

Pemodelan Prediksi Volume Penumpang Transjakarta Menggunakan Regresi Pada Algoritma Machine Learning

Ilham Maulana Wijaya*, Andi Taufik

Fakultas Teknik dan Informatika, Program Studi Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1,*}iammaulana447@gmail.com, ²andi.iuf@bsi.ac.id

Email Penulis Korespondensi: iammaulana447@gmail.com

Abstrak—Pertumbuhan populasi dan urbanisasi di Jakarta menimbulkan tantangan besar terhadap penyediaan transportasi publik yang efektif, khususnya pada layanan Transjakarta yang menghadapi fluktuasi volume penumpang sehingga menyulitkan pengelolaan kapasitas dan operasional. Penelitian ini bertujuan untuk memodelkan prediksi volume penumpang Transjakarta menggunakan metode regresi berbasis algoritma machine learning dengan membandingkan tiga model, yaitu Regresi Linear, Random Forest Regression dan Gradient Boosted Trees Regression. Data yang digunakan berupa catatan historis penumpang rute S21 (Ciputat–CSW/Tosari) dan S22 (Ciputat–Kampung Rambutan) periode Januari 2022 hingga Maret 2025. Data diolah melalui tahapan pra-pemrosesan, transformasi variabel kategorikal, pembagian data training dan testing, serta evaluasi menggunakan metrik Mean Squared Error (MSE), Root Mean Squared Error (RMSE) dan Koefisien Determinasi (R^2). Hasil penelitian menunjukkan bahwa Gradient Boosted Trees Regression memberikan kinerja prediksi terbaik dengan nilai R^2 sebesar 0,73 dan rata-rata kesalahan ± 22 ribu penumpang, lebih unggul dibandingkan Regresi Linear ($R^2 = 0,65$) maupun Random Forest Regression ($R^2 = 0,63$). Temuan ini menegaskan bahwa pendekatan ensemble boosting lebih efektif dalam menangkap pola non-linear pada data penumpang, sehingga dapat dijadikan model prediksi optimal untuk mendukung efisiensi operasional, perencanaan armada, serta pengembangan strategi transportasi publik yang adaptif dan berkelanjutan.

Kata Kunci: Transjakarta; Prediksi Volume Penumpang; Regresi; Machine Learning; Gradient Boosted Trees

Abstract—The rapid population growth and urbanization in Jakarta pose significant challenges to the provision of efficient public transportation, particularly for Transjakarta, which often experiences fluctuating passenger volumes that complicate capacity management and operational efficiency. This study aims to model and predict Transjakarta passenger volumes using regression methods within machine learning algorithms, by comparing three models: Linear Regression, Random Forest Regression, and Gradient Boosted Trees Regression. The dataset consists of historical passenger records from routes S21 (Ciputat–CSW/Tosari) and S22 (Ciputat–Kampung Rambutan) covering the period from January 2022 to March 2025. The data were processed through several stages, including preprocessing, categorical variable transformation, train-test splitting, and model evaluation using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results show that Gradient Boosted Trees Regression achieved the best predictive performance with an R^2 of 0.73 and an average error of approximately 22,000 passengers, outperforming Linear Regression ($R^2 = 0.65$) and Random Forest Regression ($R^2 = 0.63$). These findings highlight that ensemble boosting is more effective in capturing non-linear patterns in passenger data, making it the most suitable predictive model to support operational planning, fleet efficiency, and the development of adaptive and sustainable public transportation policies.

Keywords: Transjakarta; Passenger Volume Prediction; Regression; Machine Learning; Gradient Boosted Trees

1. PENDAHULUAN

Pertumbuhan populasi dan urbanisasi di kota besar menghadirkan tantangan serius dalam bidang transportasi. Jakarta sebagai ibu kota Indonesia sekaligus pusat kegiatan ekonomi, politik dan social memiliki tingkat kepadatan lalu lintas yang sangat tinggi dan salah satu kemacetan terburuk di dunia [1]. Kondisi tersebut tidak hanya menurunkan efisiensi mobilitas masyarakat, tetapi juga berdampak pada tingginya biaya ekonomi, menurunkan produktivitas, serta meningkatkan emisi karbon. Sebagai respon terhadap permasalahan tersebut, Pemerintah Provinsi DKI Jakarta meluncurkan Transjakarta pada tahun 2004 yang menjadi sistem Bus Rapid Transit (BRT) pertama di Asia Tenggara [2].

