Jurnal Profesi Insinyur Universitas Lampung (JPI)
Vol. 7 No. 1, Tahun 2026, e-ISSN: 2722-5771 https://doi.org/10.23960/jpi.v7n1.387

Penerapan *Symmetric Mean Absolute Percentage Error* Terhadap Hasil Prediksi *Multiple Linear Regression* Dan *Extreme Gradient Boosting Regression*

Neforinez Pedovanka[1], Aghiska Alvia Nisa[2], Aida Nurmagfiroh[3], Hidayanti Murtina[4], Aryo Tunjung Kusumo[5]

[1,2,3] Program Studi Sistem Informasi Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika (UBSI); Jl. Cut Mutia No.88, Sepanjang Jaya, Kec. Rawalumbu, Kota Bekasi, Jawa Barat 17113 Indonesia, Telp : (021) 82425634

Keywords:
XGBoost Regression; MLR; sMAPE, *Machine learning*; Kereta *Priority*; Prediksi

Corespondent Email:
neforinez@gmail.com

Banyaknya metode *Machine learning* dalam prediksi menyebabkan pemilihan algoritma menjadi penting karena setiap metode memiliki kemampuan berbeda dalam mengolah karakteristik data. Perbedaan performa antar metode menunjukkan perlunya pengujian untuk menentukan algoritma yang sesuai dalam menghasilkan prediksi akurat. Penelitian ini menggunakan data berupa jumlah penumpang kereta *priority* periode 2021–2025 dengan variabel pendukung berupa *load factor*, *high season*, indeks Google Trends, dan tingkat inflasi. Pengujian dilakukan dengan menggunakan metode 5-*Fold Cross Validation*. Sedangkan untuk memprediksi jumlah penumpang kereta tipe *priority* menggunakan metode *Multiple Linear Regression* (MLR) dan *Extreme Gradient Boosting Regression* (*XGBoost Regression*) dan evaluasi hasil prediksi menggunakan *Symmetric Mean Absolute Percentage Error* (sMAPE). Hasil penelitian menunjukkan bahwa MLR memiliki performa prediksi lebih baik dibandingkan *XGBoost Regression* pada kasus memprediksi jumlah penumpang kereta tipe *priority*. MLR memiliki selisih terhadap nilai aktual sebesar 1,387033912, sedangkan *XGBoost Regression* memiliki selisih sebesar 6.829,9905. Nilai rata-rata sMAPE metode MLR sebesar 0,0226%, sedangkan metode *XGBoost Regression* sebesar 16,5854%. Hasil tersebut menunjukkan MLR memiliki tingkat kesalahan lebih rendah dan lebih sesuai digunakan pada kasus ini.

The large number of Machine learning methods available for prediction makes algorithm selection important because each method has different capabilities in processing data characteristics. Differences in performance among methods indicate the need for testing to determine the appropriate algorithm for producing accurate predictions. This study uses data on the number of priority train passengers during the 2021–2025 period, with supporting variables including load factor, high season, Google Trends index, and inflation rate. Model testing was conducted using the 5-Fold Cross Validation method. The prediction of the number of priority train passengers was performed using Multiple Linear Regression (MLR) and Extreme Gradient Boosting Regression (XGBoost Regression), while the prediction results were evaluated using Symmetric Mean Absolute Percentage Error (sMAPE). The results showed that MLR achieved better prediction performance than XGBoost Regression in predicting the number of priority train passengers. MLR had a difference from the actual value of 1.387033912, while XGBoost Regression had a difference of 6,829.9905. The average sMAPE value of the MLR method was 0.0226%, whereas the XGBoost Regression method was 16.5854%. These results indicate that MLR has a lower error rate and is more suitable for use in this case.

1. PENDAHULUAN

Penggunaan *machine learning* telah menggeser metode konvensional karena kemampuannya menangkap pola data yang kompleks salah satunya dalam melakukan prediksi. Dalam analisis prediksi menggunakan *machine learning*, pemilihan metode menjadi faktor penting untuk memperoleh hasil yang akurat dan relevan. Beberapa metode yang banyak digunakan adalah *Multiple Linear Regression* dan *Extreme Gradient Boosting Regression* (*XGBoost Regression*). *Multiple Linear Regression* banyak digunakan karena memiliki pendekatan yang sederhana, serta stabil dalam memodelkan hubungan linear antar variabel. Di sisi lain, *XGBoost* hadir sebagai metode *machine learning* yang mampu menangani pola data yang lebih kompleks, termasuk hubungan non-linear. Perbandingan hasil kedua metode ini dalam melakukan prediksi menjadi krusial, untuk itu perlu dilakukan pengukuran terhadap akurasi dari hasil metode tersebut.

Berbagai penelitian telah dilakukan untuk menganalisis performa algoritma *XGBoost* dan *Multiple Linear Regression* dalam prediksi data numerik menggunakan berbagai metode evaluasi. Salah satu penelitian membandingkan algoritma *XGBoost*, Regresi Linier, dan *Random Forest* dalam memprediksi tingkat kemiskinan menggunakan evaluasi MAE dan R^2, dengan hasil menunjukkan bahwa *XGBoost* memberikan performa terbaik [1]. Penelitian lain juga membandingkan ketiga algoritma tersebut dalam prediksi harga rumah dengan hasil evaluasi MAE, RMSE, dan R^2 terbaik diperoleh oleh algoritma regresi linier [2]. Selain itu, penelitian mengenai prediksi jumlah penumpang kereta api juga telah dilakukan menggunakan metode regresi linier sederhana pada setiap kelas layanan dengan evaluasi MAE [3].

Berdasarkan penelitian terdahulu yang telah diuraikan, kebaruan (*novelty*) pada penelitian ini terletak pada fokus pengujian akurasi menggunakan metode *Symmetric Mean Absolute Percentage Error* (sMAPE) terhadap hasil prediksi *Multiple Linear Regression* dan *XGBoost Regression* dengan objek penelitian yaitu melakukan prediksi terhadap permintaan layanan kereta *priority*. Dimana pada penelitian sebelumnya *XGBoost Regression* memiliki nilai performa yang terbaik menggunakan evaluasi MAE dan R^2 dan pada pengujian MAE, RMSE, dan R^2 metode *Multiple Linear Regression* mendapatkan nilai tertinggi.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk melakukan pengujian lanjutan terhadap metode *Multiple Linear Regression* dan *XGBoost Regression* dengan melakukan pengukuran akurasi menggunakan metode sMAPE dimana objek penelitian yang akan penulis gunakan adalah permintaan layanan kereta *priority*. Pada penelitian ini, adapun variabel yang digunakan, yaitu data jumlah penumpang kereta *priority* periode tahun 2021-2025, okupansi (*load factor*), *high season*, Indeks Google Trends, dan tingkat inflasi. Sehingga hasil dari penelitian yang penulis lakukan dapat menjadi bahan pertimbangan saat menentukan metode prediksi yang akan digunakan.

