



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 3 (2025) pp: 986-993

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Penerapan Metode Regresi Linier untuk Prediksi Jumlah Penumpang Kereta Api

Septian Dwi Cahyo¹, Aang Alim Murtopo², Bayu Aji Santoso³

^{1,2}Teknik Informatika, STMIK YMI Tegal

³Sistem Informasi, STMIK YMI Tegal

cahyoseptiandwi18@gmail.com*, aang.alim@stmik-tegal.ac.id, bayu@stmik-tegal.ac.id

Abstrak

Transportasi kereta api memainkan peran penting dalam mobilitas masyarakat, namun perencanaan kapasitas yang tidak akurat sering menyebabkan inefisiensi operasional. Penelitian ini bertujuan untuk memprediksi jumlah penumpang kereta api pada masing-masing kelas layanan menggunakan metode regresi linier sederhana. Metode yang digunakan adalah pendekatan kuantitatif deskriptif-prediktif, dengan data sekunder dari PPID PT KAI tahun 2022–2024, dianalisis menggunakan Excel dan Python. Hasil penelitian menunjukkan bahwa model regresi linier mampu menangkap tren peningkatan jumlah penumpang pada hampir semua kelas layanan selama tahun 2025, dengan akurasi tertinggi pada kelas Lokal Eksekutif (MAE: 5.309) dan kesalahan terbesar pada kelas Ekonomi (MAE: 117.010), karena perbedaan volume penumpang yang signifikan. Implikasi dari penelitian ini menunjukkan bahwa pendekatan prediksi berdasarkan kelas layanan memberikan hasil yang lebih terperinci dan dapat digunakan sebagai acuan dalam perencanaan operasional yang berbasis data, serta membuka peluang pengembangan model dengan pendekatan machine learning yang lebih kompleks di masa mendatang.

Kata kunci: Data Historis, Evaluasi Akurasi Model, Prediksi Jumlah Penumpang, Regresi Linier, Transportasi Kereta Api

1. Latar Belakang

Transportasi kereta api merupakan salah satu moda transportasi massal yang memiliki peran vital dalam mendukung mobilitas masyarakat dan pertumbuhan ekonomi nasional [1]. Dengan meningkatnya aktivitas sosial, ekonomi, dan urbanisasi, kebutuhan masyarakat terhadap layanan kereta api menunjukkan tren yang semakin kompleks dan dinamis. Perencanaan operasional yang tepat berbasis data menjadi sangat penting, khususnya dalam mengelola jumlah penumpang secara efisien dan adaptif [2].

Salah satu tantangan yang dihadapi operator kereta api adalah ketidakpastian jumlah penumpang harian, yang dapat berdampak langsung terhadap ketidakseimbangan kapasitas, pemborosan sumber daya, hingga penurunan kualitas pelayanan [3]. Ketika jumlah penumpang melebihi kapasitas atau justru jauh di bawah proyeksi, kerugian baik dari sisi ekonomi maupun kepuasan pelanggan tidak dapat dihindari [4]. Oleh karena itu, diperlukan sistem prediksi yang mampu mengestimasi jumlah penumpang secara akurat untuk setiap periode, dengan mempertimbangkan karakteristik masing-masing kelas layanan kereta [5].

Penelitian sebelumnya telah mengkaji metode prediksi jumlah penumpang dengan pendekatan statistik. Penelitian yang menerapkan metode simulasi Monte Carlo untuk memproyeksikan volume penumpang harian di Stasiun Gubeng berdasarkan variasi hari dalam seminggu [6]. Sementara itu, penelitian yang memanfaatkan Regresi Linier sederhana untuk memprediksi jumlah penumpang selama periode arus mudik Lebaran di pelabuhan, dengan mempertimbangkan fluktuasi musiman [7]. Studi lain mengembangkan model Regresi Linier Berganda dengan menambahkan variabel eksternal seperti jumlah hari libur, harga tiket, dan frekuensi keberangkatan [8]. Kemudian pada penelitian yang menerapkan metode Regresi Linier Berganda di PT. Rajawali Citra Transportasi Medan untuk estimasi jumlah penumpang secara oprasional [9]. Meskipun pendekatan-pendekatan tersebut memberikan gambaran yang bermanfaat dalam konteks prediksi permintaan transportasi, sebagian besar masih bersifat agregatif dan belum mempertimbangkan segmentasi prediksi berdasarkan kelas layanan kereta api secara spesifik. Selain itu, penerapan metrik evaluasi kuantitatif seperti *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)* juga belum banyak digunakan secara sistematis untuk menilai performa prediktif model-model tersebut.

