest* adalah algoritma berbasis *ensemble learning* yang membangun sekumpulan pohon keputusan (*decision trees*) secara acak selama masa pelatihan data. Luaran prediksi final dari *Random Forest Regressor* didapatkan dari nilai rata-rata (*mean*) seluruh prediksi individu pada setiap pohon. Keunggulan utama dari algoritma ini adalah kemampuannya yang sangat baik dalam menangani interaksi fitur yang kompleks dan non-linier, seperti pola hubungan antara hari dalam seminggu dengan jarak rute bandara. Selain itu, metode ini dipilih karena sangat tangguh (*robust*) terhadap keberadaan data pencilan (*outliers*) ekstrem yang sering ditemukan dalam rekaman riwayat jadwal penerbangan komersial.

3.4. Penggabungan Model (Ensemble)

Untuk meningkatkan keakuratan dan stabilitas hasil, prediksi dari kedua model dasar (*Linear Regression* dan *Random Forest*) digabungkan menggunakan pendekatan *Ensemble Model*. Gabungan hasil prediksi (rata-rata dari kedua keluaran model) digunakan sebagai prediksi final keterlambatan penerbangan dengan tujuan meminimalkan bias dan menutupi kelemahan dari masing-masing model tunggal.

3.5. Evaluasi Kinerja Model

Tahap terakhir adalah mengevaluasi tingkat akurasi dari model *ensemble* yang telah dibangun. Evaluasi dilakukan dengan membandingkan nilai prediksi yang dihasilkan oleh model dengan nilai keterlambatan kedatangan aktual pada data uji. Metrik evaluasi yang digunakan dalam penelitian ini adalah *Mean Squared Error* (MSE) untuk melihat rata-rata kesalahan kuadrat, serta *R-Squared* (R^2) *Score* untuk mengukur seberapa baik varians dari target dapat dijelaskan oleh fitur yang ada di dalam model.

Mean Squared Error (MSE) digunakan untuk menghitung rata-rata dari kuadrat kesalahan (selisih) antara estimasi nilai yang diprediksi oleh model dengan nilai aktual di lapangan. Persamaan matematis MSE didefinisikan sebagai:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Metrik kedua, Koefisien Determinasi atau R^2 *Score*, digunakan untuk mengukur seberapa besar proporsi varians dari variabel target (keterlambatan kedatangan) yang dapat dijelaskan oleh variabel independen (fitur-fitur penerbangan) di dalam model. Rumusnya adalah:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

3.6. Lingkungan Eksperimen dan Perangkat Lunak

Eksperimen dan pelatihan model *machine learning* dalam penelitian ini dieksekusi menggunakan bahasa pemrograman *Python* versi 3.x. Manipulasi data dan komputasi *dataframe* dilakukan dengan memanfaatkan pustaka *Pandas* dan *NumPy*. Tahapan visualisasi pola data (*Exploratory Data Analysis*) dieksekusi menggunakan modul *Matplotlib* dan *Seaborn*. Sementara itu, pembangunan model *Linear Regression*, *Random Forest Regressor*, pembagian *dataset*, serta kalkulasi metrik evaluasi (MSE dan R^2) memanfaatkan pustaka *Scikit-Learn* (*sklearn*). Pemrosesan komputasi dibatasi pada 50.000 entri data untuk menjaga stabilitas memori (RAM) perangkat keras selama tahap validasi dan *hyperparameter tuning*.

