**Prediksi Harga Tiket Pesawat Menggunakan Metode Linear Regression**

**Yehezkiel Juandro Metta**

Universitas Pelita Bangsa

**Rahmat Hidayat**

Universitas Pelita Bangsa

**Hansen Julio**

Universitas Pelita Bangsa

Jl. Inspeksi Kalimalang Tegal Danas Arah Deltamas, Cibatu, Cikarang

*Korespondensi penulis: aganzzjuandrometta@gmail.com*

***Abstrak.*** *The determination of airline ticket prices is a dynamic process influenced by various factors such as airline carrier, route, booking time, and flight duration. The high price volatility often makes it difficult for consumers to make purchasing decisions. This study aims to build a predictive model for airline ticket prices by applying the Linear Regression method. In this approach, the model is developed to analyze and find the mathematical relationship between independent variables (flight features) and the dependent variable (ticket price). The results of this model are expected to provide an accurate price estimation, which can serve as a useful tool for potential passengers in planning their travels and for service providers in setting competitive pricing strategies.*

***Keywords:*** *Price Prediction, Linear Regression, Machine Learning, Airline Tickets, Data Analysis.*

**Abstrak.** Penentuan harga tiket pesawat merupakan proses dinamis yang dipengaruhi oleh berbagai faktor seperti maskapai penerbangan, rute, waktu pemesanan, dan durasi penerbangan. Tingginya volatilitas harga seringkali menyulitkan konsumen dalam membuat keputusan pembelian. Penelitian ini bertujuan untuk membangun sebuah model prediktif untuk harga tiket pesawat dengan menerapkan metode Regresi Linier. Dalam pendekatan ini, model dikembangkan untuk menganalisis dan menemukan hubungan matematis antara variabel-variabel independen (fitur penerbangan) dengan variabel dependen (harga tiket). Hasil dari model ini diharapkan dapat memberikan estimasi harga yang akurat, sehingga dapat menjadi alat bantu yang bermanfaat bagi calon penumpang dalam merencanakan perjalanan dan bagi penyedia layanan dalam menetapkan strategi harga yang kompetitif.

***Kata Kunci:*** *Prediksi Harga, Regresi Linier, Machine Learning, Harga Tiket Pesawat, Analisis Data.*

**PENDAHULUAN**

Harga tiket pesawat sering kali mengalami fluktuasi yang signifikan, yang dipengaruhi oleh berbagai faktor seperti jenis maskapai penerbangan, waktu keberangkatan, durasi penerbangan, dan jarak waktu pemesanan sebelum hari keberangkatan. Dinamika harga yang kompleks ini seringkali menciptakan ketidakpastian bagi calon penumpang dan menjadi tantangan dalam membuat keputusan pembelian tiket yang tepat waktu dan efisien. Oleh karena itu, kemampuan untuk memprediksi pergerakan harga menjadi krusial, tidak hanya bagi konsumen untuk merencanakan anggaran perjalanan, tetapi juga bagi maskapai penerbangan dalam mengelola pendapatan (*revenue management*) dan menyusun strategi penetapan harga yang kompetitif.

Untuk mengatasi tantangan ketidakpastian harga tersebut, pemanfaatan teknologi *machine learning* menawarkan solusi yang menjanjikan. Salah satu metode fundamental dalam *machine learning* untuk tugas prediksi nilai numerik adalah Regresi Linier. Metode ini bekerja dengan cara membangun model matematis untuk menemukan pola dan hubungan antara serangkaian variabel independen (fitur-fitur penerbangan) dengan satu variabel dependen (harga tiket). Dengan demikian, penelitian ini bertujuan untuk menerapkan metode Regresi Linier guna membangun sebuah model yang mampu memprediksi harga tiket pesawat secara akurat berdasarkan fitur-fitur yang relevan, yang pada akhirnya dapat menjadi alat bantu praktis bagi konsumen maupun industri penerbangan.

**KAJIAN TEORITIS**

1. **Data Mining dan Machine Learning**

Data mining adalah proses mengekstraksi informasi atau pola menarik dari kumpulan data yang besar. Dalam lingkup yang lebih modern, proses ini seringkali menjadi bagian dari *Machine Learning* (Pembelajaran Mesin), yaitu cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dari data, mengidentifikasi pola, dan membuat keputusan atau prediksi. Dalam konteks penelitian ini, *machine learning* digunakan untuk membangun model yang dapat mempelajari pola dari data historis tiket pesawat guna memprediksi harga di masa depan.