2. TINJAUAN PUSTAKA

Untuk mendukung penelitian yang dilakukan, diperlukan beberapa teori yang relevan sebagai landasan dalam pelaksanaan penelitian. Berikut merupakan beberapa teori yang digunakan sebagai dasar dalam penelitian ini:

2.1. *Machine learning*

Machine learning adalah bidang kecerdasan buatan yang melatih sistem komputer dengan memanfaatkan algoritma dan pola yang diperoleh dari data. Melalui algoritma dan model tertentu, sistem dapat melakukan prediksi maupun pengambilan keputusan tanpa harus diberikan instruksi

secara khusus untuk setiap proses yang dijalankan [4] [5].

2.2. *Supervised Learning*

Supervised Learning adalah metode pada *machine learning* yang bekerja dengan memanfaatkan data yang telah memiliki label sebagai bahan pelatihan. Melalui proses pelatihan, model akan mempelajari hubungan antara data input dan label yang diberikan sehingga dapat menghasilkan prediksi atau keputusan pada data baru secara lebih akurat [5] [6] [7].

2.3. *RapidMiner*

RapidMiner adalah aplikasi berbasis *open source* yang dapat digunakan dalam membantu proses pengolahan dan analisis data. Aplikasi ini mendukung berbagai kegiatan seperti data mining, teks mining, serta analisis prediktif untuk mengelola dan menemukan pola dari suatu data [19].

2.4. *Cross Validation*

Cross Validation metode pengujian performa model yang bekerja dengan memisahkan data menjadi beberapa bagian (*fold*). Penerapan metode ini difokuskan untuk menghindari kondisi dimana model terlalu fokus pada data latih (*overfitting*), serta memastikan model tetap memberikan hasil prediksi yang akurat terhadap data yang belum pernah digunakan sebelumnya [20]. Agar hasilnya optimal, data yang digunakan harus bersih dan valid. Oleh karena itu, tahapan pra-pemrosesan data sangat penting untuk mengubah data mentah menjadi lebih konsisten [21].

2.5. *K-Fold Cross Validation*

K-Fold Cross Validation adalah metode evaluasi dimana seluruh data penelitian dibagi menjadi beberapa kelompok (*fold*) berjumlah K dengan memiliki jumlah data yang relatif sama. Model kemudian diuji sebanyak K kali, dimana setiap tahapannya, satu kelompok data berperan sebagai data uji dan bagian lainnya difungsikan sebagai data latih. Nilai rata-rata dari seluruh tahapan tersebut kemudian dihitung untuk menentukan estimasi akhir dari performa model [20].

2.6. *Regression*

Regression atau regresi merupakan metode analisis yang untuk membangun model prediksi dengan memanfaatkan data input yang tersedia. *Regression* juga menjadi alat analisis dalam menilai seberapa erat korelasi antara variabel terikat dengan variabel bebas [8].

2.7. *Multiple Linear Regression*

Multiple Linear Regression atau Regresi Linier Berganda adalah metode analisis untuk mengetahui pengaruh beberapa variabel *independent* (X) terhadap suatu variabel *dependent* (Y) [9] [10]. Analisis ini berfungsi sebagai alat prediksi untuk memastikan ada atau tidaknya hubungan sebab-akibat (kuasal) di antara variabel-variabel tersebut [10].

2.8. *XGBoost Regression*

Extreme Gradient Boosting Regression (*XGBoost Regression*) merupakan variasi tingkat lanjut dari algoritma *gradient boosting* yang mengintegrasikan pendekatan linier dengan struktur pohon keputusan. Metode ini mampu memproses data dengan sangat cepat melalui komputasi parallel, sekaligus menyediakan fitur bawaan untuk validasi silang dan identifikasi variabel yang paling berpengaruh [11].

Secara teknis, proses pembentukan model *XGBoost* dilakukan dengan menambahkan satu pohon keputusan baru pada setiap tahapan (iterasi). Pohon baru tersebut, berfungsi untuk memprediksi nilai *error* (residual) dari kombinasi pohon-pohon sebelumnya, sehingga hasil prediksi akhir merupakan akumulasi dari seluruh pohon keputusan yang telah terbentuk [5].

2.9. Prediksi

Prediksi merupakan salah satu pendekatan untuk melihat potensi situasi yang akan datang dengan memanfaatkan informasi dan pola dari data historis [12].

Melalui prediksi, kemungkinan kondisi yang akan terjadi di masa mendatang dapat diperkirakan berdasarkan data dan informasi pada masa lalu maupun saat ini. Proses tersebut dilakukan secara sistematis untuk meminimalkan tingkat kesalahan sehingga hasil yang diperoleh dapat mendekati kondisi sebenarnya, meskipun tidak sepenuhnya bersifat pasti [13].

2.10. *Symmetric Mean Absolute Percentage Error* (sMAPE)

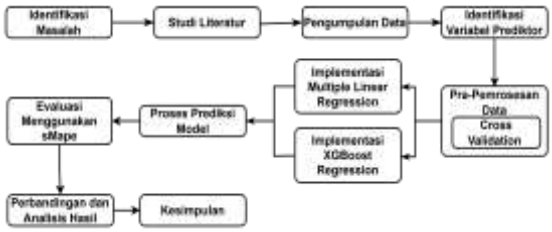
Menurut Makridakis dan Hibon (2000), *Symmetric Mean Absolute Percentage Error (sMAPE)* adalah metode evaluasi yang digunakan untuk menilai tingkat kesalahan hasil prediksi dalam bentuk persentase melalui perbandingan antara nilai aktual dan nilai prediksi. Metode ini dikembangkan dari *Mean Absolute Percentage Error (MAPE)* untuk menghasilkan pengukuran error yang lebih seimbang atau simetris serta mengurangi bias, terutama pada data dengan nilai aktual yang kecil. Dalam perhitungannya, sMAPE menggunakan rata-rata nilai aktual dan prediksi sebagai pembagi sehingga nilai error yang dihasilkan memiliki batasan nilai hingga 200% dan mampu merepresentasikan tingkat kesalahan relatif model prediksi dengan lebih baik [14] [15].

3. METODE PENELITIAN

Penelitian ini dirancang dengan menerapkan pendekatan kuantitatif karena proses analisis dilakukan berdasarkan data numerik yang diolah menggunakan metode statistik dan *machine learning*. Pendekatan kuantitatif dipilih untuk mengetahui tingkat akurasi hasil prediksi jumlah penumpang kereta tipe *priority* dengan menguji kinerja dua algoritma *Supervised Learning*, yaitu *Multiple Linear Regression* (MLR) dan *Extreme Gradient Boosting Regression* (*XGBoost Regression*) dengan melakukan pengukuran akurasi menggunakan metode sMAPE.