4. HASIL DAN PEMBAHASAN

4.1. Sumber dan Karakteristik Data

Tabel 1. Statistik Deskriptif Variabel Keterlambatan (dalam Menit)

Variabel	Nilai Min.	Nilai Maks.	Rata - Rata
DEPARTURE_DELAY	-82.00	1988.00	9.38
ARRIVAL_DELAY	-87.00	1971.00	4.40
DISTANCE (Mil)	31.00	4983.00	822.30

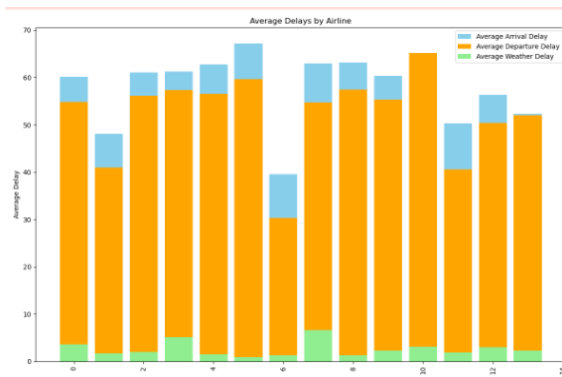
Berdasarkan Tabel 1, fenomena operasional maskapai dapat terlihat dari nilai minimum dan maksimumnya. Nilai minimum yang negatif pada *DEPARTURE_DELAY* (-82) dan *ARRIVAL_DELAY* (-87) menunjukkan bahwa terdapat penerbangan yang

mampu lepas landas atau mendarat lebih awal dari jadwal resmi. Namun, simpangan ekstrem terlihat pada nilai maksimum yang mencapai angka lebih dari 1.900 menit. Rentang data yang sangat lebar ini memvalidasi keputusan penggunaan *Random Forest* dalam penelitian ini, karena model regresi linier biasa akan sangat rentan mengalami bias (*underfitting*) jika dipaksakan memproses rentang *outlier* sejauh ini tanpa bantuan *ensemble*.

Sebagaimana dijelaskan pada Bagian 3.1, dataset yang digunakan adalah sampel 50.000 entri dari *flights.csv* (Kaggle) dengan 11 atribut terpilih. Statistik deskriptif dari variabel-variabel utama disajikan pada Tabel 1 untuk memberikan gambaran awal distribusi data sebelum pemodelan dilakukan.

4.2. Analisis Pola Keterlambatan

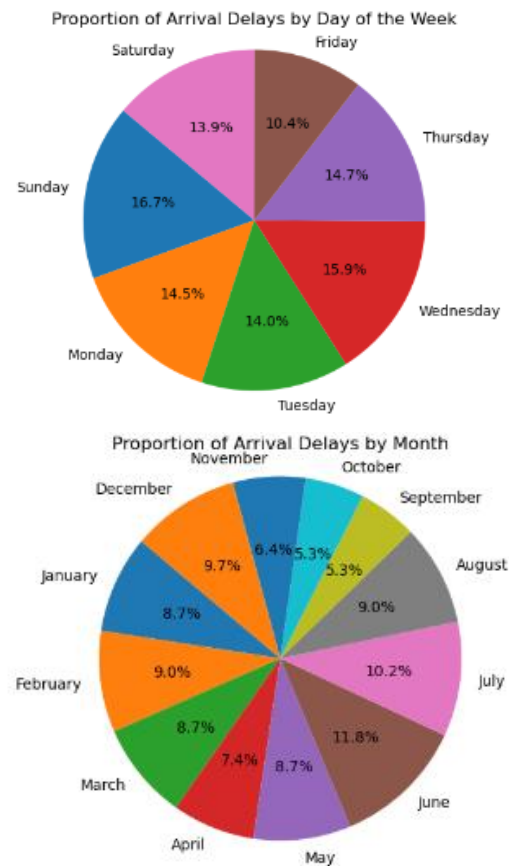
Sebelum dilakukan pemodelan *machine learning*, data dieksplorasi secara visual untuk mengidentifikasi pola keterlambatan (*delay*) bawaan dari setiap entitas penerbangan.



