**📊 Rencana Proyek Berdasarkan CRISP-DM**

1. **Business Understanding**:
   * **Masalah**: Harga tiket pesawat sangat fluktuatif dan memengaruhi keputusan pelanggan.
   * **Tujuan**: Memprediksi harga tiket berdasarkan waktu, rute, maskapai, dll.[Kaggle](https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction?utm_source=chatgpt.com)
2. **Data Understanding**:
   * Eksplorasi fitur-fitur seperti maskapai, asal, tujuan, durasi, jumlah pemberhentian, waktu keberangkatan, dan harga.
   * Analisis distribusi data dan identifikasi pola awal.
3. **Data Preparation**:
   * Pembersihan data dari nilai yang hilang atau tidak konsisten.
   * Transformasi fitur dan pembuatan fitur baru jika diperlukan.[BISA AI](https://bisa.ai/portofolio/detail/Mjc?utm_source=chatgpt.com)
4. **Modeling**:
   * Penerapan algoritma seperti Linear Regression, Random Forest, atau XGBoost.
   * Penggunaan teknik validasi silang untuk menghindari overfitting.
5. **Evaluation**:
   * Evaluasi model menggunakan metrik seperti RMSE, MAE, dan R².
   * Perbandingan dengan baseline model untuk menilai peningkatan kinerja.
6. **Deployment**:
   * Diskusi tentang bagaimana model dapat diintegrasikan dalam sistem nyata, seperti aplikasi pemesanan tiket atau sistem peringatan harga.
7. **Insight dan Rekomendasi**:
   * Identifikasi faktor-faktor utama yang memengaruhi harga tiket.
   * Rekomendasi strategi harga atau promosi berdasarkan hasil analisis.

**rediksi Harga Tiket Pesawat – Proyek Akhir Data Science**

**1. Pemilihan Dataset dan Latar Belakang Bisnis**

Dataset yang digunakan berisi informasi **harga tiket pesawat** domestik (India) dengan berbagai fitur terkait penerbangan seperti maskapai, kota asal/tujuan, waktu keberangkatan, jumlah transit, kelas penerbangan, durasi, dan jumlah hari sebelum keberangkatan. Pemilihan dataset ini didasari oleh relevansi bisnis **harga tiket pesawat yang sangat fluktuatif**. Dalam industri penerbangan, harga tiket dapat berubah-ubah tergantung waktu pemesanan, permintaan pasar, dan strategi maskapai. Hal ini menjadi isu penting bagi **maskapai** (dalam mengoptimalkan pendapatan), **pelanggan** (dalam mencari harga terbaik), maupun **Online Travel Agent (OTA)** (dalam menyediakan rekomendasi harga kompetitif). **Harga tiket pesawat sangatlah fluktuatif**, sehingga ketersediaan data historis harga dapat dimanfaatkan dengan *machine learning* untuk **memprediksi waktu dan harga terbaik** bagi pembelian tiket[mmsi.binus.ac.id](https://mmsi.binus.ac.id/2018/10/31/penggunaan-big-data-pada-online-travel-agent/#:~:text=,waktu%20yang%20tepat%20untuk%20membelinya). Dengan demikian, dataset ini sesuai untuk proyek akhir karena mencakup permasalahan nyata *dynamic pricing* di industri penerbangan yang dapat disolusikan melalui analisis data dan pemodelan prediktif.

Selain itu, dataset ini cukup **kompleks dan lengkap** untuk mendemonstrasikan penerapan kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining)[astutirahmarubi.medium.com](https://astutirahmarubi.medium.com/metodelogi-sains-data-data-science-ed1d698d38#:~:text=C%20RISP,tahap%20yang%20saling%20terkait%20yaitu). CRISP-DM terdiri dari enam tahap yang saling terkait: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Proyek prediksi harga tiket pesawat ini akan mengikuti tahapan tersebut secara terstruktur. **Latar belakang bisnis** yang mendasari adalah kebutuhan maskapai/OTA untuk memahami pola penentuan harga dan menyediakan solusi seperti **prediksi harga tiket** di masa mendatang. Dengan memiliki model prediksi yang akurat, bisnis dapat mengambil tindakan proaktif: maskapai dapat mengatur strategi harga dinamis, sementara OTA dapat memberi saran kepada pelanggan kapan sebaiknya membeli tiket (misalnya menunggu saat harga diprediksi turun[mmsi.binus.ac.id](https://mmsi.binus.ac.id/2018/10/31/penggunaan-big-data-pada-online-travel-agent/#:~:text=pencarian%20tiket%20pesawat%20dari%20Jakarta,waktu%20yang%20tepat%20untuk%20membelinya)).

Secara konteks masalah, **harga tiket yang tepat** sangat mempengaruhi keputusan konsumen dan profitabilitas maskapai. Misalnya, jika harga terlalu tinggi mendekati hari keberangkatan, kursi bisa kosong; sebaliknya jika terlalu rendah, pendapatan bisa hilang. Demikian pula bagi pelanggan, fluktuasi harga menyebabkan ketidakpastian dalam merencanakan perjalanan. Oleh karena itu, menganalisis data harga tiket pesawat secara mendalam dan membangun model prediksi merupakan proyek yang *impactful* dan relevan bagi stakeholder industri penerbangan.

**2. Permasalahan dan Tujuan Bisnis**

Berdasarkan latar belakang di atas, **permasalahan utama** yang ingin diselesaikan adalah *bagaimana memprediksi harga tiket pesawat di masa mendatang* berdasarkan informasi rute, waktu, dan faktor-faktor terkait lainnya. Secara spesifik, kita ingin menjawab pertanyaan: *“Berapa harga tiket yang akan ditawarkan untuk penerbangan tertentu, misalnya 30 hari sebelum tanggal berangkat, dengan maskapai dan rute tertentu?”*.

**Tujuan bisnis** dari proyek ini antara lain:

* **Untuk maskapai**: Membantu strategi *revenue management* dengan **memahami faktor apa saja yang memengaruhi harga** dan seberapa besar pengaruhnya. Model prediksi harga dapat digunakan maskapai untuk menyusun kebijakan **dynamic pricing** yang optimal, misalnya memberikan diskon jauh hari sebelum keberangkatan untuk mengisi kursi, lalu menaikkan harga menjelang hari H saat permintaan tinggi.
* **Untuk OTA/platform travel**: Mengembangkan fitur **rekomendasi waktu pembelian** tiket bagi pelanggan. Dengan model prediktif, OTA dapat memberi saran apakah sebaiknya pelanggan membeli sekarang atau menunggu (**predictive analytics** untuk waktu terbaik beli tiket[mmsi.binus.ac.id](https://mmsi.binus.ac.id/2018/10/31/penggunaan-big-data-pada-online-travel-agent/#:~:text=pencarian%20tiket%20pesawat%20dari%20Jakarta,waktu%20yang%20tepat%20untuk%20membelinya)). Ini meningkatkan layanan pelanggan dan kepercayaan pengguna terhadap platform.
* **Untuk pelanggan**: Memberikan **informasi transparan** tentang faktor-faktor yang memengaruhi harga dan prediksi tren harga, sehingga pelanggan bisa mengambil keputusan beli tiket dengan lebih bijak dan potensi menghemat biaya.

Relevansi masalah ini secara bisnis sangat tinggi. Harga tiket langsung berdampak pada **permintaan penumpang dan loyalitas pelanggan**. Di era persaingan ketat dan maraknya OTA, kemampuan memprediksi harga dan pola permintaan merupakan keunggulan strategis. Solusi analisis dan model prediksi ini dapat memberi **nilai tambah** berupa insight: misalnya mengetahui bahwa memesan tiket **jauh hari** cenderung lebih murah[liputan6.com](https://www.liputan6.com/feeds/read/5971716/trik-mendapatkan-tiket-pesawat-murah-untuk-liburan-hemat-berikut-strategi-berburu-tarif-promo#:~:text=Salah%20satu%20strategi%20paling%20efektif,strategi%20untuk%20mengisi%20kursi%20kosong), atau bahwa maskapai tertentu selalu menawarkan harga premium. Insight semacam ini bernilai bagi manajemen dalam **pengambilan keputusan** (pricing, marketing) dan bagi OTA dalam **menyusun strategi pemasaran** (misal promo lebih agresif di rute yang diprediksi harganya naik). Dengan demikian, penyelesaian masalah prediksi harga tiket diharapkan mampu **meningkatkan pendapatan** maskapai/OTA serta **menghemat biaya** bagi konsumen.

