BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Prediksi Harga Tiket Pesawat Domestik Rute Perjalanan Surabaya-Jakarta Menggunakan Metode Regresi Linear Berganda

Enggi Sabrilla Assara, Hamzah Setiawan, Suprianto

Fakultas Sains dan Teknologi, Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Sidoarjo, Indonesia
Email: [1]enggisabrilla08@gmail.com, [2]hamzah@umsida.ac.id, [3]suprianto@umsida.ac.id
Email Penulis Korespondensi : , hamzah@umsida.ac.id

Abstrak- Transportasi udara adalah alat transportasi yang sangat diminati karena efisiensi waktu dan kenyamanan, khususnya pada rute padat seperti Surabaya–Jakarta. Namun, fluktuasi harga tiket pesawat yang dinamis seringkali menyulitkan konsumen dalam merencanakan perjalanan. Tujuan penelitian ini untuk membangun model prediksi harga tiket pesawat rute Surabaya–Jakarta menggunakan metode regresi linear berganda. Data sebanyak 10.000 baris dianalisis dengan pendekatan statistik dan proses analitik berbasis Google Colaboratory, melalui tahapan impor data, preprocessing, transformasi variabel, pemisahan data latih dan uji, serta pengujian asumsi klasik. Model regresi yang dikembangkan dari penelitian ini memiliki performa sangat baik dengan nilai R- squared sebesar 96,4%, yang menunjukkan bahwa sebagian besar variasi harga tiket dapat dijelaskan oleh variabel- variabel independen seperti maskapai, waktu keberangkatan, lama perjalanan, kapasitas bagasi, dan jenis layanan. Beberapa pelanggaran asumsi seperti normalitas dan heteroskedastisitas berhasil diatasi melalui transformasi log dan penggunaan regresi dengan robust standard error. Selain itu, multikolinearitas berhasil diminimalkan dengan metode Ridge Regression. Evaluasi model menunjukkan tidak terjadi overfitting dan memberikan hasil prediksi yang stabil. Hanya beberapa variabel terbukti signifikan secara statistik, sehingga analisis kontribusi variabel menjadi penting untuk efisiensi model. Model prediksi harga tiket penilitian ini menghasilkan hasil yang akurat dan aplikatif, serta dapat dimanfaatkan oleh konsumen untuk perencanaan perjalanan maupun oleh maskapai untuk strategi penetapan harga yang lebih kompetitif.

Kata Kunci: Analisis statistik; Google Colab; Prediksi Harga Tiket; Regresi Linear Berganda; Surabaya–Jakarta

Abstract- Air transportation is highly favored for its time efficiency and comfort, especially on busy routes such as Surabaya–Jakarta. However, the dynamic fluctuation of airline ticket prices often makes it difficult for consumers to plan their trips. This study aims to develop a predictive model for airline ticket prices on the Surabaya–Jakarta route using the multiple linear regression method. A total of 10,000 rows of data were analyzed using statistical approaches and analytical processes based on Google Colaboratory, involving stages such as data import, preprocessing, variable transformation, data splitting (training and testing), and classical assumption testing. The resulting regression model demonstrated excellent performance with an R-squared value of 96.4%, indicating that most of the price variation could be explained by independent variables such as airline, departure time, travel duration, baggage capacity, and service type. Violations of assumptions such as normality and heteroskedasticity were addressed through logarithmic transformation and the use of regression with robust standard errors. Furthermore, multicollinearity was minimized using Ridge Regression. Model evaluation showed no signs of overfitting and produced stable prediction results. Only a few variables were statistically significant, highlighting the importance of analyzing variable contributions to enhance model efficiency. The predictive model developed in this study provides accurate and practical results, making it useful for consumers in travel planning and for airlines in developing more competitive pricing strategies.

Keywords: Statistical Analysis; Google Colab; Ticket Price Prediction; Multiple Linear Regression; Surabaya–Jakarta

1. PENDAHULUAN

Industri penerbangan mengalami pengembangan yang cepat selama periode globalisasi ini. Ini ditandai dengan peningkatan jumlah bandara baru. Dalam konteks ini, peran sumber daya manusia tidak dapat dipisahkan dari operasional bandara. Aset yang berharga yang dimiliki suatu perusahaan adalah pembelinya. Jumlah perusahaan dan maskapai yang menyediakan layanan penerbangan di beberapa perjalanan penerbangan domestik saat ini telah membuktikan perkembangan yang begitu cepat dalam industri penerbangan. Kehadiran infrastruktur dan fasilitas yang mendukung operasi penerbangan adalah salah satu faktor yang mendukung kesuksesan operasi bidang penerbangan. PT. Gapura Angkasa adalah untuk memiliki enam puluh enam kantor industri di Indonesia. PT. Gapura Angkasa sendiri berada di Bandara Juanda. Bandara ini sangat strategis ditandai dengan adanya jumlah penumpang yang sangat meningkat secara drastic disetiap tahunnya. Ini benar -benar memenuhi kebutuhan PT. Gerbang Ankasa adalah bandara terbesar ketiga di Indonesia, Bandara Internasional Juanda Surabaya[1].

Transportasi udara banyak dipilih daripada transportasi lainnya karena menghemat waktu perjalanan dan kenyamanan baik transportasi dan pelayananny. Dan umumnya digunakan oleh orang -orang dengan pendapatan tinggi. Namun, hadirnya berbagai maskapai penerbangan yang menawarkan tarif murah membuat transportasi udara tidak lagi terbatas bagi kalangan berpenghasilan tinggi, melainkan juga dapat diakses oleh masyarakat umum. Persaingan yang semakin ketat antar maskapai di Indonesia mendorong perusahaan penerbangan untuk lebih menekankan efisiensi dan efektivitas agar tetap kompetitif dan mampu bertahan di tengah persaingan tersebut[2].

Salah satu rute domestik dengan volume penumpang tertinggi di Indonesia adalah rute Surabaya–Jakarta, yang dilayani oleh berbagai maskapai dengan jadwal penerbangan yang padat setiap harinya. Rute ini tergolong strategis dan memiliki tingkat permintaan tinggi karena menghubungkan dua pusat kegiatan ekonomi, pemerintahan, dan perdagangan. Peningkatan jumlah penumpang menyebabkan perbedaan dalam kaitannya dengan harga kartu pesawat domestik yang lebih tinggi. Ini adalah salah satu faktor terpenting dalam keputusan pembelian[3].

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Harga tiket pesawat pada rute Surabaya–Jakarta dikenal sangat fluktuatif. Perubahan harga dapat terjadi beberapa kali dalam sehari, bahkan untuk maskapai dan kelas layanan yang sama. Fenomena ini sering kali membingungkan konsumen, terutama menjelang akhir pekan atau saat musim libur nasional, ketika permintaan meningkat secara signifikan. Selain faktor waktu, fluktuasi harga tiket juga dipengaruhi oleh berbagai variabel lain, seperti jenis maskapai, waktu keberangkatan dan kedatangan, durasi perjalanan, kapasitas bagasi yang disediakan, serta jenis layanan penerbangan.