Gambar 1. Rata rata keterlambatan berdasarkan maskapai

Pada Gambar 1, divisualisasikan rata-rata keterlambatan untuk setiap maskapai penerbangan yang dievaluasi. Setiap batang mewakili satu maskapai, di mana ketinggian batang mencerminkan total rata-rata keterlambatan dalam satuan menit. Visualisasi ini memecah keterlambatan menjadi tiga faktor utama: keterlambatan kedatangan (biru langit), keterlambatan keberangkatan (oranye), dan keterlambatan akibat cuaca (hijau muda). Melalui grafik ini, perbandingan performa operasional antar maskapai dapat diobservasi dengan jelas, di mana maskapai dengan batang tertinggi mengindikasikan akumulasi keterlambatan rata-rata yang paling parah.

Selanjutnya, Gambar 2 mengilustrasikan distribusi proporsional keterlambatan berdasarkan dimensi waktu. Grafik pertama menunjukkan proporsi keterlambatan kedatangan pesawat untuk setiap bulan guna mengidentifikasi tren musiman sepanjang tahun. Sementara itu, grafik kedua memperlihatkan total keterlambatan berdasarkan hari dalam satu minggu. Pola ini krusial untuk mendeteksi apakah hari-hari atau bulan tertentu memiliki lonjakan intensitas keterlambatan yang signifikan.



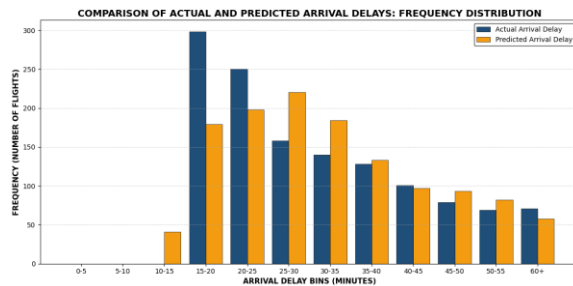
Gambar 2. Proporsi keterlambatan kedatangan berdasarkan bulan dan hari

Berdasarkan distribusi pada *pie chart* bulanan, terlihat adanya fluktuasi proporsi keterlambatan yang dipengaruhi oleh faktor musim dan tren liburan. Bulan-bulan tertentu seperti bulan liburan akhir tahun atau pertengahan tahun umumnya mencatatkan akumulasi frekuensi penerbangan yang jauh lebih padat, yang pada gilirannya memicu efek domino pada antrean di landasan pacu (*taxi-out* dan *taxi-in*). Sementara itu, dari metrik hari dalam seminggu, terlihat bahwa hari-hari menjelang akhir pekan (*weekend*) cenderung memiliki proporsi penumpukan *delay* yang patut diwaspadai oleh maskapai dibandingkan hari kerja pertengahan minggu.

4.3. Evaluasi Kinerja Model Prediksi

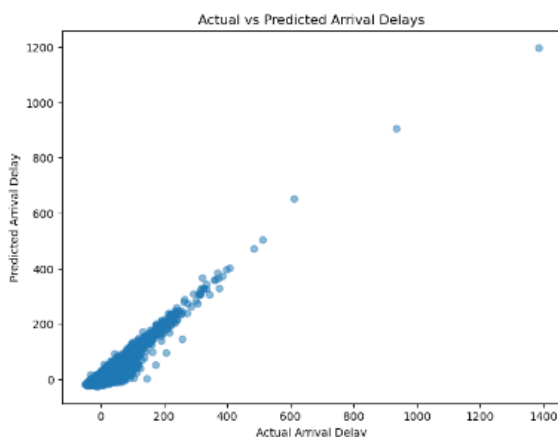
Model prediksi dibangun menggunakan pendekatan pembelajaran terawasi (*supervised learning*) yang menggabungkan algoritma *Linear Regression* dan *Random Forest* ke dalam sebuah *Ensemble Model*. Pengujian pada sampel data baru menunjukkan bahwa model secara dinamis mampu memperkirakan keterlambatan penerbangan tambahan, contohnya memprediksi angka 12,25 menit untuk sebuah rute spesifik.

