

APPLICATION OF MACHINE LEARNING IN ANALYZING BANDWIDTH USAGE PATTERNS FOR INTERNET SERVICE PROVIDERS

PENERAPAN MACHINE LEARNING DALAM ANALISIS POLA PENGGUNAAN BANDWIDTH PADA PROVIDER INTERNET

Alfin Hilmy Nurmakhlu¹, Eri Zuliarso²

^{1,2}Universitas Stikubank, Jalan Jl. Tri Lomba Juang, Kota Semarang, Jawa Tengah
email alfinhilmy0039@mhs.unisbank.ac.id¹, eri299@edu.unisbank.ac.id²

Abstract – This study aims to address bandwidth management challenges faced by Internet Service Providers (ISP) through the application of machine learning techniques for analyzing usage patterns and forecasting future demand. A key novelty of this research lies in the combined use of K-Means clustering for dynamic customer segmentation based on real-time utilization patterns, followed by accurate short-term forecasting using Random Forest regression, specifically tailored for corporate client bandwidth planning. Data was collected from 12 corporate customers over a three-month period (January–March 2025) at five-minute intervals using the PRTG Network Monitor. The analytical workflow included data preprocessing, customer segmentation using K-Means clustering, and short-term bandwidth prediction using Random Forest regression. The clustering results classified customers into three main categories: underutilized, optimal, and overutilized, with a silhouette score of 0.663 indicating good cluster separation. The regression model achieved a coefficient of determination (R^2) of 0.931, a Mean Absolute Error (MAE) of 0.036 Mbps, and a Root Mean Square Error (RMSE) of 0.062 Mbps, demonstrating high predictive accuracy for operational planning. This study is limited by the relatively short observation period and the exclusion of external variables in the modeling process. For future work, the use of deep learning methods such as Long Short-Term Memory (LSTM) or Temporal Convolutional Networks (TCN) is recommended, along with the integration of external features such as time-based traffic anomalies and customer profiles, to enhance model robustness, accuracy, and generalization.

Keywords - Machine Learning, Bandwidth Usage Patterns, PRTG, K-Means, Random Forest

Abstrak – Penelitian ini bertujuan untuk mengatasi tantangan pengelolaan bandwidth yang dihadapi oleh Penyedia Layanan Internet (ISP) melalui penerapan teknik machine learning untuk menganalisis pola penggunaan dan meramalkan permintaan di masa depan. Kebaruan penelitian ini terletak pada penerapan kombinasi metode clustering K-Means untuk segmentasi pelanggan dinamis berdasarkan pola pemanfaatan bandwidth secara real-time, diikuti dengan peramalan jangka pendek yang akurat menggunakan regresi Random Forest, khususnya untuk perencanaan bandwidth klien korporat. Data dikumpulkan dari 12 pelanggan korporat selama periode tiga bulan (Januari–Maret 2025) dengan interval lima menit menggunakan PRTG Network Monitor. Alur analisis mencakup pra-pemrosesan data, segmentasi pelanggan menggunakan clustering K-Means, dan prediksi bandwidth jangka pendek menggunakan regresi Random Forest. Hasil clustering mengklasifikasikan pelanggan ke dalam tiga kategori utama: underutilized (kurang dimanfaatkan), optimal, dan overutilized (berlebihan), dengan skor siluet sebesar 0.663 yang menunjukkan pemisahan klaster yang baik. Model regresi mencatat koefisien determinasi (R^2) sebesar 0.931, Mean Absolute Error (MAE) sebesar 0.036 Mbps, serta Root Mean Square Error (RMSE) sebesar 0.062 Mbps, yang menunjukkan akurasi prediksi tinggi untuk perencanaan operasional. Keterbatasan penelitian ini antara lain periode observasi yang relatif singkat dan tidak dimasukkannya variabel eksternal dalam proses pemodelan. Untuk pengembangan selanjutnya, disarankan penggunaan metode deep learning seperti Long Short-Term Memory (LSTM) atau Temporal Convolutional Networks (TCN), serta integrasi fitur eksternal seperti anomali lalu lintas berbasis waktu dan profil pelanggan, guna meningkatkan ketahanan, akurasi, dan generalisasi model.

Kata Kunci - Machine Learning, Pola Penggunaan Bandwidth, PRTG, K-Means, Random Forest