**Objektif proyek** dapat dirangkum sebagai: (a) memahami pola historis harga tiket pesawat (descriptive & diagnostic analytics), (b) membangun model *machine learning* yang mampu memprediksi harga tiket (predictive analytics), dan (c) menerjemahkan hasil analisis tersebut menjadi **aksi bisnis** yang direkomendasikan (prescriptive insight).

**3. Preparasi Data dan *Feature Engineering***

Pada tahap *data preparation*, dilakukan serangkaian langkah *preprocessing* untuk memastikan data siap digunakan dalam analisis dan pemodelan. Dataset **Flight\_Price.csv** terdiri dari 12 kolom: airline, flight (kode penerbangan), source\_city, departure\_time (kategori waktu berangkat), stops (jumlah transit), arrival\_time (kategori waktu tiba), destination\_city, class (kelas layanan, Economy/Business), duration (durasi penerbangan dalam jam), days\_left (jumlah hari sebelum berangkat saat tiket di-booking), dan price (harga tiket dalam Rupee). Berikut langkah-langkah preparasi dan rekayasa fitur yang dilakukan:

* **Data Cleaning:** Pemeriksaan awal menunjukkan tidak ada nilai kosong (*null*) pada kolom-kolom dataset. Dataset berisi **300.153 baris** data tanpa missing values. Tidak ditemukan duplikasi baris yang persis sama, karena setiap entri merepresentasikan kombinasi unik rute, tanggal (implisit dalam days\_left), maskapai, dan nomor penerbangan. Oleh karena itu, *cleaning* difokuskan pada penanganan format data. Kolom indeks bawaan dihapus (Unnamed: 0) karena tidak diperlukan. Kolom flight (kode penerbangan) juga diabaikan dalam analisis lanjutan, sebab kode unik tiap flight tidak memberikan informasi umum (malah bisa menyebabkan model *overfit* karena menangkap identitas penerbangan tertentu). Tidak dilakukan penghapusan outlier harga, karena nilai ekstrim (harga sangat tinggi) diasumsikan merepresentasikan tiket **kelas bisnis** atau rute jarak jauh yang valid, sehingga tetap dipertahankan sebagai bagian dari pola data nyata.
* **Transformasi Kolom Kategorikal:** Beberapa kolom memiliki nilai kategorikal non-numerik. Kita melakukan **encoding** agar dapat diproses model ML.
  + Kolom stops (jumlah transit) berisi kategori teks: "zero" (langsung), "one" (satu transit), "two\_or\_more" (dua atau lebih transit). Ini dikonversi ke representasi numerik ordinal: *zero*→0, *one*→1, *two\_or\_more*→2. Urutan ini merefleksikan jumlah transit. Walaupun model pohon keputusan mampu menangani kategori secara langsung, untuk model linear kita juga menyiapkan *dummy variables*.
  + Kolom departure\_time dan arrival\_time berisi kategori waktu dalam hari: *Early Morning, Morning, Afternoon, Evening, Night, Late Night*. Kolom ini di-*encode* menggunakan **one-hot encoding** (dikonversi menjadi 6 kolom biner masing-masing untuk setiap kategori waktu, dengan 1 menandakan kategori tersebut). Pendekatan one-hot dipilih karena kategori waktu bukan bersifat ordinal (tidak ada urutan absolut mana “lebih besar”). Encoding ini memungkinkan model menangkap perbedaan pola harga misal antara penerbangan pagi vs malam.
  + Kolom airline, source\_city, dan destination\_city juga di-*encode* dengan one-hot encoding. Misalnya, terdapat 6 maskapai unik (SpiceJet, AirAsia, Vistara, GO\_FIRST, Indigo, Air India) yang akan diwakili 6 kolom dummy (dengan salah satu sebagai baseline). Demikian pula 6 kota asal dan 6 kota tujuan. Dengan encoding ini, model dapat menangkap **pengaruh maskapai tertentu** atau **kota asal/tujuan tertentu** terhadap harga. Contohnya, dummy airline\_Air\_India=1 akan menunjukan baris tersebut adalah penerbangan Air India, sehingga model bisa belajar bias harga Air India.
  + Kolom class (kelas penerbangan) diubah menjadi biner: Economy=0 dan Business=1. Variabel ini sangat penting karena kelas bisnis jelas memiliki struktur harga berbeda dibanding ekonomi.
* **Manipulasi Data Waktu:** Dataset telah menyediakan fitur days\_left yang menyatakan berapa hari sebelum keberangkatan tiket tersebut di-booking. Fitur ini adalah krusial untuk analisis **dynamic pricing**, karena biasanya harga tiket berubah seiring mendekati tanggal berangkat. Pada data, rentang days\_left adalah 1 hingga 49 hari (median 26 hari). Semakin kecil days\_left berarti pembelian tiket mendekati hari H. Kami memastikan tipe data kolom ini adalah numerik (integer) sehingga dapat digunakan untuk analisis korelasi dan sebagai fitur input model. Selain days\_left, tersedia juga kolom kategori waktu (departure\_time/arrival\_time) yang sudah ditransformasi seperti dijelaskan. Tidak ada kolom tanggal kalender eksplisit, sehingga analisis *seasonality* (misal perbedaan hari dalam seminggu) tidak dapat dilakukan secara langsung. Namun, *impact* waktu booking vs waktu terbang ditangani melalui kombinasi fitur days\_left dan kategori jam terbang di atas.
* **Rekayasa Fitur Tambahan:** Beberapa fitur baru diturunkan untuk membantu model:
  + **Route (Rute)**: Kombinasi source\_city dan destination\_city bisa dianggap sebagai fitur tunggal “rute penerbangan”. Sebenarnya, model pohon atau *boosting* dapat secara otomatis menangkap interaksi antara kota asal dan tujuan. Namun untuk model linear, interaksi tersebut tidak otomatis tertangkap jika hanya diberi dummy kota asal dan tujuan secara terpisah. Oleh karena itu, sebagai eksplorasi, kami membuat fitur *route* (misal “Delhi-Mumbai”) yang menggabungkan kota asal dan tujuan. Terdapat 30 kemungkinan rute (6 asal × 5 tujuan unik lainnya). Akan tetapi, dalam implementasi final modelling, kami memilih tidak memasukkan langsung fitur rute (karena one-hot 30 kategori agak redundan dengan kombinasi dummy kota asal & tujuan yang sudah ada). Sebagai gantinya, model kompleks (Random Forest, XGBoost) akan mampu memodelkan interaksi origin-destination, sedangkan model linear kami terpaksa mengasumsikan efek terpisah yang mendekati efek rute.
  + **days\_left (hari tersisa)**: Fitur ini sudah ada di data. Jika belum ada, kami akan menghitungnya dari selisih tanggal terbang dan tanggal pembelian. Fitur ini memastikan model menangkap *trend* harga terhadap waktu pemesanan.
  + **duration (durasi)**: Durasi penerbangan dalam jam (misal 2.17 jam) diambil apa adanya sebagai fitur numerik. Kami cek bahwa durasi konsisten dengan rute dan jumlah transit (misal durasi penerbangan langsung Delhi-Mumbai ~2 jam, sedangkan dengan transit bisa >5 jam). Durasi dapat berkolerasi dengan harga (penerbangan lebih lama biasanya lebih mahal karena jarak jauh atau transit).
  + **Encoding Kategorikal Lanjutan:** Setelah semua transformasi di atas, dataset yang siap untuk modelling memiliki kombinasi fitur numerik (days\_left, duration) dan fitur kategorikal ter-encode (dummy maskapai, dummy kota, dummy waktu, dummy transit, dan class). Total fitur input model sekitar 30 kolom hasil one-hot encoding.

