

Analisis Dan Prediksi Kecepatan Jaringan Menggunakan Random Forest Regression Pada Dataset Lalu Lintas Jaringan

Arliandi Achmad^{1*}, Herdy Ashaury², Melina³

¹Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani

²Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani

³Program Studi Teknik Informatika, Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani

*Penulis Korespondensi: arliandiachmad@gmail.com

Abstract. *The increasing use of the internet demands the availability of stable and high-quality network services. Changes in network traffic over each time period cause network speed to be dynamic, requiring an approach capable of predicting network speed based on historical patterns. This study aims to analyze the characteristics of Traffic Total Speed and predict internet network speed using the Random Forest Regression algorithm based on historical network traffic data at Universitas Jenderal Achmad Yani (Unjani). The research methodology includes Exploratory Data Analysis (EDA), preprocessing, feature engineering, lag feature construction, Random Forest Regression modeling, and model evaluation using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), coefficient of determination (R^2), and 5-Fold TimeSeriesSplit. The initial dataset consisted of 722 observations, and after preprocessing and feature engineering, 716 observations were obtained. The data were then divided chronologically using a time-based split with an 80:20 ratio, consisting of 572 training data and 144 testing data. The testing results show that the model achieved an MAE of 121.03 Mbit/s, an MSE of 43,546.19 (Mbit/s)², an RMSE of 208.68 Mbit/s, a MAPE of 38.31%, and an R^2 of 0.8892. Evaluation using 5-Fold TimeSeriesSplit resulted in an average R^2 of 0.8049 $\pm$ 0.1178, an MAE of 171.86 $\pm$ 64.70 Mbit/s, and an RMSE of 286.25 $\pm$ 87.26 Mbit/s. The R^2 values for each fold were 0.5823, 0.8329, 0.8080, 0.8825, and 0.9188, respectively. Based on the feature importance analysis, Volume_lag1 was the most important feature with an importance score of 0.3512, followed by Speed_lag1 with a score of 0.1767. The results indicate that Random Forest Regression is capable of predicting Traffic Total Speed with reasonably good performance, although prediction errors remain higher during periods of significant changes in network speed, particularly during peak hours.*

Keywords: *Forest Regression; network speed prediction; Traffic Total Speed; time series; feature importance.*

Abstrak. Pertumbuhan penggunaan internet menuntut ketersediaan jaringan yang stabil dan berkualitas. Perubahan lalu lintas jaringan pada setiap periode waktu menyebabkan kecepatan jaringan bersifat dinamis sehingga diperlukan pendekatan yang mampu memprediksi kecepatan jaringan berdasarkan pola historis. Penelitian ini bertujuan untuk menganalisis karakteristik Traffic Total Speed serta memprediksi kecepatan jaringan internet menggunakan algoritma Random Forest Regression berdasarkan data historis lalu lintas jaringan di Universitas Jenderal Achmad Yani (Unjani). Metode penelitian meliputi Exploratory Data Analysis (EDA), preprocessing, feature engineering, pembentukan lag feature, pemodelan Random Forest Regression, dan evaluasi menggunakan Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), koefisien determinasi (R^2), serta TimeSeriesSplit 5-Fold. Dataset awal terdiri atas 722 observasi dan setelah proses preprocessing serta feature engineering diperoleh 716 observasi. Data kemudian dibagi secara kronologis menggunakan time-based split dengan rasio 80:20, yaitu 572 data training dan 144 data testing. Hasil pengujian pada data testing menunjukkan bahwa model memperoleh nilai MAE sebesar 121,03 Mbit/s, MSE sebesar 43.546,19 (Mbit/s)², RMSE sebesar 208,68 Mbit/s, MAPE sebesar 38,31%, dan R^2 sebesar 0,8892. Evaluasi menggunakan TimeSeriesSplit 5-Fold menghasilkan rata-rata R^2 sebesar 0,8049 $\pm$ 0,1178, MAE sebesar 171,86 $\pm$ 64,70 Mbit/s, dan RMSE sebesar 286,25 $\pm$ 87,26 Mbit/s. Nilai R^2 pada masing-masing fold

berturut-turut sebesar 0,5823, 0,8329, 0,8080, 0,8825, dan 0,9188. Berdasarkan analisis feature importance, Volume_lag1 merupakan fitur dengan tingkat kepentingan tertinggi sebesar 0,3512, diikuti Speed_lag1 sebesar 0,1767. Hasil penelitian menunjukkan bahwa Random Forest Regression mampu memprediksi Traffic Total Speed dengan performa yang cukup baik, meskipun terdapat variasi performa antarperiode dan kesalahan prediksi yang lebih tinggi pada periode dengan perubahan kecepatan jaringan yang signifikan, terutama pada jam sibuk.

Kata kunci: Random Forest Regression; prediksi kecepatan jaringan; Traffic Total Speed; *time series*; *feature importance*.

1. LATAR BELAKANG

Perkembangan teknologi digital yang semakin pesat mendorong meningkatnya kebutuhan terhadap layanan jaringan komputer yang cepat dan stabil. Aktivitas seperti streaming, konferensi daring, layanan komputasi awan (*cloud computing*), transaksi digital, serta pengiriman data berukuran besar sangat bergantung pada performa jaringan, khususnya kecepatan jaringan internet (*internet speed*). Kecepatan jaringan menjadi indikator penting dalam menilai kualitas layanan (*Quality of Service/QoS*) karena berpengaruh langsung terhadap kenyamanan, efisiensi, dan efektivitas penggunaan layanan digital (Putri, Sutomo, Informatika, Metro, & Forest, n.d.; Syarifatun, Miyarno, Fatwa, & Rizqi, 2024).