Penelitian ini bersifat prediktif dan komparatif. Bersifat prediktif karena penelitian berfokus pada proses memperkirakan jumlah penumpang pada periode tertentu berdasarkan pola data historis. Sementara itu, bersifat komparatif karena dilakukan perbandingan performa antara dua metode *machine learning* untuk mengetahui metode yang menghasilkan tingkat kesalahan prediksi paling rendah.

Adapun tahapan pelaksanaan penelitian ini ditunjukkan secara detail pada Gambar 1.



Gambar 1. Alur Penelitian
Sumber : Penelitian (2026)

3.1. Identifikasi Masalah

Pada tahap awal, penulis melakukan identifikasi permasalahan penelitian berdasarkan kesenjangan penelitian (*research gap*) yang ditemukan pada penelitian-penelitian sebelumnya. Tahap ini dilakukan untuk menentukan fokus penelitian serta metode yang digunakan. Selain itu, penulis juga mengidentifikasi kebaruan penelitian (*novelty*) berdasarkan kesenjangan yang ditemukan, termasuk penentuan instrumen evaluasi yang digunakan dalam penelitian. Hasil dari tahap ini digunakan sebagai dasar dalam penyusunan tahapan penelitian selanjutnya.

3.2. Studi Literatur

Pada tahap ini, dilakukan pengumpulan dan pengkajian mendalam terhadap berbagai literatur serta studi terdahulu yang berkaitan dengan topik

penelitian. Referensi yang digunakan dalam penelitian ini berasal dari berbagai sumber, seperti buku elektronik (e-book) dan jurnal ilmiah nasional. Buku elektronik digunakan sebagai dasar untuk memahami konsep, teori, dan prinsip kerja dari metode yang digunakan dalam penelitian, sedangkan jurnal ilmiah digunakan untuk mempelajari penelitian terdahulu, mengetahui perkembangan penelitian terkait, serta memahami penerapan variabel dan metode evaluasi yang relevan.

3.3. Pengumpulan Data

Tahap pengumpulan data dalam penelitian ini dilakukan dengan memperoleh data primer dari Pejabat Pengelola Informasi dan Dokumentasi (PPID) PT Kereta Api Pariwisata. Data yang digunakan berupa data jumlah penumpang kereta *priority* per bulan dari bulan Januari hingga Desember selama periode 2021-2025, serta data *load factor* (keterisian kursi).

Data yang digunakan selain itu berupa data sekunder, yaitu data tingkat inflasi yang diperoleh dari Badan Pusat Statistik (BPS), data indeks trend yang diperoleh melalui Google Trends, serta penentuan periode *high season* yang disesuaikan dengan Kalender Indonesia. Seluruh data yang diperoleh kemudian digunakan sebagai variabel dalam proses prediksi jumlah penumpang kereta tipe *priority*.

3.4. Identifikasi Variabel Prediktor

Berdasarkan pengumpulan data yang penulis lakukan, penulis menemukan variabel prediktor dalam penelitian ini yang meliputi tingkat inflasi, *high season*, Indeks Google Trends, serta okupansi keterisian kursi (*load factor*).

3.5. Pra-pemrosesan Data

Dalam tahap pra-pemrosesan, penulis melakukan beberapa proses untuk mempersiapkan data sebelum digunakan dalam proses pemodelan. Proses ini mencakup pemeriksaan kelengkapan data serta transformasi data kategorikal ke dalam bentuk numerik agar dapat diproses oleh algoritma *machine learning*. Dalam proses penelitian yang dilakukan, variabel *High season* yang awalnya berupa kategori berdasarkan periode liburan diubah menjadi bentuk biner, yaitu nilai 1 untuk periode *High season* dan nilai 0 untuk periode *Non-High season*.

Selanjutnya, dataset yang telah dipersiapkan digunakan dalam proses *Cross Validation* dengan metode *K-Fold* sebanyak 5 lipatan (*5-Fold Cross Validation*). Penggunaan metode ini bertujuan untuk memanfaatkan seluruh data penelitian secara lebih efektif melalui pembagian data *training* dan *testing* yang dilakukan secara bergantian pada setiap lipatan (*fold*), sehingga hasil evaluasi model dapat diperoleh secara lebih konsisten.

3.6. Implementasi *Multiple Linear Regression*

Metode *Multiple Linear Regression* (MLR), proses pemodelan dilakukan dengan menganalisis hubungan linear antara variabel prediktor terhadap jumlah penumpang kereta tipe *priority*. Metode ini mampu menunjukkan hubungan dan pengaruh masing-masing variabel terhadap hasil prediksi melalui pembentukan persamaan regresi linear berganda. Adapun persamaan umum dari model tersebut sebagai berikut [9]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \quad (1)$$

Keterangan :
Y = variabel *dependent*
β_0= konstanta
$\beta_1, \beta_2, \cdots, \beta_n$= koefisien regresi
$X_1, X_2, , \cdots, X_n$ = variable *independent*

3.7. Implementasi *XGBoost Regression*

Metode *XGBoost Regression*, model bekerja dengan membangun pohon keputusan (*decision tree*) secara bertahap berdasarkan pola dari data jumlah

penumpang kereta tipe *priority* periode 2021–2025. Model mempelajari hubungan antara variabel terhadap perubahan jumlah penumpang pada setiap periode. Metode ini digunakan karena mampu menangani pola data yang kompleks dan perubahan jumlah penumpang yang bersifat fluktuatif.

a. Fungsi prediksi

Proses optimasi model dilakukan menggunakan fungsi prediksi berikut [5]:

$$\hat{y}_i = \sum_{k=1}^{K} \quad f_k(x_i), f_k \in F \qquad (2)$$

Keterangan :
$\hat{y}_i$ = nilai prediksi untuk sampel data ke-i
K = jumlah keseluruhan pohon keputusan
F = ruang fungsi pohon keputusan (CART)
f_k = fungsi atau skor prediksi dari pohon keputusan ke-k

b. Fungsi Objektif

Untuk mengoptimalkan performa model selama proses pelatihan, dilakukan minimasi terhadap fungsi objektif (*objective function*) melalui persamaan berikut [5]:

$$Obj^{(t)} = \sum_{i=1}^{n} \quad L\left[y_i, \hat{y}_i^{(t-1)} + f_t(X_i)\right] + \Omega(f_t) \qquad (3)$$

Keterangan :
$Obj^{(t)}$ = total fungsi objektif yang diminimalkan
n = jumlah total sampel data penelitian
$l(y_i, \hat{y}_i)$ = fungsi kerugian (*loss function*) berbasis *squared error*
$\Omega(f_k)$ = fungsi regularisasi model

c. Fungsi Regulasi

Persamaan fungsi regularisasi untuk mengontrol kompleksitas model tersebut dirumuskan sebagai berikut [5]:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \quad \omega_j^2 \qquad (4)$$

Keterangan :
T = jumlah daun (*leaves*) pada pohon keputusan
ω_j = nilai bobot (*weight score*) pada daun pohon
γ = *hyperparameter penalty* untuk membatasi jumlah daun
λ = *hyperparameter* regularisasi L2 untuk mengontrol bobot daun

3.8. Proses Prediksi Model

Pada tahap ini, model yang telah dibangun menggunakan algoritma *Multiple Linear Regression* (MLR) dan *XGBoost Regression* digunakan untuk menghasilkan nilai prediksi jumlah penumpang kereta tipe *priority* berdasarkan pola hubungan antara variabel prediktor dan data historis. Proses prediksi dilakukan menggunakan aplikasi *RapidMiner* terhadap dataset yang telah melalui tahap pra-pemrosesan data. Hasil prediksi dari masing-masing algoritma kemudian digunakan sebagai dasar dalam proses evaluasi performa model.