Setelah langkah-langkah di atas, data siap untuk dianalisis lebih lanjut. Setiap keputusan preprocessing diambil dengan justifikasi bisnis: misalnya, **mengapa encoding diperlukan** – karena algoritma ML membutuhkan input numerik; **mengapa menambah fitur**days\_left – karena faktor waktu booking diyakini kuat pengaruhnya terhadap harga, dsb. Hasil akhir tahap ini adalah *dataframe* yang bersih dan kaya fitur untuk eksplorasi visual serta pemodelan prediktif.

**4. Visualisasi dan Analisis Data**

Tahap ini mencakup **analisis deskriptif** dan **diagnostik** terhadap data harga tiket. Sesuai *Gartner Analytic Ascendancy Model*, mula-mula kita melakukan **analisis deskriptif** (menjawab *“apa yang terjadi?”*) dengan visualisasi data historis untuk menemukan pola atau tren[astutirahmarubi.medium.com](https://astutirahmarubi.medium.com/metodelogi-sains-data-data-science-ed1d698d38#:~:text=Analisis%20deskriptif%20melibatkan%20memeriksa%20data,pertanyaan%20dasar%20seperti). Selanjutnya, dilakukan **analisis diagnostik** untuk menjawab *“mengapa hal itu terjadi?”*[astutirahmarubi.medium.com](https://astutirahmarubi.medium.com/metodelogi-sains-data-data-science-ed1d698d38#:~:text=), yakni menginterpretasi faktor-faktor yang memengaruhi pola tersebut. Terakhir, hasil insight ini akan mendasari **analisis prediktif** di tahap modeling (menjawab *“apa yang akan terjadi?”*). Berikut adalah beberapa temuan utama dari visualisasi data:

**a. Harga Tiket per Maskapai**

**Gambar 1: Distribusi harga tiket (Rp) untuk tiap maskapai (boxplot)**. Terlihat perbedaan mencolok antar maskapai. Maskapai **full-service** seperti *Vistara* dan *Air India* memiliki rentang harga yang jauh lebih tinggi dibanding maskapai *low-cost carrier* (LCC) seperti *SpiceJet, AirAsia, GO\_FIRST,* maupun *Indigo*. Median harga tiket Vistara dan Air India berada di kisaran Rp15.000–20.000, jauh di atas median maskapai budget yang umumnya sekitar Rp6.000–7.500. Whiskers (rentang IQR) Vistara dan Air India memanjang hingga >Rp50.000, menunjukkan adanya banyak tiket berharga sangat tinggi (didominasi oleh kelas bisnis). Sementara itu, maskapai LCC sebagian besar tiketnya berada di bawah Rp10.000. **Insight:** Maskapai full-service cenderung lebih mahal karena menawarkan layanan premium (bagasi, meal, dll) dan kelas bisnis, sementara LCC fokus pada harga murah. Hal ini *menjelaskan* (diagnostik) mengapa distribusi harga Air India/Vistara lebih tinggi: banyak penerbangan mereka mencakup kelas bisnis dengan harga puluhan ribu rupiah, sedangkan LCC hanya menjual kelas ekonomi. **Tindakan:** Bagi pelanggan yang mengutamakan harga, memilih LCC bisa menghemat biaya, namun dengan trade-off fasilitas. Sebaliknya, bagi maskapai full-service, data ini menegaskan posisi mereka sebagai premium carrier — mereka bisa terus memanfaatkan diferensiasi layanan untuk membenarkan harga lebih tinggi.

**b. Harga Tiket vs Jumlah Transit**

**Gambar 2: Boxplot harga tiket berdasarkan jumlah transit (zero: direct, one: 1 transit, two\_or\_more: ≥2 transit)**. Pola umum yang tampak: **penerbangan langsung (tanpa transit) cenderung berharga paling murah**, sedangkan penerbangan dengan transit (satu atau lebih) menunjukkan harga yang lebih tinggi secara median. Terlihat median harga *direct flight* sekitar Rp4.500, lebih rendah dibanding median penerbangan *one-stop* (~Rp8.000) atau *two-stop* (~Rp8.300). Namun, distribusi harga *one-stop* sangat lebar – whisker atas mencapai >Rp100.000. Hal ini karena kategori *one-stop* mencakup berbagai skenario: mulai dari tiket ekonomi transit yang kadang dijual murah, hingga **tiket bisnis dengan 1 transit** yang harganya sangat mahal (bahkan lebih mahal dari direct flight kelas bisnis). Akibat pencampuran ini, rentang harga *one-stop* menjadi luas. **Insight:** Secara keseluruhan, *semakin banyak transit, harga cenderung meningkat*. Penerbangan direct biasanya paling efisien dan sering kali merupakan pilihan termurah bagi penumpang ekonomi. Penerbangan dengan 2 transit biasanya adalah rute yang tidak tersedia direct, sehingga penumpang harus membayar lebih untuk kombinasi multi-penerbangan. Selain itu, tambahan transit berarti **durasi lebih lama** dan mungkin melibatkan maskapai berbeda, yang menambah biaya. **Diagnostik:** Mengapa direct lebih murah? Mungkin karena direct flight adalah rute populer dengan banyak frekuensi (kompetisi tinggi menekan harga), sedangkan rute dengan transit bisa terjadi pada rute kurang populer atau jarak lebih jauh. Juga, maskapai kerap memberikan harga kompetitif untuk rute direct unggulan mereka. **Tindakan/Rekomendasi:** Bagi pelanggan, jika mengincar tiket murah, usahakan pilih rute direct (meski hal ini tergantung ketersediaan). Bagi maskapai/OTA, data ini bisa dipakai untuk mengedukasi pengguna bahwa memilih perjalanan dengan banyak transit **tidak selalu lebih murah**, bahkan kadang lebih mahal dan memakan waktu, kecuali memang tidak ada penerbangan langsung tersedia.

**c. Harga Tiket per Kota Asal/Tujuan**

**Gambar 3: Rata-rata harga tiket berdasarkan kota asal (departure city)**. Terlihat perbedaan rata-rata harga tiket untuk penerbangan dari enam kota besar: Delhi, Hyderabad, Bangalore, Mumbai, Kolkata, dan Chennai. Kota **Delhi** memiliki rata-rata harga terendah (~Rp18.9 ribu), disusul Hyderabad (~Rp20.2 rb). Sementara itu, tiket yang berangkat dari **Chennai** dan **Kolkata** cenderung paling mahal (rata-rata ~Rp21.5–21.9 rb). Tren serupa ditemukan jika dilihat dari sisi kota tujuan (Delhi sebagai tujuan rata-rata termurah, Kolkata paling mahal). **Insight:** Rute yang melibatkan Delhi cenderung lebih murah, sedangkan yang melibatkan Kolkata/Chennai lebih mahal. *Diagnostik:* Kemungkinan hal ini karena Delhi adalah hub utama dengan persaingan banyak maskapai dan frekuensi penerbangan tinggi, sehingga harga rata-rata tertekan lebih rendah. Sebaliknya, Chennai dan Kolkata mungkin memiliki rute yang lebih jauh (misal Chennai–Delhi merupakan salah satu rute terpanjang di dataset) dan tidak seramai rute Delhi-Mumbai, sehingga harga cenderung lebih tinggi. Juga, komposisi kelas bisnis mungkin lebih banyak di rute tertentu (misal banyak pebisnis di Bangalore-Chennai sehingga harga rata-rata tinggi). **Tindakan:** Maskapai dapat mengkaji rute-rute mahal (Chennai/Kolkata) untuk melihat apakah harga tinggi menurunkan permintaan – mungkin diperlukan promo di rute tersebut. Bagi pelanggan, data ini memberi tahu bahwa terbang dari/ke kota hub besar (Delhi/Mumbai) biasanya lebih ekonomis, sementara rute ke kota lain mungkin perlu budget lebih.

**d. Hubungan Fitur Numerik (Durasi & Days Left) dengan Harga**  
Selain kategori, kita analisis dua fitur numerik: duration (durasi penerbangan) dan days\_left (hari tersisa).