Kondisi tersebut menunjukkan perlunya pengembangan model prediksi harga tiket pesawat yang mampu membantu berbagai pihak, baik konsumen maupun penyedia jasa penerbangan, dalam mengambil keputusan yang lebih tepat dan efisien. Regresi linier berganda merukapakan model statistik yang cocok untuk mengevaluasi hubungan antara harga tiket sebagai variabel dependen dengan berbagai faktor yang bertindak sebagai variabel independen. Metode ini memungkinkan analisis hubungan secara bersamaan antara beberapa variabel, serta berfungsi dalam membangun model prediksi yang akurat dan valid secara ilmiah.

Studi oleh Setiawan et al. (2024) menunjukkan bahwa pemilihan fitur yang optimal memiliki pengaruh signifikan terhadap peningkatan akurasi model prediktif. Meskipun konteks penelitiannya berada pada klasifikasi multilabel terhadap data umpan balik mahasiswa, pendekatan pemilihan fitur berbasis filter dan metaheuristik yang digunakan berhasil meningkatkan performa model secara substansial. Hal ini menunjukkan bahwa proses pemilihan variabel yang relevan dan signifikan juga sangat penting dalam konteks regresi linier berganda, untuk memastikan akurasi dan keandalan model prediksi yang dibangun[4].

Tujuan dari penelitian ini adalah memprediksi harga kartu pesawat domestik untuk rute Surabaya Jakarta menggunakan beberapa metode regresi linier. Metode ini dipilih karena keakuratannya untuk menganalisis hubungan antara variable. Hasil analisis diharapkan mengidentifikasi faktor utama yang memengaruhi harga serta menggambarkan pola harga pada rute dengan tingkat permintaan tinggi ini. Prediksi harga tiket penting bagi penumpang untuk merencanakan perjalanan lebih efisien dan memahami pola fluktuasi harga. Sementara itu, maskapai dapat menggunakan hasil penelitian ini untuk menyusun strategi harga yang kompetitif, mengoptimalkan pendapatan, dan meningkatkan kepuasan pelanggan.

Model prediksi yang dibangun akan dievaluasi menggunakan metriks, seperti R-Squared (R^2), dan Root Relative Squared Error (RRSE). Evaluasi ini dilakukan untuk memastikan bahwa model yang dikembangkan memiliki akurasi, keandalan, dan kelayakan untuk diterapkan dalam situasi dunia nyata. Diharapkan, hasil penelitian ini dapat membantu penumpang merencanakan perjalanan dengan lebih efisien dan memberikan pemahaman mengenai fluktuasi harga tiket. Di sisi lain, maskapai penerbangan dapat memanfaatkan hasil penelitian ini untuk menyusun strategi harga yang lebih kompetitif, meningkatkan pendapatan, serta meningkatkan kepuasan pelanggan.

Analisis data dilakukan menggunakan Google Colaboratory (Google Colab) dengan langkah-langkah dimulai dari impor library, load data, dan preprocessing, termasuk mapping manual variabel kategorik seperti maskapai, kelas, dan hari terbang. Data dibagi menjadi data latih dan uji, kemudian dianalisis menggunakan regresi linear berganda dengan menggunakan statsmodels. Setelah itu, dilakukan pengujian asumsi klasik. Jika ditemukan pelanggaran, dilakukan penyesuaian seperti transformasi log. Proses analisis yang terakhir yaitu dengan meinterpretasikan akurasi regresi untuk mengetahui keterkaitan masing-masing variabel terhadap harga tiket.

Penelitian ini bertujuan agar dapat memberikan ide perkembangan ilmu pengetahuan di bidang ekonomi transportasi, data science, dan analisis statistik. Dalam konteks industri penerbangan, hasil dari penelitian ini juga memiliki nilai praktis dalam mendukung efisiensi operasional serta transparansi dalam penentuan harga tiket pesawat. Dengan demikian, diharapkan dapat dihasilkan model prediksi harga tiket pesawat yang akurat, andal, dan aplikatif untuk kebutuhan berbagai pihak terkait, baik di sektor transportasi maupun dalam pengambilan keputusan strategis di industri penerbangan.

2. METODOLOGI PENELITIAN

Metodologi ini menjelaskan tahap-tahap yang digunakan untuk memprediksi harga tiket pesawat domestik dengan metode Regresi Linear Berganda. Terdapat enam langkah yang perlu dilakukan dalam proses ini. Berikut adalah langkah-langkah yang akan diterapkan dalam metodologi yang digunakan :

Gambar 1. Tahapan Penelitian

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

2.1 Data dan Sumber Data

Data yang akan dianalisis adalah data sekunder sebanyak 5000 baris, data ini didapat dari simulasi transaksi pembelian tiket pesawat pada rute Surabaya - Jakarta. Dataset berisi berbagai atribut seperti maskapai, kelas penerbangan, jumlah transit, jumlah bagasi, tipe pengguna, metode pembayaran, jumlah penumpang, hari terbang, sisa kursi, dan selisih hari antara tanggal pembelian dan tanggal terbang.

2.2 Load Data

Data telah diunggah ke Google Sheet untuk memudahkan proses running program dan memperoleh tautan CSV-nya. Link CSV tersebut kemudian digunakan di Google Colab untuk memuat data secara langsung menggunakan pustaka pandas. Proses pemuatan data dilakukan dengan fungsi pd.read_csv() yang membaca tautan CSV dan mengubahnya menjadi dataframe sehingga siap untuk dianalisis lebih lanjut.

2.3 Pra-Pemrosesan

Data preparation, juga dikenal sebagai data preprocessing, adalah tahap yang bertujuan untuk mempersiapkan data yang telah dikenali sebelumnya untuk dianalisis dengan menggunakan teknik penambangan data. Meskipun membutuhkan banyak waktu dan usaha, data preparation sering menghabiskan sekitar 80% dari waktu yang dibutuhkan untuk proyek penambangan data. Hal ini dikarenakan data dunia nyata yang biasanya memiliki risiko seperti kesalahan atau outlier, nilai atribut yang hilang, atau hanya berisi data agregat yang tidak konsisten dalam format atau nama[5].

a. Mapping Variabel Kategori ke Numerik
   Dalam penelitian ini, dilakukan proses pra-pemrosesan data untuk mengubah variabel kategorikal menjadi bentuk numerik agar dapat digunakan dalam analisis regresi. Proses ini dilakukan secara manual dengan Google Colab. Variabel kategorikal yang dimaksud meliputi Maskapai, Kelas, dan Hari_Terbang.
b. Transdormasi Variabel Harga menjadi LogHarga
   Dalam penelitian ini, variabel Harga awalnya berbentuk data numerik bertipe rasio dengan nilai yang bervariasi secara signifikan antar observasi. Untuk mengatasi masalah distribusi data yang tidak normal dan heteroskedastisitas yang mungkin terjadi pada data harga tiket pesawat, dilakukan transformasi logaritma natural terhadap variabel Harga, sehingga diperoleh variabel baru yaitu LogHarga.
   Transformasi dilakukan untuk menstabilkan varians saat menjalankan prosedur regresi linear. Selain menangani variabel yang tidak stabil, transformasi berguna untuk memperbaiki non linearitas serta residual yang tidak terdistribusi normal[6].
   Transformasi logaritma bertujuan untuk menstabilkan varians, memperbaiki distribusi data agar mendekati normal, dan membuat hubungan antara variabel menjadi lebih linear. Rumus yang digunakan

