

# Memprediksi Kenaikan Jumlah Penumpang Kereta Api Menggunakan Metode Regresi Linier Dengan Pemrograman Python

Lucky Marsella<sup>1\*</sup>, Nurmalitasari<sup>2</sup>

<sup>1,2</sup> Sistem Informasi/Illmu Komputer  
Universitas Duta Bangsa Surakarta

<sup>1\*</sup>202020346@mhs.udb.ac.id, <sup>2</sup>nurmalitasari@mhs.udb.ac.id

*Abstrak*— Penggunaan angkutan umum sangat bermanfaat bagi pemerintah untuk mengatasi masalah kemacetan lalu lintas yang merupakan masalah umum di Indonesia. Contoh angkutan umum yang disukai penduduk adalah kereta api. Tujuan dari penelitian ini adalah memprediksi laju pertumbuhan jumlah penumpang kereta api. Penelitian ini juga menggunakan metode regresi linier dengan menggunakan bahasa pemrograman Python. Data yang diambil dari website Badan Pusat Statistika (BPS) pada tahun 2022 mulai bulan Januari hingga Desember.

*Kata kunci*— Python, Regresi Linier, Prediksi, Kereta Api, Badan Pusat Statistika.

*Abstract*— The use of public transportation is highly beneficial for the government in addressing the common problem of traffic congestion in Indonesia. One example of public transportation favored by the population is the train. The objective of this research is to predict the growth rate of train passengers. This research also utilizes linear regression method using the Python programming language. The data is obtained from the website of the Central Statistics Agency (BPS) for the year 2022, spanning from January to December.

*Keywords*— Python, Linear Regression, Prediction, Train, Central Statistics Agency.

## I. PENDAHULUAN

### 1. Latar Belakang

Menggunakan transportasi umum, sangat membantu pemerintah dalam mengatasi masalah kemacetan yang banyak terjadi di Indonesia. Salah satu contoh transportasi umum yang banyak digemari oleh masyarakat adalah kereta. Masyarakat banyak yang memilih kereta untuk menjadi transportasi umum yang ditumpangi untuk berpergian baik sekedar jalan-jalan, mudik, bahkan bekerja.

Setiap bulan dan tahunnya terjadi kenaikan jumlah penumpang kereta. Maka dari itu, peneliti merasa tertarik untuk memprediksi tingkat peningkatan jumlah penumpang kereta. Menurut Badan Pusat Statistik (BPS) telah mencatat, bahwa jumlah penumpang kereta api Indonesia sebanyak 277,12 juta orang sepanjang 2022. Jumlah tersebut meningkat 85,04% dibandingkan pada tahun 2021 yang sebanyak 149,76 juta orang.

### 2. Tinjauan Pustaka

Transportasi dikatakan sebagai pemindahan barang dan manusia dari tempat asal ke tempat tujuan (Nasution, 1996). Dalam hubungan relasi ini ada tiga hal sebagai berikut: muatan yang diangkut, tersedia kendaraan sebagai alat pengangkutnya, dan ada jalanan yang bisa dilalui.

Transportasi merupakan perpindahan manusia atau barang ke tempat dimana barang dan manusia tersebut akan dibutuhkan (Bowersoz, 1981). Menurut Soejono (1991), transportasi menyebabkan adanya perpindahan dan pergerakan yang artinya terjadi sebuah lalu lintas.

### 3. Tujuan dan Sasaran Penelitian

Dengan demikian, dengan adanya penelitian ini peneliti ingin memprediksi tingkat kenaikan jumlah penumpang kereta menggunakan metode regresi linier dengan bahasa pemrograman python. Regresi linier

merupakan metode statistika untuk mencari hubungan antara variabel terikat ( $y$ ) dengan variabel bebas ( $x$ ). Regresi linier yaitu model analisis sederhana dengan jenis data interval atau rasio. Melalui analisis ini, peneliti dapat melakukan prediksi berdasarkan data-data ada.

## II. METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah regresi linier dan diimplementasikan menggunakan bahasa pemrograman python. Metode regresi linear sendiri sebuah metode dengan teknik analisis data yang memprediksi suatu nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui.

Secara matematis dimodelkan dengan variabel yang tidak diketahui atau tergantung dan variabel yang dikenal atau independen sebagai persamaan linier. Di bawah ini merupakan tahapan – tahapan penelitian yang dilakukan :

### 1. Penentuan Tujuan Penelitian:

Penelitian ini bertujuan untuk memprediksi tingkat kenaikan jumlah penumpang kereta menggunakan metode regresi linier dengan bahasa pemrograman python.