Penelitian ini dilakukan untuk menjawab kebutuhan tersebut melalui penerapan metode regresi linier sederhana, yang bertujuan untuk memodelkan hubungan antara waktu (bulan) sebagai variabel independen dengan jumlah penumpang sebagai variabel dependen pada masing-masing kelas, seperti Eksekutif, Bisnis, Ekonomi, Lokal Eksekutif, Lokal Bisnis, dan Lokal Ekonomi. Data yang digunakan merupakan data sekunder yang diperoleh dari PPID PT Kereta Api Indonesia (Persero) Bandung untuk periode 1 Januari 2022 hingga 31 Desember 2024, yang mencakup informasi jumlah penumpang berdasarkan kelas kereta secara harian. Dataset tersebut kemudian diproses dan dianalisis untuk membentuk model prediktif yang valid dan kontekstual terhadap tren perjalanan kereta api.

Permasalahan yang ingin dipecahkan dalam penelitian ini adalah bagaimana membangun model prediksi jumlah penumpang yang akurat dengan regresi linier, serta mengukur kinerja model tersebut menggunakan tiga metrik evaluasi umum yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*.

Penelitian ini dilakukan untuk mengisi celah (*research gap*) dari studi-studi sebelumnya yang hanya menggunakan prediksi total penumpang tanpa mempertimbangkan diferensiasi kelas layanan atau validasi model secara menyeluruh. Mengadopsi pendekatan berbasis *machine learning* supervisi sederhana melalui regresi linier, serta dengan dukungan teknologi komputasi *Python* dan *Visual Studio Code*, penelitian ini memberikan kontribusi nyata terhadap pengembangan sistem prediksi berbasis data yang dapat diimplementasikan langsung oleh pihak operator.

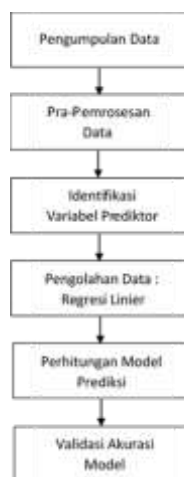
Inovasi yang diusulkan dalam studi ini adalah pemisahan model prediksi per kelas layanan kereta api serta implementasi perhitungan model ganda baik secara manual melalui *Microsoft Excel* maupun otomatis menggunakan pemrograman *Python* untuk memastikan konsistensi hasil dan validitas model.

Penelitian ini bertujuan untuk membangun model prediksi jumlah penumpang berdasarkan regresi linier sederhana untuk masing-masing kelas layanan, mengevaluasi akurasi prediksi menggunakan MAE, MSE, dan RMSE, serta memberikan rekomendasi aplikatif bagi perencanaan transportasi publik yang berbasis data. Hasil penelitian ini dapat dijadikan dasar untuk pengembangan sistem pendukung keputusan dalam manajemen operasional PT KAI Bandung.

2. Metode Penelitian

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang bertujuan untuk menggambarkan kondisi aktual dari data historis penumpang kereta api serta membangun model prediksi menggunakan metode regresi linier sederhana. Model ini memanfaatkan hubungan antara waktu (variabel independen) dengan jumlah penumpang (variabel dependen) untuk tiap kelas layanan. Analisis dilakukan secara terpisah pada masing-masing kelas: Eksekutif, Bisnis, Ekonomi, Lokal Eksekutif, Lokal Bisnis, dan Lokal Ekonomi.