Dalam praktiknya, kecepatan jaringan internet tidak selalu berada pada kondisi stabil, melainkan dapat mengalami variasi performa pada kondisi tertentu. Evaluasi kinerja jaringan menunjukkan bahwa parameter kualitas layanan seperti *throughput*, *delay*, *jitter*, dan *packet loss* dapat berubah secara dinamis tergantung pada tingkat penggunaan jaringan dan kepadatan lalu lintas data (Li, 2025; Syafrudin, Rianto, Ujianto, Informasi, & Yogyakarta, 2025). Perubahan kondisi tersebut menyebabkan performa jaringan sulit diprediksi hanya berdasarkan kondisi saat ini. Oleh karena itu, diperlukan suatu pendekatan yang mampu memanfaatkan data historis untuk memperkirakan kondisi jaringan pada waktu berikutnya sehingga pengelolaan jaringan dapat dilakukan secara lebih proaktif (Saha, Haque, & Sidebottom, 2024).

Sejumlah penelitian terdahulu menunjukkan efektivitas Random Forest dalam analisis data jaringan. Penelitian oleh Putri menunjukkan bahwa algoritma Random Forest mampu memprediksi kemacetan lalu lintas jaringan menggunakan data lalu lintas jaringan, sehingga dapat membantu meningkatkan pengelolaan dan kinerja jaringan (Putri et al., n.d.). Penelitian lain oleh Adhim dkk. menunjukkan bahwa metode Random Forest Regression mampu membangun model prediksi downlink throughput pada jaringan 4G

LTE dengan tingkat kesalahan yang rendah dan stabilitas yang tinggi (Test, 2025). Selain itu, penelitian oleh Nisa dkk. menganalisis kualitas layanan jaringan menggunakan parameter QoS berupa *throughput*, *packet loss*, *delay*, dan *jitter* berdasarkan standar TIPHON. Hasil penelitiannya menunjukkan bahwa kualitas jaringan secara keseluruhan berada pada kategori baik (Syarifatun et al., 2024).

Berdasarkan berbagai penelitian terdahulu tersebut, sebagian besar penelitian masih berfokus pada analisis kualitas jaringan (Quality of Service) (Putri et al., n.d.; Syarifatun et al., 2024), prediksi *throughput*, maupun prediksi *traffic* jaringan. Penelitian-penelitian tersebut belum banyak mengembangkan pendekatan *forecasting* kecepatan internet menggunakan data historis yang mempertimbangkan pola temporal pada berbagai parameter jaringan secara bersamaan. Selain itu, penelitian yang memanfaatkan algoritma Random Forest Regression untuk melakukan *forecasting* kecepatan internet berbasis data *time series* masih relatif terbatas.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan penggunaan algoritma Random Forest Regression untuk melakukan *forecasting* kecepatan jaringan internet berdasarkan data historis. Random Forest dipilih karena mampu menangani hubungan nonlinier antarvariabel, memiliki kemampuan generalisasi yang baik, tahan terhadap noise dan *overfitting* (Saha et al., 2024). Dengan memanfaatkan data yang memiliki informasi waktu (*timestamp*), model diharapkan mampu menghasilkan prediksi kecepatan jaringan pada periode berikutnya sehingga dapat mendukung pengelolaan jaringan yang lebih adaptif dan efisien.

Berdasarkan berbagai permasalahan yang telah diuraikan serta hasil kajian terhadap penelitian terdahulu, Penelitian ini berfokus pada analisis dan prediksi kecepatan jaringan internet berdasarkan karakteristik temporal pada data *time series* dengan memanfaatkan algoritma Random Forest Regression. Data historis lalu lintas jaringan digunakan untuk menganalisis pola perubahan kecepatan jaringan dari waktu ke waktu dengan tetap mempertahankan urutan temporal antarperiode pengamatan. Pendekatan tersebut memungkinkan model memanfaatkan informasi historis untuk menghasilkan prediksi kecepatan jaringan pada periode berikutnya. Dengan demikian, hasil penelitian ini diharapkan dapat menjadi salah satu alternatif dalam mendukung proses pemantauan evaluasi.

2. KAJIAN TEORITIS

A. Random Forest Regression

Random Forest Regression merupakan algoritma *ensemble* yang membangun banyak *decision tree* dan menggabungkan hasilnya untuk meningkatkan stabilitas, mengurangi variansi, dan menghasilkan prediksi yang lebih akurat. Pada regresi, Random Forest Regression memanfaatkan nilai rata-rata dari seluruh *decision tree* untuk memprediksi variabel numerik. Konsep tersebut pertama kali diperkenalkan oleh Breiman dalam algoritma Random Forest (Breiman, 2001). Algoritma ini memiliki keunggulan dalam menangani data berdimensi besar, fitur beragam, serta pola non-linear, sehingga mengurangi risiko *overfitting* dan meningkatkan kinerja model dalam analisis data jaringan (Rodicky & Soim, 2025). Persamaan prediksi Random Forest Regression diberikan pada Persamaan (2.3) (Breiman, 2001).

$$\hat{y}(x) = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (2.3)$$

Dimana

$\hat{y}(x)$: nilai prediksi akhir Random Forest Regression untuk data input x .

T : jumlah pohon keputusan (*decision tree*) dalam model Random Forest.

