

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan digitalisasi di berbagai sektor menyebabkan volume data yang dihasilkan terus meningkat, sehingga kebutuhan terhadap metode analisis data yang akurat dan efektif juga semakin besar. Jenis data yang paling umum di berbagai bidang adalah data deret waktu (*time series*), atau data yang dicatat berurutan pada titik waktu tertentu. Analisis dan peramalan data *time series* memiliki peranan penting dalam membantu proses perencanaan dan pengambilan keputusan karena mampu memberikan gambaran mengenai situasi di masa depan berdasarkan pola historis data. Peramalan yang baik dapat digunakan sebagai dasar dalam penyusunan strategi, pengalokasian sumber daya, serta evaluasi kebijakan pada berbagai sektor, seperti ekonomi, energi, kesehatan, industri, dan transportasi.

Salah satu sektor yang memerlukan analisis peramalan adalah sektor transportasi udara. Jumlah penumpang pesawat internasional merupakan salah satu indikator yang menunjukkan tingkat mobilitas masyarakat dan aktivitas transportasi antarnegara. Di sektor transportasi udara, pengambilan langkah strategis terkait rancangan operasional, manajemen kapasitas fasilitas, dan penilaian kebijakan dapat didukung secara optimal oleh adanya informasi estimasi jumlah penumpang di masa yang akan datang. Data jumlah penumpang pesawat internasional umumnya memiliki pola deret waktu yang dipengaruhi oleh tren, fluktuasi, maupun kemungkinan pola musiman, sehingga memerlukan metode peramalan yang mampu menangkap karakteristik data tersebut dengan baik. Meskipun demikian, penelitian mengenai peramalan pada data transportasi udara masih relatif lebih sedikit dibandingkan penelitian pada sektor keuangan, pasar saham, maupun energi.

Dalam analisis data *time series*, metode peramalan secara umum dapat dibedakan menjadi pendekatan statistika klasik dan pendekatan berbasis *machine learning*. *Autoregressive Integrated Moving Average* (ARIMA) adalah metode statistika klasik yang populer. Model ARIMA memiliki kemampuan yang baik untuk memodelkan pola linier dan ketergantungan temporal pada data *time series*. Namun, data *time series* pada kondisi nyata tidak selalu hanya mengandung pola linier, melainkan juga dapat memiliki pola nonlinier dan fluktuasi yang kompleks

sehingga tidak seluruh informasi pada data dapat direpresentasikan oleh model ARIMA [2]. Akibatnya, residual hasil pemodelan ARIMA masih berpotensi mengandung pola yang belum berhasil dijelaskan oleh model linier dan dapat dimodelkan lebih lanjut.

Perkembangan metode *machine learning* turut mendorong pemanfaatan algoritma yang mampu mempelajari hubungan data yang rumit dan sulit direpresentasikan oleh model konvensional. *Extreme Gradient Boosting* (XGBoost), yaitu algoritma *ensemble* berbasis *gradient boosting* yang membangun pohon keputusan secara bertahap dan aditif [2]. Pada setiap iterasi, XGBoost membentuk pohon baru untuk memperbaiki kesalahan prediksi dari iterasi sebelumnya melalui proses optimasi fungsi loss berbasis gradien. Mekanisme ini memungkinkan model secara bertahap meminimalkan error prediksi dan menangkap pola nonlinier pada data. Akan tetapi, penerapan XGBoost secara langsung pada data *time series* memiliki keterbatasan dalam menangkap struktur linier dan ketergantungan temporal yang umumnya dimodelkan secara baik oleh metode statistika klasik seperti ARIMA. Oleh karena itu, penggunaan satu model saja sering kali belum mampu merepresentasikan seluruh karakteristik data secara optimal.

Keterbatasan tersebut mendorong berkembangnya pendekatan *hybrid* yang mengombinasikan metode statistika dan *machine learning*. Salah satu pendekatan tersebut yaitu *hybrid* ARIMA-XGBoost. Pendekatan ini didasarkan pada asumsi bahwa ARIMA terlebih dahulu memodelkan komponen linier pada data, sedangkan informasi yang belum berhasil dijelaskan oleh ARIMA masih tersimpan pada residualnya. Oleh karena itu, residual ARIMA digunakan sebagai masukan bagi XGBoost untuk memodelkan komponen nonlinier yang masih tersisa [3]. Dengan demikian, proses *hybrid* memanfaatkan informasi yang masih terkandung pada residual sehingga hasil peramalan akhir diharapkan lebih akurat dibandingkan penggunaan model tunggal.