1. **Supervised Learning dan Regresi**

*Supervised Learning* (pembelajaran terarah) merupakan pendekatan *machine learning* di mana algoritma dilatih menggunakan kumpulan data yang telah diberi label, artinya setiap data input (fitur) memiliki pasangan output (target) yang sudah diketahui. Regresi adalah salah satu jenis utama dari *supervised learning* yang bertujuan untuk memprediksi nilai output yang bersifat kontinu atau numerik, seperti harga, suhu, atau berat. Tujuan utama analisis regresi adalah untuk memahami dan memodelkan hubungan antara satu atau lebih variabel independen (prediktor) dengan satu variabel dependen (target).

1. **Metode Regresi Linier (Linear Regression)**

Regresi Linier adalah salah satu algoritma regresi yang paling fundamental dan banyak digunakan. Algoritma ini bekerja dengan mengasumsikan adanya hubungan linier (berbentuk garis lurus) antara variabel independen (X) dan variabel dependen (Y). Tujuannya adalah untuk menemukan garis lurus terbaik (*best-fit line*) yang dapat memodelkan hubungan tersebut dengan meminimalkan selisih (error) antara nilai prediksi dan nilai aktual. Dalam kasus dengan lebih dari satu variabel independen, metode ini disebut Regresi Linier Berganda dan dimodelkan melalui persamaan:

**Y= β0 ​+ β1​X1 ​+ β2​X2 ​+ ... + βn​Xn ​+ ϵ**

Di mana:

* **Y** adalah variabel dependen (harga tiket yang diprediksi).
* **X1**​, **X2**​, **...** ,**Xn**​ adalah variabel-variabel independen (fitur seperti durasi, jumlah transit, dll.).
* **β0**​ adalah *intercept*, yaitu nilai dasar dari **Y** ketika semua variabel **X** bernilai nol.
* **β1**​, **...** ,**βn**​ adalah koefisien regresi, yang mengukur seberapa besar pengaruh setiap variabel **X** terhadap **Y**.
* **ϵ** adalah *error term* atau residu, yang merepresentasikan selisih antara nilai aktual dan nilai prediksi.

**METODE PENELITIAN**

1. **Jenis Penelitian**

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan prediktif menggunakan metode machine learning.

1. **Data dan Sumber Data**

Data yang digunakan berasal dari dataset penerbangan yang memuat informasi seperti maskapai, asal dan tujuan kota, kelas penerbangan, durasi, waktu keberangkatan, dan harga tiket. Dataset ini telah dibersihkan untuk menghilangkan data duplikat atau tidak lengkap.