Gambar 1. Proses Analisis Data

2.2. Pengumpulan Data

Data yang digunakan merupakan data sekunder yang diperoleh melalui permohonan resmi dari laman E-PPID PT Kereta Api Indonesia (Persero) Bandung. Dataset mencakup jumlah penumpang harian dari tanggal 1 Januari 2022

hingga 31 Desember 2024 untuk masing-masing kelas layanan. Masing-masing *record* mencakup atribut tanggal, hari, kelas, dan volume penumpang. Data tahun 2024 digunakan sebagai basis utama untuk pembuatan model, dan prediksi dilakukan untuk tahun 2025 (bulan ke-13 hingga ke-24).

Tabel 1. Dataset Jumlah Penumpang Kereta Api Tahun 2024

HARI	TANGGAL	KELAS						TOTAL VOLUME
		EKS	BIS	EKO	LOK EKS	LOK BIS	LOK EKO	
Senin	01/01/2024	42.512	8.377	116.521	2.022	0	28.023	197.455
Selasa	02/01/2024	49.793	9.259	118.012	1.767	0	24.260	203.091
Rabu	03/01/2024	39.329	7.391	96.342	1.230	0	22.932	167.224
Kamis	04/01/2024	31.760	6.346	83.285	1.029	0	21.020	143.440
Jumat	05/01/2024	30.565	5.522	80.792	1.206	0	21.203	139.288
Sabtu	06/01/2024	35.810	6.271	89.615	1.549	0	24.492	157.737
⋮					⋮			⋮
⋮					⋮			⋮
⋮					⋮			⋮
Kamis	24/10/2024	28.148	4.041	70.940	509	376	14.551	118.565
Jumat	25/10/2024	36.084	4.552	87.977	1.278	388	16.403	146.682
Sabtu	26/10/2024	38.883	4.715	93.890	1.601	383	20.715	160.187
Minggu	27/10/2024	34.705	5.117	93.823	1.741	417	23.379	159.182
Senin	28/10/2024	39.247	5.075	92.813	1.157	383	15.939	154.614
Selasa	29/10/2024	25.316	3.977	69.943	403	327	13.334	113.300

Tabel 1. menampilkan data jumlah penumpang dari tahun 2024, data tahun 2024 akan digunakan sebagai data acuan untuk menghasilkan prediksi penumpang kereta di tahun 2025. Data yang diperoleh terdapat 9 atribut seperti atribut HARI yang mewakili nama hari, kemudian ada atribut TANGGAL yang mewakili tanggal, bulan dan tahun, selanjutnya ada atribut KELAS yang memiliki 6 jenis kelas yaitu EKS untuk kelas Eksekutif, EKO untuk Kelas Ekonomi, BIS untuk kelas Bisnis, kemudian LOK EKS untuk kelas Lokal Eksekutif, LOK EKO untuk kelas Lokal Ekonomi dan LOK BIS untuk Lokal Bisnis.

2.3. Pra-pemrosesan Data

Langkah ini mencakup penghapusan data duplikat, penanganan *missing values* dengan rata-rata, serta agregasi data harian menjadi data bulanan. Hasilnya adalah dataset yang bersih, rapi, dan siap digunakan untuk proses pemodelan.

2.4. Implementasi Regresi Linier

Regresi linier memodelkan hubungan variabel independen dan variabel dependen, proses prediksi regresi linier dengan memasukkan nilai-nilai variabel independen ke persamaan regresi [10], [11]. Pada penelitian ini, regresi linier yang akan digunakan adalah regresi linier sederhana. Regresi linier sederhana menggunakan satu variabel dependen (Y) dan satu variabel independen (X) [12]. Rumus regresi linier dapat dituliskan pada persamaan (1)

$$Y = a + bX \tag{1}$$

Y adalah variabel dependen, a adalah konstanta, X adalah variabel independen dan b adalah koefisien Regresi. Pada penelitian ini yang variabel dependen (Y) adalah jumlah penumpang kereta api, jumlah penumpangnya diambil dari jumlah penumpang berdasarkan kelas gerbong kereta untuk bulan tertentu. Sedangkan yang menjadi variabel independen (X) adalah waktu yang berupa bulan. Rumus mencari konstanta *a* dapat dituliskan pada persamaan (2)

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \tag{2}$$

$\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data. Rumus mencari koefisien b dapat dituliskan pada persamaan (3)

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (3)$$

$\sum y$ adalah total variabel dependen, $(\sum x^2)$ adalah hasil pangkat dua variabel independen, $\sum x$ adalah total variabel independen, $\sum xy$ adalah total hasil perkalian antara variabel dependen dan variabel independen, n adalah jumlah data.