$$LogHarga = \ln(Harga) \qquad (1)$$

   Di mana:
   Harga adalah nilai harga tiket asli dalam satuan rupiah.
   ln adalah logaritma natural (basis eee).
c. Split Data
   Proses Split data menggunakan dala latih dan dana uji. Perbandingan split data adalah 70:30 yaitu data latih dan data uji. Pembagian dataset menggunakan operator split data, sementara pengujian dilakukan dengan operator apply data[7].
d. Standarisasi
   Standardisasi atau feature scaling dilakukan untuk menormalisasi nilai-nilai fitur dalam dataset sehingga nilai-nilai tersebut memiliki skala yang seragam. Standardisasi adalah teknik transformasi data yang mengubah data sehingga memiliki mean 0 dan standar deviasi 1[8]. Ini berarti data akan memiliki distribusi yang mengikuti distribusi normal standar. Mencegah fitur dengan skala lebih besar mendominasi hasil model.
   Sebelum dilakukan proses pemodelan regresi linear berganda, seluruh variabel dalam dataset telah melalui tahap standarisasi menggunakan teknik standard scaler. Proses ini bertujuan untuk menyetarakan skala antar variabel numerik sehingga memiliki kontribusi yang seimbang terhadap model, terutama karena setiap fitur memiliki satuan dan rentang nilai yang berbeda.

2.4 Modeling dengan statsmodels

Statsmodels dirancang untuk ilmu data, analisis data, dan tujuan statistik. Berdasarkan numpy terintegrasi ke dalam panda untuk manajemen data, ini digunakan untuk memeriksa data menggunakan statistik, memperkirakan model statistik, dan melakukan uji[9]. Model yang dikembangkan adalah regresi linear berganda, yang berfungsi untuk menggambarkan hubungan antara berbagai variabel independen, seperti waktu pembelian tiket, maskapai, dan kelas penerbangan, dengan harga tiket pesawat sebagai variabel dependen.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Regresi Linear Berganda

Regresi Linear berganda merupakan teknik analisis statistik yang menguji hubungan antara variabel dependen (y) dan dua atau lebih variabel independent (x). Tujuan utamanya adalah untuk memahami dan memprediksi bagaimana perubahan dalam variabel dependen mempengaruhi perubahan dalam satu atau lebih variabel x. Dalam beberapa regresi, hubungan antara variabel -variabel ini ditunjukkan dalam bentuk persamaan linier. Persamaan digunakan untuk menentukan efek relatif dari setiap variabel independent pada variabel dependent dan untuk memprediksi nilai variabel dependen[10]. Rumus Regresi Linear Berganda:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta nXn + \varepsilon \quad (2)$$

Keterangan:

Y	: variabel dependen
X1, ... Xn	: variabel independen
β0	: intercept
β1, ... βn	: koef regresi untuk variabel independen
n	: variabel ke-n
ε	: kesalahan acak

Regresi linear berganda memiliki keuntungan dalam memberikan pemahaman tentang hubungan antara variabel dalam penelitian kuantitatif, karena bisa mempertimbangkan pengaruh dari beberapa variabel independen sekaligus[10].

2.5 Uji Asumsi Klasik

a. Uji Normalitas
Uji ini bertujuan memeriksa apakah residual dalam model regresi berdistribusi normal. Distribusi normal residual penting agar uji t dan F valid. Jika tidak terpenuhi, hasil uji statistik bisa tidak valid. Ada dua cara untuk melihat apakah residu didistribusikan secara normal, yaitu dua cara untuk memeriksa melalui analisis grafis dan pengujian statistik. Secara umum, kesehatan dapat diketahui dengan mengamati data pada diagonal grafis atau dengan penyebaran atau histogram residual titik. Model regresi diasumsikan sesuai dengan uji normalitas ketika data didistribusikan di sekitar diagonal, atau untuk menunjukkan distribusi yang sesuai dalam histogram[11].

b. Uji Multikolinearitas
Uji multikolinearitas digunakan mendeteksi hubungan kuat antar variabel independen. Jika korelasi terlalu tinggi, estimasi koefisien menjadi tidak akurat karena standard error membesar atau tak terhingga[10]. Multikolinearitas dapat dideteksi menggunakan Variance Inflation Factor (VIF) dengan rumus:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (3)$$

$R2$ adalah R-squared dari regresi variabel independen ke-i terhadap variabel independen lainnya. Nilai VIF > 10 menunjukkan adanya indikasi multikolinearitas tinggi.

c. Uji Heteroskedastisitas
Uji heteroskedastisitas dilakukan untuk mengetahui apakah varian residual bersifat konstan (homoskedastisitas) atau tidak. Deteksi dilakukan melalui dua pendekatan: metode grafik dan metode statistik. Metode grafik melihat pola sebar antara nilai prediksi dan residual, sedangkan metode statistik seperti uji Glejser digunakan untuk menguji secara kuantitatif. Dalam penelitian ini, kedua metode tersebut digunakan[11].

d. Uji Auto Korelasi
Uji autokorelasi bertujuan mengidentifikasi apakah residual pada periode t berkorelasi dengan residual pada periode sebelumnya (t-1). Karena efek satu periode dapat memengaruhi periode berikutnya, residual menjadi tidak independen, sehingga autokorelasi sering terjadi pada data time series. Sebaliknya, autokorelasi jarang muncul pada data cross-section. Model regresi yang baik dan akurat seharusnya bebas dari autokorelasi. Deteksinya dapat dilakukan melalui beberapa metode pengujian statistik[12].

Uji Durbon Watson

Uji Durbin Watson adalah cara untuk mendeteksi adanya autokorelasi dalam regresi linear berganda, metode ini adalah salah satu metode yang sering dipakai. Suatu model regresi dinyatakan terindikasi adanya autokorelasi apabila:

$$d_u < d < 4 - d_u \quad (4)$$

Di mana:
d adalah Nilai DW hitung
du adalah Nilai batas atas/upper DW[13]

Uji Partial Regresi

Uji Partial (Uji T)

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Uji t bertujuan untuk menguji apakah suatu variabel independen secara signifikan memengaruhi variabel dependen, Menurut Ghozali (2016). Jika nilai t hitung lebih besar dari t tabel pada tingkat signifikansi < 5%, maka variabel tersebut berpengaruh signifikan. Sebaliknya, jika t hitung lebih kecil dari t tabel, maka pengaruhnya tidak signifikan[14].

Uji Simultan F (Uji F)

Uji F (simultan) bertujuan untuk mengetahui apakah seluruh variabel independen memiliki pengaruh signifikan terhadap variabel dependen. Dalam penelitian ini, leverage, profitabilitas, dan likuiditas diuji terhadap LogHarga dengan 5% signifikansi. Jika H_0 ditolak, maka nilai signifikansi <0,05, yang berarti model regresi signifikan dan ketiga variabel tersebut secara simultan memengaruhi LogHarga. Sebaliknya, jika signifikansi >0,05, model dianggap tidak signifikan atau tidak layak[15].

2.6 Evaluasi

Evaluasi model bertujuan untuk menilai seberapa baik model mampu memberikan prediksi yang tepat dan konsisten berdasarkan data pelatihan yang digunakan. Langkah ini memastikan bahwa model berfungsi sesuai dengan tujuan analisis. Dalam konteks regresi linear berganda, evaluasi dilakukan untuk menegaskan bahwa model secara akurat merepresentasikan hubungan antara variabel independen dan dependen yang diteliti.

a. R Square (R^2)

Menggambarkan variansi dalam variabel dependen yang dapat dijelaskan oleh semua variabel independen dalam model. Rumus:

$$R^2 = \frac{\text{Jumlah Kuadrat Regresi (SSR)}}{\text{Jumlah Kuadrat Total (SST)}} \qquad (5)$$

Penjelasan:

SSR (Sum of Squares Regression): jumlah kuadrat yang dijelaskan oleh model regresi.