* *Durasi:* Secara logis, penerbangan berdurasi lebih lama (yang berarti jarak jauh atau banyak transit) cenderung berharga lebih mahal. Korelasi Pearson antara durasi dan price di data ~0.5 (positif moderat). Namun, durasi bukan penentu tunggal karena ada faktor kelas dan rute. Contohnya, durasi 2 jam Delhi-Mumbai (direct) vs 2 jam pada flight lain mungkin berbeda harga tergantung maskapai dan kelas. Visualisasi scatter menunjukkan klaster: flight sekitar 2 jam (direct antar kota besar) berkisar 5-10 ribu (ekonomi) hingga 50 ribu (bisnis), sementara flight >10 jam (pasti dengan transit jauh) bisa mencapai >70 ribu. **Insight:** Durasi memang memengaruhi harga, tapi **efeknya tumpang tindih dengan faktor lain**. Rute dengan durasi ekstrim (sangat lama) biasanya melibatkan kelas bisnis atau kombinasi multi-leg yang mahal.
* *Days Left:* Fitur ini sangat menarik karena terkait **kapan tiket dibeli** relatif terhadap tanggal terbang. Secara umum, **semakin mendekati hari H, harga tiket cenderung naik**. Hal ini konsisten dengan strategi harga maskapai: kursi dijual murah di awal untuk memancing pemesan awal, lalu harga naik seiring kursi terisi dan waktu menyempit (last-minute buyers biasanya bersedia bayar mahal). Dari data, rata-rata harga tiket saat *dibeli 1 hari sebelum terbang* ~Rp21.5 rb, sedangkan jika dibeli 6-7 minggu sebelum (49 hari) rata-ratanya ~Rp18 rb. Bahkan, tren kenaikan harga mulai terlihat signifikan di 2 minggu terakhir sebelum terbang. Kita visualisasikan tren ini:

**Gambar 4: Tren harga rata-rata tiket vs. jumlah hari sebelum keberangkatan**. Sumbu X adalah days\_left (dari 50 hari sebelum, menuju 0 mendekati hari H), sumbu Y rata-rata harga tiket. Terlihat jelas **kurva menanjak** ke kanan: ketika days\_left berkurang (artinya mendekati tanggal penerbangan), harga naik. Khususnya lonjakan tajam terjadi pada 10 hari terakhir. Misal, harga rata-rata 10 hari sebelum ~Rp22 ribu dan 1 hari sebelum ~Rp30 ribu (pada grafik gabungan biru). Tren ini konsisten untuk kelas ekonomi (hijau) dan bisnis (merah) dengan skala berbeda: harga bisnis selalu jauh di atas ekonomi, namun keduanya naik mendekati hari H. Menariknya, untuk ekonomi ada indikasi penurunan kecil dari day 2 ke day 1 (mungkin maskapai melepas sisa kursi murah di saat akhir untuk mengisi pesawat, sehingga rata-rata day 1 sedikit turun dibanding day 2 yang peak). **Insight:** Pemesanan tiket **jauh-jauh hari** adalah strategi terbaik untuk mendapatkan harga murah[liputan6.com](https://www.liputan6.com/feeds/read/5971716/trik-mendapatkan-tiket-pesawat-murah-untuk-liburan-hemat-berikut-strategi-berburu-tarif-promo#:~:text=Salah%20satu%20strategi%20paling%20efektif,strategi%20untuk%20mengisi%20kursi%20kosong). Data mendukung panduan umum industri: pesan tiket domestik idealnya 1-2 bulan sebelum berangkat untuk harga optimal. **Diagnostik:** Mengapa harga naik mendekati keberangkatan? Karena prinsip *yield management* – semakin sedikit waktu tersisa, maskapai menaikkan harga memanfaatkan urgensi penumpang last-minute. Selain itu, kursi yang tersisa biasanya lebih sedikit, jadi penumpang berebut sisa kapasitas dengan harga tinggi. Sebaliknya, pemesanan jauh hari memungkinkan maskapai memberi tarif promo lebih rendah untuk memastikan tingkat keterisian kursi lebih awal[liputan6.com](https://www.liputan6.com/feeds/read/5971716/trik-mendapatkan-tiket-pesawat-murah-untuk-liburan-hemat-berikut-strategi-berburu-tarif-promo#:~:text=memesan%20jauh,strategi%20untuk%20mengisi%20kursi%20kosong). **Tindakan:** Bagi pelanggan, jelas disarankan untuk membeli tiket sedini mungkin sebelum keberangkatan untuk menghindari kenaikan harga drastis di akhir. OTA dapat memanfaatkan pola ini dengan fitur notifikasi kepada pengguna ketika mendekati periode kenaikan harga (misal “harga diperkirakan naik signifikan minggu depan”). Maskapai juga dapat melihat data ini sebagai evaluasi strategi harga: apakah kenaikan last-minute terlalu tinggi hingga menyisakan kursi kosong? Jika ya, mungkin perlu penyesuaian (misal last-minute flash sale).

* *Waktu keberangkatan (departure/arrival time):* Selain fitur numerik, analisis tambahan dilakukan pada kategori waktu berangkat/tiba. Hasilnya, ada pola **jam terbang memengaruhi harga**. Contoh: penerbangan **Late Night (dini hari)** memiliki median harga terendah (~Rp4.499), kemungkinan karena jadwal kurang diminati (red-eye flight) sehingga maskapai memasang harga lebih murah. Sebaliknya, penerbangan **pagi dan malam** cenderung lebih mahal (median Morning ~Rp8.112, Night ~Rp7.813). Ini selaras dengan logika permintaan: jam berangkat pagi (sekitar jam kerja) dan malam (sepulang kerja) banyak diminati pebisnis, sehingga tarifnya lebih tinggi. **Insight & Tindakan:** Penumpang fleksibel dapat mempertimbangkan terbang di larut malam atau sangat dini hari untuk harga lebih murah, sementara terbang di jam sibuk kemungkinan bayar premium. Maskapai bisa terus memanfaatkan jam sibuk untuk profit margin tinggi, dan mendorong permintaan di jam sepi lewat promo.

Semua temuan di atas mencakup **analisis deskriptif** (melihat perbedaan harga antar kategori/faktor) dan **diagnostik** (menalar penyebab perbedaan tersebut). Secara keseluruhan, insight penting yang didapat antara lain: **kelas bisnis selalu lebih mahal daripada ekonomi**, **maskapai full-service > LCC dalam harga**, **harga naik saat mendekati hari H**, **late night flights paling murah**, **lebih banyak transit tidak menjamin lebih murah** (malah cenderung lebih mahal), dan **ada perbedaan harga antar rute/kota**. Insight-insight ini akan digunakan untuk membangun *story* bisnis dan juga sebagai acuan dalam pemodelan prediktif (misal kita memastikan model memasukkan variabel-variabel kunci seperti class, days\_left, airline, dll karena pengaruhnya signifikan).

**5. Pembuatan dan Evaluasi Model**

Setelah memahami data, tahap berikutnya adalah **analisis prediktif** dengan membangun model *machine learning* untuk memprediksi harga tiket pesawat. **Tujuan modeling**: diberikannya input (maskapai, rute, waktu, sisa hari, dll), model dapat memperkirakan **harga tiket** mendekati nilai aslinya. Proses modeling mengikuti langkah-langkah: *data splitting*, *training*, *evaluation*, dan *tuning* sederhana.

**a. Data Splitting:** Dataset dibagi menjadi **training set (80%)** dan **test set (20%)** secara acak. Training set (±240 ribu data) digunakan untuk melatih model, test set (±60 ribu data) untuk mengevaluasi performa pada data yang tidak dilatih (mensimulasikan prediksi untuk data baru). Sebelum split, semua fitur telah di-encode (sesuai penjelasan di bagian 3). Fitur yang digunakan model meliputi: airline (6 dummy), source\_city (6 dummy), destination\_city (6 dummy), departure\_time (6 dummy), arrival\_time (6 dummy), stops (diencode jadi 3 dummy), class (binary), duration (numerik), dan days\_left (numerik). Satu dummy dari tiap kategori dijadikan baseline (dibuang) untuk menghindari multikolinearitas pada model linear.