$h_t(x)$: nilai prediksi yang dihasilkan oleh pohon keputusan ke- t terhadap input x .

t : indeks pohon keputusan, dengan $t = 1, 2, \dots, T$.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional (*computational experiment*) dalam kerangka *supervised machine learning*. Penelitian ini menggunakan data primer yang diperoleh dari hasil monitoring historis lalu lintas jaringan di Universitas Jenderal Achmad Yani (UNJANI). Tahap penelitian diawali dengan data preprocessing yang meliputi pemeriksaan kualitas data, penyesuaian nilai dan tipe data, penanganan outlier, serta feature engineering. Selanjutnya dilakukan Exploratory Data Analysis (EDA) untuk memahami karakteristik data serta mengidentifikasi pola perubahan kecepatan jaringan berdasarkan waktu pengamatan sebagai dasar proses pemodelan. Model prediksi dibangun menggunakan *Random Forest Regression* yang mampu memodelkan hubungan nonlinier antarvariabel serta memiliki kemampuan generalisasi yang baik melalui pendekatan *ensemble learning*.

Selanjutnya, performa model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *koefisien determinasi* (R^2).

4. HASIL DAN PEMBAHASAN

A. Pengambilan Data

Pada penelitian ini dilakukan analisis dan prediksi kecepatan jaringan menggunakan algoritma Random Forest Regression berdasarkan data historis lalu lintas jaringan yang diperoleh dari hasil pemantauan jaringan di Universitas Jenderal Achmad Yani (Unjani). Dataset yang digunakan memiliki karakteristik *time series*, di mana setiap observasi direkam berdasarkan interval waktu tertentu sehingga setiap data memiliki keterkaitan dengan observasi pada periode sebelumnya. Data tersebut merepresentasikan kondisi lalu lintas jaringan selama periode pengamatan dan digunakan sebagai dasar dalam membangun model prediksi kecepatan jaringan pada satu jam berikutnya. Dataset awal terdiri atas 722 observasi dengan sembilan variabel utama, yaitu Date Time, Traffic Total (volume), Traffic Total (speed), Traffic In (volume), Traffic In (speed), Traffic Out (volume), Traffic Out (speed), Downtime, dan Coverage.

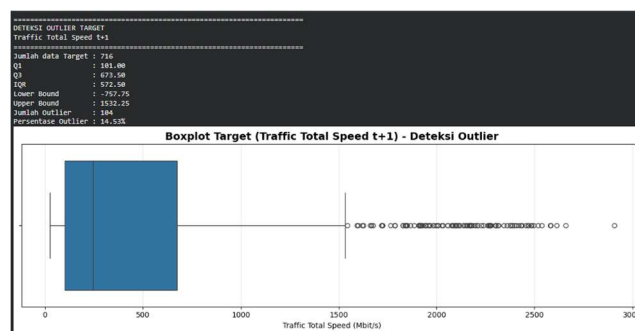
Tabel 1 Deskripsi Data

No	Variabel	Deskripsi
1	Date Time	Menunjukkan periode waktu pengamatan (<i>timestamp</i>) yang merepresentasikan waktu pencatatan kondisi lalu lintas jaringan. Variabel ini menjadi acuan dalam analisis data <i>time series</i> serta digunakan untuk membentuk fitur temporal seperti <i>Hour</i> , <i>Day</i> , dan <i>DayOfWeek</i> .
2	Traffic Total (volume)	Total volume lalu lintas data yang melewati jaringan selama satu periode pengamatan. Nilai ini merupakan akumulasi data masuk (<i>Traffic In</i>) dan data keluar (<i>Traffic Out</i>) yang dinyatakan dalam satuan megabyte (MB).
3	Traffic Total (speed)	Kecepatan rata-rata total lalu lintas jaringan pada setiap periode pengamatan yang dinyatakan dalam satuan megabit per detik (Mbit/s). Variabel ini digunakan sebagai variabel target (output) dalam proses pembangunan model prediksi.
4	Traffic In (volume)	Total volume data yang diterima (<i>incoming traffic</i>) oleh jaringan selama satu periode pengamatan, dinyatakan dalam satuan megabyte (MB).
5	Traffic In (speed)	Kecepatan rata-rata lalu lintas data yang masuk ke jaringan pada setiap periode pengamatan, dinyatakan dalam satuan megabit per detik (Mbit/s).
6	Traffic Out	Total volume data yang dikirim (<i>outgoing traffic</i>) dari jaringan

	(volume)	selama satu periode pengamatan, dinyatakan dalam satuan megabyte (MB).
7	Traffic Out (speed)	Kecepatan rata-rata lalu lintas data yang keluar dari jaringan pada setiap periode pengamatan, dinyatakan dalam satuan megabit per detik (Mbit/s).
8	Downtime	Persentase waktu ketika jaringan mengalami gangguan atau tidak dapat beroperasi selama periode pengamatan.
9	Coverage	Persentase tingkat ketersediaan (<i>availability</i>) atau cakupan layanan jaringan selama periode pengamatan.

B. Exploratory Data Analysis (EDA)

Berdasarkan hasil Exploratory Data Analysis (EDA), variabel target Traffic Total Speed ($t+1$) yang digunakan dalam penelitian ini terdiri dari 716 data hasil proses preprocessing dan pembentukan target. Berdasarkan metode Interquartile Range (IQR) diperoleh nilai Q1 sebesar 101,00 Mbit/s, Q3 sebesar 673,50 Mbit/s, sehingga menghasilkan IQR sebesar 572,50 Mbit/s. Dari hasil tersebut diperoleh batas bawah (lower bound) sebesar -757,75 Mbit/s dan batas atas (upper bound) sebesar 1532,25 Mbit/s. Visualisasi boxplot menunjukkan terdapat 104 data outlier atau sekitar 14,53% dari keseluruhan data. Sebagian besar outlier berada pada nilai Traffic Total Speed yang tinggi sehingga distribusi data cenderung right-skewed (miring ke kanan). Outlier tersebut tidak dihapus karena masih merepresentasikan kondisi lalu lintas jaringan yang sesungguhnya, seperti lonjakan trafik pada waktu-waktu tertentu, sehingga tetap dipertahankan agar model Random Forest Regression mampu mempelajari variasi pola kecepatan jaringan secara lebih realistis.