SST (Sum of Squares Total): total jumlah kuadrat dari variabel dependen terhadap rata-ratanya. R^2 mengukur seberapa besar proporsi variasi data yang berhasil dijelaskan oleh model[16].

b. Adjusted R Square

Adjusted R-squared merupakan indikator kecocokan model yang telah dikoreksi terhadap jumlah variabel bebas yang digunakan. Semakin besar proporsi variabel dependen yang dapat dijelaskan oleh variabel independen dalam model regresi, semakin tinggi nilai adjusted R-squared[17].

Rumus:

$$Adjusted\ R^2 = 1(\frac{1-R2)(n-1)}{n-k-1}) \qquad (6)$$

Penjelasan:

n : jumlah dataset

k : jumlah variabel x.

Adjusted R^2 akan turun jika penambahan variabel independen tidak meningkatkan model secara signifikan.

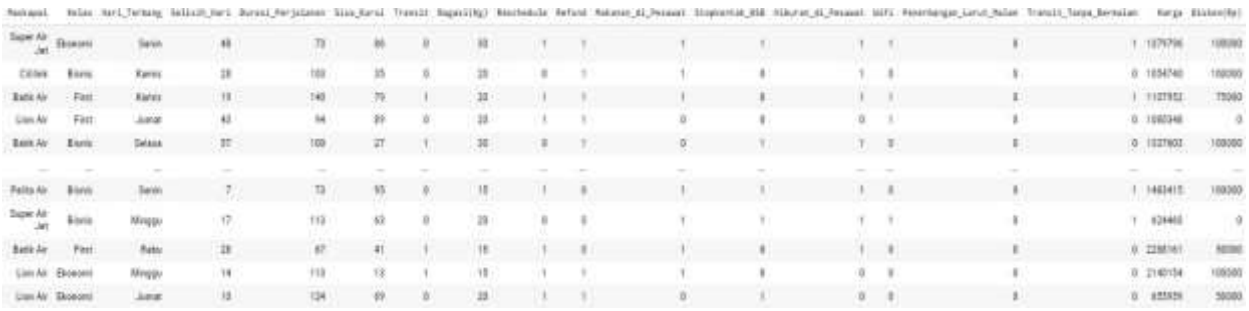
3. HASIL DAN PEMBAHASAN

Metode penelitian yang diterapkan melibatkan beberapa tahapan penting dalam pengembangan dan evaluasi model prediksi harga tiket pesawat menggunakan Regresi Linear Berganda. Berikut hasil dan pembahasan dari proses.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

3.1 Import Data

Data penelitian ini diupload ke Google Sheets yang diakses langsung dalam format CSV melalui tautan dengan parameter export?format=csv. Data dibaca menggunakan fungsi `pd.read_csv()`. Library pandas untuk mempermudah pemrosesan secara otomatis tanpa unduhan manual. Pemeriksaan awal dilakukan dengan perintah `data.head()` yang menampilkan lima baris pertama untuk melihat struktur dan isi dataset. Pendekatan ini meningkatkan efisiensi akuisisi data dan mendukung analisis lebih lanjut.



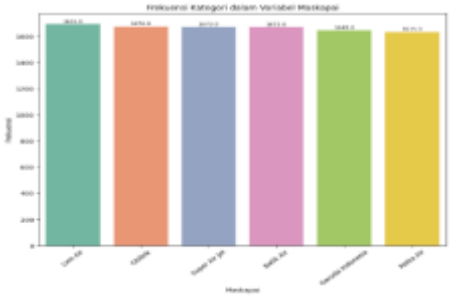
Gambar 2. Records Dataset

Dataset yang diteliti dalam penelitian ini terdiri dari 10.000 entri data penerbangan dengan 18 kolom variabel. Tiga kolom pertama yaitu Maskapai, Kelas, dan Hari_Terbang bertipe data object atau kategorikal, yang menunjukkan informasi non-numerik seperti nama maskapai, kelas penerbangan, dan hari keberangkatan. Sementara itu, 15 kolom lainnya bertipe int64 atau numerik, yang mencakup variabel-variabel kuantitatif seperti Selisih_Hari (jarak hari pemesanan dengan hari keberangkatan), Durasi_Perjalanan, Sisa_Kursi, serta berbagai fitur layanan penerbangan seperti Bagasi(Kg), Reschedule, Refund, Makanan_di_Pesawat, Stopkontak_USB, Hiburan_di_Pesawat, Wifi, hingga informasi terkait waktu seperti Penerbangan_Larut_Malam dan Transit_Tanpa_Bermalam. Kolom Harga berfungsi sebagai variabel dependen yang akan diprediksi, sedangkan Diskon(Rp) merupakan salah satu variabel independen yang bernilai numerik. Seluruh kolom dalam dataset ini memiliki jumlah data lengkap tanpa adanya nilai kosong (non-null = 10.000), yang menunjukkan bahwa data siap untuk dianalisis lebih lanjut dalam proses regresi linear berganda. Struktur ini menunjukkan kualitas data yang baik untuk keperluan pemodelan prediktif.

3.2 Pra-Pemrosesan Data

Hasil dari metode Regresi Linear Berganda menunjukkan bahwa model memiliki train score sebesar 0.81 dan test score sebesar 0.81. Ini berarti model mampu menjelaskan 81% variabilitas data pada set pelatihan maupun set pengujian. Skor yang hampir identik pada data pelatihan dan pengujian mengindikasikan bahwa model tidak mengalami overfitting atau underfitting, sehingga dapat diandalkan untuk melakukan prediksi pada data yang belum terlihat sebelumnya. Model ini memberikan kinerja yang memadai dalam memperkirakan harga berdasarkan variable yang digunakan sebagai prediktor.

Mapping Variabel Kategorikal ke Numerik Variabel Maskapai



Gambar 3. Frekuensi Kategori Maskapai

Variabel Maskapai merupakan data kategorikal yang menunjukkan nama maskapai penerbangan. Karena tipe data ini adalah object, maka perlu diubah menjadi nilai numerik agar dapat digunakan dalam analisis regresi.
Mapping dilakukan secara manual sebagai berikut:

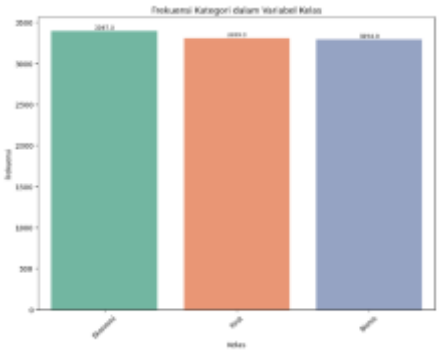
Tabel 1. Variabel Maskapai ke Numerik

Kategori	Numerik

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Batik Air	1
Citilink	2
Garuda Indonesia	3
Lion Air	4
Pelita Air	5
Super Air Jet	6