Studi sebelumnya menunjukkan bahwa metode *hybrid* mampu menghasilkan hasil peramalan yang baik. Odhiambo dkk. (2024) memperoleh tingkat akurasi prediksi yang tinggi menggunakan model *hybrid* ARIMA-XGBoost pada data transaksi *mobile money* di Kenya [2]. Penelitian lain pada peramalan jumlah kasus baru HIV di Indonesia juga menunjukkan bahwa model *hybrid* ARIMA-XGBoost mampu menghasilkan nilai MAPE sebesar 9.75%, sehingga menunjukkan akurasi peramalan yang baik pada data tersebut [3]. Selain itu, Nasiri dan Ebadzadeh (2023) menerapkan model *Recurrent Fuzzy Neural Network* yang dikombinasikan dengan *Variational Mode Decomposition* pada horizon peramalan *1-step ahead*,

3-step ahead, dan *5-step ahead*. Hasil penelitian tersebut menunjukkan bahwa tingkat akurasi model dapat berbeda pada setiap horizon peramalan dan cenderung menurun seiring bertambahnya panjang horizon prediksi. Penelitian tersebut juga menunjukkan bahwa pendekatan *hybrid* memberikan performa yang lebih baik dibandingkan model tunggal pada berbagai horizon peramalan [4].

Meskipun penelitian mengenai model *hybrid* telah banyak dilakukan, kajian mengenai perbandingan kinerja antara ARIMA, XGBoost, dan *hybrid* ARIMA–XGBoost pada data transportasi udara masih sangat terbatas. Padahal, perbandingan tersebut diperlukan untuk mengetahui apakah penggunaan model *hybrid* benar-benar memberikan peningkatan akurasi dibandingkan model tunggal pada karakteristik data transportasi udara. Selain itu, penelitian mengenai *hybrid* ARIMA–XGBoost umumnya masih berfokus pada satu horizon peramalan, sehingga analisis pengaruh horizon peramalan terhadap tingkat akurasi model masih belum banyak dibahas. Meskipun demikian, penggunaan berbagai horizon, seperti horizon *1-step*, *3-step*, dan *5-step*, dapat memberikan gambaran tentang kemampuan model untuk melakukan prediksi jangka pendek serta prediksi untuk masa depan yang lebih jauh.

Dengan demikian, penelitian ini menerapkan model *hybrid* ARIMA-XGBoost pada data jumlah kedatangan penumpang pesawat internasional di Bandara Soekarno-Hatta. Selanjutnya dilakukan evaluasi menggunakan horizon peramalan *1-step ahead*, *3-step ahead*, dan *5-step ahead* untuk menentukan horizon dengan tingkat akurasi terbaik pada masing-masing model. Kinerja model *hybrid* kemudian dibandingkan dengan model ARIMA dan XGBoost menggunakan ukuran RMSE, MAE, dan MAPE guna mengetahui efektivitas pendekatan *hybrid* dalam meningkatkan akurasi peramalan pada data transportasi udara.

1.2 Rumusan Masalah

Dari latar belakang tersebut, masalah yang akan dikaji dalam penelitian ini adalah sebagai berikut.

1. Bagaimana hasil pemodelan *hybrid* ARIMA–XGBoost pada data jumlah kedatangan penumpang pesawat internasional di Bandara Soekarno–Hatta?
2. Horizon peramalan manakah (*1-step ahead*, *3-step ahead*, atau *5-step ahead*) yang menghasilkan tingkat akurasi terbaik pada masing-masing model ARIMA, XGBoost, dan *hybrid* ARIMA–XGBoost?

3. Apakah model *hybrid* ARIMA–XGBoost memberikan kinerja peramalan yang lebih baik dibandingkan model ARIMA dan XGBoost pada horizon peramalan terbaik model *hybrid* berdasarkan nilai RMSE, MAE, dan MAPE?