Sejak implementasinya, Transjakarta berperan penting dalam menyediakan moda transportasi massal yang efisien, terjangkau dan berkelanjutan. Dengan jaringan koridor yang semakin meluas dan integrasi antarmoda, Transjakarta menjadi tulang punggung mobilitas harian masyarakat Jakarta [3]. Selain itu, Transjakarta berkontribusi signifikan dalam mengurangi ketergantungan masyarakat terhadap kendaraan pribadi dan menekan tingkat polusi udara. Namun, di balik keberhasilan tersebut, sistem transportasi ini menghadapi tantangan berupa fluktuasi volume penumpang yang sulit diprediksi dan seringkali tidak sejalan dengan kapasitas layanan yang disediakan [4].

Fluktuasi volume penumpang dipengaruhi oleh berbagai faktor, seperti pola musiman, kondisi ekonomi, kebijakan pemerintah, peristiwa sosial, bahkan kondisi cuaca. Misalnya, pada awal tahun dan menjelang akhir tahun, terjadi lonjakan jumlah penumpang, sedangkan pada kuartal kedua biasanya terjadi penurunan signifikan [2]. Fenomena ini menuntut adanya prediksi jumlah penumpang yang akurat untuk mendukung pengambilan keputusan terkait penjadwalan armada, pengelolaan kapasitas, serta peningkatan kualitas layanan [5]. Tanpa prediksi yang tepat, risiko *overcapacity* (kelebihan armada) atau *undercapacity* (kekurangan armada) dapat meningkat, yang pada akhirnya berdampak pada efisiensi operasional maupun kepuasan pengguna.

Dalam bidang ilmu data, metode prediksi telah banyak digunakan untuk mengestimasi kebutuhan transportasi publik. Regresi Linear sering dijadikan baseline karena sederhana dan mudah diinterpretasikan [6]. Namun, model ini terbatas dalam menangkap pola non-linear yang kerap muncul pada data penumpang dengan variabilitas tinggi. Perkembangan teknologi *machine learning* menghadirkan alternatif yang lebih adaptif, seperti Random Forest Regression dan Gradient Boosted Trees Regression, yang terbukti unggul dalam menangani data kompleks serta mengurangi

kesalahan prediksi [7]. Menurut [8] menunjukkan bahwa penerapan algoritma Random Forest dan Extreme Gradient Boosting mampu memberikan akurasi prediksi hingga 99,5% dalam konteks prediksi kelulusan siswa. Hasil ini menegaskan potensi algoritma berbasis pohon dalam menangani data kompleks dengan fitur multidimensi. Selain itu, [9] juga menekankan bahwa pendekatan regresi modern dapat dimanfaatkan untuk mengidentifikasi pola data yang tidak terdeteksi oleh metode konvensional, sehingga dapat meningkatkan nilai akurasi dan keandalan hasil prediksi dalam berbagai konteks analisis berbasis data.

Di sisi lain, Syakrani dan Athaya (2025) menekankan pentingnya konsistensi model regresi pada populasi besar dengan menggunakan pendekatan multi variabel linear regression [10]. Mereka membuktikan bahwa akurasi dan koefisien determinasi (R^2) dapat tetap stabil pada sampel representatif, selama data dinormalisasi dan dibagi secara sistematis antara data latih dan uji. Prinsip ini relevan dalam konteks penelitian ini yang juga menggunakan pembagian data training–testing dan metrik evaluasi seperti RMSE dan R^2 .

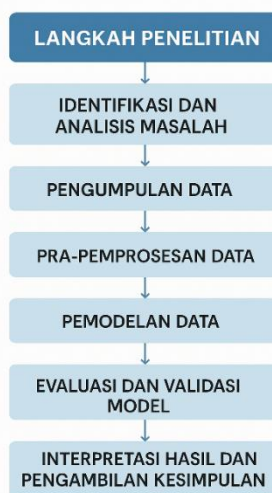
Sejumlah penelitian terdahulu menegaskan relevansi penggunaan model prediktif dalam konteks transportasi. Penelitian [2] menunjukkan bahwa validitas data operasional merupakan faktor kunci dalam perencanaan layanan Transjakarta. Penelitian [7] menemukan bahwa Random Forest Regression memiliki tingkat kesalahan prediksi yang lebih rendah dibanding Regresi Linear. Sementara itu, penelitian [4] menyoroti pentingnya pemilihan fitur seperti waktu, rute dan jenis layanan dalam meningkatkan akurasi model. Kajian lain menekankan pentingnya metrik evaluasi yang tepat, seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE) dan R^2 Score, untuk menilai kualitas model secara obyektif [11].

