Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

Analisis Karakteristik Pola Trafik Jaringan Berbasis Data Time-Series Untuk Identifikasi Zona Temporal, Korelasi Variabel dan Deteksi Anomali

Analysis of Network Traffic Pattern Characteristics Based on Time-Series Data To Identification of Temporal Zones, Variable Correlations, and Anomaly Detection

Yudi Sutanto[1], Hendra Kurniawan[2], Tristanto Ari Aji[3]
[1,3] Informatika, Universitas AMIKOM Yogyakarta
[2] Sistem Informasi, Universitas AMIKOM Yogyakarta
Email : [1]yudisuta@amikom.ac.id

ABSTRACT

The rapid growth in computer network traffic volume over the past few years has posed serious challenges for information technology infrastructure managers, particularly in capacity management, bandwidth planning, and ensuring quality of service (QoS). Most previous studies have jumped straight into deep learning-based predictive modeling without first conducting in-depth statistical characterization, even though an understanding of distributions, temporal patterns, correlations between variables, and the presence of anomalies constitutes the methodological foundation that determines the reliability of predictive models. This study proposes a four-dimensional statistical characterization framework—comprising temporal segmentation, weekly patterns, multivariate correlations, and anomaly detection—that is objectively validated, rather than merely descriptive, as an explicit foundation before predictive modeling is developed.
This study employs a descriptive-quantitative time-series analysis approach on 8,640 data points collected over 30 days at 5-minute intervals, covering three main variables: throughput (Mbps), packet rate (pps), and active sessions. The analysis was conducted using eight complementary statistical methods: descriptive statistics, temporal zone segmentation based on t-tests between hours, weekly pattern analysis, assumption tests (normality and homogeneity of variance), independent t-tests (Welch), Pearson correlation, linear regression, and Z-score-based anomaly detection—all of which were implemented using Python.
Statistical validation of the zone boundaries (rather than mere visual inspection) identified three—not four—statistically distinct temporal zones: Early Morning (11:00 PM–6:59 AM) with 14.4% utilization, Business Hours (7:00 AM–5:59 PM) at 62.9%, and Evening-Night (6:00 PM–10:59 PM) at 39.9%. Weekday traffic averaged 88.01 Mbps, which is 72.0% higher than on weekends, which averaged only 51.16 Mbps (Welch's t-test, $t = 44.83$; $p < 0.001$), after Levene's test confirmed that the variances of the two groups were not homogeneous. The Pearson correlation between the three traffic variables showed a very high value ($r > 0.998$; $p < 0.001$). Z-score anomaly detection did not find any extreme outliers in the throughput data; the only anomalies identified were 10 rows (0.12%) of negative values in packet_rate and active_sessions, which were categorized as instrument artifacts.
These findings provide a statistically validated quantitative basis for network administrators to design time-based QoS policies and develop capacity plans, as well as a replicable methodological framework for the development of predictive network traffic models in subsequent research.

Keywords : anomaly detection; network management; network traffic; pearson correlation; temporal zones; time-series analysis

ABSTRAK

Pertumbuhan volume trafik jaringan komputer yang pesat dalam beberapa tahun terakhir telah menimbulkan tantangan serius bagi pengelola infrastruktur teknologi informasi, khususnya dalam manajemen kapasitas, perencanaan bandwidth, dan penjaminan kualitas layanan (Quality of Service/QoS). Sebagian besar penelitian terdahulu langsung menuju pemodelan prediktif berbasis *deep learning* tanpa didahului karakterisasi statistik yang mendalam, padahal pemahaman distribusi, pola temporal, korelasi antar-variabel, dan keberadaan anomali merupakan fondasi metodologis yang menentukan keandalan model prediktif. Penelitian ini mengajukan kerangka kerja karakterisasi statistik empat dimensi yaitu segmentasi temporal, pola

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

mingguan, korelasi multivariat, dan deteksi anomali yang divalidasi secara objektif, bukan sekadar deskriptif, sebagai fondasi eksplisit sebelum pemodelan prediktif dibangun.
Penelitian ini menggunakan pendekatan analisis *time-series* deskriptif-kuantitatif terhadap 8.640 titik pengukuran yang dikumpulkan selama 30 hari dengan interval perekaman 5 menit, mencakup tiga variabel utama yaitu *throughput (Mbps), packet rate (pps),* dan *active sessions*. Analisis dilakukan menggunakan delapan metode statistik yang saling melengkapi yaitu statistik deskriptif, segmentasi zona temporal berbasis uji-*t* antar jam, analisis pola mingguan, uji asumsi (normalitas dan homogenitas varians), uji *t* independen (Welch), korelasi Pearson, regresi linier, dan deteksi anomali berbasis Z-*score* yang seluruhnya diimplementasikan menggunakan Python.
Validasi statistik atas batas-batas zona (bukan sekadar inspeksi visual) mengidentifikasi tiga, bukan empat, zona temporal yang berbeda secara signifikan: Dini Hari (23.00-06.59) dengan utilisasi 14,4%, Jam Bisnis (07.00-17.59) 62,9%, dan Sore-Malam (18.00-22.59) 39,9%. Trafik pada hari kerja (*weekday*) tercatat rata-rata 88,01 *Mbps*, lebih tinggi 72,0% dibandingkan hari libur (*weekend*) yang hanya 51,16 *Mbps* (uji-*t* Welch, $t = 44{,}83$; $p < 0{,}001$), setelah uji Levene mengonfirmasi varians kedua kelompok tidak homogen. Korelasi Pearson antar ketiga variabel trafik menunjukkan nilai sangat tinggi ($r > 0{,}998$; $p < 0{,}001$). Deteksi anomali Z-*score* tidak menemukan *outlier* ekstrem pada data *throughput*, satu-satunya anomali yang teridentifikasi adalah 10 baris (0,12%) nilai negatif pada *packet_rate* dan *active_sessions* yang dikategorikan sebagai artefak instrumen.
Temuan-temuan ini memberikan landasan kuantitatif yang tervalidasi secara statistik bagi pengelola jaringan untuk merancang kebijakan QoS berbasis waktu, menyusun perencanaan kapasitas, serta menjadi kerangka metodologis yang dapat direplikasi bagi pengembangan model prediktif trafik jaringan pada tahap penelitian selanjutnya.

Kata kunci: deteksi anomali; manajemen jaringan; trafik jaringan; korelasi Pearson; zona temporal; analisis *time-series*

I. PENDAHULUAN

Pertumbuhan penggunaan *internet* dan layanan berbasis jaringan yang pesat dalam satu dekade terakhir telah menempatkan manajemen kapasitas jaringan sebagai salah satu tantangan teknis yang krusial bagi para pengelola infrastruktur teknologi informasi. Volume trafik *internet* global terus meningkat dengan laju tahunan yang signifikan, didorong oleh proliferasi perangkat terhubung, adopsi layanan *cloud*, *streaming* multimedia, dan komunikasi data berbasis aplikasi [1]. Dalam konteks ini, pemahaman mendalam tentang karakteristik dan pola trafik jaringan bukan sekadar kebutuhan akademis, melainkan prasyarat operasional yang nyata.

