

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

Oleh:

Enggi Sabrilla Assara

Hamzah Setiawan

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Mei, 2025



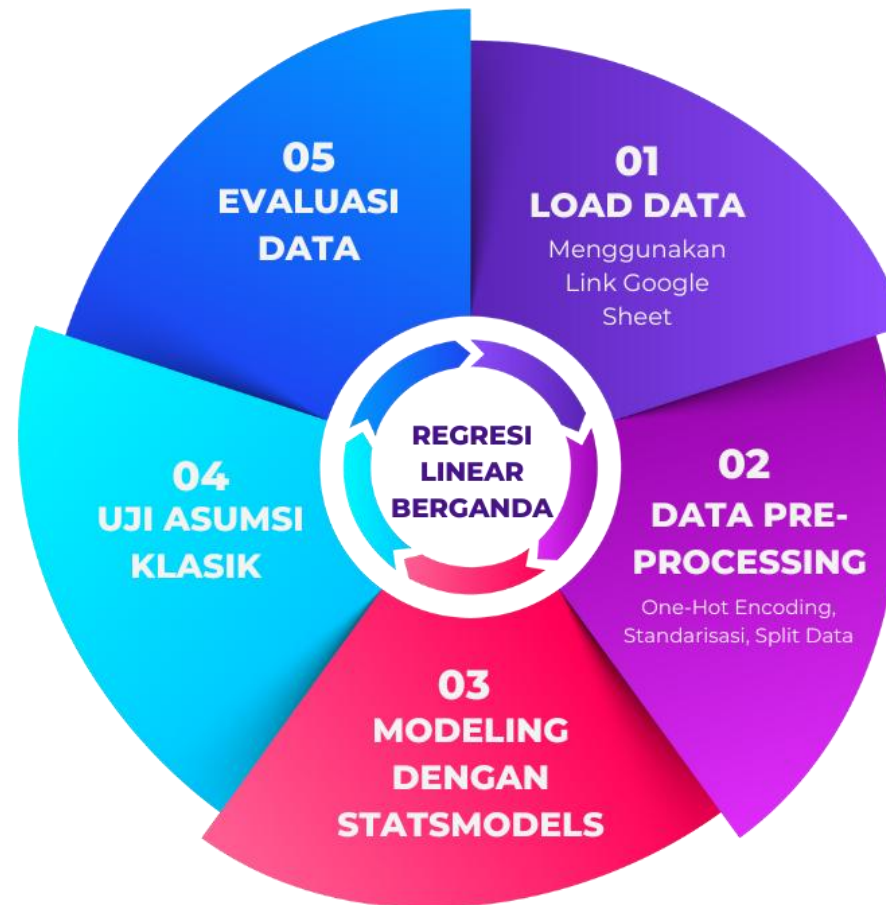
# Pendahuluan

Industri penerbangan yang terus berkembang menyebabkan persaingan antar maskapai semakin ketat, terutama pada rute padat seperti Surabaya–Jakarta yang mengalami fluktuasi harga tiket signifikan akibat berbagai faktor, seperti maskapai, waktu terbang, dan jenis layanan. Kondisi ini menyulitkan konsumen dalam merencanakan perjalanan serta menyulitkan maskapai dalam menetapkan strategi harga yang tepat. **Penelitian ini bertujuan** membangun model prediksi harga tiket menggunakan regresi linier berganda untuk mengidentifikasi faktor-faktor yang memengaruhi harga secara signifikan, sehingga hasilnya dapat dimanfaatkan untuk pengambilan keputusan yang lebih efisien dan strategis oleh berbagai pihak terkait.

# Pertanyaan Penelitian (Rumusan Masalah)

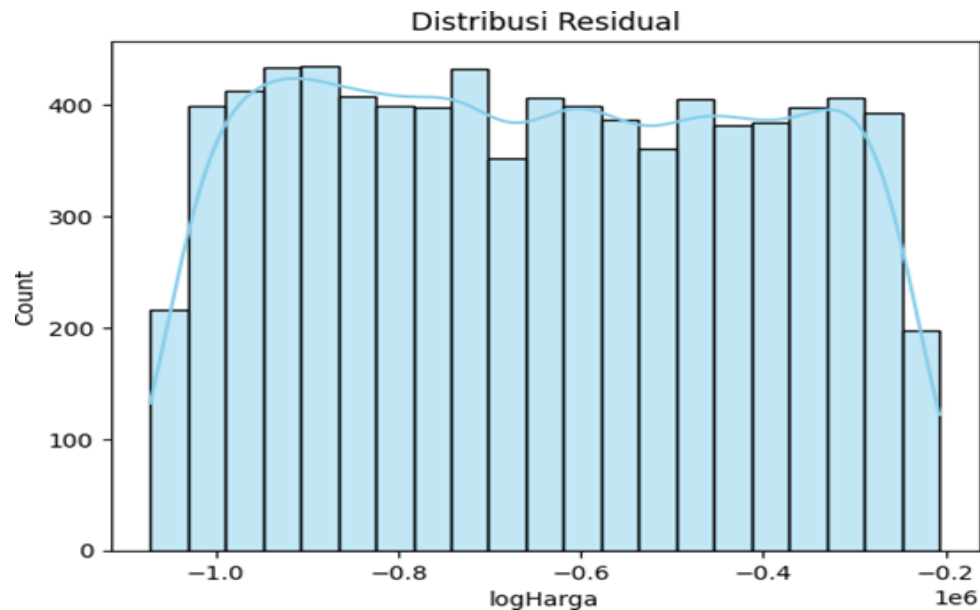
- Bagaimana membangun model prediksi harga tiket pesawat rute Surabaya– Jakarta yang akurat menggunakan regresi linier berganda?
- Apa saja variabel independen yang paling berpengaruh terhadap harga tiket pesawat pada rute tersebut?
- Bagaimana mengatasi permasalahan statistik seperti non-normalitas, heteroskedastisitas, dan multikolinearitas dalam model regresi harga tiket pesawat?
- Sejauh mana performa model regresi linier berganda dalam memprediksi harga tiket pesawat, ditinjau dari nilai R-squared dan uji data uji (test data)?
- Bagaimana hasil dari model prediksi ini dapat digunakan oleh konsumen, maskapai penerbangan, dan pemangku kepentingan lainnya dalam pengambilan keputusan yang lebih efisien?