Variabel Kelas



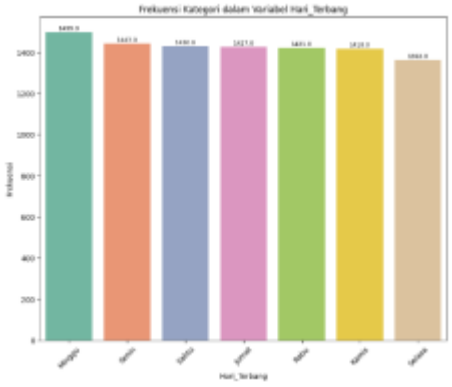
Gambar 4. Frekuensi Kategori Kelas

Variabel Kelas menggambarkan kelas layanan penerbangan (First, Bisnis, Ekonomi). Untuk kepentingan regresi, kelas dikodekan sebagai berikut:

Tabel 2. Variabel Kelas ke Numerik

Kategori	Numerik
First	1
Bisnis	2
Ekonomi	3

Variabel Hari_Terbang



Gambar 5. Frekuensi Kategori Hari Terbang

Variabel Hari_Terbang menunjukkan hari dalam seminggu ketika penerbangan dijadwalkan. Nilai kategorikal diubah menjadi kode numerik sebagai berikut:

Tabel 3. Variabel Hari Terbang ke Numerik

Kategori	Numerik
Senin	1
Selasa	2
Rabu	3

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Kamis	4
Jumat	5
Sabtu	6
Minggu	7

Setelah dilakukan proses mapping, seluruh variabel yang semula bertipe *object* telah berhasil dikonversi ke bentuk numerik bertipe *int64*. Dengan demikian, dataset telah memenuhi syarat untuk dianalisis menggunakan regresi linear berganda.

Mengubah Harga ke LogHarga

Tabel 4. Variabel Harga ke LogHarga

No.	Harga	LogHarga
1	1379796	14,137446
2	1054740	13,868805
3	1107952	13,918024
…	…	….
10.000	655959	13.393855

Variabel LogHarga digunakan sebagai variabel dependen utama dalam model regresi linear berganda, karena lebih memenuhi asumsi yang dibutuhkan dalam analisis regresi.

Normalisasi dengan StandardScaler

Gambar 6. Normalisasi dengan StandardScaler

Setelah standarisasi, setiap kolom fitur dalam data pelatihan (X_train) dan data pengujian (X_test) memiliki nilai rerata mendekati nol dan deviasi standar mendekati satu. Hal ini terlihat pada contoh data X_train di mana nilai-nilai untuk fitur seperti *Maskapai*, *Durasi_Perjalanan*, *Selisih_Hari*, hingga fitur-fitur biner seperti *Wifi* dan *Makanan_di_Pesawat* telah berubah ke dalam bentuk nilai z-score. Misalnya, nilai pada kolom *Maskapai* berkisar antara -1.4 hingga 1.4, menunjukkan bahwa variabel ini telah disesuaikan terhadap rata-ratanya. Demikian pula fitur *Durasi_Perjalanan* yang awalnya dalam satuan waktu, kini berada dalam rentang nilai sekitar -1.4 hingga 1.4, menandakan telah distandarisasi.

Split Data

```
Ukuran X train  8000
Ukuran X test  2000
Ukuran y train  8000
Ukuran y test  2000
```

Gambar 7. Split Data

Split data dilakukan untuk mengukur akurasi model terhadap data baru. Dari 10.000 data, 80% (8.000) digunakan sebagai data latih dan 20% (2.000) sebagai data uji, dibagi secara dari scikit-learn agar distribusi tetap seimbang. Data latih digunakan untuk mempelajari pengaruh Maskapai, Durasi, Kelas, dan Diskon terhadap Harga tiket, sedangkan data uji digunakan untuk mengevaluasi performa model. Pembagian ini penting untuk mencegah overfitting dan memastikan hasil prediksi yang akurat.

3.3 Model Regresi Linear Berganda

Tabel 5. Hasil Train dan Test

Model Regresi Linear	
Train Score	96,40%

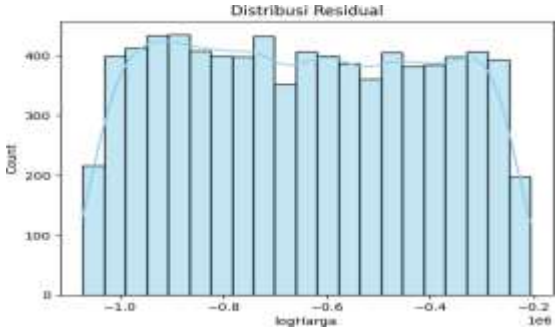
BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Test Score	96,40%

Model regresi linear berganda menunjukkan kinerja yang sangat baik, berdasarkan evaluasi pada data pelatihan dan data uji. Hasil pengujian menunjukkan nilai train score sebesar 96,40%, yang berarti model dapat menjelaskan sekitar 96,40% variasi harga tiket pada data pelatihan. Ini mengindikasikan bahwa model sangat efektif dalam mempelajari hubungan antara variabel independen dengan harga tiket. Begitu juga dengan nilai test score yang sama, yaitu 96,40%, yang menunjukkan kemampuan model yang sangat baik dalam memprediksi harga tiket pada data yang belum pernah dipelajari sebelumnya. Konsistensi yang tinggi antara train score dan test score menunjukkan bahwa model tidak mengalami overfitting dan berhasil menjaga keseimbangan antara kesesuaian dengan data pelatihan dan kemampuannya untuk menggeneralisasi pada data uji. Dengan nilai tersebut, model regresi linear berganda ini menunjukkan performa yang sangat andal dan dapat digunakan untuk memprediksi harga tiket pesawat dengan tingkat akurasi yang tinggi.

3.4 Uji Asumsi Klasik

a. Uji Normalitas



Gambar 8. Grafik Uji Normalitas

Berdasarkan grafik distribusi residual yang dihasilkan, dapat terlihat bahwa distribusi residual tidak sepenuhnya mengikuti distribusi normal. Terlihat adanya penyimpangan dari simetri, dengan kurva yang cenderung lebih condong ke satu sisi. Hal ini menunjukkan indikasi pelanggaran terhadap asumsi normalitas residual. Untuk mengkonfirmasi hasil visual ini, dilakukan uji normalitas menggunakan uji Jarque-Bera, dengan hasil p-value yang sangat kecil (1.11e-242), yang menunjukkan bahwa residual tidak terdistribusi normal. Meskipun demikian, dengan jumlah data yang besar (10.000 baris), pelanggaran terhadap normalitas tidak memberikan dampak signifikan terhadap estimasi parameter regresi, mengingat Teorema Limit Tengah yang menyatakan bahwa distribusi estimasi parameter cenderung mendekati distribusi normal dengan ukuran sampel yang besar.