Trafik jaringan pada dasarnya merupakan data *time-series* yang memiliki struktur temporal kompleks. Pengukuran *throughput*, laju paket, dan jumlah sesi aktif yang dilakukan secara periodik menghasilkan deret waktu yang menyimpan informasi tentang perilaku pengguna, pola penggunaan layanan, dan kondisi kesehatan infrastruktur jaringan [2]. Berbeda dengan *data stasioner*, trafik jaringan umumnya memperlihatkan karakteristik *non-stasioner* dengan pola periodisitas kuat, baik pada skala harian maupun mingguan, yang mencerminkan ritme aktivitas manusia yang menggunakan jaringan tersebut [3].

Literatur tentang prediksi trafik jaringan menggunakan *deep learning*, khususnya LSTM (*Long Short-Term Memory*), telah berkembang pesat [4][5]. Namun sebagaimana ditunjukkan pada Tabel 1, mayoritas studi tersebut bergerak langsung menuju pemodelan prediktif tanpa didahului karakterisasi statistik yang eksplisit dan tervalidasi. Torres et al. [4] secara khusus menegaskan bahwa pemahaman karakteristik data merupakan prasyarat yang tidak dapat diabaikan sebelum pemodelan prediktif, namun sedikit studi yang benar-benar mengoperasionalkan prasyarat tersebut ke dalam prosedur yang dapat direplikasi. Di sinilah letak kesenjangan (*research gap*) yang dituju penelitian ini, bukan pada ketiadaan metode statistik itu sendiri yang secara individual sudah umum digunakan, melainkan pada ketiadaan kerangka kerja yang mengintegrasikan segmentasi temporal, pola mingguan, korelasi multivariat, dan deteksi anomali sekaligus, dengan validasi objektif pada setiap tahapnya, sebagai satu prosedur karakterisasi yang utuh dan dapat direplikasi sebelum pemodelan prediktif dibangun.

Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan melakukan karakterisasi komprehensif terhadap 8.640 titik data trafik jaringan selama 30 hari. Kontribusi utama penelitian ini adalah:
1. Kerangka kerja karakterisasi statistik empat dimensi yang mengintegrasikan segmentasi temporal, pola

mingguan, korelasi multivariat, dan deteksi anomali dalam satu prosedur yang dapat direplikasi.
2. Metode validasi objektif untuk segmentasi zona temporal berbasis pengujian statistik antar interval waktu, sebagai alternatif terhadap pendekatan inspeksi visual yang lazim digunakan pada studi terdahulu.
3. Membuktikan perbedaan signifikan trafik *weekday* dibanding *weekend* dengan selisih 72,0% ($p < 0{,}001$), termasuk pengujian asumsi statistik yang mendasarinya.
4. Analisis korelasi multivariat yang mengungkap hubungan sangat kuat ($r > 0{,}998$) antar tiga variabel trafik utama disertai konfirmasi absennya *outlier* ekstrem, sebagai fondasi yang dapat diaudit bagi pemodelan prediktif lanjutan.

II. LANDASAN TEORI

2.1 Trafik Jaringan Komputer

Trafik jaringan komputer didefinisikan sebagai jumlah data yang berpindah melalui suatu jaringan dalam satuan waktu tertentu. Menurut Tanenbaum dan Wetherall, trafik jaringan merupakan aliran paket data yang ditransmisikan melalui medium komunikasi dari sumber ke tujuan menggunakan protokol tertentu [6]. Pemahaman mengenai karakteristik trafik jaringan menjadi dasar yang esensial dalam perancangan, pengelolaan, dan optimasi infrastruktur jaringan modern.

2.2 Analisis *Time-Series*

Analisis *time-series* (deret waktu) adalah metode statistik yang digunakan untuk menganalisis serangkaian titik data yang dikumpulkan atau dicatat dalam urutan waktu yang berurutan. Menurut Box et al., *time-series* merupakan himpunan observasi *xt*, masing-masing direkam pada waktu spesifik *t*, yang membentuk suatu rangkaian temporal yang memiliki keterkaitan antar observasi [7].

2.3 Statistik Deskriptif

Statistik deskriptif adalah cabang statistik yang bertujuan meringkas, mendeskripsikan, dan menyajikan karakteristik utama dari suatu kumpulan data secara sistematis dan terstruktur. Menurut Montgomery dan Runger, statistik deskriptif menyediakan metode untuk mengorganisasi dan merangkum informasi dari data sampel guna memperoleh gambaran komprehensif mengenai distribusi dan variabilitas data tersebut [8].

2.4 Korelasi Pearson

Korelasi Pearson, atau koefisien korelasi momen produk Pearson (*Pearson product-moment correlation coefficient*), adalah ukuran statistik yang mengkuantifikasi kekuatan dan arah hubungan linear antara dua variabel kontinu. Koefisien ini pertama kali dikembangkan oleh Karl Pearson pada akhir abad ke-19 dan hingga kini merupakan ukuran korelasi yang paling banyak digunakan dalam penelitian kuantitatif [9].

Koefisien korelasi Pearson (r) didefinisikan sebagai:

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] \,/\, \sqrt{[\Sigma(x_i - \bar{x})^2 \cdot \Sigma(y_i - \bar{y})^2]}$$

di mana x_i dan y_i adalah pasangan observasi, $\bar{x}$ dan $\bar{y}$ adalah rata-rata masing-masing variabel. Nilai r berkisar antara -1 hingga +1, di mana nilai +1 menunjukkan korelasi linear positif sempurna, nilai -1 menunjukkan korelasi linear negatif sempurna, dan nilai 0 menunjukkan tidak adanya hubungan linear [9].

2.5 Uji Asumsi: Normalitas dan Homogenitas Varians

Sebelum penerapan uji-*t* independen, dua asumsi utama perlu diperiksa, normalitas distribusi pada masing-masing kelompok dan homogenitas varians antar-kelompok. Uji normalitas D'Agostino-Pearson K^2 digunakan untuk sampel besar sebagai alternatif uji Shapiro-Wilk yang kurang sesuai pada $n > 5.000$ [10]. Uji Levene digunakan untuk menguji homogenitas varians, ketika asumsi ini dilanggar, uji-*t* Welch yang tidak mengasumsikan varians sama lebih sesuai digunakan dibandingkan uji-*t* *Student* konvensional [11].

2.6 Uji *t* Independen (*Independent Samples t-test*)

Uji *t* independen adalah uji statistik parametrik yang digunakan untuk membandingkan rata-rata dua kelompok yang tidak saling berhubungan guna menentukan apakah terdapat perbedaan signifikan secara statistik di antara keduanya. Metode ini dikembangkan oleh William Sealy Gosset dan dipublikasikan pada tahun 1908 di bawah nama

samaran "*Student*", sehingga sering disebut pula *Student's t-test* [12]. Varian Welch dari uji ini digunakan ketika asumsi homogenitas varians tidak terpenuhi, dengan penyesuaian derajat kebebasan sehingga tetap valid pada kondisi tersebut [11].

2.7 Deteksi Anomali Berbasis *Z-Score*

Deteksi anomali (*anomaly detection*) adalah proses identifikasi pola data yang menyimpang secara signifikan dari perilaku normal atau yang diharapkan. Dalam konteks keamanan dan manajemen jaringan, anomali trafik dapat mengindikasikan serangan siber, kegagalan perangkat, atau perubahan perilaku pengguna yang tidak normal [13].