2.5. Metrik Evaluasi

Metrik evaluasi memberikan gambaran mengenai performa model prediksi dalam menangkap pola dan hubungan yang ada dalam data [13]. Penelitian ini menggunakan tiga metrik evaluasi *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*.

Mean Absolute Error (MAE) memberikan gambaran tentang seberapa besar kesalahan prediksi secara keseluruhan [11]. Rumus *Mean Absolute Error (MAE)* dapat dituliskan pada persamaan (4)

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (4)$$

n adalah jumlah sampel dalam data, y^i adalah aktual dan \hat{y}^i adalah nilai prediksi.

Mean Squared Error (MSE) memberikan bobot lebih besar pada kesalahan yang lebih besar [14]. Rumus *Mean Squared Error (MSE)* dapat dituliskan pada persamaan (5)

$$MSE = \sum \frac{(\hat{y} - y)^2}{n} \quad (5)$$

n adalah jumlah sampel dalam data, y^i adalah aktual dan \hat{y}^i adalah nilai prediksi.

Root Mean Squared Error (RMSE) memiliki interpretasi yang serupa dengan MSE, namun lebih mudah dipahami karena berada dalam satuan yang sama dengan data asli [15]. Rumus *Root Mean Squared Error (RMSE)* dapat dituliskan pada persamaan (6)

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (6)$$

n adalah jumlah data, i adalah urutan data pada database, y_i adalah aktual dan \hat{y}_i adalah nilai prediksi.

3. Hasil dan Diskusi

Sebelum dilakukan proses pemodelan, data melalui tahap pra-pemrosesan guna memastikan kualitas dan konsistensi data yang digunakan. Tahapan ini meliputi pemeriksaan data duplikat, penanganan nilai kosong, serta agregasi data harian menjadi data bulanan. Hasil dari pra-pemrosesan ini menjadi dasar dalam membangun model regresi linier pada masing-masing kelas layanan kereta api.

Tabel 2. Data setelah Pra Pemrosesan Data

Bulan	EKS	BIS	EKO	LOK EKS	LOK BIS	LOK EKO
Januari	948.135	144.176	2.312.745	27.354	0	535.250
Februari	954.246	138.589	2.278.201	28.502	0	521.035
Maret	838.998	120.600	2.042.541	25.166	0	463.658
April	1.313.826	215.214	3.215.226	44.064	0	628.850
Mei	1.042.646	153.274	2.533.240	36.939	0	601.596
Juni	1.117.387	165.919	2.688.739	38.618	0	560.373
Juli	1.176.864	170.592	2.747.838	37.888	1.210	589.523
Agustus	1.028.623	143.137	2.432.010	30.794	10.743	466.806
September	1.011.733	133.237	2.421.370	32.144	10.758	486.720
Oktober	999.134	132.384	2.473.632	30.529	10.811	512.368
November	1.029.574	148.704	2.477.272	32.783	3.483	526.039

Bulan	EKS	BIS	EKO	LOK EKS	LOK BIS	LOK EKO
Desember	1.074.185	155.295	2.586.082	34.512	3.980	545.447

Tabel 2. menampilkan hasil pra-pemrosesan data jumlah penumpang kereta api tahun 2024 yang telah dikelompokkan per bulan dan per kelas kereta, mencakup Kelas Eksekutif, Kelas Bisnis, Kelas Ekonomi, Kelas Lokal Eksekutif, Kelas Lokal Bisnis, dan Kelas Lokal Ekonomi. Serta telah dilakukan proses penghapusan data duplikat, penanganan *missing values* dengan rata-rata.

Data yang telah melalui tahap pra pemrosesan data, kemudian di olah menggunakan regresi linier. Proses ini dilakukan melalui dua pendekatan, yaitu perhitungan secara manual menggunakan Microsoft Excel dan secara otomatis menggunakan bahasa pemrograman *Python*.