### 2. Pengumpulan Data:

Data penelitian diambil dari website Badan Pusat Statistik (BPS) untuk diprediksi tingkat kenaikan penumpang kereta api.

### 3. Pembersihan dan Penyediaan Data:

Data yang ada perlu dipreproses sebelum digunakan dalam analisis. Tahapan preprocessing meliputi pembersihan data dari nilai yang hilang atau tidak valid, pengubahan format data ke format yang sesuai, dan normalisasi data jika diperlukan.

### 4. Eksplorasi Data:

Analisis dan eksplorasi data untuk memahami hubungan antara variabel penjelas (misalnya, Kartu Elektronik, Kartu Multi Trip, Tiket Harian Berjaminan) dan variabel target (Jumlah Penumpang dan Rata – rata penumpang per hari).

### 5. Pembagian Data:

Bagi data menjadi dua bagian: data pelatihan (training data) dan data pengujian (testing data). Data pelatihan akan digunakan untuk mengembangkan model regresi linier, sedangkan data pengujian akan digunakan untuk menguji kinerja model.

### 6. Pembangunan Model:

Bahasa pemrograman Python digunakan untuk mengimplementasikan model regresi linier. Dapat menggunakan pustaka seperti scikit-learn atau statsmodels untuk membangun model tersebut.

### 7. Evaluasi Model:

Penelitian ini menggunakan data pengujian untuk mengevaluasi kinerja model. Hitung metrik evaluasi seperti mean squared error (MSE), mean absolute error (MAE), atau coefficient of determination (R-squared) untuk mengukur seberapa baik model dapat memprediksi jumlah penumpang.

## III. HASIL DAN PEMBAHASAN

Untuk membahas hasil prediksi kenaikan penumpang kereta api diperlukan informasi yang lebih tepat dan contoh implementasi yang lebih luas. Di sini saya memberikan contoh data untuk memberikan gambaran tentang cara menganalisis dan menginterpretasikan hasil prediksi.

Data ini saya diperoleh dari situs web <https://bogorkota.bps.go.id/statictable/2016/01/05/34/jumlah-penumpang-kereta-api-stasiun-bogor-2014.html>. Badan Pusat Statistik atau disingkat (BPS) yang mana merupakan lembaga pemerintah non-kementerian yang bertanggung jawab langsung kepada Presiden.

Saya memiliki data historis penumpang kereta api pada tahun 2022 mulai bulan Januari hingga Desember. Setelah membangun model regresi linier dan melakukan prediksi, berikut adalah hasil yang didapatkan:

a. Data yang Diperoleh

Mengambil data dari website Badan Pusat Statistika (BPS) yang disimpan dalam bentuk file excel. Total data yang ada sebanyak 60 data.

1	Bulan	Tiket Harian Berjaminan	Kartu Multi Trip	Kartu Elektronik	Jumlah Penumpang	Rata - rata penumpang per hari
2	Januari	708612	708612	708612	708612	708612
3	Februari	628874	443458	11654	1083986	38714
4	Maret	777888	478813	23130	1279831	41285
5	April	735678	468677	28940	1233295	41110
6	Mei	823673	461809	28735	1314217	42394
7	Juni	814748	476229	26570	1317547	43918
8	Juli	794375	418758	32594	1245727	40185
9	Agustus	811920	459387	42294	1313601	42374
10	September	697760	499499	44880	1242139	41405
11	Oktober	737714	521526	80384	1339624	43214
12	November	731505	49364	80181	1309050	43635
13	Desember	820130	492206	88568	1400904	45190

Gambar 1 data penumpang kereta api periode 2022

b. Pemrosesan Data

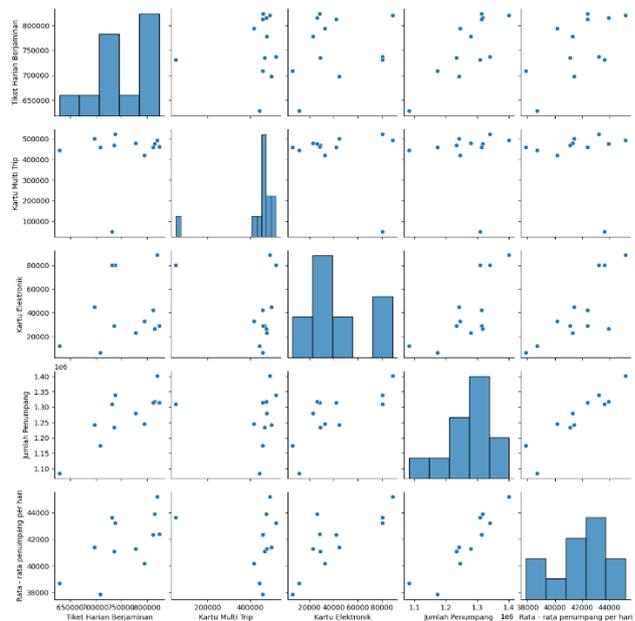
Setelah data diunduh dari website BPS maka selanjutnya data diproses agar kemudian dapat dianalisis. Pada proses ini data diubah ke dalam format yang sesuai agar data dapat dianalisis yang mendapatkan hasil yang bermutu dan data diubah hanya 5 bulan saja yg ditampilkan.