2.8 Tinjauan Pustaka

2.8.1 Karakterisasi Trafik Jaringan Berbasis *Time-Series*

Tiga karakteristik metodologis penting muncul secara konsisten pada studi karakterisasi trafik berbasis *time-series*. Pertama, dari sisi model statistik, Purnawansyah et al. [14] menunjukkan ARIMA (1,0,2) sebagai model terbaik dengan komponen musiman *additive* maupun *multiplicative*, sementara Millan et al. [15] mengonfirmasi perilaku fraktal (Hurst *exponent* 0,5-1) yang justru menunjukkan model Markov konvensional tidak memadai untuk trafik kecepatan tinggi, kedua temuan ini bersama-sama menyiratkan bahwa struktur temporal trafik jaringan tidak dapat diasumsikan seragam antar konteks jaringan, sehingga karakterisasi awal yang spesifik konteks seperti pada penelitian ini tetap diperlukan sebelum pemilihan model prediktif. Kedua, dari sisi kompleksitas variabel, Di Mauro et al. [16] menunjukkan bahwa pendekatan *multivariate time-series* pada trafik VoIP mengungguli model statistik *univariat* justru karena adanya ketergantungan antar variabel QoS yang kuat, pola yang juga ditemukan pada penelitian ini ($r > 0{,}998$ antar *throughput*, *packet rate*, dan *active sessions*), namun dengan interpretasi yang berbeda, pada skala data menengah dengan korelasi setinggi ini, kompleksitas *multivariat* justru dapat disederhanakan menjadi pemantauan *univariat* tanpa kehilangan informasi substansial, sebuah nuansa yang belum banyak dibahas literatur yang berfokus pada skala besar.

Sebagian besar studi ini beroperasi pada skala data yang sangat besar (miliaran paket) atau bergerak langsung ke pemodelan *machine learning* tanpa tahap karakterisasi statistik transparan yang terpisah dan dapat diaudit. Penelitian ini memposisikan diri secara berbeda, pada skala menengah (30 hari, resolusi 5 menit, tiga variabel), menggunakan metode statistik yang sepenuhnya transparan dan *reproducible*, dengan setiap tahap karakterisasi, termasuk segmentasi zona divalidasi secara statistik eksplisit, bukan diasumsikan.

2.8.2 Pola Temporal dan Perilaku Pengguna

Wang et al. [17] membuktikan bahwa pola trafik mingguan terbagi secara alami menjadi *weekday*, Sabtu, dan Minggu, dengan setiap hari terdiri atas komponen temporal yang terdistribusi normal. Studi ini konsisten dengan temuan Wang et al. dalam hal signifikansi perbedaan *weekday-weekend*, namun berbeda pada satu aspek metodologis penting, penelitian ini menemukan bahwa distribusi *throughput* tidak sepenuhnya normal (uji D'Agostino, $p < 0{,}001$), sehingga inferensi keduanya tetap valid bukan karena normalitas data melainkan karena ukuran sampel besar (*Central Limit Theorem*). Kepner et al. [18], pada analisis 100 miliar paket jaringan, menemukan berbagai *scaling relation* baru yang berguna sebagai dasar *anomaly detection* dan rekayasa fitur AI, skala tersebut jauh melampaui penelitian ini, namun prinsip bahwa pola skala besar dapat diturunkan menjadi aturan operasional yang lebih sederhana turut mendasari pendekatan zona temporal pada penelitian ini. Studi selama pandemi COVID-19 [19] juga mengonfirmasi bahwa pola *weekday-weekend* merupakan karakteristik struktural trafik yang bersifat universal, memperkuat generalisasi temuan pada Bagian 4.3.

2.8.3 Deteksi Anomali pada Trafik Jaringan

Ahmed et al. [20] mengklasifikasikan teknik *anomaly detection* ke dalam berbagai kategori, termasuk pendekatan berbasis statistik dan *machine learning*, dan menyatakan bahwa pendekatan statistik *non-signature* lebih fleksibel karena mampu mendeteksi anomali baru tanpa pengetahuan *prior*. Zehra et al. [21] menegaskan bahwa metode statistik seperti *Z-score* tetap relevan dan komplementer terhadap pendekatan *machine learning* yang lebih

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

canggih dalam *pipeline monitoring* jaringan. Kedua studi ini mendukung penggunaan Z-*score* pada penelitian ini, namun juga menyiratkan keterbatasannya, Z-*score* efektif mendeteksi *outlier* titik (*point anomaly*) tetapi kurang sensitif terhadap anomali kontekstual atau kolektif, keterbatasan yang secara eksplisit diakui pada Bagian 5.2 sebagai arah penelitian lanjutan.

2.8.4 Prediksi Trafik dengan *Deep Learning*

Torres et al. [4], dalam survei komprehensif, menyatakan LSTM konsisten menunjukkan performa terbaik pada data sekuensial, dan menekankan bahwa pemahaman karakteristik data merupakan prasyarat yang tidak dapat diabaikan sebelum pemodelan prediktif. *Scoping review* terkini pada prediksi trafik 5G [22] mengonfirmasi bahwa pola *weekday-weekend* terdokumentasi konsisten dalam berbagai *dataset*, dan karakterisasi temporal yang akurat merupakan faktor penentu keberhasilan model. Kedua sumber ini secara langsung memotivasi kontribusi utama penelitian ini, mengoperasionalkan "karakterisasi yang akurat" tersebut menjadi prosedur konkret dan tervalidasi, alih-alih pernyataan normatif semata.

2.8.5 Posisi Penelitian

Tabel 1. Rangkuman Tinjauan Pustaka

No.	Penulis & Tahun	Metode	Temuan Utama	Keterkaitan dengan Penelitian ini
1	Purnawansyah et al. [14] (2018)	Dekomposisi, ARIMA	ARIMA(1,0,2) terbaik; pola musiman teridentifikasi	Menguatkan perlunya karakterisasi pola sebelum pemodelan
2	Millan et al. [15] (2021)	Analisis fraktal, DFA	Perilaku fraktal (Hurst 0,5-1); model Markov tidak memadai	Menunjukkan struktur trafik spesifik konteks, bukan generik
3	Wang et al. [17] (2023)	Pemodelan UBB	Pola trafik terbagi alami: *weekday*, Sabtu, Minggu	Konsisten dengan temuan Bagian 4.3, beda pada uji normalitas
4	Di Mauro et al. [16] (2024)	*Multivariate time-series, Machine Learning*	Ketergantungan temporal antar variabel QoS terungkap lewat analisis *multivariat*	Sejalan dengan r > 0,998; interpretasi disederhanakan pada skala menengah
5	Torres et al. [4] (2021)	Survei *deep learning*	LSTM unggul; karakterisasi wajib sebelum pemodelan	Motivasi utama kontribusi penelitian ini
6	Kepner et al. [18] (2020)	Analisis *multi-temporal*	*Scaling relation* baru dari 100 miliar paket	Prinsip skala besar diturunkan ke aturan operasional zona
7	Zehra et al. [21] (2023)	Survei *Machine Learning anomaly detection*	Z-*score* tetap relevan komplementer terhadap ML	Mendasari pemilihan metode pada Bagian 4.5
8	Ahmed et al. [20] (2016)	Survei *anomaly detection*	Pendekatan *non-signature* efektif tanpa *prior knowledge*	Mendasari justifikasi Z-score; keterbatasan diakui di 5.2