# Metode



# Pembahasan

## UJI NORMALITAS



Berdasarkan grafik residual, terlihat penyimpangan dari distribusi normal, dengan p-value sangat kecil ( $1,11e-242$ ). Ini menunjukkan adanya pelanggaran asumsi normalitas. Namun, karena jumlah data besar (10.000 baris), dampaknya terhadap estimasi regresi dapat diabaikan sesuai Teorema Limit Tengah.

# Pembahasan

## UJI MULTIKOLINEARITAS

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.

<u>Feature</u>	<b>VIF</b>
Maskapai	4.83
Kelas	6.52
Hari Terbang	4.75
Selisih Hari	3.87
Durasi Perjalanan	13.31
Sisa Kursi	3.76
Transit	1.98
Bagasi (Kg)	13.14
Reschedule	1.97
Refund	1.99
Makanan di Pesawat	1.96
Stopkontak USB	1.99
Hiburan di Pesawat	2.04
Wifi	1.97
Penerbangan Larut Malam	1.14
Transit Tanpa Bermalam	1.99
Harga	7.42
Diskon (Rp)	2.90



# Pembahasan

## UJI 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]
```

Ridge Regression dengan alpha 1000 **berhasil mengurangi multikolinearitas** dengan mengecilkan koefisien variabel yang saling berkorelasi, sehingga model menjadi lebih stabil dan akurat. Meski ada penalti, variabel penting tetap signifikan.

# Pembahasan

## 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  $\leq 0.05$ )

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



# Pembahasan

```
=====
                        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
=====
```

## UJI HETEROSKEDASTISITAS

### Solusi untuk Mengatasi Heteroskedastisitas

Nilai F-statistic 9086 dengan  $p < 0,001$  menunjukkan 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.

# Pembahasan

## UJI AUTOKORELASI

=== Uji Autokorelasi (Durbin-Watson) ===  
Durbin-Watson Statistic: 1.9582  
Tidak terdapat autokorelasi ( $1.5 \leq DW \leq 2.5$ )

Nilai Durbin-Watson sebesar 1,958 berada dalam kisaran 1,5–2,5, yang mengindikasikan **tidak terjadi autokorelasi**. Artinya, residual bersifat independen, sehingga asumsi kemandirian error terpenuhi dan model regresi dapat dianggap sah dalam aspek ini.

# Hasil Pembahasan

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.

```
=====
                        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:                   04:54:11      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
=====
```

# Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi harga tiket pesawat rute Surabaya–Jakarta menggunakan regresi linear berganda dengan 10.000 data terstruktur. Model menunjukkan performa sangat baik ( $R^2 = 96,4\%$ ) dan konsisten pada data uji. Meskipun terdapat pelanggaran asumsi normalitas dan heteroskedastisitas, solusi seperti transformasi log dan regresi dengan standar error robust menjaga validitas model. Masalah multikolinearitas juga berhasil dikurangi dengan Ridge Regression. Hanya beberapa variabel seperti Harga dan Diskon yang signifikan, menunjukkan perlunya penyederhanaan model. Model ini akurat, stabil, dan dapat digunakan sebagai dasar pengambilan keputusan serta diterapkan pada rute lain dengan data serupa.

# Temuan Penting Penelitian

- Model prediksi yang sangat akurat
- Model tidak mengalami overfitting
- Pelanggaran asumsi klasik berhasil ditangani
- Model dapat diaplikasikan secara praktis
- Potensi pengembangan model lebih lanjut

# Manfaat Penelitian

- **Bagi Konsumen**

Model prediksi membantu penumpang merencanakan perjalanan yang lebih hemat dengan memahami pola harga dan memilih waktu serta layanan yang tepat.

- **Bagi Maskapai Penerbangan**

Model ini mendukung strategi harga yang kompetitif dengan mengidentifikasi faktor-faktor yang paling memengaruhi harga tiket.

- **Bagi Regulator Pemerintah**

Hasil penelitian dapat digunakan untuk menyusun kebijakan harga yang adil dan mengawasi dinamika harga tiket domestik.

- **Bagi Peneliti dan Akademisi**

Penelitian ini memperkaya literatur dalam analisis prediktif dan dapat menjadi dasar penelitian lanjutan di bidang data science dan transportasi.

- **Bagi Pengembang Sistem Informasi**

Model dapat diintegrasikan dalam aplikasi pencarian tiket untuk memberikan estimasi harga yang lebih akurat dan efisien.



# Referensi

- [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 Dipertimbangkan 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] H. Setiawan, C. Fatichah 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.



# Referensi

- [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.