3.9. Evaluasi Menggunakan sMAPE

Evaluasi model dilakukan menggunakan metode sMAPE guna mengetahui tingkat akurasi hasil prediksi sekaligus membandingkan performa dua algoritma *Supervised Learning*, yaitu *Multiple Linear Regression* (MLR) dan *Extreme Gradient Boosting Regression* (*XGBoost Regression*). Nilai sMAPE digunakan untuk mengetahui tingkat kesalahan prediksi terhadap data aktual, sehingga metode dengan nilai sMAPE paling rendah dianggap memiliki performa prediksi yang lebih baik. Rumus sMape dapat dinyatakan sebagai berikut [14]:

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^{n} \quad \frac{\frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}}{2} \qquad (5)$$

Keterangan:
n = Jumlah data
$\hat{y}_i$ = Nilai yang diprediksi
y_i = Nilai aktual

3.10. Perbandingan dan Analisis Hasil

Tahap akhir penelitian dilakukan dengan membandingkan performa model *Multiple Linear Regression* (MLR) dan *XGBoost Regression* berdasarkan nilai *Symmetric Mean Absolute Percentage*

Error (sMAPE) yang diperoleh dari masing-masing algoritma. Perbandingan dilakukan untuk mengetahui metode yang menghasilkan tingkat kesalahan prediksi paling rendah dalam memprediksi jumlah penumpang kereta wisata *priority*.

3.11. Kesimpulan

Pada hasil penelitian yang penulis lakukan, performa metode *Multiple Linear Regression* dan *XGBoost Regression* dibandingkan menggunakan matriks evaluasi sMAPE. Model yang menghasilkan nilai sMAPE terkecil dianggap memiliki performa prediksi yang lebih baik.

4. HASIL DAN PEMBAHASAN

Pada bab ini, penulis menyajikan seluruh hasil yang diperoleh dalam setiap tahapan penelitian. Pembahasan dimulai dari proses pengumpulan dan pra-pemrosesan data, identifikasi variabel prediktor, penerapan serta pengujian model prediksi, hingga diakhiri dengan evaluasi dan analisis hasil secara mendalam.

4.1. Pengumpulan Data

Berdasarkan proses pengumpulan data yang telah dilakukan, diperoleh data primer yang berasal dari PT Kereta Api Pariwisata melalui Pejabat Pengelola Informasi dan Dokumentasi (PPID). Rincian data jumlah penumpang yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Data Jumlah Penumpang

No	Bulan_Tahun	Jumlah Penumpang	*Load factor*
1.	Januari 2021	71	0.49%
2.	Februari 2021	413	2.84%
3.	Maret 2021	577	3.97%
...	…	…	…
59.	November 2025	8.355	57.54%
60.	Desember 2025	13.353	91.96%

Sumber : Penelitian (2026)

Selain itu, penulis juga menggunakan data sekunder yang didapat dari beberapa sumber eksternal. Data sekunder tersebut terdiri atas data inflasi, indeks Google Trends, dan *high season* yang digunakan sebagai variabel pendukung dalam proses pemodelan prediksi jumlah penumpang Kereta Wisata *Priority*. Adapun rincian data sekunder yang digunakan disajikan pada Tabel 2.

Tabel 2. Data Sekunder Penelitian

No.	Bulan_ Tahun	Inflasi	Indeks Google Trend	*High season*
1.	Januari 2021	0.26%	14	*High season*
2.	Februari 2021	0.10%	16	*Non-high season*
3.	Maret 2021	0.08%	22	*High season*
	…	…	…	...
59.	November 2025	0.17%	60	*Non-high season*
60.	Desember 2025	0.64%	84	*High season*

Sumber : [16], [17], [18]

Berdasarkan Tabel 2, adapun penjelasan mengenai masing-masing atribut yang digunakan pada penelitian ini dijabarkan sebagai berikut.

a. Inflasi sebagai indikator ekonomi yang menunjukkan persentase peningkatan harga barang dan jasa secara menyeluruh pada waktu tertentu.
b. Indeks Google Trends merupakan indeks yang menunjukkan tingkat popularitas pencarian suatu kata kunci pada mesin pencari Google.
c. *High season* merupakan periode yang ditandai dengan meningkatnya aktivitas perjalanan dan permintaan transportasi dibandingkan periode normal. Kondisi ini umumnya terjadi pada saat libur nasional, libur sekolah, maupun hari raya keagamaan.

4.2. Identifikasi Variabel Prediktor

Berdasarkan data yang telah dikumpulkan, penulis melakukan identifikasi variabel yang akan digunakan

dalam proses pemodelan prediksi. Penentuan variabel tersebut dilakukan berdasarkan ketersediaan data serta kesesuaiannya dengan tujuan penelitian. Adapun hasil identifikasi variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. Data Identifikasi Variabel

No	Atribut	Simbol	Jenis	Satuan/ Skala
1.	Jumlah_ Penumpang	Y	Target	Orang
2.	*Load_Factor*	X1	Prediktor	Persentase (%)
3.	Inflasi	X2	Prediktor	Persentase (%)
4.	Indeks_Google _Trend	X3	Prediktor	Indeks (1 Sampai 100)
5.	*High_Season*	X4	Prediktor	Biner (1 atau 0)

Sumber : Hasil Penelitian (2026)

Berdasarkan Tabel 3, jumlah penumpang ditetapkan sebagai variabel target karena penelitian ini bertujuan untuk memprediksi jumlah pengguna Kereta Wisata *Priority*. Sementara itu, *load factor*, *high season*, Indeks Google Trends, dan Inflasi digunakan sebagai variabel prediktor yang diduga memiliki pengaruh terhadap perubahan jumlah penumpang. Variabel *load factor* dipilih karena dapat menggambarkan tingkat keterisian kursi pada periode tertentu. Variabel *high season* digunakan untuk mengidentifikasi periode libur nasional, cuti bersama, dan musim liburan yang berpotensi meningkatkan aktivitas perjalanan masyarakat. Variabel Indeks Google Trends digunakan untuk menggambarkan tingkat ketertarikan masyarakat terhadap layanan Kereta *Priority* berdasarkan data pencarian "Kereta *Priority*" di Google. Sementara itu, tingkat inflasi digunakan untuk merepresentasikan kondisi ekonomi selama periode penelitian yang berpotensi memengaruhi jumlah penumpang kereta tipe *priority*.