b. Uji Multikolinearitas

Tabel 6. Interpretasi Multikolinearitas

Feature	VIF	Interpretasi
Maskapai	4.83	Nilai VIF yang relatif rendah, menunjukkan bahwa Maskapai tidak memiliki masalah multikolinearitas.
Kelas	6.52	Nilai VIF ini masih dalam batas wajar, meskipun lebih tinggi, menunjukkan adanya sedikit ketergantungan antar variabel.
Hari Terbang	4.75	Tidak menunjukkan multikolinearitas yang signifikan.
Selisih Hari	3.87	VIF yang moderat, tidak menunjukkan masalah besar dalam multikolinearitas.
Durasi Perjalanan	13.31	VIF yang sangat tinggi, menandakan bahwa Durasi Perjalanan memiliki ketergantungan yang kuat dengan variabel lain dalam model.
Sisa Kursi	3.76	Tidak menunjukkan multikolinearitas yang signifikan.
Transit	1.98	Nilai VIF rendah, mengindikasikan tidak ada masalah multikolinearitas.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Bagasi (Kg)	13.14	VIF tinggi, menunjukkan ketergantungan yang kuat dengan variabel lain, memerlukan perhatian lebih.
Reschedule	1.97	Nilai VIF rendah, tidak ada multikolinearitas.
Refund	1.99	Tidak ada multikolinearitas yang signifikan.
Makanan di Pesawat	1.96	Nilai VIF rendah, mengindikasikan bahwa variabel ini tidak terpengaruh oleh multikolinearitas.
Stopkontak USB	1.99	Tidak menunjukkan masalah multikolinearitas yang signifikan.
Hiburan di Pesawat	2.04	VIF rendah, menunjukkan ketergantungan minimal dengan variabel lain.
Wifi	1.97	VIF yang rendah, mengindikasikan bahwa Wifi tidak memiliki masalah multikolinearitas.
Penerbangan Larut Malam	1.14	VIF sangat rendah, tidak ada multikolinearitas yang signifikan.
Transit Tanpa Bermalam	1.99	VIF rendah, menunjukkan bahwa variabel ini tidak dipengaruhi oleh multikolinearitas.
Harga	7.42	VIF tinggi, tetapi masih dalam batas toleransi untuk beberapa model, meskipun perlu diperhatikan karena harga bisa dipengaruhi oleh variabel lainnya.
Diskon (Rp)	2.90	VIF yang cukup moderat, menunjukkan adanya ketergantungan ringan antara variabel.

Berdasarkan hasil analisis VIF, ditemukan adanya multikolinearitas tinggi pada beberapa variabel seperti Durasi_Perjalanan dan Bagasi(Kg). Oleh karena itu, model regresi linear berganda disesuaikan dengan menggunakan metode Ridge Regression untuk mengatasi multikolinearitas.

Solusi menggunakan metode Ridge Regression

```
Alpha terbaik: 1000.0
Koefisien: [ 3.20400783e+03 -4.66941980e+03  5.36786620e+03  9.36043108e+01
  8.43864031e+01 -1.28932238e+02 -6.44009492e+02 -1.93634847e+01
  1.73100120e+04  4.25343670e+03  2.84001058e+03  1.33401480e+04
 -1.00516000e+04 -4.86741732e+03  5.81094851e+03 -1.35903575e+04
 -9.81537012e-01]
```

Gambar 9. Hasil Solusi Uji Multikolinearitas menggunakan metode Ridge Regression

Hasil Ridge Regression dengan alpha 1000.0 menunjukkan bahwa model ini telah berhasil mengurangi masalah multikolinearitas. Koefisien variabel-variabel yang memiliki ketergantungan tinggi dengan variabel lain telah diperkecil, memungkinkan model untuk menghasilkan estimasi yang lebih stabil dan akurat. Meskipun beberapa koefisien masih tergolong besar, variabel-variabel tersebut tetap memiliki pengaruh yang signifikan terhadap prediksi harga tiket pesawat, meskipun ada penalti regulasi yang diterapkan. Ridge Regression efektif dalam mengatasi masalah multikolinearitas dan meningkatkan keandalan model regresi.

c. Uji Heteroskedastisitas

```
=== Uji Heteroskedastisitas (Breusch-Pagan) ===
LM Statistic: 992.5173
LM-Test p-value: 0.0000
F-Statistic: 61.0993
F-Test p-value: 0.0000
Terdapat indikasi heteroskedastisitas pada data (p-value <= 0.05)
```

Gambar 10. Hasil Uji Heteroskedastisitas Breusch-Pagan

Berdasarkan uji Breusch-Pagan, ditemukan adanya indikasi heteroskedastisitas (p-value < 0.05), yang mengganggu kevalidan standar error model regresi. Oleh karena itu, model diperbaiki menggunakan regresi dengan standar error robust

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

(HC3), sehingga hasil estimasi koefisien tetap dapat diandalkan dan uji signifikansi tetap valid meskipun terdapat heteroskedastisitas pada data.

Solusi untuk Mengatasi Heteroskedastisitas

Untuk mengatasi masalah heteroskedastisitas yang ditemukan dalam model regresi, penelitian ini menerapkan pendekatan regresi linier dengan standar error yang tahan terhadap heteroskedastisitas, yaitu menggunakan metode OLS atau Ordinary Least Squares dengan varian HC3. Metode ini dipilih karena dapat memberikan estimasi standar error yang lebih akurat meskipun terdapat pelanggaran terhadap asumsi homoskedastisitas.

OLS Regression Results

Dep. Variable:	logHarga	R-squared:	0.964
Model:	OLS	Adj. R-squared:	0.964
Method:	Least Squares	F-statistic:	9086.
Date:	Fri, 09 May 2025	Prob (F-statistic):	0.00
Time:	12:30:38	Log-Likelihood:	11097.
No. Observations:	10000	AIC:	-2.216e+04
Df Residuals:	9981	BIC:	-2.202e+04
Df Model:	18		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
const	13.0747	0.007	1893.236	0.000	13.061	13.088
Maskapai	-7.88e-05	0.000	-0.167	0.868	-0.001	0.001
Kelas	-0.0005	0.001	-0.537	0.592	-0.002	0.001
Hari_Terbang	-0.0005	0.000	-1.195	0.232	-0.001	0.000
Selisih_Hari	3.567e-05	4.59e-05	0.777	0.437	-5.42e-05	0.000
Durasi_Perjalanan	4.317e-05	3.03e-05	1.427	0.154	-1.61e-05	0.000
Sisa_Kursi	-3.758e-05	2.74e-05	-1.373	0.170	-9.12e-05	1.61e-05
Transit	0.0008	0.002	0.521	0.602	-0.002	0.004
Bagasi(Kg)	-8.173e-05	0.000	-0.573	0.566	-0.000	0.000
Reschedule	-0.0015	0.002	-0.958	0.338	-0.005	0.002
Refund	0.0005	0.002	0.311	0.755	-0.003	0.004
Makanan_di_Pesawat	-0.0018	0.002	-1.144	0.253	-0.005	0.001
Stopkontak_USB	-0.0023	0.002	-1.428	0.153	-0.005	0.001
Hiburan_di_Pesawat	0.0018	0.002	1.120	0.263	-0.001	0.005
Wifi	-0.0014	0.002	-0.852	0.394	-0.005	0.002
Penerbangan_Larut_Malam	0.0041	0.002	1.731	0.083	-0.001	0.009
Transit_Tanpa_Bermalam	0.0021	0.002	1.297	0.194	-0.001	0.005
Harga	7.108e-07	1.78e-09	398.508	0.000	7.07e-07	7.14e-07
Diskon(Rp)	-5.21e-08	2.24e-08	-2.323	0.020	-9.61e-08	-8.14e-09

Omnibus:	1063.072	Durbin-Watson:	1.958
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1435.843
Skew:	-0.922	Prob(JB):	0.00
Kurtosis:	3.211	Cond. No.	1.41e+07

Gambar 11. Hasil Regresi menggunakan OLS dengan varian HC3 untuk mengatasi heteroskedastisitas