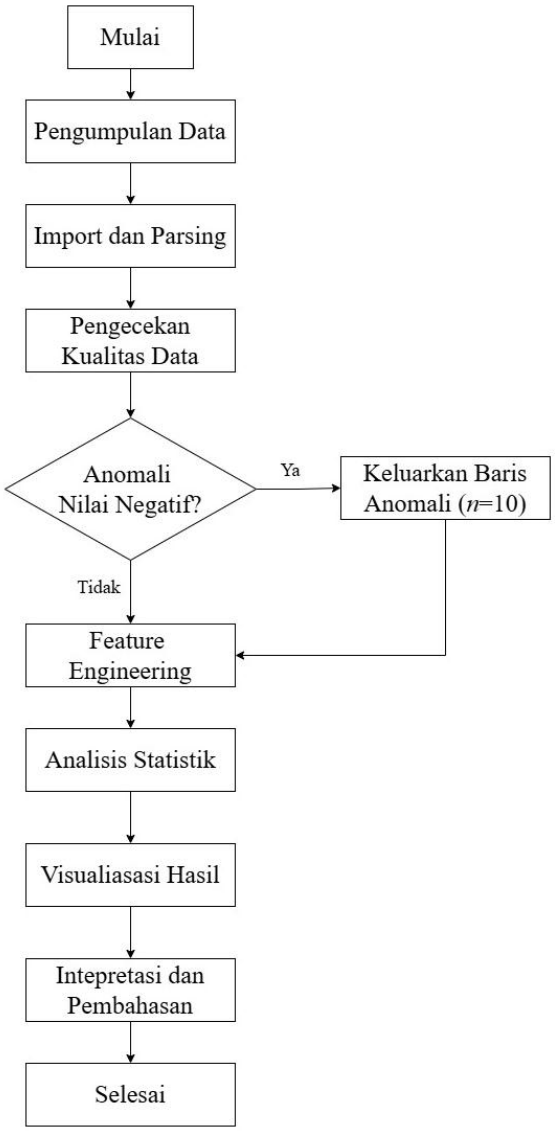
Sumber: Diolah dari berbagai sumber, 2026

Berdasarkan Tabel 1, mayoritas penelitian terdahulu berfokus pada karakterisasi skala sangat besar atau langsung menuju prediksi *machine learning*. Penelitian ini mengisi celah dengan karakterisasi mendalam skala menengah (30 hari, 5 menit, 3 variabel) menggunakan metode statistik yang transparan, *reproducible*, dan pada setiap tahap kunci tervalidasi secara objektif. Kebaruan penelitian ini bukan pada metode statistik individual yang digunakan, melainkan pada integrasi empat dimensi analisis sekaligus (zona temporal tervalidasi, pola mingguan dengan uji asumsi eksplisit, korelasi multivariat, dan deteksi anomali) menjadi satu kerangka kerja karakterisasi yang dapat direplikasi langsung sebagai prosedur pra-pemodelan.

III. METODE PENELITIAN

Penelitian ini menggunakan pendekatan deskriptif-kuantitatif berbasis analisis *time-series*. Seluruh proses, dari praproses, analisis statistik, hingga visualisasi dilaksanakan menggunakan Python. Gambar 1 menyajikan diagram alur metodologi penelitian secara keseluruhan.

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373



Gambar 1. Diagram Alur Metodologi Penelitian

3.1 Sumber dan Deskripsi Data

Data yang digunakan merupakan data trafik jaringan yang dikumpulkan secara otomatis oleh sistem monitoring selama rentang waktu 30 hari. Untuk menjaga kerahasiaan informasi infrastruktur, identitas spesifik jaringan dan periode kalender aktual tidak dicantumkan, dan seluruh nilai *timestamp* dalam *dataset* telah dianonimkan atau digeser, sesuai dengan praktik yang lazim dalam penelitian keamanan dan jaringan komputer; label tanggal pada tabel dan gambar bersifat representatif. Tabel 2 merangkum atribut utama *dataset* yang digunakan.

Tabel 2. Deskripsi *Dataset* Trafik Jaringan

Atribut	Keterangan
Sumber data	Data trafik jaringan (dianonimisasi)
Periode pengamatan	Tanggal 2 s.d 31 Januari 2026 (30 hari penuh)
Interval perekaman	5 menit (288 titik/hari)
Total titik data	8.640 baris × 4 kolom
Variabel	*timestamp, throughput_mbps, packet_rate_pps, active_sessions*
Format file	Microsoft Excel (.xlsx)
Missing values	0 (tidak ada nilai kosong)
Duplikat timestamp	0 (tidak ada duplikat)

Sumber: Data monitoring jaringan (dianonimisasi), Januari 2026

Dataset terdiri atas empat variabel: *timestamp* (cap waktu resolusi 5 menit), *throughput_mbps (Mbps), packet_rate_pps (pps),* dan *active_sessions*. Ketiga variabel numerik ini merepresentasikan kondisi jaringan dari perspektif volume data, intensitas paket, dan jumlah pengguna aktif secara bersamaan.

3.2 Praproses Data

Sebelum analisis dilakukan, data melewati serangkaian tahap praproses untuk memastikan kualitas dan konsistensi sebagaimana dirinci pada Tabel 3.

Tabel 3. Tahapan Praproses Data

No	Tahap	Tindakan	Hasil
1.	*Import and Parsing*	Membaca .xlsx, parsing *timestamp* ke *datetime64*	*DataFrame* 8.640 × 4
2.	Deteksi *missing*	Pengecekan NaN seluruh kolom	0 *missing values*
3.	Deteksi duplikat	Pengecekan *timestamp* duplikat	0 duplikat
4.	Deteksi anomali	Identifikasi nilai *negatif packet_rate* dan *sessions*	10 baris anomali teridentifikasi
5.	Pembersihan data	Mengeluarkan 10 baris anomali dari analisis zona	8.630 baris data bersih
6.	*Feature engineering*	Ekstraksi *hour, day-of-week, date,* zona waktu	4 kolom baru ditambahkan

Sumber: Diolah oleh peneliti, 2026

Ditemukan 10 baris data (0,12%) dengan nilai negatif pada *packet_rate* (5 baris, min -25 *pps*) dan *active_sessions* (6 baris, min -14 sesi, dengan satu

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

baris beririsan pada kedua variabel) yang dikategorikan sebagai artefak pengukuran dan dikeluarkan dari analisis statistik per zona dan uji-*t*. Tidak ditemukan *missing values* maupun duplikat *timestamp*, mengonfirmasi konsistensi interval perekaman selama 30 hari.

3.3 Metode Analisis

Analisis data dilakukan menggunakan delapan metode statistik sebagaimana dirangkum pada Tabel 4. Berbeda dari pendekatan visual semata, segmentasi zona temporal pada penelitian ini divalidasi secara objektif melalui pengujian statistik, rata-rata *throughput* pada setiap pasangan jam yang berdekatan, misalnya jam 06 vs jam 07, diuji menggunakan Welch's *t-test*, dan suatu jam dinyatakan sebagai batas zona (*breakpoint*) apabila perbedaan terhadap jam berikutnya signifikan ($p < 0{,}001$) dan secara praktis besar (selisih rata-rata > 10 *Mbps*). Sebelum uji hipotesis *weekday-weekend*, asumsi normalitas diperiksa dengan uji D'Agostino-Pearson K^2 dan homogenitas varians dengan uji Levene, hasil kedua uji ini menentukan varian uji-*t* independen yang digunakan (Bagian 4.3).

