

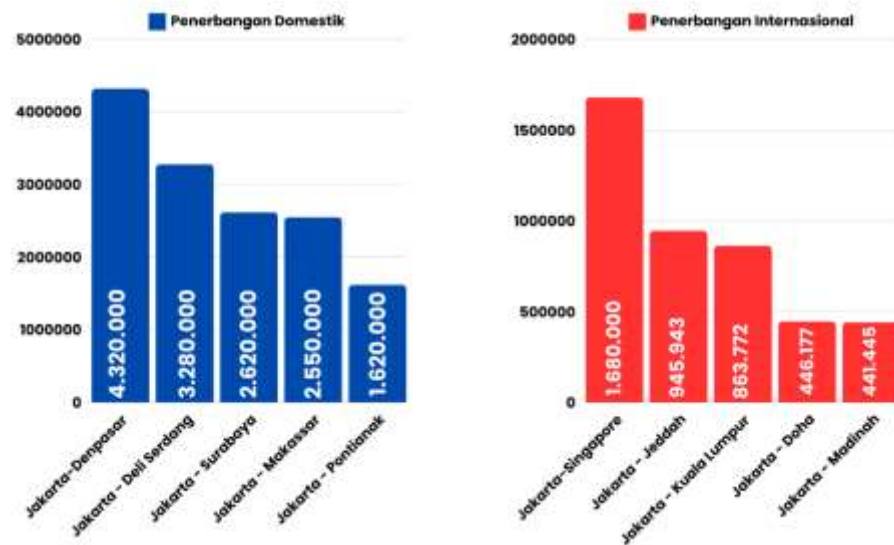
# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Indonesia merupakan negara kepulauan terbesar di dunia, dengan lebih dari 17.000 pulau yang terpisah oleh perairan luas, serta jarak yang cukup jauh antara satu pulau dengan yang lainnya [1][2]. Kondisi ini menyebabkan alat transportasi menjadi salah satu faktor krusial dalam menghubungkan berbagai wilayah di Indonesia. Terutama transportasi udara yang memiliki peran penting untuk mendukung kegiatan mobilitas masyarakat. Hal ini didukung dengan kemampuannya untuk mengatasi hambatan geografis terhadap wilayah yang sulit untuk dijangkau apabila menggunakan transportasi darat atau laut. Transportasi udara di Indonesia tidak hanya menjadi kendaraan untuk penumpang, namun juga berperan penting dalam mendukung mobilitas barang, logistik, dan pariwisata.

Badan Pusat Statistik (BPS) menyebutkan dalam publikasinya bahwa di tahun 2022 Indonesia memiliki 34 bandara untuk penerbangan internasional, dan 124 bandara untuk penerbangan domestik [3]. Berdasarkan laporan dari PT Angkasa Pura (AP) II, sepanjang bulan Januari sampai dengan bulan Desember 2022, Bandar Udara Soekarno-Hatta memiliki jumlah pelayanan penumpang pesawat sebanyak 40,54 juta. Dari informasi tersebut, diketahui rute untuk penerbangan domestik dengan jumlah penumpang terbanyak di tahun 2022 adalah rute Jakarta (CGK) - Denpasar (DPS) dengan jumlah 4,32 juta penumpang, sebagaimana ditunjukkan pada Gambar 1.1



Gambar 1. 1 Rute Penerbangan dengan Penumpang Terbanyak di Bandara Soekarno-Hatta pada Tahun 2022

*Official Airline Guide* (OAG), yang merupakan perusahaan analisis perjalanan udara berbasis di Inggris, merilis laporan yang menyebutkan bahwa rute Jakarta (CGK) – Denpasar (DPS) menempati posisi kesepuluh sebagai rute penerbangan domestik tersibuk di dunia pada tahun 2023, sebagaimana ditunjukkan pada tabel 1.1 di bawah ini.