Gambar 1 Boxplot Outlier

C. Preprocessing dan Persiapan Data

Pemeriksaan kualitas data merupakan salah satu tahapan dalam *preprocessing* yang bertujuan untuk mengidentifikasi keberadaan data duplikat (*duplicate data*) dan nilai yang hilang (*missing value*) sebelum dilakukan proses pemodelan. Tahap ini penting

untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik sehingga tidak memengaruhi kinerja model prediksi. Berdasarkan hasil pemeriksaan, dataset tidak menunjukkan adanya data duplikat maupun *missing value* pada seluruh variabel. Oleh karena itu, tidak diperlukan proses penanganan seperti penghapusan data duplikat maupun imputasi nilai yang hilang. Dengan demikian, seluruh data dapat digunakan pada tahapan *preprocessing* berikutnya, yaitu pembentukan fitur temporal, *lag feature*, dan variabel target hingga diperoleh 716 data yang siap digunakan dalam proses pelatihan (*training*) dan pengujian (*testing*) model *Random Forest Regression*.

```
*** Date Time      0
Date Time(RAW)    0
Traffic Total (volume) 0
Traffic Total (volume)(RAW) 0
Traffic Total (speed) 0
Traffic Total (speed)(RAW) 0
Traffic In (volume) 0
Traffic In (volume)(RAW) 0
Traffic In (speed) 0
Traffic In (speed)(RAW) 0
Traffic Out (volume) 0
Traffic Out (volume)(RAW) 0
Traffic Out (speed) 0
Traffic Out (speed)(RAW) 0
Downtime          0
Downtime(RAW)     0
Coverage          0
Coverage(RAW)     0
dtype: int64
0
```

Gambar 2 Data Duplikan Dan Missing Value

D. Feature Engineering

Tahap *feature engineering* dilakukan untuk menghasilkan variabel-variabel baru yang mampu merepresentasikan karakteristik temporal serta pola historis lalu lintas jaringan sehingga informasi yang digunakan dalam proses prediksi menjadi lebih representatif. Pada penelitian ini, proses *feature engineering* dilakukan melalui pembentukan fitur temporal, *lag feature*, dan variabel target berdasarkan data Traffic Total (speed) dan Traffic Total (volume).

Fitur temporal dibentuk dari atribut waktu (*timestamp*), yaitu Hour, Day, Month, DayOfWeek, dan Weekend. Fitur-fitur tersebut digunakan untuk merepresentasikan pola perubahan kecepatan jaringan berdasarkan jam pengamatan, tanggal, bulan, hari dalam seminggu, serta kategori hari kerja (*weekday*) dan akhir pekan (*weekend*). Penambahan fitur temporal bertujuan agar model dapat mengenali pola perubahan lalu lintas jaringan yang dipengaruhi oleh faktor waktu.

	Hour	Day	Month	DayOfWeek	Weekend
0	3	19	5	1	0
1	4	19	5	1	0
2	5	19	5	1	0
3	6	19	5	1	0
4	7	19	5	1	0

Gambar 3 Ekstraksi Fitur Waktu

Selanjutnya, dibentuk *lag feature* yang terdiri atas Speed_lag1, Speed_lag2, dan Speed_lag3, yang merepresentasikan nilai Traffic Total Speed pada satu, dua, dan tiga periode sebelumnya. Selain itu, dibentuk pula Volume_lag1, Volume_lag2, dan Volume_lag3 yang merepresentasikan kondisi Traffic Total Volume pada tiga periode sebelumnya. Pembentukan *lag feature* bertujuan untuk memberikan informasi historis kepada model sehingga hubungan antarperiode pada data *time series* dapat dipelajari dengan lebih baik.

Speed_lag1	Speed_lag2	Speed_lag3	Volume_lag1	Volume_lag2	Volume_lag3
71.0	93.0	131.0	30579.0	39733.0	56392.0
72.0	71.0	93.0	30813.0	30579.0	39733.0
64.0	72.0	71.0	27426.0	30813.0	30579.0
73.0	64.0	72.0	31494.0	27426.0	30813.0
346.0	73.0	64.0	148390.0	31494.0	27426.0

Gambar 4 Lag Feature

Tahap berikutnya adalah pembentukan variabel target menggunakan metode shift(-1), sehingga nilai target merepresentasikan Traffic Total Speed pada periode berikutnya (t+1). Dengan pendekatan ini, model dilatih menggunakan informasi pada waktu ke-*t* untuk memprediksi nilai kecepatan jaringan pada waktu ke-*t+1*. Proses pembentukan *lag feature* dan variabel target menghasilkan beberapa nilai kosong (*NaN*) pada awal dan akhir data. Oleh karena itu, baris yang mengandung nilai tersebut dihapus sehingga diperoleh 716 data yang selanjutnya digunakan pada proses pelatihan dan pengujian model *Random Forest Regression*.