Tabel 4. Metode Analisis yang Diterapkan dalam Penelitian

Metode	Tujuan	Ref. Subbagian
Statistik deskriptif	Menghitung *mean, std, min, max* setiap variabel	Bagian 4.1
Segmentasi zona temporal	Membagi 24 jam ke zona berdasarkan uji-*t breakpoint* antar jam	Bagian 4.2
Analisis pola mingguan	Membandingkan trafik *weekday-weekend*	Bagian 4.3
Uji asumsi statistik	Menguji normalitas (D'Agostino K^2) dan homogenitas varians (Levene)	Bagian 4.3
Uji t independen (Welch)	Menguji signifikansi perbedaan *weekday-weekend*	Bagian 4.3
Korelasi Pearson	Mengukur kekuatan hubungan linear antar variabel	Bagian 4.4
Regresi linier	Memodelkan hubungan kuantitatif antar variabel	Bagian 4.4
Deteksi anomali Z-score	Mengidentifikasi outlier dengan $\|Z\| > 3$	Bagian 4.5

Sumber: Diolah oleh peneliti, 2026

3.4 Alat dan Lingkungan Penelitian

Seluruh proses analisis dan visualisasi dilaksanakan menggunakan Python dalam lingkungan *notebook* interaktif. Penggunaan pustaka *open source* yang terdokumentasi dengan baik memastikan *reproducibility* penelitian ini. Tabel 5 merinci pustaka yang digunakan.

Tabel 5. Pustaka dan Alat yang Digunakan dalam Penelitian

Library/Tools	Versi	Fungsi dalam Penelitian
Python	3.12	Bahasa pemrograman utama seluruh analisis
Pandas	2.2.2	Manipulasi *DataFrame*, agregasi statistik, *feature engineering*
NumPy	1.26.4	Komputasi numerik, kalkulasi metrik statistik
SciPy	1.11.4	Uji statistik (independent *t*-test, regresi linier, *Z-score*)
Matplotlib	3.8.3	Visualisasi: *time-series, line plot, bar chart* dengan anotasi
Seaborn	0.13.2	Visualisasi statistik: *heatmap korelasi, scatter plot*
Scikit-learn	1.4.2	Klasterisasi K-Means untuk validasi silang segmentasi zona
Google Colab/Jupyter	-	Lingkungan eksekusi kode interaktif berbasis web

Sumber: Diolah oleh peneliti, 2026

IV. HASIL PENELITIAN DAN PEMBAHASAN

4.1 Statistik Deskriptif Data Trafik Jaringan

Statistik deskriptif terhadap 8.640 titik data disajikan pada Tabel 6. Nilai rata-rata *throughput* adalah 75,73 *Mbps* dengan standar deviasi 45,83 *Mbps* (koefisien variasi 60,5%), mengindikasikan fluktuasi yang cukup besar antara jam aktif dan tidak aktif, pola yang konsisten dengan karakteristik *non-stasioner* trafik jaringan yang dijelaskan Leland et al. [2] dan Smith [3]. Nilai minimum *throughput* sebesar 1,00 *Mbps* dan maksimum 180,64 *Mbps* menunjukkan rentang dinamis yang lebar. Anomali nilai negatif pada *packet_rate* dan *active_sessions* diidentifikasi sebagai artefak instrumen dan dibahas lebih lanjut pada Bagian 4.5.

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

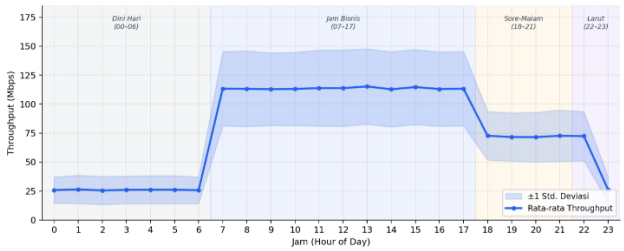
Tabel 6. Statistik Deskriptif Variabel Trafik Jaringan

Variabel	*N*	Rata-rata	*Std. Dev.*	Min	Maks
Throughput (Mbps)	8.640	75,73	45,83	1,00	180,64
Packet Rate (pps)	8.640	1.891,91	1.146,16	-25	4.548
Active Sessions	8.640	377,74	229,60	-14	913

Sumber: Hasil analisis data, 2026. Nilai minimum negatif pada Packet Rate dan Active Sessions merupakan anomali yang dibahas pada Bagian 4.5.

4.2 Analisis Pola Per Jam dan Segmentasi Zona Temporal

Analisis pola per jam mengungkapkan struktur temporal yang sangat konsisten. Gambar 2 menunjukkan rata-rata *throughput* per jam beserta *band* deviasi standar ±1σ. Berbeda dari pendekatan sebelumnya yang menetapkan batas zona berdasarkan inspeksi visual, penelitian ini menguji signifikansi statistik setiap transisi jam ke jam (Welch's *t-test*) untuk menemukan *breakpoint* yang benar-benar didukung data.



Gambar 2. Pola Rata-rata *Throughput* Trafik Jaringan per Jam (Data 30 Hari, Januari 2026)

Hasil pengujian *breakpoint* menunjukkan tiga, bukan empat, transisi yang signifikan secara statistik ($p < 0{,}001$ dan selisih rata-rata > 10 *Mbps*) yaitu antara jam 06 dan 07 (lompatan +341,7%, dari 25,65 menjadi 113,32 *Mbps*), antara jam 17 dan 18 (penurunan -35,9%, dari 113,32 menjadi 72,67 *Mbps*), dan antara jam 22 dan 23 (penurunan -63,7%, dari 72,34 menjadi 26,27 *Mbps*). Sebaliknya, tidak ditemukan transisi signifikan antara jam 21 dan 22 ($p = 0{,}85$) maupun antara jam 23 dan jam 00 ($p = 0{,}55$), kedua pasangan jam ini secara statistik tidak dapat dibedakan. Dengan demikian, jam 22 secara empiris tergabung dengan zona Sore-Malam, dan jam 23 tergabung dengan zona Dini Hari, sehingga struktur temporal trafik yang tervalidasi terdiri atas tiga zona, bukan empat sebagaimana diasumsikan pada analisis visual awal. Validasi silang menggunakan klasterisasi K-*Means* pada rata-rata *throughput* per jam mengonfirmasi hasil yang sama, metode elbow menunjukkan penurunan inertia paling tajam terjadi pada k=3, dan pengelompokan yang dihasilkan (k=3) identik dengan ketiga zona *breakpoint* di atas.

Tabel 7. Statistik *Throughput* per Zona Temporal (Tervalidasi Statistik)

Zona Waktu	Jam	Rata-rata (*Mbps*)	Std. Dev.	Min (*Mbps*)	Utilisasi (%)
Dini Hari	23.00-06.59	26,06	11,72	1,00	14,4%
Jam Bisnis	07.00-17.59	113,54	32,25	45,75	62,9%
Sore-Malam	18.00-22.59	72,14	21,28	24,03	39,9%

Sumber: Hasil analisis data, 2026. Utilisasi dihitung sebagai rata-rata zona dibagi throughput maksimum tercatat (180,64 Mbps)

Zona Dini Hari (23.00-06.59) mencatat *throughput* terendah (26,06 *Mbps*, utilisasi 14,4%) dengan deviasi yang relatif stabil (σ = 11,72 *Mbps*), mencerminkan aktivitas jaringan minimal saat mayoritas pengguna tidak aktif. Zona Jam Bisnis (07.00-17.59) merupakan zona dengan trafik tertinggi (113,54 Mbps, utilisasi 62,9%), transisi dari Dini Hari ke Jam Bisnis terjadi secara *abrupt* pada jam 06-07, mencerminkan aktivitas terstruktur yang memengaruhi penggunaan jaringan secara serempak pola serupa dilaporkan Wang et al. [17] pada trafik seluler. Zona Sore-Malam (18.00-22.59) menunjukkan trafik menengah (72,14 *Mbps*, utilisasi 39,9%), mencakup periode transisi setelah jam operasional hingga menjelang tengah malam.