4.3. Pra-pemrosesan Data

Berdasarkan data yang telah dikumpulkan, penulis melakukan tahap pra-pemrosesan untuk mempersiapkan data yang digunakan sebelum proses pemodelan dilakukan. Tahap ini mencakup pemeriksaan kelengkapan data dan transformasi variabel yang masih berbentuk kategorikal. Hasil pemeriksaan menunjukkan bahwa seluruh data penelitian lengkap sehingga tidak diperlukan penanganan *missing value*. Selain itu, variabel *High season* ditransformasikan ke dalam bentuk numerik menggunakan metode biner, yaitu nilai 1 untuk periode *High season* dan nilai 0 untuk periode *Non-High season*. Hasil transformasi data yang dilakukan dalam penelitian ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil transformasi variabel *High season*

Bulan	Tahun	High Season
1	2021	1
2	2021	0
3	2021	1
…	….	….
10	2025	0
11	2025	0
12	2025	1

Sumber : Hasil Penelitian (2026)

Setelah proses transformasi selesai, dataset siap digunakan dalam tahap pemodelan. Sebelum model dibangun, data terlebih dahulu dibagi menggunakan metode 5-*Fold Cross Validation*. Pada metode ini, seluruh data penelitian yang berjumlah 60 data bulanan periode Januari 2021 hingga Desember 2025 dibagi menjadi lima kelompok (*fold*), di mana setiap *fold* berisi 12 data. Pada setiap iterasi, empat *fold* diterapkan sebagai data pelatihan (*training data*) dan satu *fold* diterapkan sebagai data pengujian (*testing data*). Tahap ini dilakukan secara bergantian hingga seluruh *fold* memperoleh kesempatan menjadi data pengujian. Apabila data diurutkan berdasarkan periode waktu, maka pembagian data pada setiap iterasi dapat dilihat pada Tabel 5.

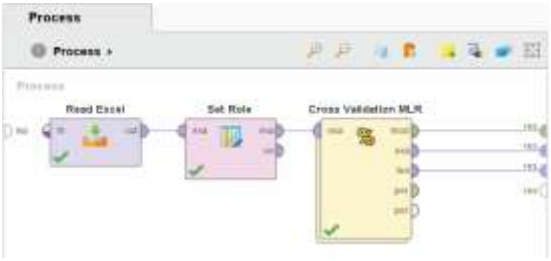
Tabel 5. Pembagian *Cross Validation*

Iterasi	Data *Training*	Data *Testing*
Fold 1	2022-2025	2021
Fold 2	2021, 2023-2025	2022
Fold 3	2021-2022, 2024-2025	2023
Fold 4	2021-2023,2025	2024
Fold 5	2021-2024	2025

Sumber : Hasil Penelitian (2026)

4.4. Implementasi *Multiple Linear Regression*

Pada tahap ini, penulis melakukan implementasi metode *Multiple Linear Regression* (MLR) menggunakan aplikasi *RapidMiner* untuk membangun model prediksi jumlah penumpang kereta tipe *priority*.

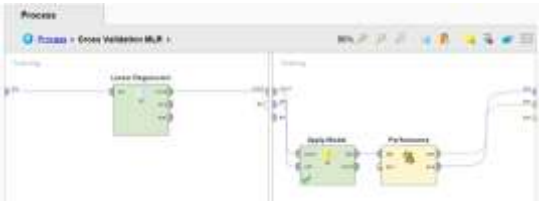


Gambar 3. Proses Implementasi MLR pada *RapidMiner*
Sumber : Hasil Penelitian (2026)

Berdasarkan Gambar 3, Proses implementasi diawali dengan memasukan dataset menggunakan operator *Read Excel*. Operator ini berfungsi untuk membaca dan mengimpor data penelitian dari file excel ke dalam *RapidMiner* sehingga dataset dapat digunakan sebagai masukan (*input*) dalam proses pemodelan. Dataset tersebut terdiri atas variabel jumlah penumpang, *load factor*, *high season*, Indeks Google Trends, dan tingkat inflasi yang digunakan dalam penelitian.

Selanjutnya, penulis menetapkan variabel jumlah penumpang sebagai variabel target (*label*) menggunakan operator *Set Role*. Operator ini berfungsi untuk menentukan variabel yang menjadi target prediksi (*label*) dan variabel yang digunakan sebagai pendukung dalam proses prediksi.

Tahap berikutnya adalah proses pemodelan menggunakan operator *Cross Validation* dengan metode *5-Fold Cross Validation.* Operator ini berfungsi untuk membagi dataset menjadi beberapa bagian (*fold*) yang digunakan secara bergantian sebagai data pelatihan (*training*) dan data pengujian (*testing*).



Gambar 4. Proses *Testing* dan *Training* MLR di dalam operator *Cross Validation*
Sumber : Hasil Penelitian (2026)

Berdasarkan Gambar 4, proses *Cross Validation* dijalankan untuk membagi tahapan pemodelan menjadi proses *training* dan *testing*. Pada bagian *training*, digunakan operator *Linear Regression* untuk membangun model prediksi. Fungsi dari operator ini adalah untuk mengetahui keterkaitan antara variabel pendukung dan variabel target berdasarkan data pelatihan sehingga menghasilkan model prediksi.

Kemudian pada bagian *testing*, model yang telah terbentuk diterapkan menggunakan operator *Apply Model.* Operator ini digunakan untuk memproses data pengujian berdasarkan model hasil pelatihan agar diperoleh nilai prediksi. Hasil prediksi tersebut kemudian dievaluasi menggunakan operator *Performance (Regression)* dengan membandingkan nilai prediksi terhadap nilai aktual. Operator *Performance (Regression*) berfungsi untuk menghitung nilai evaluasi model sehingga dapat diketahui tingkat kemampuan metode *Multiple Linear Regression* dalam memprediksi jumlah penumpang kereta tipe *priority*.