Hasil regresi linear berganda menunjukkan model sangat akurat, dengan R-squared dan Adjusted R-squared sebesar 0,964, artinya 96,4% variasi logHarga dijelaskan oleh variabel independen. Uji signifikansi menunjukkan sebagian besar variabel tidak berpengaruh signifikan ($p > 0,05$), kecuali Harga ($p < 0,001$) dan Diskon (Rp) ($p = 0,020$) yang berpengaruh nyata. Nilai F-statistic sebesar 9086 dan Prob(F) = 0,000 mengindikasikan bahwa model signifikan secara keseluruhan. Dengan demikian, penggunaan standar error robust dalam regresi ini telah berhasil mengatasi pengaruh heteroskedastisitas tanpa mengubah struktur model, sehingga hasil estimasi koefisien tetap valid dan dapat diandalkan dalam interpretasi maupun pengambilan keputusan.

d. Uji Autokorelasi

```
=== Uji Autokorelasi (Durbin-Watson) ===
Durbin-Watson Statistic: 1.9582
Tidak terdapat autokorelasi (1.5 <= DW <= 2.5)
```

Gambar 12. Hasil Uji Autokorelasi

Hasil autokorelasi dengan menggunakan statistik Durbin-Watson menghasilkan nilai sebesar 1,9582. Nilai ini berada dalam kisaran 1,5 hingga 2,5, yang secara umum mengindikasikan tidak terjadinya autokorelasi di dalam data. Dengan demikian, dapat disimpulkan bahwa model regresi tidak menunjukkan adanya hubungan antar residual, atau dengan kata

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

lain, residual bersifat independen. Kondisi ini menunjukkan bahwa salah satu asumsi dasar dalam regresi linear telah

OLS Regression Results

Dep. Variable:	logHarga	R-squared:	0.964
Model:	OLS	Adj. R-squared:	0.964
Method:	Least Squares	F-statistic:	1.198e+04
Date:	Sat, 10 May 2025	Prob (F-statistic):	0.00
Time:	04:06:59	Log-Likelihood:	8910.6
No. Observations:	8000	AIC:	-1.778e+04
Df Residuals:	7981	BIC:	-1.765e+04
Df Model:	18		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	13.0764	0.007	1746.059	0.000	13.062	13.091
Maskapai	5.271e-05	0.001	0.101	0.919	-0.001	0.001
Kelas	0.0003	0.001	0.293	0.770	-0.002	0.002
Hari_Terbang	-0.0009	0.000	-2.120	0.034	-0.002	-7.06e-05
Selisih_Hari	3.917e-05	5.09e-05	0.770	0.441	-6.06e-05	0.000
Durasi_Perjalanan	3.588e-05	3.39e-05	1.059	0.290	-3.05e-05	0.000
Sisa_Kursi	-3.111e-05	3.05e-05	-1.021	0.307	-9.08e-05	2.86e-05
Transit	0.0013	0.002	0.735	0.462	-0.002	0.005
Bagasi(Kg)	-0.0003	0.000	-1.621	0.105	-0.001	5.42e-05
Reschedule	-0.0013	0.002	-0.738	0.460	-0.005	0.002
Refund	0.0002	0.002	0.126	0.900	-0.003	0.004
Makanan_di_Pesawat	-0.0012	0.002	-0.701	0.484	-0.005	0.002
Stopkontak_USB	-0.0020	0.002	-1.127	0.260	-0.005	0.001
Hiburan_di_Pesawat	0.0028	0.002	1.576	0.115	-0.001	0.006
Wifi	-0.0021	0.002	-1.193	0.233	-0.006	0.001
Penerbangan_Larut_Malam	0.0052	0.003	1.940	0.052	-5.54e-05	0.011
Transit_Tanpa_Bermalam	0.0028	0.002	1.599	0.110	-0.001	0.006
Harga	7.108e-07	1.54e-09	463.006	0.000	7.08e-07	7.14e-07
Diskon(Rp)	-3.069e-08	2.52e-08	-1.218	0.223	-8.01e-08	1.87e-08

Omnibus:	830.638	Durbin-Watson:	2.015
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1114.239
Skew:	-0.909	Prob(JB):	1.11e-242
Kurtosis:	3.190	Cond. No.	1.41e+07

terpenuhi, sehingga estimasi model dapat dinilai sahih dalam hal asumsi kemandirian error.

Gambar 13. Hasil Model Statistik dengan Statsmodels

3.5 Model Statistik dengan Statsmodels

Hasil regresi linear berganda dengan logHarga sebagai variabel dependen menunjukkan performa statistik yang sangat baik. Nilai R-squared dan Adjusted R-squared sebesar 0,964 mengindikasikan bahwa 96,4% variasi harga tiket dapat dijelaskan oleh 18 variabel independen, menandakan kemampuan prediksi yang tinggi. Nilai F-statistic sebesar 11.980 dengan p-value 0,000 menunjukkan bahwa model signifikan secara keseluruhan. Namun, hanya beberapa variabel seperti Hari_Terbang dan Harga yang berpengaruh signifikan ($p < 0,05$), sedangkan sebagian besar variabel lainnya, seperti Maskapai, Kelas, Durasi, Fitur layanan, dan Diskon, tidak signifikan ($p > 0,05$). Nilai Durbin-Watson sebesar 2,015 menunjukkan tidak adanya autokorelasi, tetapi condition number yang sangat tinggi (1.41e+07) mengindikasikan potensi multikolinearitas atau masalah numerik. Meskipun uji normalitas menunjukkan sedikit penyimpangan, hal ini masih dapat ditoleransi karena ukuran data yang besar ($n = 10.000$). Secara keseluruhan, model kuat dalam menjelaskan logHarga, namun penyederhanaan model disarankan untuk meningkatkan interpretabilitas dan mengurangi risiko multikolinearitas.

4. KESIMPULAN

Kesimpulan dari penelitian prediksi harga tiket pesawat domestik ini adalah bahwa penelitian ini berhasil mengembangkan model prediksi harga tiket pesawat domestik untuk rute Surabaya–Jakarta menggunakan metode regresi linear berganda dengan memanfaatkan data penerbangan yang lengkap dan terstruktur sebanyak 10.000 data. Proses pengembangan model dilakukan melalui serangkaian tahap, mulai dari impor dan pra-pemrosesan data, transformasi variabel, normalisasi, hingga pemisahan data latih dan uji, yang semuanya berkontribusi pada evaluasi dan pengembangan

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

model yang menyeluruh. Hasil dari regresi linear berganda mengindikasikan bahwa model ini memiliki performa yang sangat baik, dengan nilai R-squared sebesar 96,4%, yang berarti bahwa sebagian besar perubahan harga tiket dapat dijelaskan oleh variabel-variabel independen yang diterapkan dalam model. Evaluasi terhadap data uji juga menunjukkan konsistensi dalam akurasi, yang menandakan bahwa model tidak mengalami overfitting. Uji asumsi klasik menunjukkan adanya pelanggaran terhadap normalitas dan heteroskedastisitas, namun langkah-langkah perbaikan seperti transformasi logaritmik dan penggunaan regresi dengan standar error robust (HC3) berhasil mempertahankan validitas model. Masalah multikolinearitas yang ditemukan pada beberapa variabel juga berhasil diminimalkan dengan metode Ridge Regression. Selain itu, tidak ditemukan autokorelasi dalam residual, yang meningkatkan keandalan hasil estimasi.Meskipun terdapat banyak variabel independen dalam model, hanya beberapa yang terbukti signifikan secara statistik, seperti Harga dan Diskon. Hal ini menunjukkan pentingnya untuk mengevaluasi kontribusi setiap variabel terhadap variabel target, yang dapat menyederhanakan model di masa depan. Secara umum, model yang dikembangkan dalam penelitian ini memiliki akurasi dan kestabilan yang tinggi, sehingga dapat dijadikan dasar untuk pengambilan keputusan dalam merumuskan strategi penetapan harga tiket pesawat. Model ini juga memiliki potensi untuk diterapkan pada prediksi harga untuk rute penerbangan lain dengan struktur data yang serupa.