3.1 Hasil Analisis dengan Microsoft Excel

Model regresi linier sederhana dihitung secara manual untuk tiap kelas dengan menggunakan rumus dasar regresi $Y = a + bX$.

Tabel 3. Perhitungan Kelas Eksekutif

Bulan	x	Penumpang (y)	x*y	x ²
Januari	1	948.135	948.135	1
Februari	2	954.246	1.908.492	4
Maret	3	838.998	2.516.994	9
April	4	1.313.826	5.255.304	16
Mei	5	1.042.646	5.213.230	25
Juni	6	1.117.387	6.704.322	36
Juli	7	1.176.864	8.238.048	49
Agustus	8	1.028.623	8.228.984	64
September	9	1.011.733	9.105.597	81
Oktober	10	999.134	9.991.336	100
November	11	1.029.574	11.325.313	121
Desember	12	1.074.185	12.890.219	144
Total	78	12.535.351	82.325.975	650

Tabel 3 menunjukkan hasil perhitungan awal pada kelas Eksekutif, yang mencakup data bulan ke-1 hingga ke-12 (Januari–Desember), jumlah penumpang per bulan (y), hasil perkalian antara bulan dan jumlah penumpang (x*y), serta kuadrat dari nilai bulan (x²).

Kemudian mencari mencari nilai konstanta *a* menggunakan rumus pada persamaan (2)

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{(12.535.351 * 650) - (78 * 82.325.975)}{12(650) - (78)^2}$$

$$a = \frac{(8.147.978.150) - (6.421.426.050)}{7.800 - 6.084}$$

$$a = 1.006.149$$

Setelah mendapatkan nilai konstanta *a*, kemudian mencari nilai Koefisien regresi *b* menggunakan rumus pada persamaan (3)

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b = \frac{12(82.325.975) - (78 * 12.535.351)}{12(650) - (78)^2}$$

$$b = \frac{(987.911.700) - (977.757.378)}{7.800 - 6.084}$$

$$b = 5.917$$

Setelah mendapatkan nilai konstanta *a* dan nilai Koefisien regresi *b*, kemudian implementasikan ke dalam rumus umum regresi linier.

$$Y = a + bX$$

$$Y = 1.006.149 + 5.917 (13)$$

$$Y = 1.083.076$$

Hasil tersebut adalah prediksi penumpang kelas Eksekutif untuk bulan Januari, untuk menghitung prediksi bulan selanjutnya, hanya mengganti nilai di variabel independennya. Pada rumus umum regresi linier, variabel dependennya (*X*) di isi dengan angka 13 karena akan memprediksi bulan Januari di tahun 2025.

Tabel 4. Hasil Prediksi Kelas Eksekutif

Bulan	Prediksi 2025
Februari	1.088.993
Maret	1.094.911
April	1.100.828
Mei	1.106.746
Juni	1.112.663
Juli	1.118.581
Agustus	1.124.498
September	1.130.416
Oktober	1.136.333
November	1.142.251
Desember	1.148.168

Tabel 4. menunjukkan nilai hasil prediksi bulan Januari 2025 sebesar 1.083.076 telah dihitung sebelumnya, sedangkan prediksi bulan Februari hingga Desember diperoleh dengan memasukkan nilai *X* = 14 hingga *X* = 24 ke dalam persamaan regresi. Hasilnya menunjukkan tren peningkatan jumlah penumpang dari bulan ke bulan sepanjang tahun 2025. Jadi, untuk kelas lainnya cukup mengikuti langkah yang sama, hanya perlu menyesuaikan nilai *a*, *b* dan angka pada variabel *X* sesuai data masing-masing.