I. PENDAHULUAN

Pertumbuhan pesat jumlah pengguna internet dalam beberapa tahun terakhir telah meningkatkan kebutuhan akan pengelolaan infrastruktur jaringan yang efisien. Menurut data dari International Telecommunication Union (ITU), jumlah pengguna internet global mencapai 5,3 miliar pada tahun 2023, naik dari 4,9 miliar pada 2021 [1]. Di Indonesia, Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) mencatat penetrasi internet sebesar 79% dengan jumlah pengguna mencapai 215 juta pada tahun 2022[2]. Lonjakan ini membawa tantangan signifikan bagi penyedia layanan internet, khususnya dalam manajemen bandwidth. Bandwidth merupakan sumber daya jaringan yang krusial karena menentukan kapasitas transmisi data. Pengelolaan bandwidth yang tidak optimal dapat menyebabkan berbagai permasalahan layanan, seperti koneksi lambat, buffering saat streaming, dan gangguan pada aplikasi real-time[1], yang pada akhirnya memengaruhi kualitas layanan dan kepuasan pelanggan. Dalam praktiknya, sering terjadi ketidakseimbangan antara kapasitas bandwidth yang dialokasikan dan pola penggunaan aktual pelanggan. Over-provisioning menyebabkan pelanggan membayar lebih untuk kapasitas yang tidak sepenuhnya digunakan, sedangkan under-provisioning berdampak pada kualitas pengalaman pengguna. Kedua kondisi ini berpotensi merugikan kedua belah pihak [3]. Untuk mengatasi permasalahan tersebut, analisis pola penggunaan bandwidth menjadi sangat penting dalam perencanaan dan pengelolaan kapasitas jaringan. Namun, kompleksitas dan volume data monitoring jaringan membuat analisis manual kurang efektif. Oleh karena itu, pendekatan machine learning dianggap sebagai solusi yang efektif untuk mengidentifikasi pola tersembunyi dalam data dan melakukan prediksi berdasarkan data historis [4].

Penelitian ini bertujuan menerapkan teknik machine learning dalam menganalisis data monitoring bandwidth pelanggan yang diperoleh dari sistem PRTG (Paessler Router Traffic Grapher). Metode yang digunakan meliputi segmentasi dengan K-Means Clustering dan klasifikasi menggunakan Random Forest untuk mengkategorikan penggunaan bandwidth ke dalam tiga kategori: rendah, ideal, atau berlebihan. Selain itu, model dilengkapi dengan modul prediksi nilai pemakaian bandwidth satu jam ke depan menggunakan Random Forest Regressor. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengambilan keputusan strategis oleh penyedia layanan internet, seperti penyesuaian kapasitas bandwidth, rekomendasi paket layanan, serta peningkatan kepuasan pelanggan melalui pendekatan berbasis data.

II. SIGNIFIKASI STUDI

Penelitian ini memberikan kontribusi signifikan dengan mengembangkan sebuah model klasifikasi penggunaan bandwidth pelanggan yang tidak hanya akurat, tetapi juga memberikan wawasan yang dapat diinterpretasikan untuk mendukung keputusan bisnis. Dari sisi nilai kebaruannya, penelitian ini terletak pada perancangan kerangka kerja yang menggabungkan analisis kluster dan klasifikasi. Metode ini memperkaya segmentasi pelanggan dengan mengintegrasikan pengelompokan pola penggunaan bandwidth menggunakan K-Means Clustering, lalu secara sistematis membangun model klasifikasi berdasarkan segmentasi tersebut menggunakan Random Forest, untuk memberikan rekomendasi bisnis yang actionable. Pendekatan ini menghasilkan model yang lebih informatif dan transparan, menjadi alternatif yang kuat bagi pendekatan machine learning tunggal yang kurang menjelaskan hubungan antar variabel. Implikasi praktis dan operasional dari penelitian ini sangat relevan, dimana dengan memanfaatkan data real-time dari PRTG Network Monitor, model ini bisa dikatakan efektif untuk berfungsi sebagai alat analisis lanjutan di layanan internet, membantu mengidentifikasi pelanggan yang berpotensi untuk peningkatan bandwidth dan menyesuaikan penawaran layanan. Hal ini berpotensi besar meningkatkan efisiensi operasional dan

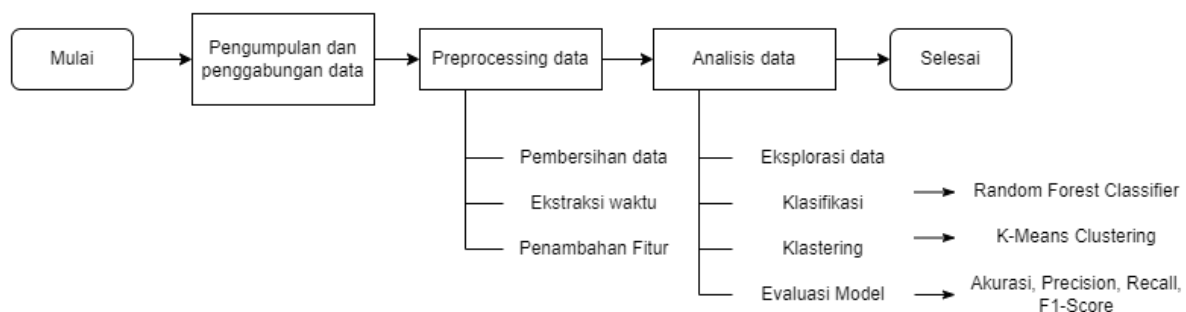
memperluas pendapatan bagi penyedia layanan internet. Karena model ini memanfaatkan data yang dikumpulkan secara periodik dan metode machine learning yang terbukti handal, sistem ini sangat mungkin diadopsi oleh ISP untuk meningkatkan kualitas layanan dan kepuasan pelanggan. Pada akhirnya, dengan memfasilitasi rekomendasi bisnis yang tepat sasaran berdasarkan pola penggunaan, penelitian ini berkontribusi langsung pada tujuan operasional penyedia layanan internet yang lebih besar, yaitu optimalisasi sumber daya jaringan dan peningkatan kualitas layanan.