Tabel 6. Perbandingan Nilai Aktual dan Nilai Hasil Prediksi MLR

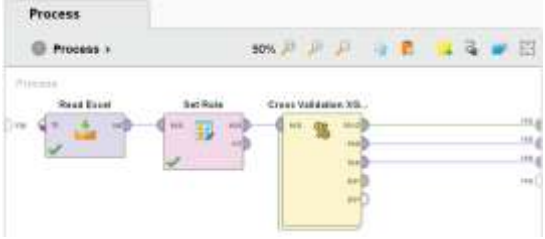
No	Bulan_ Tahun	Nilai Aktual	Nilai Prediksi
1	Januari 2021	71	71,098
2	Februari 2021	413	412,306
3	Maret 2021	577	576,385
…	….	….	….
59	November 2025	8.355	8.354,80
60	Desember 2025	13.353	13.352,62
Jumlah Nilai		222.911	222.909,613
Selisih			1,387033912

Sumber : Hasil Penelitian (2026)

Berdasarkan Tabel 6. nilai prediksi yang dihasilkan oleh metode ML menunjukkan kesesuaian yang sangat baik dengan nilai aktual jumlah penumpang Kereta *Priority* pada setiap periode pengamatan. Hal tersebut terlihat dari nilai prediksi yang berada sangat dekat dengan nilai aktual. Secara keseluruhan, jumlah nilai aktual mencapai 222.911 penumpang, sedangkan jumlah nilai prediksi sebesar 222.909,613 penumpang dengan selisih sebesar 1,387033912. Selisih yang sangat kecil tersebut menunjukkan bahwa model MLR mampu menghasilkan prediksi yang mendekati nilai aktual sehingga memiliki tingkat akurasi yang tinggi.

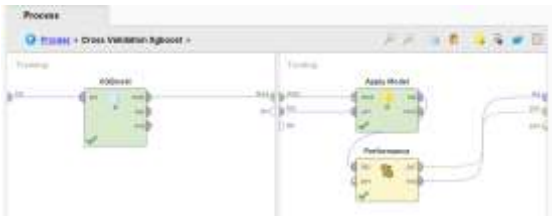
4.5. Implementasi *XGBoost Regression*

Implementasi *XGBoost Regression* pada *RapidMiner* dilakukan menggunakan parameter yang telah ditentukan sebelumnya. Tampilan proses implementasi model direpresentasikan melalui Gambar 5.



Gambar 5. Proses Implementasi *XGBoost* pada *RapidMiner*
Sumber : Hasil Penelitian (2026)

Berdasarkan Gambar 5. proses implementasi ini memiliki tahapan yang sama dengan proses pada metode MLR, yaitu dimulai dengan mengimpor dataset ke dalam *RapidMiner* menggunakan operator *Read Excel*. Selanjutnya, atribut jumlah penumpang ditetapkan sebagai label menggunakan operator *Set Role*. Pengujian model dilakukan menggunakan operator *Cross Validation* dengan metode *K-Fold* sebanyak 5 lipatan (*5-Fold Cross Validation*) yang secara otomatis membagi data menjadi data pelatihan dan data pengujian yang dilakukan secara bergantian pada setiap lipatan (*fold*).



Gambar 6. Proses *Testing* dan *Training XGBoost* di dalam operator *Cross Validation*
Sumber : Hasil Penelitian (2026)

Berdasarkan Gambar 6, di dalam operator *Cross Validation*, pada bagian *training*, digunakan operator *XGBoost Regression* untuk membangun model prediksi, sedangkan pada bagian *testing* digunakan operator *Apply Model* untuk menghasilkan nilai prediksi. Hasil prediksi kemudian dievaluasi menggunakan operator *Performance* untuk memperoleh nilai kinerja model.

Tabel 7. Perbandingan Nilai Aktual dan Nilai Hasil Prediksi *XGBoost*

No	Bulan_ Tahun	Nilai Aktual	Nilai Prediksi
1	Januari 2021	71	569,2179
2	Februari 2021	413	415,054
3	Maret 2021	577	402,1779
…	….	….	….
59	November 2025	8.355	6.637,573
60	Desember 2025	13.353	14.103,2

Jumlah Nilai	222.911	216.081,0095
Selisih		6.829,9905

Sumber : Hasil Penelitian (2026)

Tabel 7. menunjukkan perbandingan data aktual dengan hasil prediksi yang diperoleh menggunakan metode *XGBoost Regression*. Secara keseluruhan, jumlah nilai prediksi yang dihasilkan sebesar 216.081,0095 penumpang, sedangkan jumlah nilai aktual mencapai 222.911 penumpang. Terdapat selisih sebesar 6.829,9905 penumpang antara nilai prediksi dan nilai aktual. Selisih tersebut menunjukkan bahwa model *XGBoost Regression* menghasilkan selisih yang cukup jauh untuk hasil prediksi dibandingkan dengan nilai aktual.

4.6. Proses Prediksi Model

Setelah kedua model dibangun, dilakukan proses prediksi jumlah penumpang kereta tipe *priority* menggunakan masing-masing model. Hasil prediksi yang dihasilkan oleh model *Multiple Linear Regression* dan *XGBoost Regression* selanjutnya dibandingkan dengan nilai aktual jumlah penumpang. Adapun hasil prediksi dari kedua model direpresentasikan melalui Gambar 7.

Gambar 7. Diagram garis perbandingan hasil prediksi MLR dan *XGBoost* dengan nilai aktual

Sumber : Hasil Penelitian (2026)

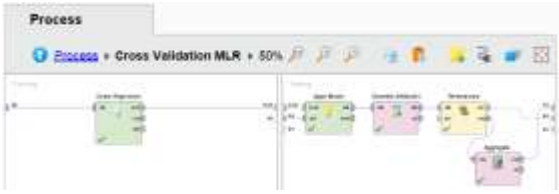
Berdasarkan Gambar 7. terlihat bahwa kedua model mampu mengikuti pola perubahan jumlah penumpang selama periode pengamatan. Namun, model *Multiple Linear Regression* menunjukkan tingkat kedekatan yang lebih tinggi terhadap data aktual dibandingkan model *XGBoost Regression*. Hal ini dapat dilihat dari kurva prediksi MLR (garis biru) yang sebagian besar berada sangat dekat dan hampir menutupi kurva data aktual (garis hijau) pada berbagai periode. Kondisi tersebut menunjukkan bahwa model MLR memiliki selisih (*error*) prediksi yang lebih rendah, sehingga lebih akurat dalam memperkirakan jumlah penumpang dibandingkan dengan model *XGBoost* (garis kuning).

Perbedaan yang signifikan ditunjukkan oleh prediksi *XGBoost* pada beberapa periode bulan tertentu, di mana garis kuning tampak menyimpang atau berada cukup jauh di bawah kurva data aktual dan MLR. Hal ini terlihat pada Desember 2023, ketika *XGBoost* tidak mampu mengikuti lonjakan tertinggi jumlah penumpang. Selain itu, pada periode April, Juni, Juli 2025 serta September hingga November 2025, prediksi *XGBoost* cenderung lebih rendah dan relatif datar, dibandingkan dengan MLR yang lebih dekat mengikuti data aktual.