Tabel 1. 1 Peringkat Rute Penerbangan Domestik dengan Jumlah Kursi Tertinggi di Dunia pada Tahun 2023

| Rank | Rute                                           | Jumlah Kursi | Rank (2019) | Rank (2022) | Persentase Perubahan 2019 ke 2023 | Persentase Perubahan 2022 ke 2023 |
|------|------------------------------------------------|--------------|-------------|-------------|-----------------------------------|-----------------------------------|
| 1    | Jeju Internasional (CJU) - Seoul Gimpo (GMP)   | 13.728.786   | 1           | 1           | -21%                              | -12%                              |
| 2    | Sapporo New Chitose (CTS) - Tokyo Haneda (HND) | 11.936.302   | 2           | 2           | -4%                               | 12%                               |

| Rank | Rute                                    | Jumlah Kursi | Rank (2019) | Rank (2022) | Persentase Perubahan 2019 ke 2023 | Persentase Perubahan 2022 ke 2023 |
|------|-----------------------------------------|--------------|-------------|-------------|-----------------------------------|-----------------------------------|
| 3    | Fukuoka (FUK) - Tokyo Haneda (HND)      | 11.264.229   | 3           | 4           | -1%                               | 8%                                |
| 4    | Hanoi (HAN) - Ho Chi Minh City (SGN)    | 10.883.555   | 4           | 3           | 6%                                | 3%                                |
| 5    | Melbourne (MEL) – Sydney (SYD)          | 9.342.312    | 5           | 5           | -6%                               | 14%                               |
| 6    | Beijing (PEK) - Shanghai Hongqiao (SHA) | 8.355.225    | 7           | 14          | 3%                                | 49%                               |
| 7    | Tokyo Haneda (HND) - Okinawa Naha (OKA) | 7.982.218    | 9           | 7           | 4%                                | 9%                                |
| 8    | Jeddah (JED) – Riyadh (RUH)             | 7.902.142    | 8           | 6           | -1%                               | 5%                                |
| 9    | Mumbai (BOM) – Delhi (DEL)              | 7.276.430    | 6           | 9           | -12%                              | 4%                                |
| 10   | Jakarta (CGK) – Denpasar (DPS)          | 7.190.961    | 12          | 8           | 8%                                | 1%                                |

Peringkat pada kolom di atas ini didasarkan dengan jumlah kursi penerbangan yang terjual sepanjang tahun untuk rute tersebut, dimana Rute Jakarta (CGK) – Denpasar (DPS) memiliki jumlah penumpang sebanyak 7.190.961 kursi. Kolom Rank (2019) dan Rank (2022) merupakan informasi peringkat rute penerbangan di tahun 2019 dan 2022, sedangkan kolom persentase perubahan 2019 dan persentase perubahan 2022 menunjukkan persentase jumlah penumpang dibandingkan dengan jumlah di tahun 2019 dan

2022, yang menginformasikan apakah jumlah penumpang pada tahun 2023 mengalami peningkatan atau penurunan dibandingkan tahun tersebut.

Di Indonesia, tipe maskapai penerbangan dibedakan menjadi tiga, yaitu *Low Cost Carrier* (LCC), *Medium Service*, dan *Full Service Airline* (FSA). LCC merupakan maskapai penerbangan yang menerapkan model bisnis dengan fokus kepada efisiensi biaya dan penawaran harga tiket yang cenderung lebih murah. Contoh maskapai LCC yang saat ini beroperasi adalah Super Air Jet, Lion Air, Citilink, TransNusa, dan masih banyak lagi. Sedangkan maskapai *Medium Service* menyediakan layanan penerbangan yang lebih baik daripada maskapai LCC, namun dengan harga tiket yang lebih terjangkau daripada maskapai FSA. Seperti adanya bagasi gratis dalam batas tertentu, makanan ringan, dan hiburan dalam penerbangan. Contoh maskapai *Medium* yang beroperasi di Indonesia adalah Pelita Air, Sriwijaya Air, dan NAM Air. Maskapai penerbangan FSA merupakan maskapai penerbangan yang menawarkan berbagai layanan tambahan kepada penumpang, termasuk makanan, minuman, bagasi gratis, dan hiburan di dalam pesawat. Contoh maskapai FSA di Indonesia saat ini adalah Garuda Indonesia dan Batik Air.