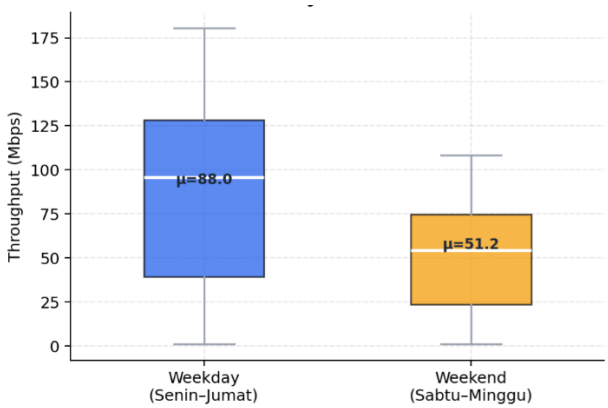
Temuan ini memiliki implikasi metodologis penting bagi studi sejenis, rata-rata dan standar deviasi tinggi yang sebelumnya dikaitkan dengan zona "Larut Malam" pada pendekatan segmentasi visual (49,31 Mbps; σ = 28,57) ternyata merupakan artefak statistik dari penggabungan dua populasi jam yang berbeda secara signifikan (jam 22 yang tergolong Sore-Malam, dan jam 23 yang tergolong Dini Hari), bukan cerminan variabilitas perilaku pengguna yang sesungguhnya. Hal ini menegaskan pentingnya validasi statistik eksplisit, bukan inspeksi visual semata, dalam segmentasi data *time-series,* sejalan dengan rekomendasi metodologis pada tinjauan pustaka Bagian 2.8.1.

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

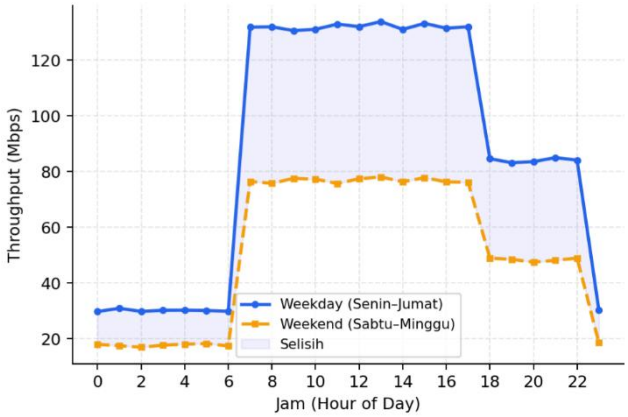
Nilai puncak absolut *throughput* tercatat 180,64 *Mbps* pada jam-jam tertentu di zona bisnis. Karena kapasitas terpasang (*installed capacity*) infrastruktur tidak berada dalam cakupan *dataset* ini, nilai tersebut disajikan sebagai batas atas empiris (*observed maximum*), bukan persentase terhadap kapasitas sebenarnya, pemantauan berkelanjutan tetap disarankan mengingat rata-rata zona bisnis (62,9% dari nilai puncak empiris) berada pada rentang yang lazim memicu evaluasi kapasitas dalam praktik manajemen jaringan [1].

4.3 Analisis Pola Mingguan *Weekday-Weekend*

Gambar 3 dan Gambar 4 menyajikan perbandingan distribusi dan pola per jam antara *weekday* dan *weekend*. Sebelum pengujian hipotesis, asumsi statistik diperiksa terlebih dahulu, uji normalitas D'Agostino-Pearson K^2 menolak hipotesis normalitas pada kedua kelompok (*weekday*: K^2 = 22.469,86; *weekend*: K^2 = 3.679,77; keduanya $p < 0{,}001$), namun nilai *skewness* (*weekday*: -0,056; *weekend*: -0,033) dan kurtosis (*weekday*: -1,305; *weekend*: -1,244) yang relatif kecil menunjukkan penyimpangan dari normalitas tidak ekstrem, dan dengan ukuran sampel besar (*n weekday* = 5.760; *n weekend* = 2.880), validitas uji-*t* tetap terjaga melalui *Central Limit Theorem*. Uji Levene menunjukkan varians kedua kelompok tidak homogen (statistik Levene = 1.172,78; $p < 0{,}001$), sehingga uji-*t* Welch yang tidak mengasumsikan varians sama digunakan sebagai pengganti uji-*t Student* konvensional.



Gambar 3. Perbandingan Distribusi dan Pola Harian *Throughput*: *Weekday-Weekend* (Distribusi *Throughput*)



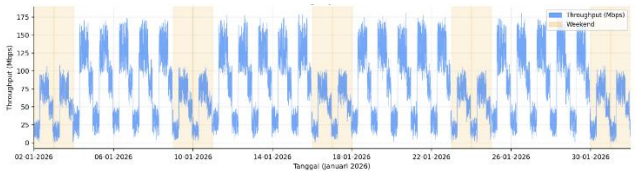
Gambar 4. Perbandingan Distribusi dan Pola Harian *Throughput: Weekday-Weekend* (Perbandingan Pola Harian)

Tabel 8. Perbandingan Statistik *Throughput Weekday-Weekend*

Tipe Hari	*N*	Rata-rata (*Mbps*)	Std. Deviasi	Nilai *t* (Welch)	*p-value*
Weekday (Senin-Jumat)	5.760	88,01	47,96	44.83	< 0,001
Weekend (Sabtu-Minggu)	2.880	51,16	28,21	-	-

Keterangan: Uji Welch's t-test (varians tidak diasumsikan sama, mengikuti hasil uji Levene); signifikan pada $p < 0{,}001$.

Rata-rata *throughput weekday* (88,01 ± 47,96 *Mbps*) lebih tinggi 72,0% dibandingkan weekend (51,16 ± 28,21 *Mbps*; $t = 44{,}83$; $p < 0{,}001$), sehingga H_0 (tidak terdapat perbedaan rata-rata antara *weekday* dan *weekend*) ditolak, dan disimpulkan terdapat perbedaan rata-rata yang signifikan secara statistik. Pola ini berlaku konsisten selama empat minggu penuh yaitu Senin 87,85 *Mbps*, Selasa 88,22 *Mbps*, Rabu 88,27 *Mbps*, Kamis 88,07 *Mbps*, Jumat 87,64 *Mbps*, Sabtu 51,23 *Mbps*, dan Minggu 51,09 *Mbps*.