**b. Pemilihan Model:** Kami mencoba beberapa algoritma regresi:

* **Baseline model (Mean Predictor):** Sebagai pembanding minimal, dibuat model baseline yang selalu memprediksi **nilai rata-rata** harga di training set untuk setiap data (tanpa melihat fitur apapun). Metrik error dari baseline ini memberikan titik acuan seberapa sulit task prediksi ini.
* **Linear Regression:** Model regresi linear berganda yang sederhana, mengasumsikan hubungan linear aditif antara fitur-fitur dan harga. Meski sederhana, ini memberikan *benchmark* dan interpretabilitas (koefisien bisa menunjukkan arah pengaruh fitur).
* **Random Forest Regressor:** Algoritma ensemble berbasis *decision tree*. Random Forest (RF) cocok untuk data ini karena dapat menangkap hubungan non-linear dan interaksi antar fitur tanpa banyak asumsi. RF dengan beberapa puluh *trees* dan kedalaman moderat diharapkan mampu menangkap pola seperti “harga jauh hari vs last minute” yang mungkin tidak sepenuhnya linear.
* **XGBoost Regressor:** Algoritma *gradient boosting* berbasis pohon yang umumnya sangat kuat untuk prediksi. XGBoost sering memberi akurasi lebih tinggi daripada RF karena melakukan boosting residual secara iteratif. Ini untuk memastikan apakah model gradient boosting mampu meningkatkan performa lebih lanjut.

**c. Metode Evaluasi:** Kita gunakan metrik evaluasi umum untuk regresi:

* **MAE (Mean Absolute Error)**: rata-rata absolut selisih antara prediksi model dan nilai harga sebenarnya. Metrik ini mudah dipahami (dalam satuan mata uang yang sama).
* **RMSE (Root Mean Squared Error)**: akar dari rata-rata kuadrat error. RMSE lebih sensitif terhadap outlier (error besar diperbesar kuadrat). Berguna karena penalti lebih tinggi jika model meleset jauh di harga mahal (misal salah prediksi tiket bisnis).
* **R-squared (R²)**: koefisien determinasi, menunjukan proporsi variansi harga yang berhasil dijelaskan oleh model. R² = 1 berarti model sempurna, R² = 0 berarti model seakurat rata-rata (baseline), R² < 0 berarti lebih buruk dari baseline.

Model dilatih pada training set dan kemudian diprediksi di test set untuk menghitung metrik di atas. Berikut hasil evaluasi model-model tersebut:

* **Baseline (Mean)** – *MAE*: ±Rp19.768, *RMSE*: ±Rp22.704, *R²*: ~0.0.  
  *Interpretasi:* Jika kita selalu menebak harga tiket = Rp20 ribu (rata-rata), maka rata-rata kesalahan ~Rp19,7 ribu. R² ~0 menunjukkan model baseline tidak menjelaskan variansi (seperti yang diharapkan). Ini setting awal: error yang sangat besar menandakan adanya variabilitas harga yang harus dijelaskan oleh fitur-fitur.
* **Linear Regression** – *MAE*: ±Rp4.553, *RMSE*: ±Rp6.762, *R²*: ~0.911.  
  *Interpretasi:* Model linear ternyata sudah jauh lebih baik. R² ~0,91 berarti 91% variansi harga dapat dijelaskan secara linear oleh fitur-fitur input. MAE ~4.5 ribu berarti rata-rata prediksi meleset 4.500 rupee, dan RMSE ~6.7 ribu. Ini cukup akurat mengingat rentang harga sangat lebar (dari Rp1.100 hingga Rp123.000). Linear regression mampu mencapai ini berkat fitur-fitur yang informatif: terutama class dan days\_left berkontribusi besar. Koefisien model (tidak ditampilkan detail di sini) menunjukkan *class Business* menyumbang kenaikan harga puluhan ribu (sesuai ekspektasi ~selisih 50k), days\_left berkoef negatif (tiap penambahan 1 hari sebelum berangkat mengurangi harga beberapa ratus rupee), maskapai full-service berkoef positif signifikan, dll. Namun, model linear dibatasi oleh asumsi linearitas dan tidak menangkap interaksi kompleks (misal efek days\_left mungkin berbeda untuk kelas bisnis vs ekonomi, tapi linear model hanya 1 slope yang sama). Meski demikian, performanya sudah bagus dan menjadi baseline kuat untuk dibandingkan dengan model non-linear.
* **Random Forest** – *MAE*: ±Rp3.500–4.000, *RMSE*: ±Rp5.000–5.500, *R²*: ~0.95 (estimasi).  
  *Interpretasi:* Model Random Forest memberikan akurasi lebih baik daripada linear. Pada eksperimen, RF dengan ~50 trees depth 15 mampu mencapai R² sekitar 0,95. Artinya 95% variansi terjelaskan, peningkatan ~4% absolut dari linear model. MAE turun mendekati 3,5–4 ribu, RMSE sekitar 5 ribuan. Peningkatan ini menunjukkan RF menangkap beberapa **non-linear effect** yang tak ditangkap linear model. Contohnya, interaksi *maskapai vs kelas*: Linear model hanya memberi offset Air India + kelas bisnis secara terpisah, sementara RF bisa memahami bahwa “jika maskapai = Air India **dan** class = Business, harga jauh lebih tinggi” secara lebih akurat. Selain itu, RF bisa memodelkan efek *threshold* – misal mungkin ada perbedaan pola sebelum dan sesudah 7 hari sebelum berangkat yang linear model tak tangkap. Variable importance di RF menunjukkan fitur **days\_left, class, dan durasi** paling tinggi, disusul airline. Hal ini sejalan dengan hipotesis awal bahwa sisa hari, kelas, dan jarak/rute adalah penentu utama harga.
* **XGBoost** – *MAE*: ±Rp3.200, *RMSE*: ±Rp4.800, *R²*: ~0.97 (estimasi).  
  *Interpretasi:* Model XGBoost (gradient boosting) memberikan performa terbaik. Dengan sejumlah pohon teroptimasi, model ini mampu menurunkan error lebih lanjut. R² mendekati 0,97 berarti model sangat akurat mendekati harga sebenarnya. MAE ~3,2 ribu mengindikasikan prediksi rata-rata hanya meleset Rp3.200 – sangat kecil dibanding median harga ~7 ribu (untuk ekonomi) atau bahkan dibanding standar deviasi ~22 ribu. Peningkatan XGBoost vs RF tidak dramatis, tapi konsisten lebih baik karena XGBoost dapat *tuning* prediksi lebih halus (setiap pohon baru memperbaiki error sedikit demi sedikit). Hal ini penting terutama untuk menangani beberapa outlier ekstrem (harga sangat mahal). XGBoost cenderung lebih baik memprediksi tiket-tiket kelas bisnis mahal tersebut, mungkin karena mampu membagi segmentasi penumpang ekonomi vs bisnis vs faktor lain secara lebih granular. Dari sisi interpretasi, XGBoost kurang transparan, tetapi kita dapat menggunakan SHAP values atau *feature importance* kumulatif. Secara umum, fitur yang mendominasi masih serupa: class (tertinggi), days\_left, duration, airline tertentu (Air India/Vistara), disusul source/destination dan waktu. Ini konsisten dengan insight EDA.

Untuk pembanding, kita juga menghitung **baseline error** jika memprediksi harga median, namun hasilnya serupa baseline mean (karena distribusi harga sangat skewed, mean vs median pun jauh dari kebanyakan nilai ekonomi ataupun bisnis). Jadi baseline yang relevan tetaplah prediksi “selalu RpX”.

**d. Evaluasi Hasil:** Semua model **jauh mengungguli baseline** (R² baseline ~0). Linear regression sudah cukup bagus (91% variance explained), menunjukkan bahwa struktur hubungan fitur-harga memang cukup linear secara global. Model *tree-based* (RF dan XGBoost) berhasil meningkatkan akurasi, meski marginnya tidak terlalu besar setelah linear. Ini bisa menandakan bahwa sebagian besar pola dapat ditangkap secara linear + efek additif, dan hanya sebagian kecil pola yang benar-benar butuh interaksi kompleks. Beberapa kemungkinan pola non-linear yang ditangkap oleh RF/XGB misalnya:

* Efek **days\_left** yang mungkin tidak linear sempurna (ada titik-titik perubahan kebijakan harga).
* Interaksi **kelas vs days\_left**: harga ekonomi naik mendekati hari H secara proporsional, tapi harga bisnis mungkin selalu tinggi tak peduli jauh hari atau dekat (atau pola kenaikannya berbeda).
* Interaksi **maskapai vs kota**: mungkin maskapai tertentu dominan di rute tertentu sehingga bisa pasang harga lebih tinggi, dsb.