	Target
0	64.0
1	73.0
2	346.0
3	1015.0
4	1838.0

Gambar 5 Feature Target

E. Pemodelan Random Forest Regressor

Berdasarkan hasil pemodelan pada model Random Forest Regression, diperoleh

perbedaan performa antara data pelatihan (*training*) dan data pengujian (*testing*). Pada data pelatihan, model menghasilkan nilai MAE sebesar 49,78 Mbit/s, MSE sebesar 9.876,48 (Mbit/s)², RMSE sebesar 99,38 Mbit/s, MAPE sebesar 14,91%, dan koefisien determinasi (R^2) sebesar 0,9815. Sementara itu, pada data pengujian diperoleh MAE sebesar 121,03 Mbit/s, MSE sebesar 43.546,19 (Mbit/s)², RMSE sebesar 208,68 Mbit/s, MAPE sebesar 38,31%, dan R^2 sebesar 0,8892.

Nilai R^2 sebesar 0,8892 pada data pengujian menunjukkan bahwa model mampu menjelaskan sekitar 88,92% variasi nilai Traffic Total Speed, sedangkan sisanya dipengaruhi oleh faktor lain yang tidak sepenuhnya dapat dijelaskan oleh model. Dibandingkan dengan data pelatihan, seluruh metrik kesalahan pada data pengujian mengalami peningkatan, yang menunjukkan bahwa akurasi model menurun ketika dihadapkan pada data yang belum pernah digunakan selama proses pelatihan. Kondisi ini merupakan hal yang umum dalam proses evaluasi model karena data pengujian digunakan untuk mengukur kemampuan generalisasi model terhadap data baru.

Nilai MAE sebesar 121,03 Mbit/s menunjukkan bahwa rata-rata selisih absolut antara hasil prediksi dan nilai aktual pada data pengujian berada pada kisaran 121,03 Mbit/s, sedangkan RMSE sebesar 208,68 Mbit/s mengindikasikan bahwa masih terdapat beberapa prediksi dengan kesalahan yang relatif lebih besar. Hal tersebut juga tercermin dari nilai MSE sebesar 43.546,19 (Mbit/s)², yang memberikan penalti lebih besar terhadap kesalahan prediksi yang tinggi. Selain itu, MAPE sebesar 38,31% menunjukkan bahwa rata-rata kesalahan prediksi relatif terhadap nilai aktual masih berada pada kisaran 38,31%, sehingga tingkat akurasi model belum merata pada seluruh rentang data.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Random Forest Regression mampu mengikuti pola perubahan Traffic Total Speed dengan baik, yang ditunjukkan oleh nilai R^2 yang masih relatif tinggi pada data pengujian. Namun demikian, peningkatan nilai MAE, MSE, RMSE, dan MAPE pada data pengujian menunjukkan bahwa tingkat kesalahan prediksi cenderung meningkat pada sebagian data, terutama pada observasi dengan variasi Traffic Total Speed yang lebih tinggi. Oleh karena itu, hasil evaluasi ini menunjukkan bahwa model telah mampu merepresentasikan pola utama pada data historis, meskipun masih terdapat ruang untuk meningkatkan kemampuan prediksi pada kondisi trafik yang lebih bervariasi.

=====								
HASIL EVALUASI MODEL RANDOM FOREST REGRESSION								
=====								
	Dataset	MAE	MAE (%)	MSE	RMSE	RMSE (%)	MAPE	R ²
0	Training	49.78	8.58 %	9,876.48	99.38	17.12 %	14.91 %	0.9815
1	Testing	121.03	23.70 %	43,546.19	208.68	40.85 %	38.31 %	0.8892

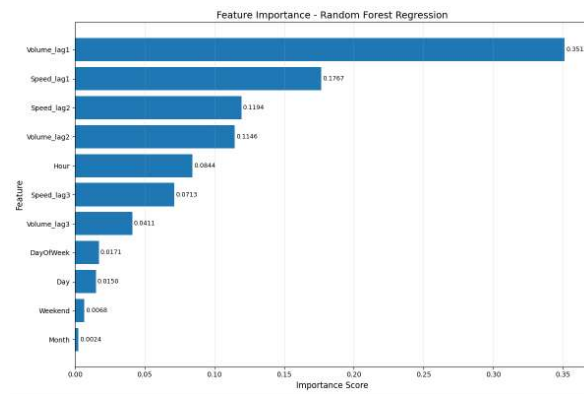
Gambar 6 Hasil Random Forest Pembahasan

F. Analisis Fitur Importance

Berdasarkan analisis feature importance pada model Random Forest Regression, diketahui bahwa Volume_lag1 merupakan fitur yang memiliki kontribusi terbesar dalam proses prediksi Traffic Total Speed, dengan nilai importance sebesar 0,3512. Hasil tersebut menunjukkan bahwa nilai volume trafik pada satu periode sebelumnya menjadi informasi yang paling banyak dimanfaatkan oleh model untuk memprediksi kecepatan jaringan pada periode berikutnya. Selain itu, fitur Speed_lag1 (0,1767), Speed_lag2 (0,1194), dan Volume_lag2 (0,1146) juga memberikan kontribusi yang cukup besar. Temuan ini menunjukkan bahwa informasi historis, baik berupa volume maupun kecepatan trafik pada satu hingga dua periode sebelumnya, memiliki peranan penting dalam membentuk pola prediksi Traffic Total Speed.

Di sisi lain, fitur pola waktu seperti Hour memiliki nilai importance sebesar 0,0844, yang menunjukkan bahwa informasi jam turut memberikan kontribusi dalam proses prediksi, meskipun pengaruhnya lebih kecil dibandingkan fitur historis. Selanjutnya, Speed_lag3 (0,0713) dan Volume_lag3 (0,0411) masih berkontribusi terhadap model, namun dengan tingkat kepentingan yang lebih rendah. Adapun fitur DayOfWeek (0,0171), Day (0,0150), Weekend (0,0068), dan Month (0,0024) memiliki nilai importance yang relatif kecil, sehingga kontribusinya terhadap pembentukan model tidak sebesar fitur-fitur historis.