A. Studi literatur

Penelitian sebelumnya secara konsisten menyoroti efektivitas pembelajaran mesin (machine learning) dalam menganalisis data jaringan dan penggunaan bandwidth, dengan banyak studi yang selaras dengan pendekatan dalam hal metode, sumber data, dan tujuan untuk mendukung keputusan operasional dan bisnis bagi penyedia layanan internet. Beberapa studi terdahulu telah memberikan kontribusi yang signifikan namun menyisakan ruang untuk perbaikan. Penelitian oleh Setiawan dan Wibowo (2023) serta Khan dan Ibrahim (2024) memang menunjukkan akurasi tinggi melalui metode Random Forest untuk deteksi anomali dan klasifikasi paket jaringan, namun tidak mengeksplorasi penggabungan dengan teknik klasterisasi untuk segmentasi awal pelanggan [5], [6], [7]. Sementara itu, penelitian oleh Arnomo dan Yulia (2021), Fauzan dkk (2025), dan Yasriady (2022) menggunakan K-Means Clustering untuk segmentasi pelanggan berdasarkan pola penggunaan, tetapi hanya sebatas pengelompokan tanpa melanjutkan ke tahap klasifikasi untuk rekomendasi yang spesifik [8], [9], [10], [11]. Di sisi lain, pendekatan penggunaan data real-time dari alat seperti PRTG Network Monitor telah terbukti efektif dalam pengumpulan data berkala untuk representasi lalu lintas dinamis, seperti yang ditunjukkan oleh Alizan dan Zolkipli (2024) serta Hendrian (2021), namun belum diterapkan secara optimal dalam model kombinasi klasterisasi dan klasifikasi untuk rekomendasi bisnis [11], [12], [13]. Inspirasi juga datang dari Samaan dan Jeiad (2023), yang menegaskan bahwa klasifikasi lalu lintas yang tepat secara signifikan membantu dalam memahami dan mengelola aliran data jaringan, meskipun tidak secara langsung digunakan dalam model kombinasi seperti yang diusulkan penelitian ini [14]. Berdasarkan hasil telaah literatur, masih terdapat sejumlah celah penting yang belum terisi secara optimal.

B. Metodologi Penelitian

Tahapan-tahapan dalam penelitian prediksi dan segmentasi penggunaan bandwidth pelanggan dapat dilihat pada gambar di bawah ini.



Gambar 1 Tahapan Penelitian

Gambar di atas merupakan tahapan penelitian berupa studi literatur, analisis masalah, pengumpulan data, preprosesing, ekstraksi fitur, klasifikasi/clustering, dan evaluasi. Berikut penjelasan tahapan penelitian.

1. Studi literatur dan Analisis masalah

Tahap awal dilakukan dengan mengkaji literatur terkini terkait pemanfaatan machine learning untuk analisis traffic jaringan, teknik segmentasi pelanggan berbasis K-Means Clustering, serta penerapan klasifikasi Random Forest dalam prediksi penggunaan bandwidth. Analisis masalah dilakukan untuk mengidentifikasi permasalahan operasional seperti ketidaksesuaian kapasitas bandwidth yang disediakan dengan pemakaian aktual pelanggan.

2. Pengumpulan data

Dataset yang digunakan diperoleh dari sistem monitoring jaringan Paessler Router Traffic Grapher (PRTG). Data dikumpulkan dari 12 pelanggan korporat yang menggunakan layanan internet dedicated 100 Mbps selama periode Januari hingga Maret 2025. Data direkam dengan interval 5 menit. Setelah penggabungan dan pembersihan data, diperoleh total 36.754 baris data siap untuk dianalisis.

3. Preprocessing

Tujuan tahap ini adalah membersihkan data dan menstandarisasi format untuk efisiensi komputasi dan kualitas analisis. Langkah-langkah preprocessing meliputi:

- Data Cleaning, Menghapus satuan 'kbit/s', mengkonversi nilai ke tipe numerik float, menghapus data duplikat, dan menangani missing value atau anomali data.
- Ekstraksi dan Transformasi Waktu: Memecah kolom Date Time menjadi komponen terpisah seperti tanggal dan jam.
- Feature Engineering, Membentuk fitur-fitur baru yang lebih representatif untuk analisis dan pemodelan.
- Persentase Penggunaan Bandwidth (Utilisasi) : Menghitung tingkat utilisasi bandwidth terhadap kapasitas maksimum (100 Mbps).

$$Utilisasi (\%) = \frac{Traffic\ Total\ (Mbps)}{Kapasitas\ Maksimal\ (100\ Mbps)} \times 100 \quad (1)$$

- Fitur Jam Kerja/Time Category: Menambahkan fitur boolean atau kategori untuk mengidentifikasi waktu kerja (07.00-18.00 WIB) vs non-jam kerja, serta kategori peak/off-peak.
- Rolling Average: Menghitung rata-rata penggunaan bandwidth dalam periode tertentu
- Burst Flag: Menandai adanya lonjakan penggunaan bandwidth berdasarkan kriteria tertentu.