1. **Alat dan Perangkat Lunak**

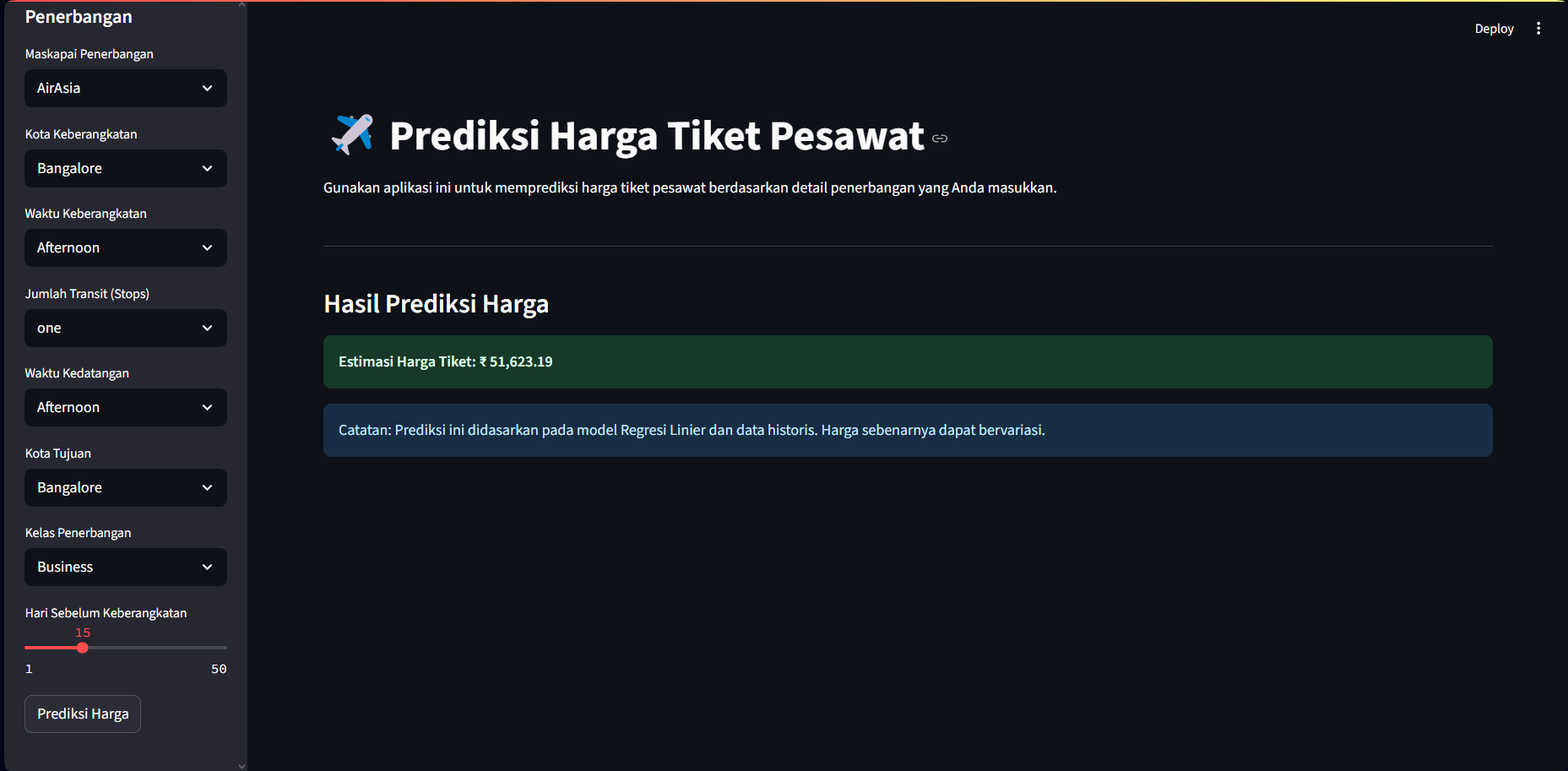
* Bahasa Pemrograman: Python
* Library: Pandas, Scikit-learn, NumPy, Matplotlib, Seaborn
* Tools Visualisasi: Jupyter Notebook

1. **Tahapan Penelitian**

Adapun tahapan penelitian dilakukan sebagai berikut:

* + **Preprocessing Data**: Membersihkan data, menghapus kolom tidak relevan, encoding fitur kategorik, dan normalisasi fitur numerik.
  + **Split Data**: Data dibagi menjadi data latih (train) dan data uji (test) dengan rasio umum 80:20.
  + **Pelatihan Model Linear Regression**: Melatih model regresi linear dengan data latih untuk mempelajari hubungan antara fitur input dan harga.
  + **Evaluasi Model**: Menggunakan metrik evaluasi seperti MAE, MSE, dan R² Score.
  + **Visualisasi Hasil**: Membandingkan hasil prediksi dan nilai aktual dalam grafik scatter atau garis.

1. **Parameter Pengujian**
   * **MAE (Mean Absolute Error)**: Rata-rata dari selisih nilai aktual dan prediksi.
   * **MSE (Mean Squared Error)**: Rata-rata dari kuadrat selisih nilai aktual dan prediksi.
   * **R² Score**: Menunjukkan seberapa baik model menjelaskan varians dari data target.
2. **Hasil Prediksi dengan Algoritma Linear Regression.**



berdasarkan input yang dimasukkan pada Streamlit diatas, berikut adalah hasil prediksi harganya menggunakan algoritma Regresi Linier.Setiap titik mewakili lokasi tempat wisata, dengan warna yang berbeda menunjukkan cluster yang berbeda.