Secara keseluruhan, hasil feature importance menunjukkan bahwa model Random Forest Regression lebih banyak memanfaatkan informasi historis (lag feature) dibandingkan fitur pola waktu dalam melakukan prediksi Traffic Total Speed. Hal ini mengindikasikan bahwa perubahan kecepatan jaringan pada suatu periode lebih dipengaruhi oleh kondisi trafik pada periode-periode sebelumnya.



Gambar 7 Feature Importance

G. Evaluasi dan Generalisasi

Berdasarkan hasil evaluasi per jam, tingkat kesalahan prediksi berbeda pada setiap periode pengamatan. Nilai MAPE tertinggi diperoleh pada pukul 07.00 (176,88%), diikuti pukul 06.00 (111,08%), 05.00 (105,04%), 04.00 (78,93%), dan 08.00 (70,82%). Pada periode tersebut, nilai MAE dan RMSE juga relatif lebih tinggi, yang menunjukkan adanya selisih prediksi yang lebih besar terhadap nilai aktual.

Sebaliknya, nilai MAPE terendah diperoleh pada pukul 19.00 (7,06%), 11.00 (11,10%), 02.00 (11,74%), 12.00 (12,93%), dan 22.00 (14,25%), sehingga hasil prediksi pada jam-jam tersebut lebih mendekati nilai aktual. Secara umum, hasil ini menunjukkan bahwa tingkat kesalahan prediksi model bervariasi pada setiap jam pengamatan dan dipengaruhi oleh karakteristik data Traffic Total Speed pada masing-masing periode.

Tabel 2 Analisis Beberapa Waktu

Kategori	Jam	MAPE (%)	Keterangan
Error Tertinggi	07.00	176.88	Selisih prediksi paling besar
	06.00	111.08	Error tinggi
	05.00	105.04	Error tinggi
	04.00	78.93	Error cukup tinggi
	08.00	70.82	Error cukup tinggi
Error Terendah	19.00	7.06	Prediksi paling mendekati aktual
	11.00	11.10	Error rendah
	02.00	11.74	Error rendah
	12.00	12.93	Error rendah
	22.00	14.25	Error rendah

Berdasarkan hasil analisis, Traffic Total Speed tertinggi terjadi pada pukul 09.00 dengan rata-rata kecepatan sebesar 1.165,17 Mbit/s, sedangkan nilai terendah terjadi pada pukul 03.00 dengan rata-rata sebesar 51,33 Mbit/s. Hal ini menunjukkan bahwa karakteristik lalu lintas jaringan berbeda pada setiap jam pengamatan, yang tercermin dari

variasi rata-rata kecepatan yang dihasilkan.

Dari sisi evaluasi prediksi, nilai absolute error terbesar diperoleh pada pukul 07.00, yaitu sebesar 564,73 Mbit/s dengan MAPE 176,88%, yang menunjukkan adanya selisih prediksi yang cukup besar dibandingkan nilai aktual pada periode tersebut. Sebaliknya, prediksi dengan tingkat kesalahan paling rendah diperoleh pada pukul 03.00, dengan absolute error sebesar 7,92 Mbit/s dan MAPE sebesar 17,68%, sehingga hasil prediksi pada jam tersebut lebih mendekati nilai aktual. Temuan ini menunjukkan bahwa tingkat akurasi model bervariasi pada setiap jam pengamatan sesuai dengan karakteristik data Traffic Total Speed yang diamati.

```
=====
JAM DENGAN TRAFFIC TERTINGGI
=====
Jam : 9
Rata-rata Speed : 1,165.17
=====
JAM DENGAN TRAFFIC TERENDAH
=====
Jam : 3
Rata-rata Speed : 51.33
=====
JAM DENGAN ERROR TERBESAR
=====
Jam : 7
Absolute Error : 564.73
MAPE : 176.88%
=====
JAM DENGAN PREDIKSI TERBAIK
=====
Jam : 3
Absolute Error : 7.92
MAPE : 17.68%
```

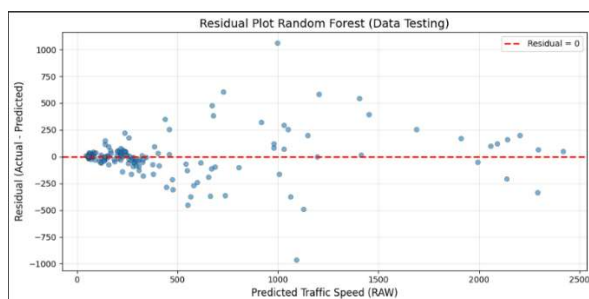
Gambar 8 Analisis Traffic

Gambar di atas menampilkan hubungan antara nilai aktual (*Actual Traffic Total Speed*) dan nilai hasil prediksi (*Predicted Traffic Total Speed*) pada data pengujian (20%). Garis merah putus-putus ($y = x$) menunjukkan kondisi ideal ketika nilai prediksi sama dengan nilai aktual. Berdasarkan visualisasi tersebut, sebagian besar titik berada di sekitar garis ideal, terutama pada rentang kecepatan rendah hingga menengah (sekitar 0–600 Mbit/s). Hal ini menunjukkan bahwa model Random Forest Regression mampu mengikuti pola data dengan cukup baik pada rentang tersebut. Namun, pada nilai *Traffic Total Speed* yang lebih tinggi, khususnya di atas 1.500 Mbit/s, terlihat beberapa titik berada lebih jauh dari garis ideal sehingga menunjukkan adanya peningkatan selisih antara nilai prediksi dan nilai aktual. Kondisi tersebut mengindikasikan bahwa akurasi model cenderung menurun ketika memprediksi nilai *Traffic Total Speed* yang tinggi.