Tabel 5. Prediksi Jumlah Penumpang per Kelas Tahun 2025

Bulan	EKS	BIS	EKO	LOK EKS	LOK BIS	LOK EKO
Januari	1.083.076	148.583	2.599.274	35.003	8.825	523.115
Februari	1.088.993	148.094	2.611.869	35.268	9.657	521.060
Maret	1.094.911	147.606	2.624.464	35.534	10.490	519.005
April	1.100.828	147.117	2.637.058	35.800	11.322	516.950
Mei	1.106.746	146.628	2.649.653	36.066	12.154	514.895
Juni	1.112.663	146.139	2.662.248	36.332	12.986	512.841
Juli	1.118.581	145.651	2.674.843	36.598	13.819	510.786
Agustus	1.124.498	145.162	2.687.438	36.864	14.651	508.731
September	1.130.416	144.673	2.700.032	37.130	15.483	506.676
Oktober	1.136.333	144.184	2.712.627	37.396	16.315	504.621
November	1.142.251	143.696	2.725.222	37.661	17.148	502.566
Desember	1.148.168	143.207	2.737.817	37.927	17.980	500.511

Tabel 5. menyajikan hasil prediksi enam kategori kelas transportasi, yaitu Eksekutif (EKS), Bisnis (BIS), Ekonomi (EKO), Lokal Eksekutif (LOK EKS), Lokal Bisnis (LOK BIS), dan Lokal Ekonomi (LOK EKO). Setiap kategori menunjukkan estimasi jumlah penumpang bulanan dari Januari hingga Desember 2025 berdasarkan analisis tren data tahun sebelumnya.

3.2 Hasil Analisis dengan Python

Melalui Python, data ditransformasikan ke dalam DataFrame dan diproses menggunakan pendekatan regresi linier dengan iterasi per kelas. Model yang dibentuk menghasilkan prediksi jumlah penumpang tahun 2025 yang konsisten dengan hasil dari Excel.

Prediksi Jumlah Penumpang per Kelas Tahun 2025								
BULAN	EKS	BIS	EKO	LOK EKS	LOK BIS	LOK EKO		
Januari	1.083.076	148.583	2.599.274	35.003	8.825	523.115		
Februari	1.088.994	148.094	2.611.869	35.269	9.658	521.060		
Maret	1.094.911	147.605	2.624.464	35.534	10.490	519.006		
April	1.100.829	147.117	2.637.059	35.800	11.322	516.951		
Mei	1.106.746	146.628	2.649.653	36.066	12.154	514.896		
Juni	1.112.664	146.139	2.662.248	36.332	12.987	512.841		
Juli	1.118.581	145.650	2.674.843	36.598	13.819	510.786		
Agustus	1.124.498	145.162	2.687.438	36.864	14.651	508.731		
September	1.130.416	144.673	2.700.033	37.130	15.484	506.676		
Oktober	1.136.333	144.184	2.712.627	37.396	16.316	504.621		
November	1.142.251	143.695	2.725.222	37.662	17.148	502.566		
Desember	1.148.168	143.206	2.737.817	37.927	17.980	500.512		

Gambar 2. Hasil Prediksi Menggunakan Python

Gambar 2. menunjukkan hasil akhir dari implementasi regresi linier menggunakan Python. Output yang ditampilkan merupakan tabel prediksi jumlah penumpang per kelas untuk tahun 2025. Hasil prediksi memperlihatkan tren peningkatan jumlah penumpang di hampir seluruh kelas layanan sepanjang tahun, dengan kelas Ekonomi dan Eksekutif mencatat volume tertinggi, sedangkan LOK BIS tetap menjadi kelas dengan jumlah penumpang terendah.

Setelah memperoleh hasil prediksi jumlah penumpang untuk masing-masing kelas layanan menggunakan regresi linier, langkah selanjutnya melakukan evaluasi terhadap performa model untuk menilai seberapa baik model dalam merepresentasikan data historis dan menghasilkan prediksi yang mendekati kondisi aktual. Evaluasi dilakukan menggunakan tiga metrik umum, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*.