1.3 Batasan Masalah

Adapun batasan masalah yang ditetapkan dalam penelitian ini adalah.

1. Data yang digunakan yaitu data *time series* jumlah kedatangan penumpang pesawat internasional di Bandara Soekarno Hatta.
2. Penelitian hanya berfokus pada model ARIMA nonmusiman sehingga pola musiman yang mungkin terdapat pada data tidak dimodelkan secara khusus melalui komponen musiman.
3. Penelitian ini dibatasi pada model *hybrid* aditif dengan pendekatan *residual modeling*, sehingga XGBoost digunakan untuk memodelkan residual hasil ARIMA, sedangkan prediksi akhir diperoleh dari penjumlahan prediksi ARIMA dan prediksi residual oleh XGBoost.
4. Fitur yang digunakan sebagai masukan pada model XGBoost dibatasi pada *lag feature* residual hasil pemodelan ARIMA, tanpa menambahkan fitur eksternal maupun fitur hasil rekayasa lainnya.
5. *Hyperparameter* XGBoost dioptimasi dengan menggunakan metode *Grid Search 5-Fold Cross Validation*.
6. *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Mean Absolute Percentage Error* (MAPE) digunakan untuk mengevaluasi kinerja model.
7. Penelitian membatasi perbandingan model *hybrid* ARIMA–XGBoost hanya terhadap model-model tunggalnya saja yaitu ARIMA dan XGBoost.
8. Perangkat lunak Python dan R digunakan untuk melakukan simulasi numerik.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut.

1. Memperoleh hasil pemodelan *hybrid* ARIMA–XGBoost pada data jumlah kedatangan penumpang pesawat internasional di Bandara Soekarno–Hatta.
2. Mengidentifikasi horizon peramalan (*1-step ahead*, *3-step ahead*, atau *5-step ahead*) yang memberikan tingkat akurasi terbaik pada masing-masing model ARIMA, XGBoost, dan *hybrid* ARIMA–XGBoost.
3. Menganalisis kinerja model *hybrid* ARIMA–XGBoost dibandingkan model ARIMA dan XGBoost pada horizon peramalan terbaik model *hybrid* berdasarkan nilai RMSE, MAE, dan MAPE.

1.5 Metodologi Penelitian

1. Studi Literatur

Langkah awal dilakukan dengan meninjau berbagai literatur tentang teori peramalan *time series*, terutama metode ARIMA, XGBoost, dan *hybrid* ARIMA–XGBoost.

2. Studi Kasus dan Analisis

Tahap ini mencakup analisis mekanisme kerja model ARIMA, XGBoost dan *hybrid* ARIMA–XGBoost. Selanjutnya dilakukan penerapan model pada data jumlah penumpang pesawat internasional di Bandara Soekarno–Hatta menggunakan perangkat lunak *Python*, yang ditujukan untuk mengevaluasi kinerja model dalam memproyeksikan data deret waktu.

1.6 Sistematika Penulisan

Penelitian ini ditulis dalam lima bab utama yang tersusun dalam sistematika sebagai berikut.

BAB I PENDAHULUAN

Bab ini tersusun atas latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi uraian mengenai teori-teori yang menjadi dasar dalam penelitian. Pembahasan difokuskan pada konsep peramalan *time series* serta teori yang mendasari metode ARIMA dan XGBoost sebagai pendekatan utama dalam pemodelan dan peramalan data.

BAB III METODE PENELITIAN

Bab ini menyajikan pembahasan mengenai metode yang digunakan, mencakup metode ARIMA, XGBoost, dan *hybrid* ARIMA-XGBoost serta langkah-langkah penerapannya.

BAB IV STUDI KASUS DAN ANALISIS

Bab ini menyajikan penerapan model *hybrid* ARIMA-XGBoost pada studi kasus, hasil peramalan yang diperoleh, serta analisis dan pembahasan perbandingan performa model.

BAB V KESIMPULAN DAN SARAN

Bab ini mencakup kesimpulan dari analisis yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.