4. Ekstraksi fitur

Ekstraksi fitur dilakukan untuk membentuk atribut-atribut yang relevan dan informatif dari data yang telah melalui preprocessing. Fitur-fitur yang dihasilkan antara lain:

- Hour (Jam): Representasi waktu dalam jam (0–23).
- Day_Type (Tipe Hari): Kategori hari (weekday atau weekend).
- Time_Category (Kategori Waktu): Segmentasi waktu seperti jam sibuk (peak) atau tidak sibuk (off-peak).
- Rolling_Avg_6h: Nilai rata-rata penggunaan bandwidth selama 6 jam terakhir.
- Burst_Flag: Indikator boolean (0 atau 1) yang menandai adanya lonjakan penggunaan bandwidth.
- Utilisasi (%): Persentase penggunaan bandwidth terhadap kapasitas maksimum.
-

5. Klasifikasi dan Clustering

Pemodelan dilakukan dalam dua skenario utama:

- Clustering (Unsupervised Learning): Menggunakan algoritma K-Means Clustering untuk mengelompokkan pelanggan berdasarkan pola penggunaan bandwidth yang serupa tanpa memerlukan label awal. Tujuannya adalah segmentasi pelanggan.
- Klasifikasi (Supervised Learning): Menggunakan algoritma Random Forest Classification untuk membangun model prediksi kategori penggunaan bandwidth (rendah, ideal, tinggi) berdasarkan fitur-fitur yang tersedia.
- Regresi (Supervised Learning) : Menggunakan algoritma Random Forest Regressor untuk memprediksi nilai penggunaan bandwidth (Traffic Total) satu jam ke depan. Random Forest dipilih karena kemampuannya menangani data non-linear, generalisasi yang baik, serta kestabilan terhadap outlier.

6. Evaluasi

Evaluasi performa model dilakukan untuk menilai keakuratan dan efektivitas model dalam melakukan klasifikasi, clustering, dan prediksi.

- Evaluasi Model Klasifikasi: Menggunakan confusion matrix dengan komponen True Positive (TP), False Positive (FP), False Negative (FN), True Negative (TN). Metrik evaluasi meliputi:

$$\circ \text{ Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$\circ \text{ Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\circ \text{ Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\circ F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

- Evaluasi Model Regresi: Menggunakan metrik
- Evaluasi Clustering: (Meskipun rumus/metrik spesifik tidak disebutkan dalam teks asal, evaluasi clustering biasanya melibatkan metrik seperti Silhouette Score, Inertia, atau visualisasi).

Hasil evaluasi metrik akan digunakan untuk membandingkan performa model dan memberikan insight untuk pengambilan keputusan, seperti optimalisasi alokasi bandwidth dan strategi bisnis.

III. HASIL DAN PEMBAHASAN

A. Dataset

Data yang digunakan dalam penelitian ini diperoleh dari sistem PRTG Monitoring dengan interval pengambilan data setiap 5 menit. Dataset mencakup data dari 12 pelanggan selama kuartal pertama tahun 2025 (Januari hingga Maret). Proses preprocessing data dilakukan untuk memastikan kualitas dan konsistensi data sebelum dianalisis lebih lanjut.

B. Penggabungan dan Pembersihan Data

Beberapa file CSV yang berasal dari masing-masing pelanggan digabungkan menjadi satu kesatuan dataset terpadu. Untuk mempermudah identifikasi data per pelanggan, kolom tambahan bernama `Customer_ID` ditambahkan dan diisi berdasarkan nama file asal. Proses pembersihan data dilakukan untuk menghilangkan karakter non-numerik yang terdapat pada kolom metrik bandwidth, seperti satuan "KByte" dan "kbit/s", sehingga data dapat diolah dalam bentuk numerik. Nilai kosong (missing values) diisi menggunakan metode interpolasi berbasis waktu (`method='time'`) guna mempertahankan kontinuitas data. Selain itu, deteksi dan penanganan nilai pencilan (outlier) dilakukan dengan menggunakan metode Interquartile Range (IQR), di mana nilai-nilai di luar batas bawah dan atas di-clip untuk menjaga konsistensi distribusi data.