Dalam penelitian ini, data historis Transjakarta pada rute S21 (Ciputat–CSW/Tosari) dan S22 (Ciputat–Kampung Rambutan) periode tahun 2022 hingga Maret 2025 digunakan sebagai basis pemodelan (PT. Transportasi Jakarta, 2025). Data tersebut diolah melalui tahapan preprocessing, encoding variabel kategorikal, pembagian data training dan testing, hingga pemodelan menggunakan Regresi Linear, Random Forest Regression dan Gradient Boosted Trees Regression. Hasil analisis empiris pada penelitian ini menunjukkan bahwa Gradient Boosted Trees Regression menghasilkan performa terbaik dengan nilai $R^2 = 0,73$ dan RMSE ± 22 ribu penumpang, lebih unggul dibanding Regresi Linear ($R^2 = 0,65$) dan Random Forest Regression ($R^2 = 0,63$). Temuan ini menegaskan bahwa model Gradient Boosting mampu menangkap pola non-linear lebih baik sehingga menghasilkan prediksi yang lebih akurat.

2. METODOLOGI PENELITIAN

2.1. Langkah Penelitian

Penelitian ini menggunakan pendekatan kuantitatif yang bertujuan untuk pemodelan dan memprediksi volume penumpang Transjakarta menggunakan metode regresi. Langkah Penelitian ini disusun Untuk memastikan sistematis dan kualitas penelitian. Langkah penelitian dibagi dalam beberapa langkah kritis sebagai berikut:



Gambar 1. Tahapan Penelitian

Berikut penjelasan lebih lanjut dari gambar 1:

- Identifikasi dan Analisis Masalah yaitu Mendefinisikan persoalan utama terkait prediksi volume penumpang dan menetapkan tujuan penelitian untuk menemukan model yang optimal.
- Pengumpulan Data yaitu Menggunakan data historis volume penumpang pada rute Transjakarta S21 dan S22 yang diperoleh dari database resmi Transjakarta.
- Pra-pemrosesan Data yaitu Meliputi proses pembersihan data dari nilai hilang atau tidak valid, transformasi variabel kategorikal ke numerik menggunakan encoding, dan pembagian data ke dalam subset training dan testing.

- d. Pemodelan Data yaitu Penerapan algoritma regresi machine learning yang relevan, termasuk regresi linier, random forest dan gradient boosting, dengan tuning hyperparameter untuk meningkatkan performa model.
- e. Evaluasi dan Validasi Model yaitu menggunakan metrik statistik seperti MSE, RMSE dan R^2 untuk menilai kualitas prediksi model. Evaluasi ini memastikan model dapat diandalkan dan tidak overfitting atau underfitting.
- f. Interpretasi Hasil dan Pengambilan Kesimpulan yaitu Menentukan model yang memiliki keseimbangan terbaik antara kompleksitas, akurasi dan generalisasi terhadap data baru sebagai hasil akhir penelitian.

Langkah – Langkah ini merujuk pada metodologi yang dijabarkan oleh [11] serta diperkuat oleh pendekatan regresi dalam penelitian [6], [7], [12] dan [13] yang juga mengadopsi kerangka kuantitatif Berbasis analisis regresi linier.

Metode kuantitatif dengan pendekatan eksplanatori dan empiris, yang berfokus pada pengujian dan pemodelan hubungan sebab-akibat antara variabel bebas dan variabel terikat. Metode kuantitatif dipilih karena data yang dimanfaatkan berupa data numerik, seperti volume penumpang, yang dapat diukur secara objektif dan dianalisis menggunakan alat statistik dan *machine learning*. Pendekatan ini sangat sesuai untuk melakukan pemodelan prediktif karena bersifat sistematis, terukur dan berorientasi pada hasil yang dapat diulang.

Explanatori Dalam bahasa Inggris, "penjelasan" berarti penjelasan atau hal-hal yang berkaitan dengan penjelasan. Ini dapat mencakup menjelaskan peristiwa atau keadaan saat ini atau yang akan datang [14]. Dengan arti Pendekatan eksplanatori ini berfungsi untuk menjelaskan mekanisme hubungan antar variabel yang mempengaruhi volume penumpang, sehingga tidak hanya sekadar memprediksi tapi juga memahami pola dan faktor penyebab. Sementara itu, pendekatan empiris diterapkan untuk mengembangkan model berdasarkan data asli (real-world data) tanpa manipulasi teoritis yang rumit, melainkan penyesuaian melalui trial dan error model dalam machine learning.