Adanya persaingan antara maskapai penerbangan menjadi salah satu faktor kenaikan dan penurunan harga tiket. Selain itu, penggunaan strategi penetapan harga dinamis berdasarkan variabel tersembunyi oleh perusahaan penerbangan juga mempersulit konsumen untuk memprediksi harga tiket di masa depan [4][5]. Harga, kualitas pelayanan, dan kepercayaan konsumen memberikan pengaruh kepada pembeli dalam memutuskan untuk membeli tiket atau tidak [6]. Selain itu, pengaruh harga dengan promosi secara bersamaan juga mempengaruhi minat pelanggan untuk membeli tiket pesawat, khususnya pada maskapai LCC yang menawarkan pilihan harga yang lebih beragam. Harga tiket pesawat yang terus berubah-ubah menjadi salah satu faktor pertimbangan terbesar bagi pelanggan. Pemantauan dan observasi saja tidak memungkinkan seseorang memprediksi harga penerbangan dengan tepat, diperlukan sistem prediksi yang mampu memprediksi harga dengan akurat. Prediksi dapat dilakukan dengan *data mining*, yaitu proses ekstraksi

pengetahuan yang berguna, atau mencari pola yang tersembunyi dari kumpulan data. Beberapa contoh aplikasi *data mining* adalah regresi, klasifikasi, klasterisasi, asosiasi, dan lain sebagainya.

Dalam penelitian ini, prediksi harga tiket pesawat dilakukan dengan menggunakan algoritma yang dirancang untuk menangani hubungan yang kompleks antara variabel, termasuk hubungan non-linear yang umum terjadi pada data *real*. Beberapa algoritma yang sesuai untuk karakteristik data tersebut adalah *Extreme Gradien Boosting Regression (XGBoost Regression)*, *Categorical Boosting Regression (CatBoost Regression)*, *Random Forest*, dan algoritma *ensemble learning* lainnya.

*XGBoost* merupakan salah satu teknik *ensemble* yang digunakan untuk mengatasi masalah klasifikasi dan regresi. Teknik ini cukup populer digunakan karena kemampuannya untuk menangani data yang kompleks, termasuk data yang memiliki *noise* atau kesalahan[7]. Model *XGBoost* mendukung *parallel processing*, yang memungkinkan pelatihan model yang lebih cepat pada dataset besar. *XGBoost* memiliki keunggulan dalam menangani data yang kompleks dan non-linear, serta tidak memerlukan asumsi normalitas, linearitas, atau homoskedastisitas. Terdapat penelitian sebelumnya yang menunjukkan bahwa algoritma *XGBoost* memberikan performa yang lebih baik dibandingkan algoritma *Linear Regression* dan *Random Forest Regression* [5]. Hal ini menunjukkan keunggulan *XGBoost* yang memiliki fleksibilitas tinggi dalam menangani berbagai jenis data dan mampu mengatasi *missing values* secara otomatis.

Metode boosting lainnya yang juga memanfaatkan pohon keputusan sebagai prediktor dasar adalah *Categorical Boosting (CatBoost)*. *CatBoost* merupakan algoritma *boosting* yang didasarkan pada algoritma peningkatan gradien menggunakan pohon keputusan untuk regresi. Berbeda dengan *XGBoost* dan *LightGBM* yang menggunakan pohon tidak simetris, *CatBoost* menggunakan pohon simetris untuk mencapai prediksi yang optimal. Algoritma ini dirancang khusus untuk menangani data kategorikal, mampu mengatasi nilai yang hilang, dan bekerja dengan baik pada volume data yang

besar. Penelitian sebelumnya menunjukkan performa *CatBoost* yang mampu menungguli *XGBoost* dari hasil evaluasinya dalam mengklasifikasikan penyakit jantung [7].

Selain *XGBoost* dan *CatBoost*, salah satu metode *ensemble learning* yang juga populer digunakan dalam regresi adalah metode *Random Forest* yang menggunakan pendekatan *bagging*. Pendekatan ini menggabungkan banyak model independen (pohon keputusan) untuk mengurangi variansi dan meningkatkan akurasi. *Random Forest* mampu menangani data yang kompleks, lebih stabil, dan memiliki kinerja yang baik tanpa memerlukan tuning parameter yang rumit. Selain itu, *Random Forest* juga memungkinkan pengukuran tingkat kepentingan setiap fitur dalam membuat prediksi, sehingga memberikan wawasan tabahan terkait kontribusi variabel terhadap hasil prediksi. Penelitian sebelumnya yang membandingkan algoritma *Random Forest* dengan *Logistic Regression* dalam memprediksi cacat *software* menunjukkan bahwa algoritma *Random Forest* menghasilkan kinerja yang paling baik.