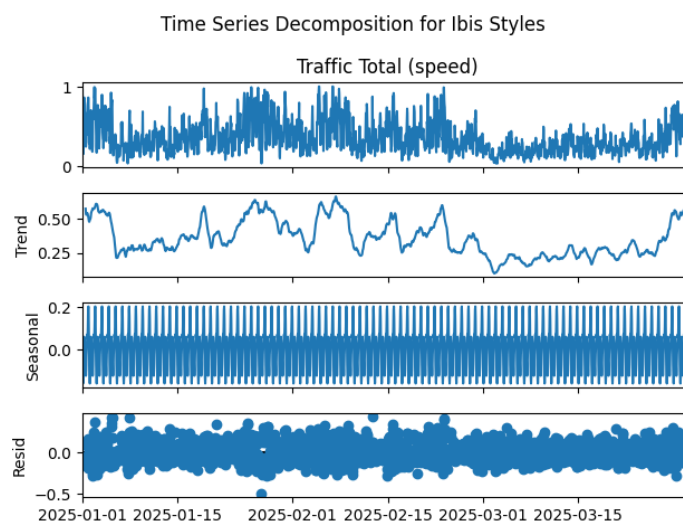
Kolom Date Time dikonversi ke dalam format datetime yang sesuai untuk pemrosesan waktu. Selanjutnya, kolom ini dipecah menjadi fitur-fitur temporal yang lebih spesifik, seperti jam (Hour) dan tipe hari (Day_Type), yang dikategorikan menjadi weekday dan weekend untuk analisis lebih lanjut.

C. Analisis Time Series

Analisis terhadap pola penggunaan bandwidth dilakukan melalui pendekatan dekomposisi time series serta visualisasi dalam bentuk heatmap. Tujuan dari analisis ini adalah untuk mengidentifikasi karakteristik pola temporal, baik dalam bentuk tren jangka panjang, pola musiman harian, maupun fluktuasi acak. Proses dekomposisi dilakukan menggunakan fungsi `seasonal_decompose` dari pustaka `statsmodels`, dengan asumsi periode musiman harian (daily seasonality) selama 24 jam (`period=24`). Dekomposisi ini menghasilkan tiga komponen utama, yaitu:

- **Tren (Trend):** Menunjukkan adanya kecenderungan peningkatan penggunaan bandwidth selama jam operasional, yaitu sekitar pukul 08.00 hingga 21.00.
- **Musiman (Seasonal):** Menggambarkan pola berulang yang konsisten setiap hari, dengan puncak aktivitas bandwidth biasanya terjadi pada tengah hari.
- **Residual:** Mewakili fluktuasi acak yang tidak dijelaskan oleh komponen tren maupun musiman.

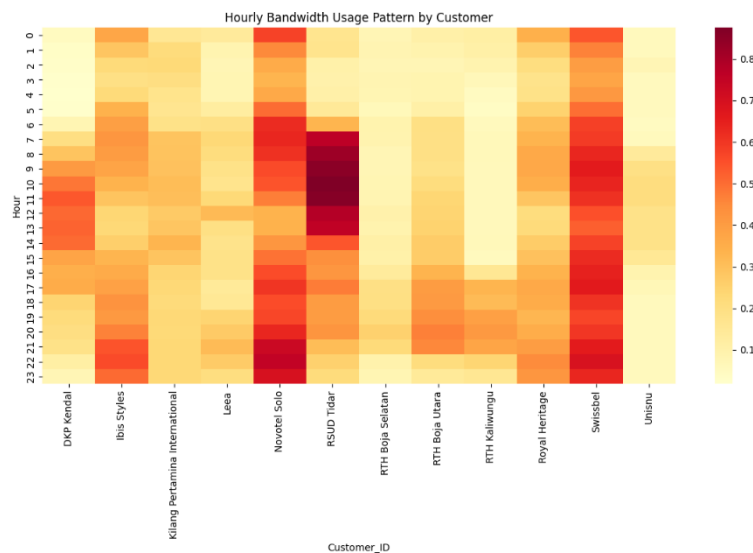
Sebagai ilustrasi, dekomposisi time series untuk pelanggan "Novotel Solo" menunjukkan ketiga komponen tersebut secara jelas, di mana aktivitas bandwidth meningkat selama jam sibuk dan menurun pada malam hari.



Gambar 2 Grafik dekomposisi time series

D. Analisis Time Series

Visualisasi dalam bentuk heatmap digunakan untuk memetakan variasi penggunaan bandwidth per jam pada masing-masing pelanggan selama periode observasi. Representasi visual ini memudahkan identifikasi pola temporal antar pelanggan. Sebagai contoh, pelanggan seperti "RSUD Tidar" menunjukkan intensitas penggunaan bandwidth yang tinggi pada jam kerja, sejalan dengan aktivitas operasional pelayanan publik. Sebaliknya, pola penggunaan bandwidth pada pelanggan "Swissbel" tampak lebih stabil sepanjang hari, mencerminkan karakteristik operasional hotel yang berjalan selama 24 jam.



Gambar 3 Heatmap pola harian

E. Klasifikasi Pola Penggunaan

Klasifikasi pola penggunaan bandwidth dilakukan menggunakan algoritma K-Means Clustering untuk mengelompokkan pelanggan berdasarkan karakteristik penggunaan jaringan. Pendekatan ini bertujuan untuk mengidentifikasi segmen pelanggan dengan perilaku penggunaan yang serupa, guna mendukung analisis efisiensi pemanfaatan bandwidth.