2.2. Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari laporan resmi PT. Transportasi Jakarta mengenai jumlah pelanggan yang menggunakan layanan Transjabodetabek. Data tersebut diberikan dalam format Microsoft Excel (.xlsx) dengan nama file "Pelanggan S21 dan S22.xlsx". file ini berisi catatan jumlah pelanggan tahunan dan bulanan pada rute S21 (Ciputat – CSW/Tosari) dan S22 (Ciputat – Kampung Rambutan) selama periode 2022 sampai 2025 bulan Maret. Data yang digunakan merupakan data yang berkarakteristik *time series* memungkinkan analisis pola musiman, tren jangka panjang, dan fluktuasi yang terjadi secara periodik, sangat relevan dalam konteks transportasi publik yang dipengaruhi oleh variabel waktu dan faktor eksternal lainnya. Pemahaman karakteristik ini menjadi kunci dalam pemilihan algoritma dan teknik pengolahan data yang tepat [15]. Dibawah ini merupakan sampel data pada file "Pelanggan S21 dan S22.xlsx" periode tahun 2022.

Tabel 1. Sampel Data Pelanggan Tahun 2022

Tahun	Bulan	Kode Rute	Nama Rute	Jumlah Pelanggan
2022	Jan	S21	Ciputat - CSW	4699
2022	Feb	S21	Ciputat - CSW	20749
2022	Mar	S21	Ciputat - CSW	38214
2022	Apr	S21	Ciputat - CSW	39039
2022	May	S21	Ciputat - CSW	32039
2022	Jun	S21	Ciputat - CSW	48111
2022	Jul	S21	Ciputat - CSW	47908
2022	Aug	S21	Ciputat - CSW	54814
2022	Sep	S21	Ciputat - CSW	69939
2022	Sep	S22	Ciputat - Kampung Rambutan	8427
2022	Oct	S21	Ciputat - CSW	105705
2022	Oct	S22	Ciputat - Kampung Rambutan	44454
2022	Nov	S21	Ciputat - Tosari	106957
2022	Nov	S22	Ciputat - Kampung Rambutan	47783
2022	Dec	S21	Ciputat - Tosari	103847
2022	Dec	S22	Ciputat - Kampung Rambutan	48842

2.3. Proses Pengolahan Data

Pengolahan data merupakan tahap penting untuk memastikan kualitas input yang maksimal dalam pemodelan. Proses diawali dengan:

- a. Pembersihan Data: Menghilangkan data yang tidak lengkap, duplikat atau outlier yang dapat merusak hasil analisis.
- b. Transformasi Data: Seperti kode rute diubah menggunakan Label Encoding karena bersifat kategorikal nominal yang sederhana, sedangkan variabel waktu (bulan) dikonversi menjadi nilai numerik (1–12) agar dapat diproses oleh model regresi. Teknik ini dipilih karena efisien, mudah diimplementasikan, dan sesuai dengan karakteristik dataset time series yang digunakan.

- c. Pembagian Data: Data dibagi menjadi dua subset, yaitu data training (70–80%) untuk melatih model, dan data testing (20–30%) untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat.
- d. Normalisasi/Standarisasi: Menerapkan transformasi standar atau normalisasi untuk menyesuaikan skala variabel agar mempercepat konvergensi dan hasil pelatihan model lebih optimal.

Melalui pendekatan pengolahan data yang terstruktur dan Berbasis software ini, hasil penelitian menjadi lebih dapat dipertanggung jawabkan dan terbebas dari bias manual. Strategi ini juga berjalan dengan praktik analitik modern sebagaimana diterapkan dalam studi [16] yang melakukan pemodelan data menggunakan google colab.

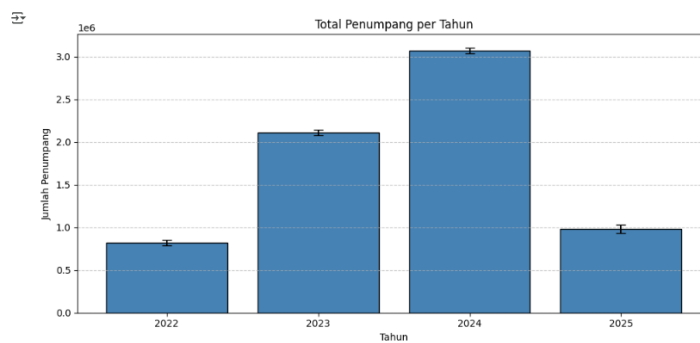
Google Colab adalah platform berbasis cloud yang memungkinkan pengguna menjalankan program Python secara interaktif tanpa perlu instalasi perangkat lunak lokal. Platform ini menyediakan lingkungan yang lengkap dengan berbagai pustaka data science seperti Pandas, Matplotlib dan Scikit-learn yang memudahkan proses analisis data dan pengembangan model *machine learning* [17].

3. HASIL DAN PEMBAHASAN

Penelitian ini berfokus pada pemodelan prediksi jumlah penumpang Transjakarta dengan memanfaatkan pendekatan regresi pada algoritma machine learning. Sebelum dilakukan pemodelan, data melalui serangkaian proses pengumpulan data, pra-pemrosesan, eksplorasi dan analisis deskriptif yang bertujuan untuk memahami struktur data sekaligus mengidentifikasi pola yang relevan.