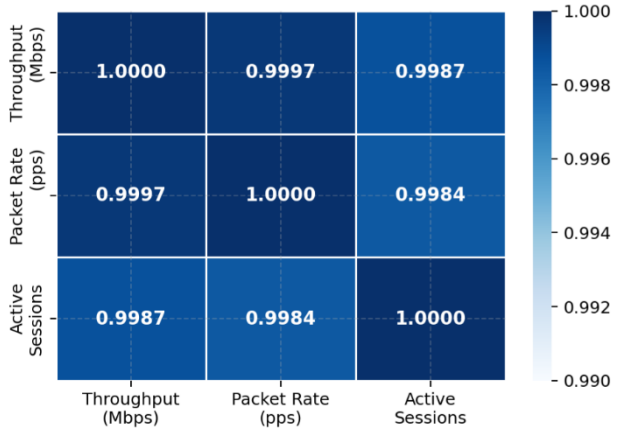


Gambar 5. *Time-Series Throughput* Trafik Jaringan Selama 30 Hari (Januari 2026)

Jurnal EEMISAS, Vol. 5, No.2, Desember 2026, Hal. 01 - 13
Journal of Electrical, Electronic, Mechanical, Informatic and Social Applied Science
ISSN 2964-1373

Gambar 5 memperlihatkan *weekly seasonality* yang sangat konsisten sepanjang seluruh periode pengamatan, sejalan dengan temuan Moreira et al. [19] bahwa pola *weekday-weekend* merupakan karakteristik struktural trafik yang bersifat universal. Konsistensi pola mingguan ini memiliki implikasi praktis pada jadwal pemeliharaan jaringan, pembaruan sistem, atau aktivitas intensif *bandwidth* sebaiknya dijadwalkan pada periode *weekend* untuk meminimalkan gangguan terhadap pengguna aktif.

4.4 Analisis Korelasi Antar Variabel Trafik



Gambar 6. Analisis Korelasi Antar Variabel Trafik Jaringan-Matriks Korelasi Pearson Antar Variabel Trafik

Tabel 9. Matriks Korelasi Pearson Antar Variabel Trafik Jaringan

Variabel	*Throughput*	*Packet Rate*	*Active Sessions*
Throughput (Mbps)	1,0000	0,9997**	0,9987**
Packet Rate (pps)	0,9997**	1,0000	0,9984**
Active Sessions	0,9987**	0,9984**	1,0000

Keterangan: Hasil analisis data, 2026. ** signifikan pada $p < 0{,}001$ (two-tailed)

Ketiga pasang variabel menunjukkan korelasi Pearson yang sangat tinggi ($r > 0{,}998$; $p < 0{,}001$). Korelasi antara *throughput* dan *packet rate* mencapai $r = 0{,}9997$, sementara korelasi antara *throughput* dan *active sessions* adalah $r = 0{,}9987$. Hubungan *throughput-packet rate* dapat dinyatakan secara regresi linier sebagai *Packet Rate* = 25,00 x *Throughput* - 1,18 ($R^2 = 0{,}9994$).

Nilai korelasi yang mendekati 1,0 ini mengindikasikan ketiga variabel merespons sumber variasi yang sama, yaitu intensitas aktivitas pengguna jaringan sejalan dengan temuan Di Mauro et al. [16] pada trafik VoIP. Perlu dicatat bahwa korelasi setinggi ini juga mengandung implikasi praktis ganda yaitu dari sisi efisiensi monitoring operasional, pengukuran satu variabel dapat menjadi proksi yang memadai untuk kondisi umum jaringan pada *dataset* dan konteks infrastruktur ini, namun dari sisi pemodelan *machine learning multivariat*, kondisi ini berpotensi menimbulkan multikolinearitas yang dapat memengaruhi stabilitas dan interpretabilitas koefisien model prediktif jika ketiga variabel digunakan bersamaan sebagai prediktor independen.

4.5 Deteksi dan Analisis Anomali Data

Penerapan metode Z-*score* ($|Z| > 3$) terhadap 8.640 titik data *throughput* tidak menemukan satu pun *outlier* statistik. Variasi *throughput* yang lebar (1,00-180,64 *Mbps*) merupakan refleksi pola penggunaan periodik yang telah dikonfirmasi pada Bagian 4.2 - 4.3, bukan kejadian anomali. Satu-satunya anomali yang teridentifikasi adalah 10 baris nilai negatif (0,12%) pada variabel *packet_rate* dan *active_sessions*, yang secara fisik tidak mungkin terjadi pada trafik jaringan nyata dan karenanya dikategorikan sebagai artefak instrumen, bukan gangguan jaringan yang sesungguhnya.

Temuan ini mengindikasikan bahwa data trafik yang dianalisis merupakan sinyal periodik yang secara substansial dapat dijelaskan oleh faktor temporal, yaitu pola harian dan mingguan yang telah divalidasi pada Bagian 4.2 dan 4.3, dengan tingkat anomali titik yang sangat rendah pada resolusi pengamatan ini. Kondisi ini merupakan prasyarat yang baik untuk pemodelan prediktif berbasis *time-series*, karena rendahnya proporsi *outlier* ekstrem berpotensi mengurangi kebutuhan *preprocessing* yang intensif, perlu dicatat pula bahwa metode Z-*score* sensitif terhadap *outlier* titik namun kurang sensitif terhadap anomali kontekstual, misalnya pola jam sibuk yang bergeser, sehingga kesimpulan ini bersifat spesifik terhadap jenis anomali yang diuji, sebagaimana diakui pada Bagian 5.2.

4.6 Implikasi untuk Manajemen dan Perencanaan Kapasitas Jaringan

Keempat implikasi berikut diturunkan langsung dari temuan tervalidasi pada Bagian 4.2 - 4.5, yaitu:

1. Implikasi terhadap QoS (*Quality of Service*). Utilisasi 62,9% di zona bisnis, dengan nilai puncak empiris 180,64 *Mbps*, mengindikasikan perlunya mekanisme *traffic shaping* dan prioritisasi layanan kritis berbasis *time-of-day policy* pada jam 07.00-17.59 di hari kerja.

2. Implikasi terhadap perencanaan kapasitas. Selisih *weekday-weekend* sebesar 72,0% (tervalidasi Welch's *t*-test) memberikan landasan kuantitatif bagi tim jaringan untuk merencanakan kebutuhan kapasitas berdasarkan dua skenario penggunaan yang berbeda.

3. Implikasi terhadap penjadwalan pemeliharaan. Zona Dini Hari (23.00-06.59) dengan utilisasi hanya 14,4% merupakan jendela waktu paling optimal untuk pemeliharaan jaringan dan pembaruan sistem. Periode *weekend* secara keseluruhan juga menawarkan kondisi jaringan yang jauh lebih longgar.

4. Implikasi terhadap pemodelan prediktif. Struktur temporal tiga zona yang tervalidasi, pola harian konsisten, dan *weekly seasonality* yang kuat menjadikan *dataset* ini kandidat yang relevan untuk pemodelan LSTM atau SARIMA sebagai langkah penelitian selanjutnya, dengan zona sebagai fitur kategorikal yang telah teruji validitasnya.

V. KESIMPULAN

5.1 Kesimpulan

Penelitian ini mengajukan dan menerapkan kerangka kerja karakterisasi statistik empat dimensi terhadap trafik jaringan berbasis *time-series* selama 30 hari dengan resolusi 5 menit, menggunakan delapan metode statistik yang salah satu kontribusi metodologis utamanya adalah validasi objektif bukan visual terhadap segmentasi zona temporal. Empat temuan utama dihasilkan:

1. Trafik terpola dalam tiga zona temporal yang berbeda secara statistik dan tervalidasi melalui uji *breakpoint* antar jam serta klasterisasi K-*Means* (utilisasi 14,4%-62,9%), mengoreksi asumsi segmentasi empat zona berbasis visual yang lazim digunakan pada pendekatan konvensional.

2. Trafik *weekday* (88,01 *Mbps*) lebih tinggi 72,0% dibanding *weekend* (51,16 *Mbps*; $t = 44{,}83$; $p < 0{,}001$, uji Welch mengikuti hasil uji Levene) dengan *weekly seasonality* konsisten selama empat minggu.