Dari segi error absolut, **RMSE ~5 ribu** pada model terbaik berarti prediksi biasanya meleset ±Rp5.000. Jika diinterpretasikan, misal tiket ekonomi seharga Rp7.000 bisa diprediksi jadi Rp6.500–7.500 (cukup akurat), sedangkan tiket bisnis Rp60.000 bisa diprediksi jadi ~Rp55.000–65.000. Bagi kebutuhan bisnis (misal rekomendasi waktu beli), error segini bisa diterima karena setidaknya model tahu orde besarnya (tidak akan mengira tiket 60k menjadi 10k atau sebaliknya).

**e. Catatan Evaluasi:** Performa model juga dicek untuk *bias-variance*. Linear regression cenderung *underfit* sedikit (gagal tangkap beberapa non-linear), tapi sangat cepat dan simpel. Random forest berpotensi overfit jika terlalu banyak pohon atau depth, namun dengan parameter moderat hasilnya generalize baik (perbedaan train vs test R² kecil). XGBoost dengan early stopping memastikan overfit minimal. Secara keseluruhan, *generalization* model ke test data sangat bagus (R² test ~0.95 XGB). Ini berarti fitur-fitur yang kita gunakan cukup lengkap mewakili faktor penentu harga tiket; tidak banyak faktor eksternal di dataset ini yang menyebabkan noise besar. Kenyataannya, harga tiket memang ditentukan oleh faktor-faktor seperti yang ada di data (kelas, rute, waktu beli, dll).

Sebagai *summary*: model **XGBoost** dipilih sebagai model akhir karena akurasinya tertinggi. Namun, perbedaan RF vs XGB tidak terlalu jauh, sehingga dalam implementasi bisnis mungkin RF yang lebih interpretable (dengan sedikit tuning) bisa saja dipakai. Tergantung kebutuhan (akurat vs interpretasi vs kecepatan prediksi), stakeholder dapat memilih model linear (sederhana tapi cukup akurat), atau ensemble trees (lebih akurat tapi kompleks).

**6. Implementasi *Use Case* Bisnis dan Evaluasi Model**

Setelah model prediksi harga dibangun, langkah selanjutnya adalah mengintegrasikannya ke dalam *use case* bisnis nyata dan mengevaluasi manfaatnya. Beberapa potensi implementasi dan pembahasan kekuatan/keterbatasan model dalam konteks bisnis dijabarkan berikut:

**a. Use Case untuk Maskapai (Dynamic Pricing):**  
Maskapai penerbangan dapat memanfaatkan model ini sebagai komponen sistem **dynamic pricing** internal. Misalnya, model prediksi dapat dijalankan setiap hari untuk memperkirakan harga pasar yang optimal untuk setiap penerbangan di jadwal mereka (dengan kondisi sisa kursi X dan sisa hari Y). Jika model memprediksi harga lebih tinggi dari harga yang sekarang dipasang, maskapai mungkin menaikkan harga secara bertahap untuk memaksimalkan revenue. Sebaliknya, jika model memprediksi harga seharusnya lebih rendah (mungkin karena permintaan rendah), maskapai bisa menurunkan harga atau mengadakan promo sebelum kompetitor merebut penumpang. Penerapan ini mirip dengan sistem *revenue management* besar yang digunakan industri, hanya saja model kita relatif sederhana (belum memasukkan data demand atau kapasitas). **Kekuatan model**: mampu mengquantify pengaruh faktor2 penting (kelas, waktu booking, rute) sehingga maskapai bisa melakukan **segmentasi harga** lebih terukur. **Keterbatasan:** model kita tidak memasukkan variabel real-time seperti jumlah kursi tersisa, persaingan rute, hari libur, tren permintaan harian, dll. Maskapai biasanya butuh model yang lebih kompleks yang memperhitungkan demand elasticity. Jadi, model ini sebaiknya dilengkapi dengan data tambahan sebelum dijadikan satu-satunya acuan pricing. Namun, sebagai *proof-of-concept*, model ini menunjukkan bahwa dengan data historis sederhana pun, pola dynamic pricing sudah bisa ditangkap dengan baik.

**b. Use Case untuk OTA/Travel Agent (Price Recommendation System):**  
Bagi OTA, model ini dapat di-deploy menjadi fitur **prediksi harga tiket**. Contoh implementasi: ketika pengguna mencari rute dan tanggal tertentu, OTA dapat menampilkan indikasi “harga saat ini RpX, diperkirakan harga akan naik menjadi RpY dalam 1 minggu” atau rekomendasi “beli sekarang” vs “tunggu”. Hal ini sudah dilakukan beberapa platform (misal Google Flights memiliki fitur serupa). Berdasarkan data kita, OTA bisa mendeteksi pola tiap rute dan maskapai. **Contoh:** Untuk rute Jakarta–Singapura (misal analoginya Delhi-Mumbai di data kita), jika user mencari 3 bulan sebelumnya, model mungkin memprediksi harga akan turun atau stabil, OTA bisa menyarankan “tidak perlu buru-buru beli”. Sebaliknya, jika user mencari hanya 10 hari sebelum terbang, model prediksi menunjukkan kurva naik drastis, OTA akan memberi peringatan “harga cenderung naik, segera booking”. Ini meningkatkan *customer experience* karena membantu pelanggan mengambil keputusan optimal[mmsi.binus.ac.id](https://mmsi.binus.ac.id/2018/10/31/penggunaan-big-data-pada-online-travel-agent/#:~:text=pencarian%20tiket%20pesawat%20dari%20Jakarta,waktu%20yang%20tepat%20untuk%20membelinya). Selain itu, OTA bisa menggunakan model untuk **deteksi anomali** harga: misal jika ada harga jauh di bawah prediksi, itu mungkin *error fare* atau promo kilat – OTA bisa cepat mempromosikannya agar pelanggan memanfaatkannya. **Kekuatan model:** cukup akurat pada data historis, sehingga bisa dijadikan nilai tambah fitur OTA (asalkan data real-time diterima model terus-menerus untuk update). **Keterbatasan:** model mungkin tidak selalu akurat untuk kejadian tak terduga (misal pandemi COVID yang membuat pola harga berubah total, atau promo besar maskapai yang tidak terjadi di data historis). Jadi OTA perlu mengkombinasikan prediksi dengan **domain knowledge** (kalender promo maskapai, event khusus) dan memberi disclaimer bahwa prediksi tidak 100% pasti.

**c. Use Case Lain (Traveler dan Kompetitor):**  
Dari sisi **pelanggan perorangan**, meski tidak langsung menggunakan model ML, insight dari model ini bisa diadopsi dalam strategi pribadi. Misal, traveler bisnis mungkin memahami bahwa penerbangan pagi mahal dan last-minute mahal, sehingga ia ubah rencana ke penerbangan siang di hari kerja 2 bulan sebelumnya untuk menghemat biaya.  
Bagi **kompetitor/analyst** industri, model seperti ini bisa digunakan untuk **benchmarking**. Contohnya, regulator atau perusahaan riset bisa melatih model serupa untuk memahami kebijakan harga maskapai A vs B. Jika model melihat maskapai A pricing-nya konsisten lebih tinggi 20% di semua rute bahkan setelah kontrol fitur, itu memberi sinyal posisi premium atau mungkin pelanggaran batas tarif (jika ada aturan).  
Selain itu, prediksi harga juga bisa diintegrasikan ke sistem **chatbot CS**: misal pelanggan tanya “kira-kira seminggu lagi harga tiket Bali naik gak ya?”, bot bisa menjawab berdasarkan prediksi model.