**Detail Penerbangan Anda :**

* **Maskapai:** Air\_India
* **Rute:** Mumbai ke Delhi
* **Jumlah Transit:** Satu (one)
* **Kelas Penerbangan:** **Business**
* **Waktu Pembelian:** **15 hari** sebelum keberangkatan

**Hasil Prediksi Estimasi Harga Tiket: ₹ 51,623.19**

**Penjelasan Hasil**

Harga prediksi yang tinggi ini sangat masuk akal dan disebabkan oleh dua faktor utama:

1. **Kelas Penerbangan (Business):** Ini adalah faktor paling signifikan. Harga tiket kelas bisnis secara drastis lebih mahal dibandingkan kelas ekonomi.
2. **Waktu Pembelian (2 Hari):** Memesan tiket sangat dekat dengan tanggal keberangkatan (H-2) menyebabkan harga melonjak tinggi karena permintaan yang biasanya meningkat dan ketersediaan kursi yang menipis.

Kombinasi antara kemewahan **kelas bisnis** dengan urgensi **pemesanan mendadak** membuat model memprediksi harga di angka yang premium.

1. **Interpretasi Clustering**

Berdasarkan analisis hasil clustering, diperoleh interpretasi sebagai berikut:

|  |  |
| --- | --- |
| Aspek | Interpretasi |
| Tujuan Proyek | Membangun aplikasi untuk memprediksi harga tiket pesawat di India menggunakan algoritma Regresi Linier. |
| Faktor Harga Utama | **1. Kelas Penerbangan:** Pembeda harga paling signifikan (Bisnis vs. Ekonomi). **2. Waktu Pemesanan:** Harga naik drastis jika memesan dekat dengan tanggal keberangkatan. **3. Maskapai:** Ada hierarki harga yang jelas (Vistara termahal, AirAsia termurah). |
| Kinerja Model | **Baik dalam memahami tren umum** (misal: harga kelas bisnis > ekonomi), namun **kurang presisi** untuk prediksi angka yang akurat. |
| Visualisasi Kunci | **1. Bar Charts:** Menunjukkan perbandingan harga rata-rata antar maskapai dan kota keberangkatan. **2. Scatter Plot:** Mengonfirmasi bahwa model menangkap tren tetapi prediksinya sering meleset dari harga asli. |
| Cara Kerja Prediksi | Model menggabungkan "bobot" dari setiap input (maskapai, kelas, rute, waktu) untuk menghasilkan satu estimasi harga akhir. |
| Kesimpulan Akhir | Proyek berhasil sebagai demonstrasi, namun untuk akurasi yang lebih tinggi, diperlukan algoritma yang lebih canggih daripada Regresi Linier. |

1. **Pembahasan**

Berdasarkan hasil analisis dan pemodelan, dapat ditarik beberapa kesimpulan kunci mengenai faktor-faktor yang memengaruhi harga tiket pesawat domestik di India serta kinerja dari model Regresi Linier yang digunakan.

**Faktor Penentu Harga**

Analisis data menunjukkan bahwa beberapa faktor memiliki pengaruh signifikan terhadap harga tiket:

* **Kelas Penerbangan:** Merupakan faktor dengan dampak terbesar. Harga tiket kelas **Bisnis** secara konsisten jauh lebih mahal dibandingkan kelas **Ekonomi**.
* **Maskapai:** Terdapat stratifikasi harga yang jelas antar maskapai. Maskapai *full-service* seperti **Vistara** dan **Air India** menempati posisi harga teratas, sementara maskapai *low-cost* seperti **AirAsia** berada di posisi terbawah.
* **Waktu Pemesanan:** Harga tiket berbanding terbalik dengan jumlah hari tersisa sebelum keberangkatan. Pemesanan yang dilakukan mendekati tanggal penerbangan (misal, H-2) menunjukkan lonjakan harga yang signifikan.

**Evaluasi Kinerja Model**

Model Regresi Linier yang dibangun menunjukkan kinerja sebagai berikut:

* **Kelebihan:** Model berhasil menangkap **tren dan hubungan umum** dalam data. Model secara logis dapat memprediksi harga yang lebih tinggi untuk kelas bisnis atau pemesanan mendadak.
* **Kekurangan:** Akurasi prediksi model tergolong **rendah**. Grafik *scatter plot* antara harga aktual dan prediksi menunjukkan penyebaran data yang luas, yang mengindikasikan bahwa banyak prediksi yang nilainya jauh dari harga sebenarnya.