4.7. Evaluasi Menggunakan sMAPE

Setelah diperoleh hasil prediksi dari metode *Multiple Linear Regression* dan *XGBoost Regression*, tahap selanjutnya adalah melakukan evaluasi performa model menggunakan metode *Symmetric Mean Absolute Percentage Error* (sMAPE). Penggunaan sMAPE diterapkan guna mengevaluasi tingkat kesalahan prediksi dengan membandingkan nilai aktual terhadap nilai prediksi yang dihasilkan oleh setiap model. Semakin kecil nilai sMAPE yang diperoleh, semakin baik kemampuan model dalam menghasilkan prediksi yang mendekati data aktual.

Untuk mengetahui tingkat akurasi model yang dihasilkan, dilakukan perhitungan nilai sMAPE pada *RapidMiner*. Perhitungan tersebut dilakukan melalui rangkaian operator yang terdapat pada bagian *testing subprocess* dalam *Cross Validation*. Proses perhitungan tersebut direpresentasikan pada Gambar 8.



Gambar 8. Proses perhitungan sMape menggunakan operator *Generate Atributte*s di dalam operator *Cross Validation*
Sumber : Hasil Penelitian (2026)

Berdasarkan Gambar 8. Perhitungan sMAPE dilakukan menggunakan operator *Generate Attributes* yang berfungsi untuk membuat atribut atau kolom baru berdasarkan perhitungan tertentu. Pada operator ini, kolom *Column Name* diisi dengan nama atribut baru yaitu sMAPE yang digunakan untuk menyimpan hasil perhitungan sMAPE pada setiap data. Selanjutnya, pada kolom *Function Expression* dimasukkan formula sMAPE yang menggunakan nilai aktual dan nilai prediksi sebagai dasar perhitungan. Formula sMAPE yang dimasukkan ke dalam operator dapat dilihat pada Gambar 9.

Gambar 9. Formula sMape pada kolom *Function Expression*
Sumber : Hasil Penelitian (2026)

Formula tersebut digunakan untuk menghitung nilai sMAPE pada setiap data hasil prediksi. Hasil perhitungan yang diperoleh kemudian disimpan sebagai atribut baru dan selanjutnya digunakan dalam proses perhitungan rata-rata sMAPE menggunakan operator *Aggregate*. Operator ini berfungsi untuk melakukan perhitungan ringkasan data, seperti menghitung nilai rata-rata dari atribut tertentu. Adapun hasil dari sMAPE untuk setiap *K Fold* dapat dilihat pada tabel 8.

Tabel 8. Hasil perbandingan akurasi berdasarkan *K Fold* 5 sMape

Iterasi	SMAPE MLR	SMAPE XGBOOST
Fold 1	0.003 %	16.209 %
Fold 2	0.049 %	40.968 %
Fold 3	0.039 %	8.548 %
Fold 4	0.002 %	10.316 %
Fold 5	0.020 %	6.886 %
Average	0.0226 %	16.5854 %

Sumber : Hasil Penelitian (2026)

4.8. Perbandingan dan Analisis Hasil

Berdasarkan hasil pengujian yang telah dilakukan terhadap kedua algoritma dalam penelitian ini, didapatkan hasil bahwa metode *Multiple Linear Regression* (MLR) menunjukkan performa prediksi yang lebih unggul dibandingkan dengan algoritma *XGBoost Regression*. Keunggulan performa ini ditunjukkan secara signifikan melalui nilai tingkat kesalahan prediksi *Symmetric Mean Absolute Percentage Error* (sMAPE) model MLR yang lebih kecil, yaitu sebesar 0.0226 %, sedangkan *XGBoost Regression* menghasilkan nilai error yang lebih tinggi, yaitu sebesar 16.5854%.

Tabel 9. Hasil perbandingan akurasi sMAPE

Model	Nilai SMAPE
Multiple Linear Regression	0.0226 %
Extreme Gradient Boosting Regression	16.5854 %

Sumber : Hasil Penelitian

Berdasarkan Tabel 9. model *Multiple Linear Regression* memperoleh nilai sMAPE yang paling rendah dibandingkan model *XGBoost Regression*. Hasil tersebut menunjukkan bahwa *Multiple Linear Regression* mampu memberikan tingkat akurasi prediksi yang lebih baik pada dataset penelitian ini. Performa tersebut menunjukkan bahwa pola hubungan antara variabel prediktor dan jumlah penumpang

dapat direpresentasikan dengan baik menggunakan pendekatan regresi linear, sehingga model mampu menghasilkan prediksi yang lebih mendekati nilai aktual.

Di sisi lain, *XGBoost Regression* merupakan algoritma berbasis *ensemble learning* yang dirancang untuk menangani hubungan data yang lebih kompleks melalui kombinasi beberapa *decision tree*. Namun pada penelitian ini, kompleksitas model tersebut tidak menghasilkan peningkatan performa prediksi. Hal ini terlihat dari nilai sMAPE yang diperoleh *XGBoost Regression* yang masih lebih tinggi dibandingkan *Multiple Linear Regression*.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah penulis lakukan mengenai prediksi jumlah penumpang kereta tipe *priority* dengan membandingkan performa dua algoritma *Supervised Learning*, yaitu *Multiple Linear Regression* (MLR) dan *Extreme Gradient Boosting Regression* (*XGBoost Regression*) dengan melakukan pengukuran akurasi menggunakan metode sMAPE, diperoleh kesimpulan sebagai berikut:

a. Penelitian ini memanfaatkan data primer berupa data jumlah penumpang (Y) Kereta *Priority* per bulan selama periode 2021–2025 serta data *load factor* (X_1). Selain itu, digunakan data sekunder berupa tingkat inflasi (X_2), indeks Google Trends (X_3), dan *high season* (X_4).
b. Pra pemrosesan data pada penelitian ini mencakup pengujian data *training* dan data *testing* dengan menggunakan model validasi 5-*Fold Cross Validation*.
c. Hasil penelitian ini menunjukan nilai prediksi keseluruhan untuk metode *Multiple Linear Regression* sebesar 222.909,613 sedangkan metode *XGBoost Regression* 216.081,0095.
d. Hasil selisih untuk metode metode *Multiple Linear Regression* sebesar 1,387033912 sedangkan metode *XGBoost Regression* 6.829,9905 dari nilai aktual sebesar 222.911.
e. Hasil pengujian metode prediksi menggunakan sMAPE didapat angka untuk metode metode *Multiple Linear Regression* sebesar 0,0226%, sedangkan metode *XGBoost Regression* 16,5854% yang artinya penerapan metode *machine learning* dalam melakukan prediksi untuk menentukan jumlah penumpang yang lebih cocok adalah metode *Multiple Linear Regression* dikarenakan mendekati nilai 0 dan memberikan tingkat akurasi yang lebih tinggi dibandingkan dengan metode *XGBoost Regression* .
f. Penelitian selanjutnya dapat dilakukan dengan menambahkan jumlah data dan variabel prediktor yang lebih beragam untuk meningkatkan pengujian model dalam melakukan prediksi.
g. Dapat dilakukan penelitian lanjutan berupa melakukan pengujian terhadap algoritma *machine learning* lainnya seperti *Random Forest Regression*, *Support Vector Regression* (SVR), atau algoritma berbasis *deep learning* untuk memperoleh perbandingan performa yang lebih luas dalam memprediksi jumlah penumpang kereta tipe *priority*.