Bulan	Tiket Harian Berjaminan	Kartu Multi Trip	Kartu Elektronik	Jumlah Penumpang	Rata - rata penumpang per hari	
0	Januari	708612	459178	6234	1174024	37872
1	Februari	628874	443458	11654	1083986	38714
2	Maret	777888	478813	23130	1279831	41285
3	April	735678	468677	28940	1233295	41110
4	Mei	823673	461809	28735	1314217	42394

Gambar 2 data yang sudah diubah dan ditampilkan hanya bulan Januari-Mei

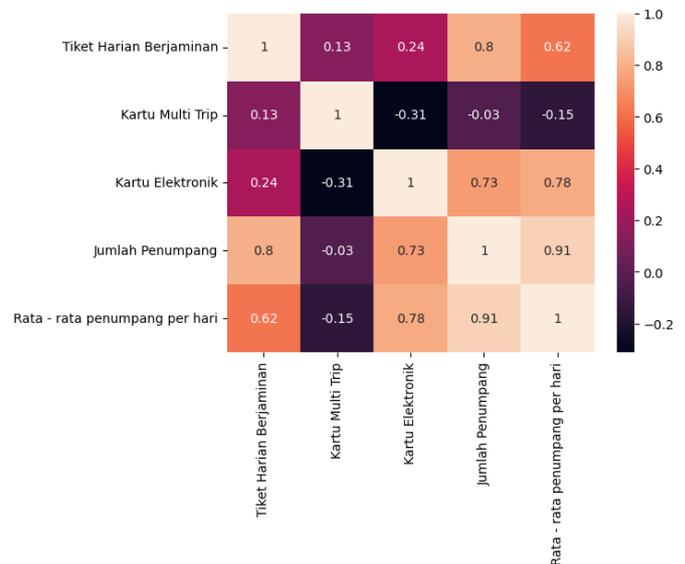
c. Data Dalam Bentuk Scatterplot

Scatterplot untuk masing masing variabel berfungsi untuk melihat korelasi beberapa data terhadap kolom poin untuk memudahkan dalam visualisasi dalam kasus ini saya menggunakan library dari seaborn agar lebih efektif dan efisien.