Tabel 6. Evaluasi Model

KELAS	MAE	MSE	RMSE
EKS	117010,2574	17918692448	133860,7203
BIS	16895,02139	581593412,4	24116,24789
EKO	264079,7233	97537029577	312309,1891
LOK EKS	5309,335227	36003713,68	6000,309465
LOK BIS	9987,186424	111279737,5	10548,92115
LOK EKO	44201,09222	3041174785	55146,84746

Tabel 6. Menunjukkan hasil validasi akurasi model regresi linier yang dihitung secara manual menggunakan Microsoft Excel menunjukkan bahwa kelas Ekonomi (EKO) memiliki tingkat error tertinggi dengan MAE sebesar 117.010, MSE sebesar 1,79 miliar, dan RMSE sebesar 133.860, yang wajar mengingat volume penumpangnya paling besar. Sebaliknya, kelas Lokal Eksekutif (LOK EKS) menunjukkan akurasi terbaik dengan MAE hanya 5.309 dan RMSE sebesar 6.000. Kelas Bisnis (BIS) berada pada tingkat error menengah dengan MAE 16.895, sedangkan LOK BIS dan LOK EKO menunjukkan error moderat dengan MAE masing-masing 9.987 dan 44.201.

Hasil penelitian menunjukkan bahwa penerapan regresi linier sederhana mampu memodelkan tren jumlah penumpang kereta api secara efektif untuk masing-masing kelas layanan. Proses pemodelan dilakukan melalui dua pendekatan, yaitu secara manual menggunakan Microsoft Excel dan otomatis menggunakan Python, dengan hasil

yang konsisten di antara keduanya. Prediksi jumlah penumpang untuk tahun 2025 menunjukkan tren peningkatan pada hampir seluruh kelas, terutama pada kelas Eksekutif dan Ekonomi yang mencatat volume tertinggi. Sementara itu, kelas Lokal Bisnis menunjukkan jumlah penumpang terendah, tetapi tetap mengalami kenaikan dari bulan ke bulan. Evaluasi model menggunakan MAE, MSE, dan RMSE menunjukkan bahwa kelas Lokal Eksekutif memiliki tingkat akurasi tertinggi dengan MAE terendah sebesar 5.309, sedangkan kelas Ekonomi mencatat kesalahan prediksi tertinggi karena volume penumpangnya yang besar dan fluktuatif. Temuan ini mengindikasikan bahwa model regresi linier lebih akurat untuk kelas dengan pola data yang stabil. Secara keseluruhan, pendekatan pemodelan per kelas memberikan hasil prediksi yang lebih terperinci dan dapat digunakan sebagai dasar pengambilan keputusan dalam perencanaan operasional transportasi publik berbasis data.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode Regresi Linier sederhana efektif digunakan untuk memprediksi jumlah penumpang kereta api pada masing-masing kelas layanan. Prediksi yang dihasilkan menunjukkan tren peningkatan penumpang secara konsisten pada tahun 2025, dengan tingkat akurasi tertinggi pada kelas Lokal Eksekutif dan error terbesar pada kelas Ekonomi. Pendekatan per kelas terbukti memberikan hasil yang lebih terperinci dan informatif dibandingkan model agregat. Untuk penelitian selanjutnya, disarankan agar ditambahkan variabel eksternal seperti hari libur, harga tiket, dan faktor musiman, serta mempertimbangkan penggunaan metode machine learning lain yang lebih kompleks seperti *Random Forest* atau *Decision Tree* untuk meningkatkan akurasi model prediksi.