**KESIMPULAN**

Aplikasi yang dikembangkan berhasil memberikan estimasi kasar harga tiket yang logis berdasarkan faktor-faktor utama. Namun, algoritma Regresi Linier terbukti terlalu sederhana untuk menangani kompleksitas variabel dalam penetapan harga tiket pesawat. Untuk mendapatkan prediksi yang lebih akurat dan andal, disarankan untuk menggunakan algoritma yang lebih canggih seperti Gradient Boosting atau Random Forest pada penelitian selanjutnya.

**DAFTAR PUSTAKA**

1. Tan, P.-N., Steinbach, M., & Kumar, V. (2019). Introduction to Data Mining (2nd ed.).Pearson. <https://www.pearson.com/store/p/introduction-to-data-mining/P100000647643>
2. Han, J., Pei, J., & Kamber, M. (2021). Data Mining: Concepts and Techniques (4th ed.). Morgan Kaufmann. <https://www.elsevier.com/books/data-mining/han/978-0-12-381479-1>
3. Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow(2nded.).O’ReillyMedia. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
4. Chollet, F. (2021). Deep Learning with Python (2nd ed.). Manning Publications. <https://www.manning.com/books/deep-learning-with-python-second-edition>
5. Raschka, S., & Mirjalili, V. (2020). Python Machine Learning (3rd ed.). Packt Publishing. <https://www.packtpub.com/product/python-machine-learning-third-edition/9781789955750>
6. Aggarwal, C. C. (2020). Machine Learning for Text. Springer. <https://link.springer.com/book/10.1007/978-3-030-18305-9>
7. Zhang,Z.(2021).MachineLearning:Fundamentals.Springer. <https://link.springer.com/book/10.1007/978-3-030-71650-9>
8. Shalev-Shwartz, S., & Ben-David, S. (2020). Understanding Machine Learning: FromTheorytoAlgorithms.CambridgeUniversityPress. <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/>
9. Scikit-learn Developers. (2024). Scikit-learn Documentation. <https://scikit-learn.org>
10. Kaggle.(2025).FlightFarePredictionDataset. <https://www.kaggle.com/datasets/nikhilmittal/flight-fare-prediction-mh>
11. Alpaydin, E. (2021). Introduction to Machine Learning. MIT Press. <https://mitpress.mit.edu/9780262043793/introduction-to-machine-learning/>
12. Witten, I. H., Frank, E., & Hall, M. A. (2021). Data Mining: Practical Machine Learning Tools and Techniques. Elsevier. <https://www.elsevier.com/books/data-mining/witten/9780123748560>
13. Kurniawan, R., & Hartati, S. (2023). Prediksi harga e-commerce dengan regresi linier. Jurnal Teknologi dan Sistem Komputer, 11(1), 85–92. <https://doi.org/10.14710/jtsiskom.11.1.85-92>
14. Wahyuni, D., & Putra, F. (2024). Prediksi harga tiket pesawat dengan regresi. Jurnal Sistem dan Teknologi Informasi, 12(1), 43–52. <https://jsti.unib.ac.id/index.php/jsti/article/view/365>
15. Lubis, M. F., & Hanafiah, N. (2022). Regresi linier untuk prediksi harga hotel online. Jurnal Sains dan Informatika, 8(1), 60–68. <https://ejournal.unsri.ac.id/index.php/jsi/article/view/14864>
16. Putri, A. I., & Prasetyo, E. (2022). Prediksi harga dengan regresi linear pada data properti. Jurnal Sistem Informasi, 18(3), 212–220. <https://journal.uii.ac.id/JSI/article/view/21695>
17. Ramadhani, M., & Sari, Y. N. (2023). Penerapan Linear Regression untuk Prediksi Harga Tiket Kereta Api. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(1), 100–108. <https://jtiik.ub.ac.id/index.php/jtiik/article/view/3122>
18. Syahputra, R. H., & Nugroho, A. S. (2020). Prediksi Harga dengan Regresi pada Data E-Commerce. Jurnal Ilmiah Teknik Informatika, 9(2), 55–63. <https://journal.ittelkom-pwt.ac.id/index.php/jiti/article/view/104>
19. Hastie, T., Tibshirani, R., & Friedman, J. (2021). The Elements of Statistical Learning (2nd ed.). Springer. <https://link.springer.com/book/10.1007/978-0-387-84858-7>
20. Murtagh, F., & Devlin, K. (2020). Data Science Foundations. Springer. <https://link.springer.com/book/10.1007/978-3-030-22475-2>