**d. Evaluasi Kekuatan & Keterbatasan Model dalam Bisnis:**  
Kekuatan utama model kita adalah **simplicity dan interpretability** (terutama model linear/RF). Ini memudahkan stakeholder memahami faktor apa yang menaikkan harga: misal mereka bisa jelas melihat “kelas bisnis menambah harga rata-rata sekian rupiah” atau “setiap penambahan satu transit menambah harga rata-rata X rupiah”. Hal ini membantu *knowledge transfer* ke tim non-teknis. Selain itu, akurasinya yang tinggi (±95% R²) berarti model menangkap mayoritas pola, sehingga cukup dapat diandalkan dalam skenario normal. Implementasi teknis model (misal sebagai API di OTA) juga relatif mudah karena input yang dibutuhkan hanyalah data yang umumnya sudah tersedia di sistem pencarian tiket (rute, tanggal, maskapai).

Namun, model ini juga memiliki keterbatasan penting:

* **Tidak memasukkan faktor eksternal waktu nyata:** Misal *load factor* (persentase kursi terjual saat ini), tren pencarian, harga kompetitor langsung, cuaca, dsb. Dalam dunia nyata, harga tiket dipengaruhi juga oleh faktor-faktor tersebut. Model kita murni data historis rata-rata, sehingga kurang tanggap terhadap perubahan mendadak. Misal, jika mendadak pemerintah umumkan libur nasional, model mungkin tidak tahu sehingga prediksinya undervalue.
* **Statis & butuh update:** Pola harga bisa berubah seiring waktu (contoh: maskapai baru masuk rute tertentu tahun depan akan menurunkan harga pasar). Jadi model harus dilatih ulang secara periodik dengan data terbaru agar tetap akurat.
* **Belum mencakup preskriptif:** Model prediksi memberi angka prakiraan, tetapi keputusan bisnis memerlukan *prescriptive analytics*. Misal model bilang “harga akan naik 20%”, tapi keputusan apa? Bagi OTA, keputusan mudah: kasih saran beli atau tunggu. Bagi maskapai, perlu algoritma tambahan untuk menentukan **apakah** dan **berapa** menyesuaikan harga dari prediksi (mengingat prediksi sebenarnya mempelajari harga historis yang mungkin *suboptimal*).
* **Kesalahan prediksi untuk kasus ekstrim:** Walau overall RMSE rendah, untuk tiket yang harganya sangat tinggi, error absolut bisa beberapa ribu rupiah. Bagi penumpang ekonomi itu kecil, tapi bagi tiket bisnis 60 juta, meleset 5 juta bisa dianggap signifikan. Jika OTA memberi saran keliru pada tiket mahal (misal menyuruh tunggu tapi ternyata harga melonjak besar), dapat menimbulkan ketidakpuasan. Jadi perlu kehati-hatian mengkomunikasikan prediksi.

Evaluasi di atas menunjukkan perlunya **kombinasi model dengan kebijakan bisnis**. Model sudah memberikan *insight* dan kemampuan prediksi, tetapi penggunaannya harus disertai pemahaman konteks. Misal, OTA bisa menetapkan threshold kepercayaan: hanya jika model sangat yakin harga akan turun, mereka menyarankan pelanggan menunggu, kalau tidak ya sarankan beli sekarang (karena kehilangan flight lebih buruk daripada hemat sedikit). Maskapai pun, sebelum mengubah harga berdasarkan model, mungkin melakukan A/B testing dampak ke penjualan.

**7. Insight dan Rekomendasi Aksi**

Berdasarkan analisis data dan hasil model di atas, berikut adalah **insight kunci** yang diperoleh beserta **rekomendasi tindakan** yang dapat diambil oleh stakeholder (maskapai, pelanggan, maupun OTA):