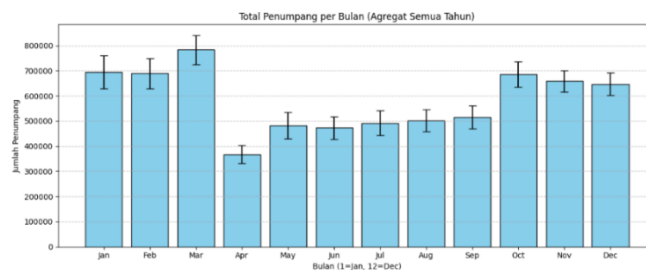
3.1. Analisis Ekploratif Data

a. Distribusi Penumpang per Tahun



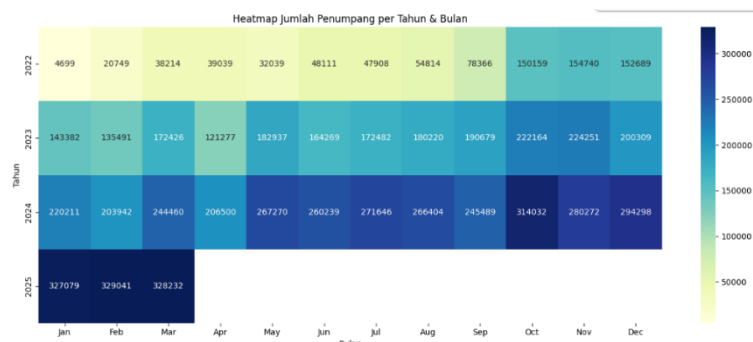
Gambar 2. Total Penumpang per Tahun

b. Distribusi Penumpang per Bulan



Gambar 3. Total Penumpang per Bulan (Agregat Semua Tahun)

c. Distribusi Penumpang per Tahun dan Bulan



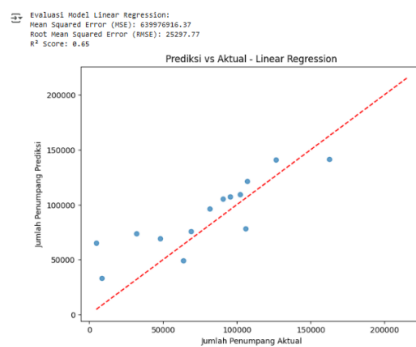
Gambar 4. Heatmap Jumlah Penumpang per Tahun dan Bulan

Pada gambar 5, gambar 6 dan gambar 7 menjelaskan Jumlah penumpang Transjakarta menunjukkan tren peningkatan yang signifikan dalam kurun waktu 2022 hingga 2025. Pada tahun 2022, jumlah penumpang tercatat sebesar 0,8 juta, kemudian meningkat tajam menjadi 2,1 juta pada tahun 2023 atau mengalami kenaikan sekitar 162%. Pertumbuhan ini berlanjut pada tahun 2024 dengan capaian lebih dari 3 juta penumpang, yang berarti terdapat kenaikan sekitar 45% dibandingkan tahun sebelumnya. Sementara itu, data awal tahun 2025 (Januari–Maret) telah menunjukkan jumlah penumpang sekitar 1 juta, yang secara total sudah melampaui capaian sepanjang tahun 2022. Apabila tren pertumbuhan bulanan tersebut relatif stabil, maka pada akhir tahun 2025 jumlah penumpang diproyeksikan berada pada kisaran 3,5 hingga 3,8 juta penumpang, sehingga melampaui rekor tertinggi pada tahun 2024.

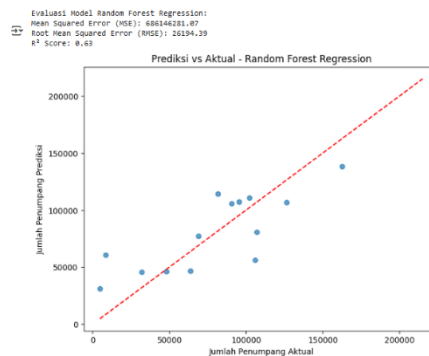
Dengan demikian, baik dari perspektif tahunan maupun musiman, data historis menunjukkan adanya pertumbuhan permintaan yang berkelanjutan. Temuan ini dapat dijadikan landasan dalam merumuskan proyeksi kebutuhan transportasi publik di masa depan sekaligus memperkuat kebijakan perencanaan strategis yang lebih responsif terhadap perubahan pola mobilitas masyarakat.