3. Korelasi Pearson antar tiga variabel trafik sangat tinggi ($r > 0{,}998$; $p < 0{,}001$), yang pada konteks *dataset* ini mendukung efisiensi pemantauan satu variabel namun juga menandakan risiko multikolinearitas bagi pemodelan multivariat lanjutan.

4. Tidak ditemukan *outlier* titik ekstrem pada throughput (Z-*score*: 0 titik $|Z| > 3$), mengonfirmasi data sebagai sinyal periodik yang secara substansial dapat dijelaskan oleh faktor temporal, sebagai fondasi bagi pemodelan prediktif lanjutan.

Kontribusi ilmiah penelitian ini terletak pada pengintegrasian keempat temuan tersebut ke dalam satu prosedur karakterisasi yang dapat direplikasi dan diaudit, khususnya metode validasi *breakpoint* untuk segmentasi zona dan pengujian asumsi eksplisit sebelum uji hipotesis yang dapat diterapkan pada *dataset* trafik jaringan lain sebagai tahap pra-pemodelan yang lebih rigor dibandingkan pendekatan visual konvensional.

Penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan pada interpretasi dan replikasi. Pertama, data bersumber dari satu segmen jaringan selama satu periode 30 hari, sehingga generalisasi ke infrastruktur atau musim lain memerlukan validasi lanjutan. Kedua, metode Z-*score* yang digunakan efektif untuk anomali titik namun tidak dirancang untuk mendeteksi anomali kontekstual

atau kolektif. Ketiga, kapasitas terpasang infrastruktur tidak berada dalam cakupan *dataset*, sehingga interpretasi utilisasi terbatas pada nilai puncak empiris, bukan persentase kapasitas sebenarnya.

5.2 Saran dan Rekomendasi

Penelitian lanjutan disarankan untuk:

1. Mengimplementasikan model LSTM atau SARIMA dengan zona temporal tervalidasi dan *weekly seasonality* sebagai fitur kategorikal dan eksogen.
2. Memperluas periode pengamatan ke lebih dari 90 hari, mencakup lebih dari satu segmen jaringan, dan menambahkan variabel jenis protokol (*protocol*).
3. Melakukan analisis komparatif antar segmen jaringan untuk menguji generalisasi struktur tiga zona yang ditemukan.
4. Memformalisasi temuan zona temporal ke dalam *early warning system* otomatis berbasis *time-of-day policy*.
5. Membandingkan efektivitas *Z-score* dengan metode deteksi anomali kontekstual atau kolektif seperti *Isolation Forest* atau *Local Outlier Factor* (LOF), khususnya untuk menangkap jenis anomali yang tidak terdeteksi metode saat ini.

REFERENSI

[1] Cisco Systems, "*Cisco Annual Internet Report* (2018-2023)," Cisco Public, 2020. [Online]. *Available:* https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "*On the self-similar nature of Ethernet traffic (extended version)*," IEEE/ACM Trans. Netw., vol. 2, no. 1, pp. 1-15, 1994, doi: 10.1109/90.282603.

[3] R. D. Smith, "*Internet traffic periodicities andoscillations: A brief review*," arXiv preprint arXiv:0904.3797, 2009.

[4] J. F. Torres, D. Hadjout, A. Sebaa, F. MartinezAlvarez, and A. Troncoso, "*Deep learning for time series forecasting: A survey," Big Data*, vol. 9, no. 1, pp. 3-21, 2021, doi: 10.1089/big.2020.0159.

[5] A. Azzouni and G. Pujolle, "*A long short-termmemory recurrent neural network framework for network traffic matrix prediction*," arXiv preprint arXiv:1705.05690, 2017.

[6] A. S. Tanenbaum and D. J. Wetherall, *ComputerNetworks*, 5th ed. Boston, MA, USA: Pearson, 2011.

[7] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G.M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[8] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 7th ed. Hoboken, NJ, USA: Wiley, 2018.

[9] R. A. Fisher, *Statistical Methods for ResearchWorkers*, 5th ed. Edinburgh, Scotland: Oliver and Boyd, 1934.

[10] R. B. D'Agostino, A. Belanger, and R. B. D'Agostino Jr., "*A suggestion for using powerful and informative tests of normality*," Am. Stat., vol. 44, no. 4, pp. 316-321, 1990, doi: 10.1080/00031305.1990.10475751.

[11] B. L. Welch, "*The generalization of 'Student's' problem when several different population variances are involved*," Biometrika, vol. 34, no. 1-2, pp. 28-35, 1947, doi: 10.1093/biomet/34.1-2.28.

[12] "Student" [W. S. Gosset], "*The probable error of a mean*," Biometrika, vol. 6, no. 1, pp. 1-25, 1908.

[13] V. Chandola, A. Banerjee, and V. Kumar, "*Anomaly detection: A survey*," ACM Comput. Surv., vol. 41, no. 3, Art. no. 15, 2009, doi: 10.1145/1541880.1541882.

[14] Purnawansyah, Haviluddin, R. Alfred, and A. F. O. Gaffar, "*Network traffic time series performance analysis using statistical methods*," Knowl. Eng. Data Sci., vol. 1, no. 1, pp. 1-7, 2018, doi: 10.17977/um018v1i12018p1-7.

[15] G. Millan, G. Lefranc, R. Osorio-Comparan, and V. Lomas-Barrie, "*Time series analysis of computer network traffic in a dedicated link aggregation*," arXiv preprint arXiv:2107.05484, 2021.
[16] M. Di Mauro, G. Galatro, F. Postiglione, W. Song, and A. Liotta, "*Multivariate time series characterization and forecasting of VoIP traffic in real mobile networks*," IEEE Trans. Netw. Serv. Manag., vol. 21, no. 1, pp. 851-865, 2024, doi: 10.1109/TNSM.2023.3295748.
[17] L. Wang, J. Zhang, Z. Zhang, and J. Zhang, "*Analytic network traffic prediction based on user behavior modeling*," IEEE Commun. Lett., 2023, doi: 10.1109/LCOMM.2023.3278100.
[18] J. Kepner et al., "*Multi-temporal analysis and scaling relations of 100,000,000,000 network packets*," in Proc. 2020 IEEE High Perform. Extreme Comput. Conf. (HPEC), 2020, doi: 10.1109/HPEC43674.2020.9286235.
[19] J. D. O. Moreira, A. P. Pasarkar, W. Chen, W. Hu, J. Janak, and H. Schulzrinne, "*Social distancing and the internet: What can network performance measurements tell us?*," arXiv preprint arXiv:2012.09850, 2020.
[20] M. Ahmed, A. N. Mahmood, and J. Hu, "*A survey of network anomaly detection techniques*," J. Netw. Comput. Appl., vol. 60, pp. 19-31, 2016, doi: 10.1016/j.jnca.2015.11.016.
[21] S. Zehra et al., "*Machine learning-based anomaly detection in NFV: A comprehensive survey*," Sensors, vol. 23, no. 11, Art. no. 5340, 2023, doi: 10.3390/s23115340.
[22] J. Ferreira, S. I. Lopes, and A. Curado, "*Data traffic prediction for 5G and beyond: Emerging trends, challenges, and future directions*," Electronics, vol. 14, no. 23, Art. no. 4611, 2025, doi: 10.3390/electronics14234611.