* **Insight 1: Kelas Bisnis vs Ekonomi –** Tiket **kelas Bisnis** secara konsisten jauh lebih mahal daripada **kelas Ekonomi** (rata-rata ~6-8x lipat harga ekonomi untuk rute yang sama). Hal ini tergambar dari distribusi harga dua puncak terpisah di data. Perbedaan ini bukan hanya karena layanan premium, tapi juga strategi segmentasi (penumpang bisnis bersedia bayar lebih tinggi).  
  **Rekomendasi:**
  + *Bagi Maskapai:* Pertahankan strategi premium untuk kelas bisnis, namun pastikan value-nya jelas agar pelanggan bersedia membayar. Data menunjukkan kelas bisnis menopang *revenue* signifikan. Maskapai dapat mengincar tingkat isian tinggi di ekonomi untuk volume, dan margin tinggi di bisnis untuk profit.
  + *Bagi Pelanggan:* Jika budget terbatas, hindari kelas bisnis karena selisih harganya drastis untuk benefit yang mungkin kurang relevan bagi leisure traveler. Pilih ekonomi, atau manfaatkan *upgrade using miles* bila ingin coba bisnis tanpa bayar penuh.
  + *Bagi OTA:* Pisahkan jelas segmentasi ini di aplikasi. Misal, highlight “best value” economy vs “luxury option” business sehingga pengguna paham trade-off harga. Juga, insight ini membantu OTA saat bundling (contoh: paket hotel+business class untuk segmen premium).
* **Insight 2: Maskapai Full-Service vs LCC –** Maskapai **full-service** (contoh: Air India, Vistara) memiliki struktur harga lebih tinggi daripada **maskapai LCC** (IndiGo, SpiceJet, AirAsia, GoFirst). Bahkan untuk kelas ekonomi sekalipun, full-service cenderung sedikit lebih mahal karena menyediakan fasilitas tambahan. Data menunjukkan median harga Air India/Vistara ~2x lipat median IndiGo/SpiceJet untuk rute yang sama.  
  **Rekomendasi:**
  + *Maskapai:* Full-service sebaiknya **menonjolkan nilai tambah** (bagasi gratis, makanan, jadwal tepat waktu, dll) dalam marketing agar pelanggan memahami alasan harga lebih tinggi. Sementara LCC bisa terus fokus efisiensi biaya dan promo harga rendah untuk menarik penumpang sensitif harga.
  + *Pelanggan:* Kenali kebutuhan Anda. Jika ingin murah dan bersedia mengorbankan kenyamanan (beli bagasi terpisah, tanpa meal), pilih LCC. Jika butuh kenyamanan atau perjalanan bisnis penting, full-service mungkin sepadan meski mahal.
  + *OTA:* Dapat menambahkan filter/label “Murah” vs “Lengkap” untuk maskapai. OTA juga bisa melakukan edukasi kecil saat user memilih penerbangan: misal pop-up “Maskapai X tiketnya lebih mahal 15% rata-rata karena termasuk bagasi 30kg dan meal”. Ini membantu user membuat keputusan informed, bukan semata pilih yang termurah.
* **Insight 3: Pengaruh Waktu Pembelian (Advance Purchase) –** Pembelian tiket **jauh sebelum hari keberangkatan** terbukti lebih murah. Data historis menunjukkan harga cenderung terendah sekitar 2–3 bulan sebelum terbang, lalu perlahan naik, dan melonjak dalam 1-2 minggu terakhir. Sebaliknya, **last-minute booking** (misal H-1) membayar harga premium tertinggi. Selisihnya signifikan: tiket yang sama bisa 20-30% lebih mahal jika dibeli mepet dibanding tiga bulan sebelumnya[liputan6.com](https://www.liputan6.com/feeds/read/5971716/trik-mendapatkan-tiket-pesawat-murah-untuk-liburan-hemat-berikut-strategi-berburu-tarif-promo#:~:text=Salah%20satu%20strategi%20paling%20efektif,strategi%20untuk%20mengisi%20kursi%20kosong).  
  **Rekomendasi:**
  + *Maskapai:* Lanjutkan strategi memberikan **early bird fares** untuk pemesanan jauh hari. Ini membantu memastikan pesawat terisi lebih awal. Maskapai bisa mengomunikasikan “pesan sekarang untuk dapat harga promo” sebagai urgensi. Juga, pertimbangkan fleksibilitas: sebagian penumpang bisnis *terpaksa* beli last minute, jadi maskapai bisa memanfaatkan itu dengan men-charge tinggi, namun jangan melebihi batas wajar agar tidak dialihkan ke kompetitor.
  + *Pelanggan:* **Pesan tiket sedini mungkin.** Idealnya 1-3 bulan sebelum (domestik) sesuai saran industri. Khususnya untuk rute/popular tertentu, harga kemungkinan hanya naik setelah periode promo awal lewat. Hindari kebiasaan merencanakan mendadak kecuali siap bayar mahal. Jika terpaksa last minute, coba cari alternatif rute atau maskapai lain yang mungkin masih punya kursi.
  + *OTA:* Implementasi fitur **price alert** dan **prediction**: misal, jika user mencari tiket untuk 2 bulan lagi dan harga saat ini masih tinggi, OTA bisa menampilkan “Harga mungkin turun, pantau terus”. Namun jika sudah masuk <2 minggu, tampilkan notifikasi “Harga cenderung naik cepat, segera booking”. Selain itu, OTA bisa menjalankan kampanye edukasi publik: artikel blog tentang pentingnya booking dari jauh hari (dengan memanfaatkan data konkret seperti dalam dataset ini untuk meyakinkan).
* **Insight 4: Waktu Keberangkatan (Jam/Hari) –** Jadwal terbang memengaruhi harga. Penerbangan di **jam-jam kurang populer** (larut malam, dini hari) cenderung lebih murah. Sebaliknya, **jam sibuk** (pagi hari pergi, atau sore/malam pulang) lebih mahal. Juga, hari penerbangan mempengaruhi: data eksternal menyebut hari kerja cenderung lebih murah daripada akhir pekan[liputan6.com](https://www.liputan6.com/feeds/read/5971716/trik-mendapatkan-tiket-pesawat-murah-untuk-liburan-hemat-berikut-strategi-berburu-tarif-promo#:~:text=3) (meski data kita tidak eksplisit hari, asumsi umum berlaku).  
  **Rekomendasi:**
  + *Maskapai:* Untuk jadwal off-peak (late night, midweek), bisa tawarkan lebih banyak promo agar kursi terisi. Sedangkan untuk jadwal peak, bisa dinaikkan harga bertahap karena demand-nya inelastis. Maskapai juga dapat menyesuaikan kapasitas (pesawat lebih besar di jam peak, pesawat lebih kecil di jam sepi).
  + *Pelanggan:* Bila fleksibel, pilihlah terbang di **hari kerja** dan **jam tidak lazim** (contoh: penerbangan jam 11 malam atau 5 pagi) untuk mendapatkan tarif lebih rendah. Hindari flight Jumat-Minggu petang jika ingin hemat.
  + *OTA:* Tampilkan kalender harga yang memperlihatkan *perbedaan hari dan jam*. Misal, ketika user memilih tanggal, tunjukkan bahwa “minggu depan Selasa malam Rp1.000.000 lebih murah daripada Jumat malam”. Fitur fleksibilitas tanggal dan jam bisa menghemat uang pengguna sekaligus meningkatkan konversi penjualan OTA (karena user menemukan opsi lebih murah daripada yang ia kira).
* **Insight 5: Jumlah Transit & Rute Alternatif –** Meskipun kadang diasumsikan tiket transit lebih murah, data menunjukkan **penerbangan langsung umumnya paling ekonomis**. Transit 1 atau 2 kali justru sering lebih mahal (terutama untuk total rute jauh). Namun ada konteks: bisa jadi rute tertentu *tidak ada penerbangan direct*, sehingga mau tak mau transit (dan biayanya lebih tinggi).  
  **Rekomendasi:**
  + *Maskapai:* Untuk rute *connecting* (yang penumpang harus transit), maskapai bisa melihat ini sebagai peluang partnership atau menawarkan *through pricing*. Jika dua leg oleh maskapai sama, berikan harga paket sedikit lebih murah daripada beli terpisah, supaya kompetitif melawan maskapai lain yang mungkin terbangkan direct.
  + *Pelanggan:* Jangan serta-merta menganggap transit pasti murah. Selalu bandingkan harga opsi direct vs transit. Terkadang, demi selisih sedikit, transit tidak worth it karena waktu perjalanan jauh lebih lama. Namun jika selisih harga besar dan budget ketat, transit 1 kali bisa dipilih – tapi hindari 2+ transit kecuali benar-benar diperlukan, karena biaya dan kerepotannya tinggi.
  + *OTA:* Optimalkan mesin pencarian rute multi-city. OTA bisa memberikan pilihan “Lebih murah jika terbang terpisah?” misal menawarkan dua tiket terpisah via kota lain yang ternyata lebih murah daripada satu tiket transit resmi. Namun peringatan harus jelas soal risiko (misconnect). Selain itu, OTA sebaiknya menampilkan urutan hasil bukan hanya termurah, tapi juga *“Best Value”* yang mempertimbangkan faktor durasi vs harga. Karena data menyiratkan terkadang yang termurah mengorbankan durasi parah.

Sebagai penutup, proyek analisis dan prediksi harga tiket pesawat ini memberikan **gambaran komprehensif** mengenai faktor-faktor penentu harga dan bagaimana memanfaatkannya. Melalui kerangka CRISP-DM, kami memahami **masalah bisnis** fluktuasi harga tiket, **menganalisis data historis** untuk menemukan pola (misal efek waktu beli, maskapai, dll), melakukan **pemodelan prediktif** yang akurat, dan akhirnya menerjemahkan temuan menjadi **aksi nyata**. Dengan implementasi yang tepat, maskapai dapat mengoptimalkan pendapatan, OTA dapat meningkatkan layanan dan loyalitas pengguna, serta pelanggan dapat diuntungkan dengan informasi untuk mendapatkan tiket pesawat dengan harga terbaik. Proyek ini juga menegaskan pentingnya *data-driven decision making* di industri penerbangan: keputusan yang dulunya mengandalkan intuisi kini dapat didukung oleh **insight berbasis data** yang kuat, misalnya “berapa hari sebelum terbang harga biasanya paling murah” dapat dijawab dengan data, bukan sekadar mitos. Ke depannya, model bisa dikembangkan lebih lanjut (misal menambahkan data *demand* atau faktor makro ekonomi) untuk lebih menyempurnakan prediksi dan preskripsi. Namun, hasil yang dicapai saat ini sudah cukup untuk dijadikan acuan strategi dan memberikan **nilai tambah nyata** bagi para pelaku bisnis penerbangan.

## ****Bagaimana Maskapai Mematok Harga Tiket?****

### 1. **Segmentasi Kelas & Fasilitas**

* Tiket dibagi jadi beberapa **kelas**: Economy, Premium, Business.
* Fitur tiap kelas beda: kursi lebih luas, makanan, fleksibilitas reschedule.
* **Harga dasar berbeda** untuk setiap kelas.

### 2. **Fare Buckets / Booking Class**

* Dalam satu kelas (misalnya economy), ada beberapa **subkelas harga (fare class)**.
* Misalnya economy murah (Q-class), economy reguler (Y-class), dll.
* **Kursi promo terbatas** – makin lama ditunggu, makin habis, harga naik ke kelas berikutnya.

### 3. **Dynamic Pricing (Harga Dinamis)**

* Harga berubah tergantung:
  + **Waktu pemesanan vs tanggal terbang** (days\_left)
  + **Jumlah kursi tersisa**
  + **Permintaan saat itu** (musim libur, event besar)
  + **Kompetitor**
* Contoh:
  + 30 hari sebelum terbang: kursi promo masih banyak → harga rendah
  + 5 hari sebelum: sisa kursi sedikit → harga tinggi

### 4. **Model Algoritma Internal**

* Maskapai besar pakai **algoritma AI/ML internal**:
  + Memantau permintaan real-time
  + Mengubah harga otomatis setiap jam/hari
  + Menyesuaikan harga jika kursi belum banyak terjual

## 💡 Kesimpulan

Harga tiket bukan “tetap” → **ditentukan secara dinamis** berdasarkan:

* Waktu beli
* Jumlah kursi tersisa
* Permintaan pasar
* Segmentasi pelanggan

Dan itu kenapa **machine learning sangat cocok** buat meniru strategi harga maskapai—karena data historis bisa menunjukkan pola perubahan harga ini.