Pengukuran kinerja gabungan (*Averaging predictions*) secara matematis menghasilkan tingkat galat *Mean Squared Error* (MSE) sebesar 178.71 dan skor koefisien determinasi (R^2) sebesar 0.906.



Gambar 3. Distribusi frekuensi keterlambatan kedatangan aktual dan prediksi

Distribusi akurasi tersebut divisualisasikan pada Gambar 3 melalui grafik histogram berkelompok (*grouped bar chart*). Batang berwarna biru mewakili frekuensi data keterlambatan kedatangan pesawat secara aktual, sedangkan batang berwarna oranye merepresentasikan frekuensi nilai yang diprediksi oleh algoritma yang dikelompokkan berdasarkan rentang waktu (*bins*). Ketinggian kedua batang yang diletakkan secara bersebelahan terlihat saling mengikuti dan memiliki proporsi yang relatif sejajar pada hampir seluruh rentang waktu. Hal ini membuktikan bahwa model mampu menangkap tren distribusi data alami dengan baik dan menyajikannya tanpa adanya tumpang tindih visual yang membingungkan pembaca.



Gambar 4. Plot sebaran nilai prediksi terhadap nilai aktual

Untuk mengevaluasi presisi tebakan model pada setiap entri data tunggal, Gambar 4 menyajikan *scatter plot* antara nilai aktual pada sumbu X dan prediksi pada sumbu Y. Terlihat bahwa titik-titik sampel membentuk formasi yang berdekatan dan mengikuti garis diagonal. Hal ini merepresentasikan korelasi linier yang sangat kuat antara kenyataan dengan hasil prediksi. Dengan skor R^2 mencapai 90,6%, dapat disimpulkan bahwa kombinasi *Random Forest* dan *Linear Regression* sangat kompeten dalam menjelaskan dan memprediksi varians keterlambatan penerbangan komersial berdasarkan fitur-fitur historisnya.

4.4. Analisis Faktor Penentu (*Feature Importance*)

Salah satu keunggulan utama dari penerapan algoritma *Random Forest* di dalam *Ensemble Model* ini adalah kemampuannya dalam mengekstraksi *Feature Importance* (Tingkat Kepentingan Fitur). Proses ini memetakan seberapa besar kontribusi setiap variabel independen dalam menentukan terjadinya hasil prediksi keterlambatan kedatangan.

Berdasarkan hasil pelatihan model, variabel keterlambatan keberangkatan (*DEPARTURE_DELAY*) secara konsisten muncul sebagai fitur dengan bobot kepentingan absolut yang paling dominan. Hal ini memvalidasi hukum logis dalam aviasi bahwa pesawat yang terlambat lepas landas memiliki probabilitas matematis yang nyaris mutlak untuk mengalami keterlambatan saat mendarat, kecuali rute penerbangan tersebut memiliki kelebihan durasi *air-time* yang bisa dipersingkat. Fitur lain yang memberikan kontribusi signifikan adalah *DISTANCE* (jarak tempuh). Penerbangan dengan jarak tempuh yang lebih pendek cenderung memiliki margin toleransi waktu yang lebih sempit, sehingga gangguan cuaca minor atau keterlambatan *boarding* akan langsung berdampak pada jadwal kedatangannya secara keseluruhan. Analisis fitur ini sangat berharga bagi pihak maskapai untuk menentukan variabel operasional mana yang harus diprioritaskan dalam manajemen harian.

4.5. Diskusi dan Implikasi Manajerial

Penggunaan *Ensemble Model* yang menggabungkan *Linear Regression* dan *Random Forest* terbukti mampu menekan limitasi dari masing-masing algoritma tunggal. Nilai Koefisien Determinasi (R^2) sebesar 0.906 membuktikan bahwa model ini mampu menjelaskan 90,6% varians data keterlambatan kedatangan pesawat secara presisi. Jika dikomparasikan dengan studi terdahulu yang hanya mengandalkan algoritma KNN atau klasifikasi tunggal, pendekatan *ensemble* gabungan ini menawarkan stabilitas komputasi yang jauh lebih relevan untuk diimplementasikan.