UCAPAN TERIMA KASIH

Penelitian ini dapat diselesaikan berkat dukungan, kerja sama, serta kontribusi dari berbagai pihak. Sehubungan dengan hal tersebut, penulis menyampaikan apresiasi dan terima kasih kepada:

a. Pejabat pengelola informasi dan dokumentasi (PPID) yaitu Bima Laputro dan Arief Rahman serta seluruh insan PT KAI Wisata atas izin risetnya.
b. Ibu Hidayanti Murtina, M.Kom dan Bapak Aryo Tunjung Kusumo, M.Kom selaku dosen pembimbing yang senantiasa memberikan arahan,

bimbingan, serta berbagai masukan selama penyusunan dan pelaksanaan penelitian ini.

c. Seluruh anggota kelompok penelitian yang telah berkolaborasi dan berkontribusi dalam mendukung penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] Z. Bunga *et al.*, "Perbandingan Metode Machine Learning (Linear Regression , Random Forest , dan *XGBoost*) dalam memprediksi Kemiskinan di Jawa Tengah Tahun 2024 Comparison of *Machine learning* Methods (Linear Regression , Random Forest , and *XGBoost*) for Predicting Poverty in Central Java in 2024," vol. 14, pp. 2492–2499, 2025.

[2] "Analisis Perbandingan Kinerja Algoritma Linear Regression , Random Forest , dan *XGBoost* dalam Prediksi Harga Rumah," vol. 4, no. 4, pp. 1541–1548, 2025.

[3] S. D. Cahyo, A. A. Murtopo, and B. A. Santoso, "Penerapan Metode Regresi Linier untuk Prediksi Jumlah Penumpang Kereta Api," vol. 4, no. 3, pp. 986–993, 2025.

[4] R. S. Nurhalizah and R. Ardianto, "Analisis Supervised dan Unsupervised Learning pada Machine Learning : Systematic Literature Review," vol. 4, no. 1, pp. 61–72, 2024.

[5] I. G. S. M. D. Dyasa, F. Umam, V. H. Satria, H. Sukri, and F. Adiputra, *MACHINE LEARNING*. Malang: Media Nusa Creative, 2026.

[6] W. Andriyani, R. Purnomo, S. A. Hendrawan, A. I. Irvani, A. Sujarno, and Y. N. Asri, *Data Sebagai Fondasi Kecerdasan Buatan*. Makassar: CV. Tohar Media, 2024.

[7] A. N. Khudori, W. T. Kusuma, P. Korespondensi, and P. Manusia, "PERBANDINGAN METODE SUPERVISED MACHINE LEARNING UNTUK PREDIKSI PREVALENSI STUNTING DI PROVINSI JAWA TIMUR SUPERVISED MACHINE LEARNING METHODS COMPARISON ON STUNTING," vol. 9, no. 7, pp. 0–5, 2022, doi: 10.25126/jtiik.202296744.

[8] R. Swatika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *IMPLEMENTASI DATA MINING (CLASTERING, ASSOCIATION, PREDICTION, ESTIMATION, CLASSIFICATION)*. Indramayu: CV. Adanu Abimata, 2023.

[9] C. Rosanti, F. A. Artanto, R. A. Saputra, and E. Subowo, *Machine learning dengan Regresi untuk Analisa Ekonomi Syariah dari Sentimen Media Sosial*. Penerbit NEM, 2024.

[10] M. Hasanah, N. H. Harani, and N. Riza, *IMPLEMENTASI BARCODE DAN ALGORITMA REGRESI LINIER UNTUK MEMPREDIKSI DATA PERSEDIAAN BARANG*. Bandung: Kreatif Industri Nusantara, 2020.

[11] I. R. Hendrawan and E. Utami, *Natural Languange Processing (Eksplorasi Sentimen Masyarakat dalam Evaluasi Produk Lokal Indonesia menggunakan Algoritma Bag of Words*. Yogyakarta: Penerbit ANDI, 2023.

[12] M. Imam, N. Pasaribu, U. H. Medan, L. P. Penduduk, and R. Linier, "Implementasi Data Mining Dengan Metode Regresi Linier Untuk Memprediksi Laju Pertumbuhan Penduduk Kota Binjai," no. September, pp. 405–415, 2025.

[13] A. F. Rozy, N. Wayan, and S. Wardhani, "PENINGKATAN AKURASI METODE WEIGHTED FUZZY TIME SERIES FORECAST ING MENGGUNAKAN ALGORITMA EVOLUSI DIFFERENSIAL DAN FUZZY C-MEANS IMPROVING THE ACCURACY OF WEIGHTED FUZZY TIME SERIES FORECASTING METHOD USING DIFFERENTIAL EVOLUTION ALGORITHM AND FUZZY C-MEANS," vol. 10, no. 5, pp. 1047–1054, 2023, doi: 10.25126/jtiik.2023107505.

[14] Nurjaya, *FUNDAMNETAL MACHINE LEARNING*. Purbalingga: EUREKA MEDIA ASKARA, 2025.

[15] J. E. Boylan and A. A. Syntetos, *INTERMITTENT DEMAND FORECASTING (CONTEXT, METHODS AND APPLICATION)*. West Sussex: John Wiley & Sons Ltd, 2021.

[16] B. P. Statistik, "Inflasi Bulanan (M-to-M) (Persen) 2021-2025," *Badan Pusat Statistik*. https://www.bps.go.id/id/statistics-table/2/MSMy/inflasi-month-to-month--april-2026.html

[17] Google, "Google Trends Kereta *Priority*," *Google Trend*. https://trends.google.com/explore?q=kereta%2520*priority*&date=today 5-y&geo=ID

[18] Kalenderku.id., "Kalender Tahun 2021–2025," *Kalenderku*. https://kalenderku.id/2025#google_vignette

[19] P. H. Marpaung, M. Darwis, Khairunnisah, and M. Pardamean, *RapidMiner:Solusi Belajar Data Mining*. Padangsisimpuan: CV. Alfa Pustaka, 2024.

[20] Marji, E. Santoso, M. M. Arifin, and E. W. Handamari, *APLIKASI MACHINE LEARNING PEMROGRAMAN PYTHON*. Malang: UB Press, 2024.

[21] R. M. Adhim, V. N. Sulistyawan, and S. Sukamta, "Prediksi Downlink Throughput Menggunakan Metode Random Forest Regression Pada Jaringan 4G LTE Berdasarkan Data," vol. 6, no. , 2025.