Gambar 3 data yang diubah ke dalam Scatterplot

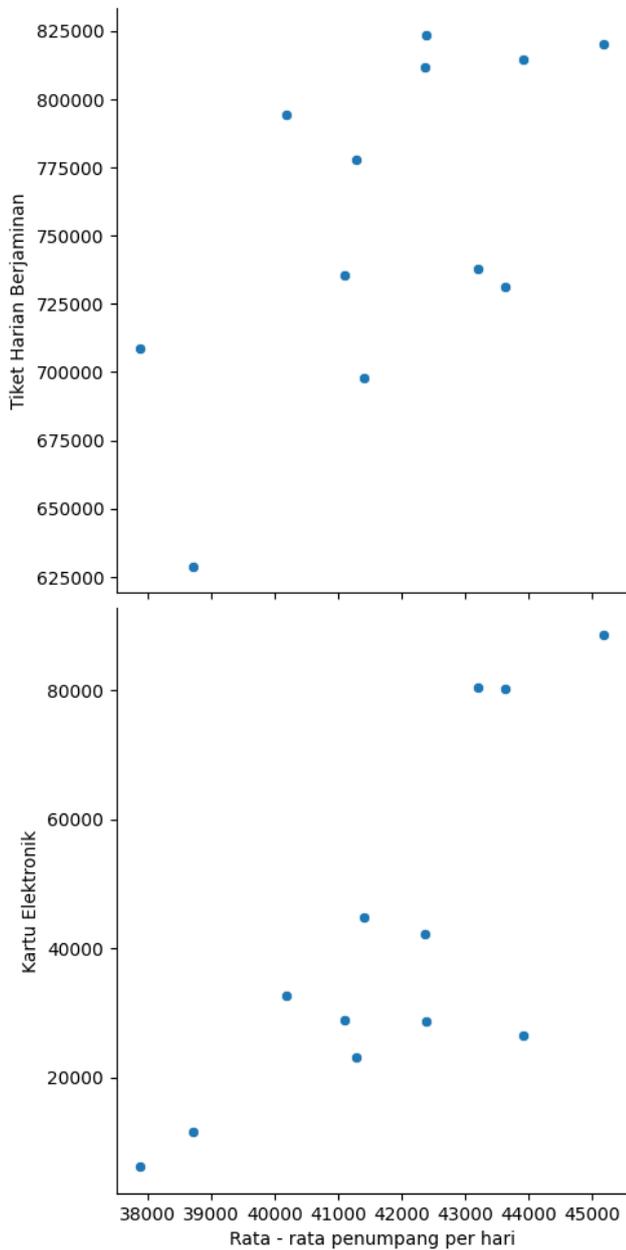
d. Adanya Korelasi Data



Gambar 4 heatmap yang mengindikasikan suatu korelasi

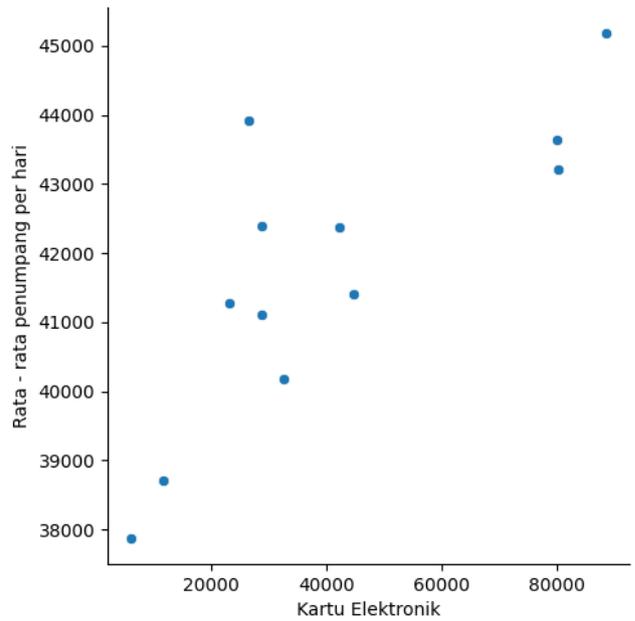
Setelah melihat heatmap diatas, yang memungkinkan untuk berkorelasi adalah warna cerah, karena semakin cerah akan mendekati 1, yang artinya sangat berkorelasi. Terdapat 2 korelasi yang lumayan cerah dengan tingkat korelasi 0,78 dan 0,91.

Kemudian didapatkan lagi korelasi yang lebih spesifik antara "Rata-rata penumpang per hari" dengan "Kartu Elektronik dan Tiket Harian Berjaminan".



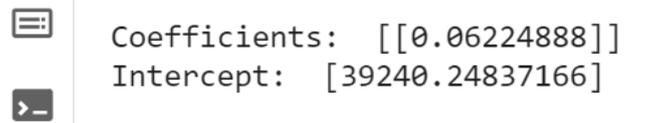
Gambar 5 korelasi antara "Rata-rata penumpang per hari" dengan "Kartu Elektronik dan Tiket Harian Berjaminan"

Selanjutnya saya memilih untuk mengambil data "Kartu Elektronik" dengan "Rata-rata penumpang per hari" untuk diuji seberapa kuat tingkat korelasi.



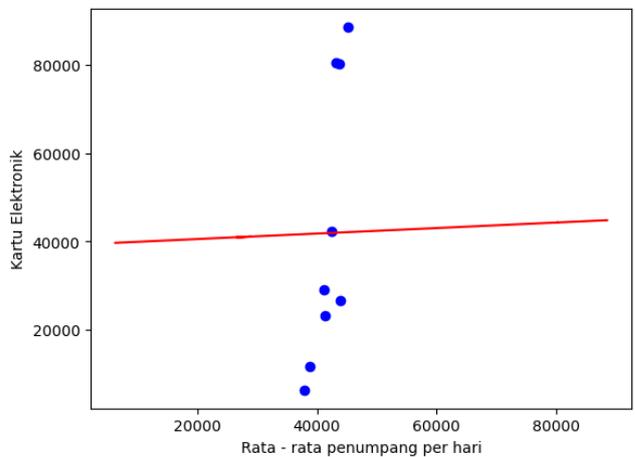
Gambar 6 korelasi antara "Kartu Elektronik" dengan "Rata-rata penumpang per hari"