Dari kacamata industri penerbangan (implikasi manajerial), hasil prediksi dari pemodelan ini—seperti contoh estimasi *delay* 12,25 menit pada penerbangan baru—dapat diintegrasikan secara *real-time* ke dalam sistem informasi bandara. Maskapai dengan catatan rata-rata *delay* tertinggi (seperti yang terlihat pada visualisasi Gambar 1) dapat memanfaatkan prediksi dini ini untuk mengelola alokasi kru cadangan, mengatur ulang *gate* kedatangan, atau memberikan notifikasi keterlambatan kepada penumpang lebih awal demi meminimalisasi klaim kompensasi dan menjaga kepuasan pelanggan.

5. KESIMPULAN DAN SARAN

Analisis mendalam terhadap data penerbangan menunjukkan bahwa ada variasi yang signifikan dalam tingkat keterlambatan antar maskapai penerbangan, serta pola yang konsisten dari segi hari dalam

seminggu. Model prediksi yang telah dikembangkan menawarkan tingkat akurasi yang memadai dalam memprediksi keterlambatan kedatangan penerbangan. Dari grafik yang dihasilkan, terlihat bahwa beberapa maskapai memiliki tingkat keterlambatan yang lebih tinggi daripada yang lain, menyoroti kemungkinan area di mana peningkatan operasional dapat dilakukan. Disarankan agar penelitian selanjutnya melakukan pemahaman yang lebih mendalam tentang penyebab keterlambatan, seperti faktor cuaca, masalah teknis, atau manajemen lalu lintas udara, sebagai fokus untuk pengembangan model prediksi yang lebih canggih.

DAFTAR PUSTAKA

- [1] F. Raharjo, A. Wiratno, K. Nugroho, dan E. Ardianto, "Prediksi Durasi Keterlambatan Penerbangan Menggunakan Algoritma Random Forest," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 10, no. 2, Apr. 2026.
- [2] A. J. Winata, "Prediksi Keterlambatan Sebuah Maskapai Menggunakan Metode K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 13, no. 1, Jan. 2025.
- [3] V. Y. P. Ardhana *et al.*, "Prediksi Flight Delay Berbasis Algoritma Neural Network," *Journal of Informatics, Electrical and Electronics Engineering*, vol. 2, no. 1, 2022.
- [4] N. A. Putri, "Pengaruh Faktor Cuaca Terhadap Perencanaan Penerbangan dan Dampaknya Pada Keterlambatan Penerbangan," *Jurnal Penelitian*, vol. 10, no. 1, 2025.
- [5] J. Kautsarina, "Metode Random Survival Forest dengan Aturan Pemisahan Log-Rank dan Log-Rank Score untuk Prediksi Delay Penerbangan Akibat Cuaca," *Skripsi*, Universitas Pendidikan Indonesia, Bandung, 2022.
- [6] F. Adinarta, "Predicting Flight Delay Using Random Forest Algorithm, XGBoost Algorithm, and Stacking Ensemble Method," *Proxies - Jurnal UNIKA Soegijapranata*, vol. 4, no. 1, 2025.
- [7] R. Kothari *et al.*, "Selection of Best Machine Learning Model to Predict Delay in Passenger Airlines," *IEEE Access*, vol. 11, 2023.
- [8] [Abhishek Kumar Singh, "Flight Delay Prediction Using Machine Learning Algorithms," *EasyChair Preprint*, 2025.
- [9] Ishaya, N. J, "Flight Delay Analysis and Prediction Using Hybrid Shallow Machine Learning Algorithms," *ISA Publisher*, Mar. 2026.
- [10] Sarah Ahmed A, "Flight Delay Prediction: Evaluating Machine Learning Algorithms for Enhanced Accuracy," *Semantic Scholar Repository*, Des. 2025.