3.2. Evaluasi Model

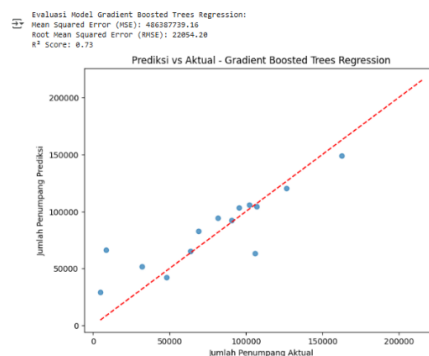
Tiga model diuji, yakni Linear Regression, Random Forest Regression dan Gradient Boosting Trees Regression, dengan menggunakan metrik evaluasi MSE, RMSE dan R2.



Gambar 5. Hasil Linear Regression



Gambar 6. Hasil Random Forest Regression



Gambar 7. Hasil Gradient Boosted Trees Regression

Pada gambar 10, Model Gradient Boosted Trees Regression memperlihatkan kinerja terbaik dengan nilai koefisien determinasi ($R^2 = 0,73$), Mean Squared Error ($MSE = 486.387.739,16$), serta Root Mean Squared Error ($RMSE = 22.054,20$). Nilai R^2 yang relatif tinggi mengindikasikan bahwa model ini mampu menjelaskan sekitar 73% variasi data aktual, dengan tingkat kesalahan prediksi yang lebih rendah dibandingkan model lainnya.

Pada gambar 8 dan gambar 9, model Random Forest Regression dan Linear Regression menunjukkan performa yang lebih rendah. Random Forest Regression menghasilkan nilai $R^2 = 0,63$, MSE sebesar 686.146.281,07, dan RMSE sebesar 26.194,39, yang menandakan bahwa model ini hanya mampu menjelaskan sekitar 63% variasi data dan memiliki tingkat kesalahan prediksi yang lebih besar dibandingkan Gradient Boosted Trees. Sementara itu, Linear Regression memiliki nilai $R^2 = 0,65$, MSE sebesar 639.976.916,37, dan RMSE sebesar 25.297,77, yang sedikit lebih baik dibandingkan Random Forest.

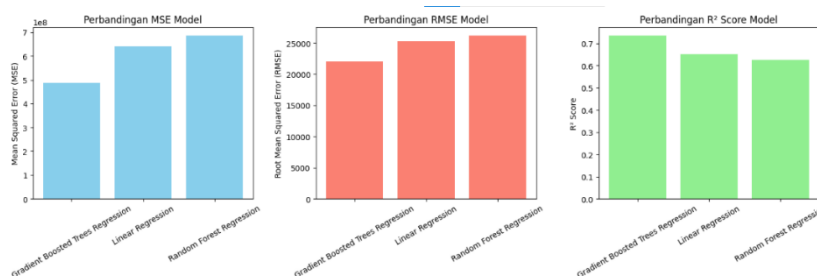
Dari pengujian tiga algoritma regresi, diperoleh hasil bahwa Gradient Boosted Trees Regression memiliki performa prediksi terbaik dengan nilai $R^2 = 0,73$ dan RMSE ± 22.054 penumpang. Model ini mampu menjelaskan sekitar 73% variasi data aktual, lebih tinggi dibandingkan Linear Regression ($R^2 = 0,65$, RMSE ± 25.298) maupun Random Forest Regression ($R^2 = 0,63$, RMSE ± 26.194). Temuan ini sejalan dengan penelitian oleh [18] yang juga menunjukkan bahwa regresi dengan kompleksitas model lebih tinggi mampu menangkap hubungan non-linear dan menghasilkan nilai error yang lebih rendah. Hal ini memperkuat argumentasi bahwa model ensemble seperti Gradient Boosted Trees lebih adaptif terhadap variasi data kompleks.

3.5. Perbandingan Model

Tabel 2. Perbandingan Kinerja Model Regresi

Model	MSE	RMSE	R^2
Linear Regression	6.399.76916.37	25.297,77	0,65
Random Forest Regression	6,86146281.07	26.194,39	0,63
Gradient Boosted Trees Regression	4,86387739.16	22.054,20	0,73

Tabel 2 ini menunjukkan hasil evaluasi tiga model regresi menggunakan metrik MSE, RMSE, dan R^2 . Dari hasil yang diperoleh, Gradient Boosted Trees Regression memiliki performa terbaik dengan nilai MSE paling rendah (4.863.87739,16), RMSE terkecil (22.054,20), serta nilai R^2 tertinggi (0,73). Hal ini menunjukkan bahwa model ini mampu memberikan prediksi jumlah penumpang yang lebih akurat dibandingkan dua model lainnya. Sementara itu, Linear Regression memperoleh nilai R^2 sebesar 0,65 dengan MSE dan RMSE yang relatif lebih tinggi, sedangkan Random Forest Regression menghasilkan nilai R^2 terendah (0,63) serta kesalahan prediksi yang lebih besar. Terdapat juga model bar chart dari setiap model Regresi di tunjukan pada gambar 11 ini.