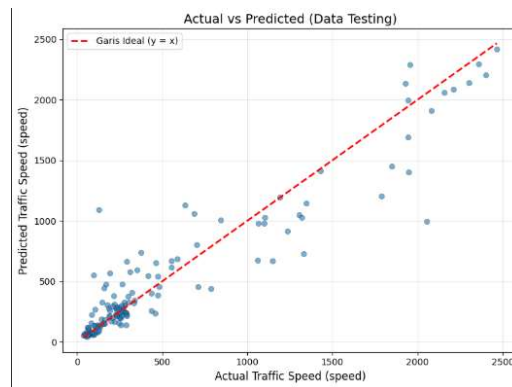
Temuan tersebut diperkuat oleh Gambar 9 yang menampilkan Residual Plot pada data pengujian. Sebagian besar residual tersebar di sekitar garis nol, terutama pada nilai prediksi yang rendah hingga menengah, sehingga menunjukkan bahwa sebagian besar hasil prediksi memiliki selisih yang relatif kecil terhadap nilai aktual. Akan tetapi, pada

nilai prediksi yang lebih tinggi, penyebaran residual terlihat semakin lebar dengan rentang residual sekitar -970 Mbit/s hingga $+1.060$ Mbit/s. Residual positif menunjukkan bahwa model menghasilkan nilai prediksi yang lebih rendah daripada nilai aktual (*underestimate*), sedangkan residual negatif menunjukkan bahwa model menghasilkan nilai prediksi yang lebih tinggi daripada nilai aktual (*overestimate*). Meskipun masih terdapat beberapa residual yang cukup besar, penyebaran residual tidak membentuk pola tertentu sehingga tidak menunjukkan adanya kecenderungan sistematis pada kesalahan prediksi.

Hasil visualisasi tersebut sejalan dengan hasil evaluasi model pada data pengujian yang memperoleh MAE sebesar 121,03 Mbit/s, MSE sebesar 43.546,19 (Mbit/s)², RMSE sebesar 208,68 Mbit/s, MAPE sebesar 38,31%, dan koefisien determinasi (R^2) sebesar 0,8892. Nilai R^2 menunjukkan bahwa model mampu menjelaskan sekitar 88,92% variasi *Traffic Total Speed* pada data pengujian. Sementara itu, nilai MAE dan RMSE menunjukkan bahwa rata-rata kesalahan prediksi masing-masing sebesar 121,03 Mbit/s dan 208,68 Mbit/s. Di sisi lain, nilai MAPE sebesar 38,31% mengindikasikan bahwa tingkat kesalahan relatif model masih cukup tinggi pada beberapa observasi, terutama ketika memprediksi nilai *Traffic Total Speed* yang besar. Oleh karena itu, berdasarkan kedua visualisasi tersebut dapat disimpulkan bahwa model telah mampu merepresentasikan pola umum perubahan *Traffic Total Speed*, namun akurasi prediksi masih cenderung menurun pada data dengan nilai kecepatan yang tinggi sehingga masih terdapat ruang untuk peningkatan performa model pada rentang tersebut.



Gambar 9 Resudal Analisis



Gambar 10 Prediksi Dan Actual

Berdasarkan hasil *TimeSeriesSplit* 5-Fold Cross Validation, model Random Forest Regression memperoleh nilai rata-rata koefisien determinasi (R^2) sebesar $0,8049 \pm 0,1178$, nilai Mean Absolute Error (MAE) sebesar $171,86 \pm 64,70$ Mbit/s, serta Root Mean Squared Error (RMSE) sebesar $286,25 \pm 87,26$ Mbit/s. Nilai rata-rata R^2 tersebut menunjukkan bahwa model mampu menjelaskan sekitar 80,49% variasi *Traffic Total Speed* pada proses validasi silang. Sementara itu, nilai standar deviasi sebesar 0,1178 menunjukkan adanya variasi performa model antar-*fold*. Variasi tersebut terlihat dari nilai R^2 pada masing-masing *fold*, yaitu 0,5823, 0,8329, 0,8080, 0,8825, dan 0,9188, yang menunjukkan bahwa kemampuan model dalam menjelaskan variasi *Traffic Total Speed* berbeda pada setiap periode validasi dalam skema *time series*.

Apabila ditinjau berdasarkan nilai R^2 pada setiap fold, nilai tersebut menunjukkan kecenderungan peningkatan performa model pada fold berikutnya. Kondisi ini mengindikasikan bahwa bertambahnya data historis pada setiap proses pelatihan melalui mekanisme *TimeSeriesSplit* memberikan informasi yang lebih lengkap sehingga model mampu mengenali pola data dengan lebih baik pada periode selanjutnya.

Selain itu, perbandingan antara RMSE data pelatihan dan data pengujian menghasilkan rasio sebesar 2,0998. Nilai tersebut menunjukkan bahwa kesalahan prediksi pada data pengujian lebih besar dibandingkan data pelatihan, sehingga terdapat indikasi bahwa kemampuan model pada data baru mengalami penurunan dibandingkan saat proses pelatihan. Namun demikian, kondisi tersebut tidak secara langsung menunjukkan overfitting yang kuat. Hal ini didukung oleh nilai koefisien determinasi pada data pengujian sebesar 0,8892 serta rata-rata nilai R^2 hasil cross-validation sebesar 0,8049, yang masih menunjukkan kemampuan model dalam merepresentasikan sebagian besar variasi data. Oleh karena itu, model dapat dikatakan memiliki kemampuan

generalisasi yang cukup baik, meskipun akurasi prediksi cenderung menurun pada observasi dengan nilai Traffic Total Speed yang relatif tinggi.

```
=====
EVALUASI TAMBAHAN MODEL RANDOM FOREST
=====
MAPE Training      : 14.91%
MAPE Testing       : 38.31%
-----
Rasio Overfitting (RMSE) : 2.0998
-----
HASIL CROSS VALIDATION (TimeSeriesSplit 5-Fold)
R² : 0.8049 ± 0.1178
MAE : 171.86 ± 64.70
RMSE : 286.25 ± 87.26
-----
Nilai R² setiap Fold
Fold 1 : 0.5823
Fold 2 : 0.8329
Fold 3 : 0.8080
Fold 4 : 0.8825
Fold 5 : 0.9188
=====
```