Berdasarkan karakteristik data yang kompleks dan adanya kemungkinan hubungan non-linear antara variabel, penelitian ini menggunakan tiga algoritma regresi berbasis ensemble, yaitu *XGBoost Regression*, *CatBoost Regression*, dan *Random Forest Regression*. Ketiga algoritma ini dipilih karena keunggulannya dalam menangani data yang kompleks, robust terhadap noise, serta fleksibilitasnya dalam mengolah data dengan karakteristik beragam. Dengan menerapkan ketiga algoritma ini, penelitian diharapkan dapat memodelkan prediksi harga tiket pesawat secara akurat serta mengidenfitikasi variabel yang mempengaruhi fluktuasi harga. Perbandingan hasil prediksi dari ketiga algoritma tersebut akan memberikan wawasan yang lebih komprehensif mengenai kinerja masing-masing metode dalam konteks tugas prediksi harga tiket pesawat.

## 1.2 Rumusan Masalah

Fluktuasi harga tiket pesawat yang sulit diprediksi melalui pengamatan dan observasi manual menyulitkan para calon penumpang dalam memprediksi harga tiket pesawat dengan tepat. Kurangnya informasi mengenai faktor-faktor yang mempengaruhi tinggi rendahnya harga tiket pesawat menjadi tantangan lain bagi calon penumpang. Selain itu, penerapan strategi penetapan harga dinamis oleh perusahaan, yang didasari pada variabel tersembunyi, semakin memperumit upaya konsumen untuk memprediksi harga tiket di masa depan. Oleh karena itu, diperlukan model prediksi yang akurat menggunakan algoritma regresi untuk membantu calon penumpang mengantisipasi fluktuasi harga tiket pesawat dengan memanfaatkan variabel-variabel penerbangan yang relevan.

## 1.3 Pertanyaan Penelitian

Penelitian ini bertujuan untuk menjawab pertanyaan yang berkaitan dengan prediksi harga tiket pesawat sebagai berikut:

1. Berapa nilai performa *XGBoost Regression*, *CatBoost Regression*, dan *Random Forest Regression* untuk memprediksi harga tiket pesawat pada rute penerbangan Jakarta (CGK) – Denpasar (DPS)?
2. Variabel apa saja yang mempengaruhi fluktuasi harga tiket pesawat dalam rute penerbangan Jakarta (CGK) – Denpasar (DPS)?

## 1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut:

1. Membandingkan kinerja model algoritma *XGBoost Regression*, *CatBoost Regression*, dan *Random Forest Regression* dalam memprediksi harga tiket pesawat.
2. Mengidentifikasi variabel-variabel yang signifikan mempengaruhi fluktuasi harga tiket pesawat.

### 1.5 Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan untuk memperjelas ruang lingkup penelitian agar tetap terarah dan sesuai dengan tujuan yang ingin dicapai. Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data penelitian yang digunakan merupakan data sekunder yang berasal dari website pemesanan tiket pesawat, Traveloka Indonesia.
2. Dataset diambil dengan menggunakan aplikasi perangkat lunak Octoparse 8.
3. Dataset yang digunakan merupakan data harga untuk rute penerbangan Jakarta (CGK) – Denpasar (DPS).
4. Periode pengambilan dataset dari tanggal 1 Juli 2024 hingga 7 September 2024.
5. Algoritma yang digunakan pada penelitian adalah *XGBoost Regression*, *CatBoost Regression*, dan *Random Forest Regression*.

### 1.6 Manfaat Penelitian

Penelitian ini dilakukan untuk dapat memberikan beberapa manfaat sebagai berikut:

1. Memberikan wawasan mendalam mengenai variabel-variabel yang mempengaruhi perubahan harga tiket pesawat.
2. Berkontribusi pada pengembangan pengetahuan dan teknologi di bidang *Machine Learning* melalui penggunaan algoritma regresi.
3. Membantu pelanggan untuk membuat keputusan yang lebih baik dalam merencanakan perjalanannya.