Gambar 8. Perbandingan MSE, RMSE dan R^2

Hasil ini mengindikasikan bahwa Gradient Boosted Trees mampu menangkap pola non-linear dalam data jumlah penumpang Transjakarta, yang tidak dapat sepenuhnya dijelaskan oleh Linear Regression. Temuan ini konsisten dengan penelitian [7] yang menunjukkan keunggulan metode ensemble dalam menangani data dengan variabilitas tinggi. Sementara itu, performa Random Forest Regression dalam penelitian ini lebih rendah dibanding penelitian serupa, kemungkinan karena keterbatasan jumlah variabel independen (hanya waktu dan rute) sehingga kedalaman pohon tidak menghasilkan keunggulan signifikan. Secara praktis, hasil ini menunjukkan bahwa penggunaan model boosting dapat membantu PT. Transportasi Jakarta dalam merencanakan kapasitas armada secara lebih presisi, mengurangi risiko kelebihan atau kekurangan armada, serta meningkatkan kepuasan penumpang.

Hasil prediksi yang diperoleh dalam penelitian ini menunjukkan tingkat akurasi yang dapat diterima untuk penerapan di bidang praktis. Temuan ini sejalan dengan penelitian oleh [19], yang menyatakan bahwa pendekatan prediksi berbasis simulasi dapat memberikan tingkat akurasi di atas 80% pada kasus nyata. Dengan demikian, model regresi yang digunakan dalam penelitian ini juga dapat dianggap memiliki performa prediksi yang baik untuk data yang dianalisis. Hal ini juga konsisten dengan hasil penelitian oleh [20] yang menunjukkan bahwa XGBoost mencapai nilai R^2 hingga 0,7746 dan nilai Mean Squared Error terendah, menandakan keunggulan algoritma berbasis boosting dalam mengolah data tabular [18].

4. KESIMPULAN

Berdasarkan hasil penelitian mengenai Pemodelan Prediksi Volume Penumpang Transjakarta Menggunakan Metode Regresi pada Algoritma Machine Learning, dapat disimpulkan bahwa data historis penumpang Transjakarta periode 2022 sampai dengan Maret 2025 menunjukkan tren peningkatan yang konsisten dengan variasi musiman yang stabil, terutama

pada awal dan akhir tahun. Hasil evaluasi menunjukkan bahwa di antara tiga model regresi yang diuji (Linear Regression, Random Forest Regression, dan Gradient Boosted Trees Regression), model Gradient Boosted Trees Regression memberikan performa terbaik dengan nilai $R^2 = 0,73$ dan tingkat kesalahan prediksi rata-rata ± 22 ribu penumpang. Hal ini membuktikan kemampuan model dalam menangkap hubungan non-linear antar variabel secara lebih akurat dibandingkan model lainnya. Dengan demikian, pendekatan machine learning berbasis regresi dapat menjadi metode efektif dalam memprediksi volume penumpang transportasi publik, serta bermanfaat sebagai dasar pengambilan keputusan strategis dalam perencanaan kapasitas armada, jadwal operasional, dan kebijakan transportasi adaptif di masa mendatang. Untuk pengembangan penelitian selanjutnya, disarankan memperluas cakupan data ke lebih banyak rute dan periode waktu yang lebih panjang, menambahkan variabel kontekstual seperti cuaca, tarif, dan kondisi lalu lintas, serta menguji algoritma lain seperti XGBoost atau model berbasis deep learning. Selain itu, hasil penelitian ini dapat dikembangkan menjadi dashboard prediksi interaktif guna mendukung keputusan operasional secara real-time.