REFERENCES

[1] A. C. Revandha And A. Syaputra, "Pengaruh Kinerja Petugas Check-In Counter Terhadap Kepuasan Penumpang Maskapai Pelita Air Pt. Gapura Angkasa Bandar Udara Internasional Juanda Surabaya," *Indonesian Journal Of Aviation Science And Engineering*, Vol. 1, No. 4, P. 7, Jun. 2024, Doi: 10.47134/Pjase.V1i4.2794.

[2] E. Lastriani, D. Jurusan, M. Stie, And D. Putra, "Analisis Faktor-Faktor Yang Dipertimbangakan Konsumen Dalam Pemilihan Maskapai Penerbangan Sebagai Alat Transportasi Udara Di Kota Pekanbaru."

[3] D. Triyana, M. Muharrom, A. Haromainy, And H. Maulana, "Implementasi Metode Ensemble Majority Vote Pada Algoritma Naive Bayes Dan Random Forest Untuk Analisis Sentimen Twitter Harga Tiket Pesawat Domestik," 2024.

[4] C. H. Setiawan, . Fatichah And, And A. Saikhu, "Feature Selection For Multilabel Classification Of Student Feedback Using Filter And Metaheuristic Methods," International Conference On Intelligent Cybernetics Technology & Applications (Icicyta). [Online]. Available: Https://Ieeexplore.Ieee.Org/Document/10913038

[5] D. Miftahul Huda, G. Dwilestari, And A. Rizki Rinaldi, "Jurnal Informatika Dan Rekayasa Perangkat Lunak Prediksi Harga Mobil Bekas Menggunakan Algoritma Regresi Linear Berganda".

[6] T. Nurmansyah, R. Kurniawan, Y. A. Wijaya, P. Studi, T. Informatika, And I. Cirebon, "Jurnal Informatika Dan Rekayasa Perangkat Lunak Analisis Data Stok Alat Kesehatan Menggunakan Metode Regresi Linier Berdasarkan Nilai Rmse," Vol. 6, No. 1, Pp. 177–182, 2024.

[7] M. Sholeh, E. Kumalasari Nurnawati, And U. Lestari, "Penerapan Data Mining Dengan Metode Regresi Linear Untuk Memprediksi Data Nilai Hasil Ujian Menggunakan Rapidminer," 2023. [Online]. Available: Https://Archive.Ics.Uci.Edu/Ml/Datasheets.Php.

[8] F. Ramdani And I. Q. Utami, *Pengantar Data Science*. 2022.

[9] R. Kaestria And E. F. Himmah, "Implementasi Bahasa Pemrograman Python Untuk Path Analysis," *Jurnal Komputasi*, Vol. 11, No. 2, Pp. 105–117, 2023, Doi: 10.23960/Komputasi.V11i2.6634.

[10] A. Wardhana And Z. Iba, "Analisis Regresi Dan Analisis Jalur Untuk Riset Bisnis Menggunakan Spss 29.0 & Smart-Pls 4.0," Pp. 1–65, Jul. 2024.

[11] M. Regi Abdi Putra Amanta *Et Al.*, "Analisis Pengaruh Motivasi Dan Disiplin Kerja Terhadap Kinerja Dosen Di Institut Teknologi Sumatera Dengan Metode Regresi Linier Berganda," *Prosiding Seminar Nasional Sains Dan Teknologi Seri Iii Fakultas Sains Dan Teknologi*, Vol. 2, No. 1, 2025.

[12] Dyah. N. Arum. Janie, *Statistik Deskriptif & Regresi Linier Berganda Dengan Spss*. Semarang University Press.

[13] A. Fania And F. Sri Handayani, "Analisis Usabilitas Aplikasi Magang Rri Palembang Menggunakan Metode Regresi Linier Berganda," *Teknomatika*, Vol. 14, No. 01, 2024.

[14] B. Teta, "Pengaruh Kelalaian Karyawan Terhadap Produktivitas Di Tambak Udang Venambak," Vol. 6, No. 3, Pp. 613–623, 2024.

[15] S. W. Ningtias, "Pengaruh Leverage, Profitabilitas Dan Likuiditas Terhadap Respon Investor Akhmad Riduwan Sekolah Tinggi Ilmu Ekonomi Indonesia (Stiesia) Surabaya."

[16] Hermansyah, A. Abdullah, And P. Y. Utami, "Penerapan Metode Regresi Linier Berganda Untuk Memprediksi Panen Kelapa Sawit," Vol. 20, Pp. 1–15, Feb. 2024, Accessed: May 04, 2025. [Online]. Available: Https://Jurnalmahasiswa.Stiesia.Ac.Id/Index.Php/Jira/Article/View/6211/6264

[17] S. Mulyani, Y. Fitri, S. Selvia, N. Rahmadani, D. Lestari, And W. Meka, "Prediksi Potensi Timbulan Limbah Ampas Kopi Sebagai Sumber Penghasil Senyawa Bioaktif Di Kota Pekanbaru," *Jurnal Ilmu Lingkungan*, Vol. 22, No. 6, Pp. 1412–1423, Nov. 2024, Doi: 10.14710/Jil.22.6.1412-1423.

[18] B. Bangun and A. K. Karim, "Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining," Jurnal Media Informatika Budidarma, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.

[19] Karim, "Penerapan Algoritma Entropy dan Aras Menentukan Desa Terbaik Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 1, pp. 33–43, 2022.

[20] Karim, "Implementation of the Multi-Objective Optimization Method on the Basic of Ratio Analysis (MOORA) and Entropy Weighting in New Employee Recruitment," vol. 5, no. 2, pp. 704–712, 2024, doi: 10.47065/josh.v5i2.4859.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 45 - 58
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

[21] Karim, "Sistem Pendukung Keputusan Penerimaan Analis Di Pusat Penelitian Kelapa Sawit Menggunakan Metode Complex Proportional Assessment (Copras)," Buletin Ilmiah Informatika Teknologi, vol. 2, no. 1, pp. 32–42, [Online]. Available: https://ejurnal.amikstiekomsu.ac.id/index.php/BIIT

[22] Karim, "Clusterisasi Tingkat Pengangguran Terbuka Menurut Provinsi di Indonesia Menggunakan Algoritma K-Medoids," 2024, doi: 10.47065/bits.v6i3.6198.

[23] Abdul Karim, "Implementasi Metode Multi-Objective Optimization On The Basis Of Ratio Analysis dalam Seleksi Mahasiswa Program Indonesia Pintar," Bulletin of Computer Science Research, vol. 3, no. 5, pp. 351–356, 2023, doi: 10.47065/bulletincsr.v3i5.283.