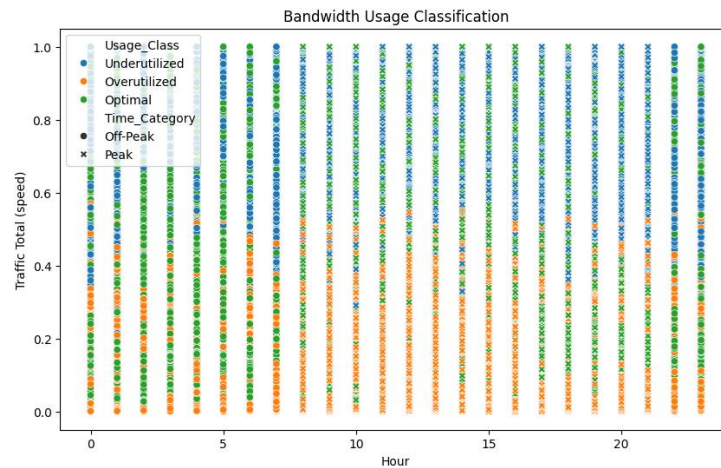
Tiga fitur utama digunakan sebagai dasar dalam proses klasterisasi, yaitu:

- **Utilization:** Rasio antara bandwidth yang digunakan terhadap kapasitas kontrak (dalam hal ini 100 Mbps). Fitur ini merepresentasikan tingkat pemanfaatan kapasitas jaringan.
- **Burst_Flag:** Indikator lonjakan lalu lintas (traffic spike) yang ditentukan berdasarkan nilai bandwidth yang melebihi 1,5 kali rata-rata bergerak dalam 6 jam terakhir.
- **Rolling_Avg_6h:** Rata-rata bergerak selama 6 jam yang digunakan untuk mendeteksi pola penggunaan yang stabil atau mengalami lonjakan.
- **Rolling_std_6h:** Fitur ini membantu model mengidentifikasi periode dengan fluktuasi tinggi yang mungkin memerlukan prediksi lebih dinamis.
- **Delta Last Hour:** Perubahan jam-ke-jam memberikan sinyal temporal jangka pendek yang tidak tertangkap oleh rata-rata bergerak.

Proses clustering menghasilkan tiga kelompok (kluster) utama berdasarkan pola penggunaan, yaitu:

- **Kluster 1 – Underutilized:** Kelompok dengan pemanfaatan bandwidth di bawah 30% dari kapasitas kontrak.
- **Kluster 2 – Optimal:** Kelompok dengan pemanfaatan bandwidth antara 30% hingga 70%, yang mencerminkan penggunaan yang seimbang dan efisien.
- **Kluster 3 – Overutilized:** Kelompok dengan pemanfaatan lebih dari 70% kapasitas, yang mengindikasikan potensi kebutuhan akan peningkatan kapasitas atau penyesuaian kontrak.

Evaluasi kualitas kluster dilakukan menggunakan nilai Silhouette Score, yang diperoleh sebesar 0,663. Nilai ini menunjukkan bahwa pemisahan antar kluster cukup baik dan masing-masing kluster memiliki batasan yang jelas.



Gambar 4 Scatter plot klasifikasi

F. Prediksi Bandwidth dengan Random Forest

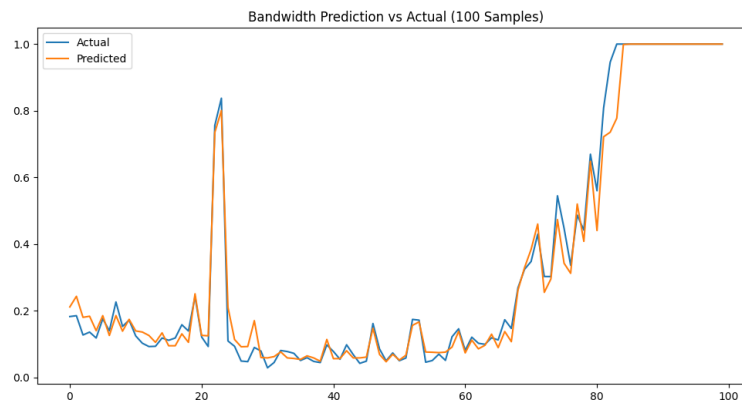
Model prediksi dikembangkan menggunakan algoritma Random Forest Regressor dengan tujuan untuk memproyeksikan kebutuhan bandwidth satu jam ke depan. Pendekatan ini memungkinkan peramalan berbasis data historis dengan mempertimbangkan pola temporal dan indikator lalu lintas jaringan. Model dibangun dengan menggunakan sejumlah fitur prediktor yang relevan, yaitu:

- Hour: Jam dalam format 24-jam yang merepresentasikan aspek waktu penggunaan.
- Day_Type: Kategori hari, diklasifikasikan sebagai weekday atau weekend.
- Time_Categor: Segmentasi waktu dalam kategori (misalnya pagi, siang, malam) untuk menangkap pola perilaku penggunaan.
- Rolling_Avg_6h: Rata-rata bergerak selama 6 jam terakhir, yang mencerminkan tren lokal dalam penggunaan bandwidth.
- Burst_Flag: Indikator terjadinya lonjakan trafik dalam kurun waktu pendek.
- Traffic Total (speed): Kecepatan total lalu lintas bandwidth pada jam berikutnya, dalam satuan Mbps. Adapun variable target yang diprediksi

Kinerja model dievaluasi menggunakan beberapa metrik regresi, yang hasilnya ditunjukkan sebagai berikut:

- Mean Absolute Error (MAE) : 0,036
- Root Mean Squared Error (RMSE) : 0,062
- Koefisien Determinasi (R^2) : 0,931, yang menunjukkan bahwa model memiliki akurasi prediksi yang tinggi serta mampu menjelaskan sebagian besar variabilitas dalam data target.

Visualisasi perbandingan antara nilai prediksi dan data aktual menunjukkan bahwa model mampu mengikuti pola temporal penggunaan bandwidth dengan cukup baik. Deviasi kecil hanya terjadi pada beberapa titik lonjakan (spike), yang merupakan kondisi ekstrim dan tidak umum. Grafik hasil prediksi ini menunjukkan potensi model dalam mendukung perencanaan kapasitas jaringan dan pengelolaan sumber daya secara proaktif.



Gambar 5 Visualisasi prediksi

G. Pembahasan

Hasil klasifikasi dan prediksi bandwidth memberikan wawasan strategis yang dapat dimanfaatkan oleh penyedia layanan atau pengelola jaringan dalam pengambilan keputusan, di mana pelanggan dalam kluster overutilized menunjukkan tingkat pemanfaatan bandwidth yang melampaui kapasitas kontrak secara konsisten sehingga mengindikasikan perlunya evaluasi dan kemungkinan peningkatan kapasitas untuk mencegah penurunan kualitas layanan, sementara pola penggunaan harian yang teridentifikasi melalui dekomposisi time series dan visualisasi heatmap memungkinkan optimasi dalam alokasi bandwidth seperti penerapan kebijakan manajemen trafik pada jam sibuk atau pengaturan prioritas layanan. Meskipun hasil yang diperoleh menunjukkan potensi yang menjanjikan, terdapat beberapa keterbatasan dalam ruang lingkup penelitian ini, di mana cakupan data hanya terbatas pada periode kuartal pertama tahun 2025 (Januari–Maret) sehingga belum mencerminkan variasi musiman tahunan secara menyeluruh dan diperlukan validasi lanjutan dengan rentang waktu yang lebih panjang agar model menjadi lebih generalis, sementara model prediksi belum mempertimbangkan faktor eksternal seperti event khusus, libur nasional, atau kondisi cuaca yang dapat memengaruhi perilaku penggunaan bandwidth secara signifikan. Sebagai pengembangan dari penelitian ini, beberapa saran dapat diajukan untuk meningkatkan akurasi dan cakupan model, di mana eksplorasi metode prediksi berbasis deep learning seperti LSTM atau TCN dapat menjadi alternatif untuk mengatasi kompleksitas data deret waktu non-linier dan jangka panjang, sementara eksplorasi dengan faktor lain seperti latency, jitter, packet loss, atau faktor eksternal lainnya yang dapat digunakan sebagai fitur tambahan untuk menangkap dinamika penggunaan bandwidth secara lebih akurat.

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pola penggunaan bandwidth pelanggan pada layanan internet dedicated dengan menerapkan metode machine learning berbasis data monitoring dari sistem PRTG. Berdasarkan hasil analisis dan pemodelan yang telah dilakukan, pendekatan ini terbukti memberikan kontribusi metodologis yang signifikan dalam pemahaman dan prediksi pola bandwidth. Dataset monitoring dengan resolusi 5 menit dari 12 pelanggan selama Januari–Maret 2025 memberikan informasi temporal yang kaya. Proses preprocessing data, termasuk konversi waktu, perhitungan rata-rata bergerak, dan penandaan burst, berhasil menyiapkan data yang representatif. Pengelompokan pelanggan menggunakan algoritma K-Means Clustering secara efektif memisahkan pelanggan berdasarkan karakteristik penggunaan bandwidth yang berbeda, seperti penggunaan tinggi di jam kerja atau penggunaan konstan sepanjang hari. Hal ini memberikan wawasan penting untuk strategi layanan berbasis profil pelanggan. Selanjutnya, penerapan model Random Forest Regressor untuk memprediksi penggunaan bandwidth satu jam ke depan menunjukkan performa prediktif yang sangat baik, dengan Mean Absolute Error (MAE) sebesar 0,036, Root Mean Squared Error (RMSE) sebesar 0,062, dan koefisien determinasi (R^2) sebesar 0,931. Nilai R^2 yang tinggi menunjukkan bahwa model mampu menjelaskan sebagian besar variabilitas data aktual.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi metode K-Means untuk segmentasi data dan Random Forest Regressor untuk prediksi mampu menghasilkan wawasan yang akurat mengenai pola penggunaan bandwidth pelanggan. Hasil tersebut membuka peluang aplikasi nyata, antara lain dalam merencanakan kapasitas jaringan secara lebih efisien, melakukan penyesuaian alokasi bandwidth secara proaktif, serta memberikan rekomendasi upselling berbasis data pelanggan. Namun demikian, penelitian ini memiliki keterbatasan pada jumlah sampel pelanggan dan durasi pengambilan data yang relatif singkat. Untuk mengatasi hal tersebut dan meningkatkan kualitas prediksi, langkah pengembangan selanjutnya sebaiknya dimulai dengan perluasan cakupan data historis misalnya mengumpulkan data penggunaan pelanggan selama enam bulan hingga satu tahun serta menambah variabel pendukung seperti jenis layanan yang digunakan, sektor industri pelanggan, dan lokasi geografis.