REFERENCES

- [1] F. Jauregui-Fung, "BRT Transjakarta: Phasing in, Performing and Expanding a New System within a consolidated urban area," 2022, doi: 10.23661/r6.2022.
- [2] S. Hidayat and R. Mesra, "Studi Kasus Pelayanan dan Kualitas Busway Transjakarta," *COMTE J. Sociol. Res. Educ.*, vol. 1, no. 1, pp. 26–31, 2024, doi: 10.64924/yg5xpr71.
- [3] Y. Aprilia, "Analisis Peningkatan Efektivitas Transportasi Umum (Transjakarta)," *Pros. SEMDIKJAR (Seminar Nas.)*, vol. 6, pp. 1598–1601, 2023, doi: 10.29407/1wyk5h36.
- [4] T. Nurholipah, R. Kurniawan, and Y. A. Wijaya, "Evaluasi Performa Model Regresi Linear Dengan Rmse Pada Jumlah Penumpang Bus Transjakarta," *JIKA (Jurnal Inform.)*, vol. 8, no. 2, p. 180, 2024, doi: 10.31000/jika.v8i2.10405.
- [5] E. Hasibuan *et al.*, "Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web," *J. Ilm. Komputasi*, vol. 21, no. 4, pp. 595–602, 2022, doi: 10.32409/jikstik.21.4.3327.
- [6] N. Almuntazah, N. Azizah, Y. L. Putri, and D. C. R. Novitasari, "Prediksi Jumlah Mahasiswa Baru Menggunakan Metode Regresi Linier Sederhana," *J. Ilm. Mat. Dan Terap.*, vol. 18, no. 1, pp. 31–40, 2021, doi: 10.22487/2540766x.2021.v18.i1.15465.
- [7] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 58–64, 2023, doi: 10.52158/jacost.v4i1.491.
- [8] A. F. A. Naibaho and A. Zahra, "Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3.3056.
- [9] N. K. Zainal, "Prediksi Harga Real Estate Menggunakan Metode Regresi Linear Berbasis Machine Learning," *J. Artif. Intell. Appl.*, vol. 1, no. 1, pp. 19–27, 2024, [Online]. Available: <https://jurnal.mutiaraamaliyah.com/index.php/jaia/article/view/5%0Ahttps://jurnal.mutiaraamaliyah.com/index.php/jaia/article/download/5/9>
- [10] N. Syakrani and N. A. S. R., "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Konsistensi Model Regresi Empat Variabel Pada Populasi dan Sampel untuk Prediksi Temperatur," vol. 6, no. 1, pp. 9–16, 2025, doi: 10.52158/jacost.v6i1.971.
- [11] I. Amansyah, J. Indra, E. Nurlaelasari, and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia," *J. Soc. Sci. Res.*, vol. 4, no. 4, pp. 1199–1216, 2024, doi: 10.31004/innovative.v4i4.12735.
- [12] M. R. Athallah and A. F. Rozi, "Implementasi Data Mining Untuk Prediksi Peramalan Penjualan Produk Hj Karpet Menggunakan Metode Linear Regression," *J. Sains dan Teknol.*, vol. 2, no. 3, pp. 180–187, 2023, doi: 10.47233/jsit.v2i3.550.
- [13] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.
- [14] D. Assayakurrohim, D. Ikhrum, R. a Sirodj, and M. W. Afgani, "Jurnal pendidikan sains dan komputer metode studi kasus dalam penelitian kualitatif jurnal pendidikan sains dan komputer," *J. Pendidik. sains dan Komput.*, vol. 3, no. 1, pp. 1–9, 2023, doi: 10.47709/jpsk.v3i01.1951.
- [15] P. F. A. Tambuwun, N. Nainggolan, and Y. A. R. Langi, "d ' CartesiaN Jurnal Matematika dan Aplikasi Peramalan Banyaknya Penumpang Bandar Udara Internasional Sam Ratulangi Manado Dengan Metode Winter ' s Exponential Smoothing dan Seasonal ARIMA," vol. 12, no. 1, pp. 14–20, 2023, doi: 10.35799/dc.12.1.2023.48066.
- [16] Muhammad Faiq Abdi and Yonhendri, "Implementasi Sistem Prediksi Saham Real-Time dengan Integrasi Yahoo Finance API dan Machine Learning di Google Colab," *El-Mujtama J. Pengabd. Masyarakat*, vol. 5, no. 3, pp. 25–31, 2025, doi: 10.47467/elmujtama.v5i3.7379.
- [17] M. Sholeh, Y. Rachmawati, and E. N. Cahyo, "Penerapan Regresi Linear Ganda Untuk Memprediksi Hasil Nilai Kuesioner Mahasiswa Dengan Menggunakan Python," *J. Din. Inform.*, vol. 11, no. 1, pp. 13–24, 2022.
- [18] A. T. Nurani, A. Setiawan, and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *J. Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.
- [19] K. Alfikrizal, S. Defit, and Y. Yunus, "Simulasi Monte Carlo dalam Prediksi Jumlah Penumpang Angkutan Massal Bus Rapid Transit Kota Padang," *J. Inform. Ekon. Bisnis*, vol. 3, no. 2, pp. 78–83, 2020, doi: 10.37034/infeb.v3i2.72.
- [20] B. W. Sari and D. Prabowo, "Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest , Gradient Boosting , dan XGBoost," vol. 04, no. 01, pp. 42–51, 2025.