Gambar 11 Analisis Cross Validation

Tabel 3 Tabel Croos Validation

Fold	Nilai R ²
Fold 1	0.5823
Fold 2	0.8329
Fold 3	0.8080
Fold 4	0.8825
Fold 5	0.9188

Tabel 4 ringkasan hasil cross validation

Statistik	Nilai
Mean R ²	0.80490
Standar Deviasi R ²	0.11780
Nilai R ² Minimum	0.5823
Nilai R ² Maksimum	0.9188
Rentang (Max – Min)	0.3365

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, Traffic Total Speed memiliki karakteristik distribusi right-skewed dengan 104 data outlier atau sebesar 14,53%. Variasi kecepatan jaringan berbeda pada setiap periode waktu, dengan kesalahan prediksi tertinggi terjadi pada rentang pukul 05.00–08.00, khususnya pukul 07.00 dengan MAPE sebesar 176,88%. Hal ini menunjukkan bahwa perubahan Traffic Total Speed dipengaruhi oleh pola waktu dan kondisi lalu lintas pada periode sebelumnya. Penerapan Random Forest Regression dilakukan dengan membentuk fitur temporal dan lag feature serta target Traffic Total Speed pada periode berikutnya (t+1). Dari 716 data hasil preprocessing, diperoleh 572 data training dan 144 data testing melalui time-based split dengan rasio 80:20. Hasil pengujian menunjukkan bahwa model memperoleh MAE sebesar 121,03 Mbit/s, RMSE sebesar 208,68 Mbit/s, MAPE sebesar 38,31%, dan R² sebesar 0,8892. Nilai tersebut

menunjukkan bahwa model mampu menjelaskan 88,92% variasi Traffic Total Speed pada data pengujian. Hasil TimeSeriesSplit 5-Fold menghasilkan rata-rata R^2 sebesar 0,8494, sehingga model memiliki kemampuan prediksi yang cukup baik meskipun performanya bervariasi antarperiode. Berdasarkan analisis feature importance, Volume_lag1 merupakan fitur paling dominan dengan nilai importance sebesar 0,3512, diikuti Speed_lag1 sebesar 0,1767, Speed_lag2 sebesar 0,1194, dan Volume_lag2 sebesar 0,1146. Hasil ini menunjukkan bahwa informasi historis volume dan kecepatan lalu lintas memiliki kontribusi lebih besar dibandingkan fitur temporal dalam memprediksi Traffic Total Speed. Secara keseluruhan, Random Forest Regression dapat digunakan untuk memprediksi Traffic Total Speed berdasarkan data historis dengan performa yang cukup baik. Namun, model masih memiliki keterbatasan dalam memprediksi perubahan kecepatan yang tinggi, terutama pada periode jam sibuk.

DAFTAR REFERENSI

- Breiman, L. (2001). *Random Forests*.
- Li, E. (2025). *Evaluasi Kinerja Jaringan WiFi Berdasarkan Kecepatan dan Kualitas di Lingkungan Rumah*. 1(1), 35–41.
- Putri, D. W., Sutomo, B., Informatika, J. T., Metro, K., & Forest, R. (n.d.). *Prediksi Kemacetan Lalu Lintas Jaringan Menggunakan Algoritma Random Forest Berbasis Data Mikrotik*. 75–86.
- Rodicky, N., & Soim, S. (2025). *SELEKSI FITUR DENGAN GENETIKA ALGORITHM UNTUK OPTIMASI RANDOM FOREST DALAM MENGLASIFIKASI TENTANG JENIS PENGGUNAAN INTERNET Abstrak Pendahuluan Metode*. (3), 185–190. <https://doi.org/10.14710/transmisi.27.3.185-190>
- Saha, S., Haque, A., & Sidebottom, G. (2024). *Multi-Step Internet Traffic Forecasting Models with Variable Forecast Horizons for Proactive Network Management †*. 1–29. <https://doi.org/10.1109/ISNCC55209.2022.9851774>.Citation
- Syafrudin, T., Rianto, R., Ujianto, E. I. H., Informasi, M. T., & Yogyakarta, U. T. (2025). *Analisis Kualitas Layanan Jaringan Wlan Berdasarkan Parameter Throughput , Delay , Jitter , dan Packet Loss di Universitas X Analysis of Wlan Network Service Quality Based on Throughput , Delay , Jitter , and Packet Loss Parameters at University X*. 5(8), 2143–2151.
- Syarifatun, I., Miyarno, R., Fatwa, T., & Rizqi, A. (2024). *Analisis Quality of Service (QoS) Menggunakan Standar Parameter Tiphon pada Jaringan Internet Berbasis Wi-Fi Kampus 1 Unjaya*. 17(1), 1–9.
- Test, D. (2025). *Prediksi Downlink Throughput Menggunakan Metode Random Forest Regression Pada Jaringan 4G LTE Berdasarkan Data*. 6(2).