#### e. Menganalisis Data



Gambar 7 hasil koefisien dan konstanta dari data "Rata-rata penumpang per hari" dengan "Kartu Elektronik"

#### f. Hasil Grafik Regresi Linier



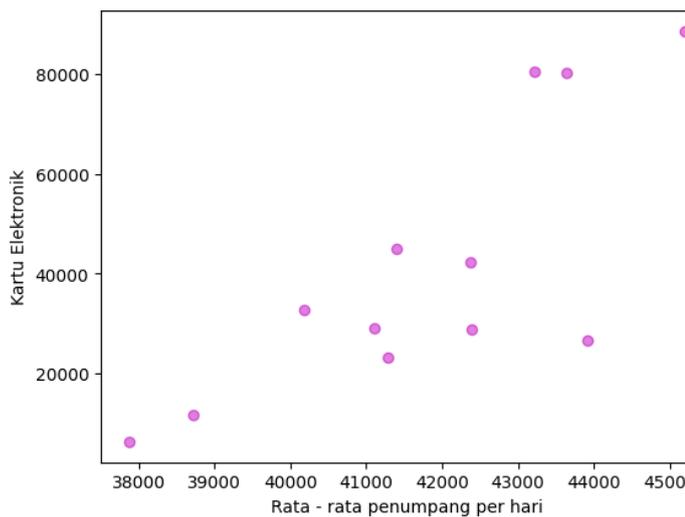
Gambar 8 hasil grafik regresi linier

g. Mode Evaluasi

```
Mean absolute error: 8451.43  
Residual sum of squares (MSE): 88609798.64  
R2-score: -28014.49
```

Gambar 9 Hasil pengujian MSE, MAE, R-Squared

h. Hasil Pengujian Regresi Nonlinier



Gambar 10 hasil pengujian regresi non linier

i. Hasil dari MSE, MAE, R-Squared.

```
Sigmoid  
Mean absolute error: 0.10  
Residual sum of squares (MSE): 0.01  
R2-score: -531.23
```

```
Exponens  
Mean absolute error: 0.13  
Residual sum of squares (MSE): 0.02  
R2-score: -1.41
```

```
Qubic  
Mean absolute error: 0.12  
Residual sum of squares (MSE): 0.02  
R2-score: -1.22
```

Gambar 11 hasil MSE, MAE, R- Squared

#### IV. KESIMPULAN

Untuk memberikan kesimpulan tentang prediksi kenaikan penumpang kereta api menggunakan regresi linier, diperlukan informasi tambahan berupa hasil analisis dan evaluasi model regresi linier yang telah dilakukan. Namun, secara umum, kesimpulan dapat dibuat berdasarkan evaluasi metrik evaluasi seperti MSE, MAE, dan R-squared, serta interpretasi koefisien regresi dan signifikansi statistiknya.

Jika MSE dan MAE rendah, artinya model memiliki tingkat kesalahan prediksi yang rendah, sehingga dapat dikatakan bahwa model secara umum cukup baik dalam memprediksi kenaikan penumpang kereta api. Selain itu, jika R-squared mendekati 1, hal ini menunjukkan bahwa variabilitas dalam data dapat dijelaskan dengan baik oleh model regresi linier.

#### REFERENSI

- [1] Rahayu, Y. (2014). Peramalan Jumlah Penumpang Kereta Api Menggunakan Regresi Linier dengan Metode Backward Elimination. *Jurnal Gaussian*, 3(4), 477-482.
- [2] Budi, A. S., & Ernawati, E. (2016). Analisis Regresi Linier pada Peramalan Penumpang KA Rongkasbitung-Tanahabang. *Jurnal Matematika*, 3(2), 126-134.
- [3] Sari, R. P., & Wahyuni, I. (2018). Peramalan Jumlah Penumpang Kereta Api Menggunakan Regresi Linier pada PT. Kereta Api Indonesia (Persero) Daerah Operasi III Palembang. *Jurnal Matematika*, 4(1), 59-70.
- [3] Widhiastuti, A. N., & Sukarno, S. (2019). Analisis Peramalan Jumlah Penumpang KA Sidoarjo-Jakarta dengan Metode Regresi Linier. *Jurnal Matematika*, 3(2), 27-36.
- [4] Yusuf, M. M., & Qadafi, M. (2020). Peramalan Penumpang Kereta Api Menggunakan Regresi Linier pada PT. Kereta Api Indonesia (Persero) Daerah Operasi VI Yogyakarta. *Jurnal Simetris*, 11(2), 567-574.