Referensi

- [1] M. Firdausi, D. Firman, and Y. Putra, "ANALISIS PEMILIHAN MODA TRANSPORTASI UMUM ANTARA BUS DAN KERETA API TRAYEK KOTA SURABAYA-KOTA YOGYAKARTA," *Jurnal Rekayasa Teknik Sipil Universitas Madura*, vol. 6, no. 2, 2021, doi: 10.53712/rjrs.v6i2.1447.
- [2] Musakirawati, Jemmy, F. Anngriawan, A. F. Triansyah, Akib, and A. Tahir, "PEMANFAATAN PLATFORM RAPOR PENDIDIKAN INDONESIA TERHADAP PERENCANAAN BERBASIS DATA," *Jurnal Dinamika Manajemen Pendidikan*, vol. 7, Apr. 2023, doi: 10.26740/jdmp.v7n2.p201-10.
- [3] T. Ade Amelia and I. Muslim Karo Karo, "ANALISIS PREDIKSI JUMLAH PENUMPANG KERETA API MENGGUNAKAN METODE MONTE CARLO (STUDI KASUS: PT.KAI WILAYAH SUMATERA BARAT)," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 4, 2024, doi: 10.36040/jati.v8i4.9935.
- [4] M. Y. Zidan and A. Susanti, "Redesain Stasiun Indro dan Stasiun Kandungan dengan Peningkatan Fasilitas Pelayanan Penumpang Kereta Api," *Jurnal Media Publikasi Terapan Transportasi*, vol. 2, no. 3, p. 313, 2024, doi: 10.26740/mitrans.v2n3.p313-325.
- [5] A. Setia Budi and P. H. Susilo, "Sistem Prediksi Jumlah Penumpang di Bandar Udara Juanda Surabaya dengan Metode Double Exponential Smoothing," *Generation Journal*, vol. 6, no. 1, pp. 2580–4952, 2022, doi: 10.29407/gj.v6i1.16668.
- [6] A. K. Bayu Viargo, T. Saifudin, and N. Chamidah, "Prediksi Jumlah Penumpang Kereta Api Stasiun Surabaya Gubeng dengan Metode Monte Carlo," *Limits: Journal of Mathematics and Its Applications*, vol. 20, no. 3, p. 275, Nov. 2023, doi: 10.12962/limits.v20i3.16123.
- [7] S. Amelliah, R. Kusumawati, and F. Fatimah, "PREDIKSI JUMLAH PENUMPANG LEBARAN PELABUHAN TANJUNG PERAK MENGGUNAKAN REGRESI LINIER," *Jurnal Pembangunan Daerah*, vol. 2, no. 1, 2023, doi: 10.62389/bina.v2i1.49.
- [8] Y. Goktua Siadari and D. Saripuna, "DATA MINING UNTUK MENGESTIMASI JUMLAH PENUMPANG PADA PT. PINEM LAU GUNA MEDAN DENGAN MENGGUNAKAN METODE REGRESI LINEAR BERGANDA," *Jurnal CyberTech*, vol. x, No.x, 2020, doi: 10.53513/jct.v4i2.1998.
- [9] E. G. Bangun, Y. H. Syahputra, and R. I. Ginting, "Data Mining Untuk Mengestimasi Jumlah Penumpang Pada PT Rajawali Citra Transportasi Medan Dengan Menggunakan Metode Regresi Linear Berganda," 2020. doi: 10.53513/jct.v4i1.2062.
- [10] A. Adnan Rusdy, "Penerapan Metode Regresi Linear pada Prediksi Penawaran dan Permintaan Obat Studi Kasus Aplikasi Point of Sales," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 3, no. 2, pp. 121–126, 2022, doi: 10.33096/busiti.v3i2.1130.
- [11] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear Studi Kasus pada Industri Otomotif di Indonesia," *Innovative: Journal of Social Science Research*, vol. 4, no. 4, 2024, doi: 10.31004/innovative.v4i4.12735.
- [12] A. Arisandi and U. Ependi, "Analisis Peramalan Penjualan Produk Pada PT.Enseval Putera Megatrading TBK Menggunakan Metode Regresi Linear Sederhana," *Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 15, no. 1b, p. 317, 2023, doi: 10.5281/5318/15.jupiter.2023.04.
- [13] R. N. Silalahi and M. Muljono, "Perbandingan Kinerja Metode Linear Regression, LSTM dan GRU Untuk Prediksi Harga Penutupan Saham Coco-Cola," *Komputika: Jurnal Sistem Komputer*, vol. 13, no. 2, pp. 201–211, Oct. 2024, doi: 10.34010/komputika.v13i2.12265.
- [14] M. F. Almaliki, I. Isnawaty, M. Satyadharma, and H. Hado, "Perbandingan Metode Exponential Smoothing dan Moving Average pada Arus Barang Bongkar," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 14, no. 2, pp. 125–134, Jun. 2024, doi: 10.34010/jamika.v14i2.12828.
- [15] A. N. Afiati and F. Frazna Az-Zahra, "IMPLEMENTASI ALGORITMA REGRESI LINEAR DALAM PREDIKSI PERSEDIAAN VOUCHER DI RAFFA CELL SUKABUMI," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 5, 2024, doi: 10.36040/jati.v8i5.11146.