REFERENSI

- [1] A. Alimuddin, "PENGUNAAN INTERNET DAN PELUANG BERWIRAUSAHA DI INDONESIA," *Jurnal Kewirausahaan dan Bisnis*, vol. 26, no. 2, p. 112, Dec. 2021, doi: 10.20961/jkb.v26i2.50913.
- [2] Tedyyana, Agus, Osman Ghazali, and Onno W. Purbo. "Machine learning for network defense: automated DDoS detection with telegram notification integration." *Indonesian Journal of Electrical Engineering and Computer Science* 34.2 (2024): 1102.
- [3] E. Purba, "Analisis Faktor-Faktor yang Mempengaruhi Kepuasan Konsumen Terhadap Layanan Internet pada Warnet," 2023.
- [4] S. J. Ahn and S. H. Lee, "The effect of consumers' perceived value on acceptance of an internet-only bank service," *Sustainability (Switzerland)*, vol. 11, no. 17, Sep. 2019, doi: 10.3390/su11174599.
- [5] K. Setiawan and A. Wibowo, "DATA MINING IMPLEMENTATION FOR DETECTION OF ANOMALIES IN NETWORK TRAFFIC PACKETS USING OUTLIER DETECTION APPROACH," *JIKO (Jurnal Informatika dan Komputer)*, vol. 6, no. 2, Aug. 2023, doi: 10.33387/jiko.v6i2.6092.
- [6] F. A. Khan and A. A. Ibrahim, "Machine Learning-based Enhanced Deep Packet Inspection for IP Packet Priority Classification with Differentiated Services Code Point for Advance Network Management," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 16, no. 2, pp. 5–12, Jun. 2024, doi: 10.54554/jtec.2024.16.02.002.
- [7] S. S. Samaan and H. A. Jeiad, "Architecting a machine learning pipeline for online traffic classification in software defined networking using spark," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 861–873, Jun. 2023, doi: 10.11591/ijai.v12.i2.pp861-873.
- [8] S. A. Arnomo and Y. Yulia, "Clustering the potential bandwidth upgrade of FTTH broadband subscribers," *ILKOM Jurnal Ilmiah*, vol. 13, no. 1, pp. 51–57, Apr. 2021, doi: 10.33096/ilkom.v13i1.805.51-57.
- [9] A. Fauzan, N. Suarna, I. Ali, and H. Susana, "PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENINGKATKAN MODEL PENGELOMPOKAN DAN KINERJA JARINGAN WI-FI SECARA OPTIMAL," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6272.
- [10] D. Yasriady, "Klasterisasi Penggunaan Trafik Internet Menggunakan K-Mean Clustering," *Jurnal Sistim Informasi dan Teknologi*, pp. 112–117, Aug. 2022, doi: 10.37034/jsisfotek.v4i3.141.
- [11] A. Nariesha Alizan and M. F. Zolkipli, "A Comparative Study Between Wireshark and Paessler Router Traffic Grapher (PRTG) in Network Monitoring and Analysis," 2024.
- [12] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.
- [13] Y. Hendrian, "Sistem Monitoring Local Area Network Pelanggan PT. PHE ONWJ Pada PT. Patrakom Jakarta," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 1, 2021, doi: 10.31294/jtk.v4i2.
- [14] K. P. Murphy, *Probabilistic machine learning: an introduction*. MIT press, 2022.
- [15] A. Bilal *et al.*, "Improved Support Vector Machine based on CNN-SVD for vision-threatening diabetic retinopathy detection and classification," *PLoS One*, vol. 19, no. 1 January, pp. 1–27, 2024, doi: 10.1371/journal.pone.0295951.
- [16] Y. K. Bintang and H. Imaduddin, "Pengembangan Model Deep Learning Untuk Deteksi Retinopati Diabetik Menggunakan Metode Transfer Learning," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1442–1455, 2024, doi: 10.29100/jupi.v9i3.5588.
- [17] S. Kundu *et al.*, "LBP Feature Extraction for Early Detection of Diabetic Eye Diseases," *IJRASET*, no. June, pp. 2903–2907, 2023.