

Analisis Pemanfaatan *Resource Server* pada *Cloud Computing* Menggunakan *EDA* dan *K-Means Clustering*

Aura Zahra Ramadhani¹, Arip Solehudin², Apriade Voutama³

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

¹aurazahra20053@gmail.com, ²arip.solehudin@unsika.ac.id, ³apriade.voutama@staff.unsika.ac.id

Abstract

This study aims to analyze server resource utilization patterns in a cloud computing environment using an Exploratory Data Analysis (EDA) approach and clustering techniques. The data used is the Cloud Datacenter Workload dataset, consisting of 135,873 observations and 11 variables.

The analysis results show that CPU usage averages 2.17%, with a distribution dominated by low values and a right-skewed pattern. Trend analysis indicates that resource usage is fluctuating with several spikes under certain conditions, but generally remains at a low level. Correlation results indicate a moderate positive relationship between CPU and memory, while the relationship with other resources is relatively weak. Clustering analysis using the K-Means method grouped server workloads into three categories: low, medium, and high, with a dominance in the low-load category. Overall, this study indicates the potential for underutilization of server resources, necessitating more adaptive resource management strategies, such as autoscaling, to improve resource utilization efficiency in cloud systems.

Keywords: Cloud Computing, Resource Utilization, Exploratory Data Analysis, Clustering, Workload Analysis

Abstrak

Penelitian ini bertujuan untuk menganalisis pola pemanfaatan resource server pada lingkungan *cloud computing* menggunakan pendekatan *Exploratory Data Analysis* (EDA) dan teknik clustering. Data yang digunakan merupakan dataset *Cloud Datacenter Workload* yang terdiri dari 135.873 data observasi dan 11 variabel.

Hasil analisis menunjukkan bahwa penggunaan CPU memiliki rata-rata sebesar 2,17% dengan distribusi yang didominasi oleh nilai rendah serta pola *right-skewed*. Analisis tren menunjukkan bahwa penggunaan resource bersifat fluktuatif dengan beberapa lonjakan pada kondisi tertentu, namun secara umum berada pada tingkat rendah. Hasil korelasi menunjukkan adanya hubungan positif dengan tingkat sedang antara CPU dan memori, sementara hubungan dengan resource lain relatif lemah. Analisis clustering menggunakan metode K-Means mengelompokkan workload server ke dalam tiga kategori, yaitu beban rendah, sedang, dan tinggi, dengan dominasi pada kategori beban rendah. Secara keseluruhan, penelitian ini menunjukkan adanya potensi *underutilization* pada resource server, sehingga diperlukan strategi pengelolaan resource yang lebih adaptif seperti autoscaling untuk meningkatkan efisiensi penggunaan resource pada sistem cloud.

Kata kunci: Cloud Computing, Resource Utilization, Exploratory Data Analysis, Clustering, Workload Analysis

© 2026 Author
Creative Commons Attribution 4.0 International License



1. Pendahuluan

Perkembangan teknologi informasi dalam beberapa tahun terakhir mengalami peningkatan yang sangat pesat, khususnya pada pemanfaatan teknologi *cloud computing* sebagai infrastruktur komputasi modern. Cloud computing memungkinkan pengguna untuk mengakses sumber daya komputasi seperti server, penyimpanan data, jaringan, maupun aplikasi melalui internet tanpa harus membangun infrastruktur fisik secara mandiri. Hal ini memberikan fleksibilitas, skalabilitas, serta efisiensi biaya dalam pengelolaan sistem informasi dan layanan digital [1], [2].

Dalam lingkungan cloud, berbagai resource seperti CPU, memori, penyimpanan, dan jaringan digunakan secara bersama oleh banyak pengguna dan aplikasi. Oleh karena itu, pengelolaan dan pemantauan penggunaan resource menjadi aspek penting untuk menjaga performa sistem agar tetap optimal. Monitoring server memungkinkan administrator untuk mengetahui kondisi penggunaan resource secara *real-time* serta mendeteksi potensi permasalahan seperti overload atau bottleneck pada sistem [3], [4].

Implementasi sistem monitoring server telah banyak dilakukan untuk mengawasi pemanfaatan resource seperti CPU, RAM, dan penyimpanan. Sistem monitoring yang baik mampu memberikan informasi kondisi server secara berkala serta memberikan notifikasi ketika terjadi kondisi abnormal pada sistem [5], [6]. Hal ini menjadi penting mengingat beban kerja pada cloud computing bersifat dinamis dan dapat berubah-ubah sesuai kebutuhan pengguna.

Selain itu, pengelolaan resource yang tidak optimal dapat menyebabkan pemborosan kapasitas maupun penurunan kualitas layanan. Oleh karena itu, diperlukan analisis terhadap data monitoring server untuk memahami pola penggunaan resource secara lebih mendalam [7], [8]. Pemanfaatan cloud computing juga semakin luas, termasuk pada bidang Internet of Things (IoT) dan analisis data, sehingga efisiensi penggunaan resource menjadi semakin krusial dalam mendukung kinerja sistem secara keseluruhan [9].

Beberapa penelitian sebelumnya menunjukkan bahwa pendekatan berbasis data seperti *Exploratory Data Analysis* (EDA) dapat digunakan untuk memahami karakteristik data, seperti distribusi, tren, serta hubungan antar variabel dalam dataset [10], [11], [12]. Selain itu, teknik data mining seperti clustering juga dapat digunakan untuk mengelompokkan data berdasarkan kemiripan pola, sehingga mempermudah dalam mengidentifikasi kategori workload server [13], [14].

Meskipun demikian, sebagian besar penelitian masih berfokus pada implementasi sistem monitoring, dan belum banyak yang secara khusus mengeksplorasi

data hasil monitoring untuk menganalisis pola pemanfaatan resource secara komprehensif. Oleh karena itu, penelitian ini dilakukan untuk menganalisis data monitoring server menggunakan pendekatan EDA serta metode clustering K-Means guna mengidentifikasi pola workload dan tingkat pemanfaatan resource pada lingkungan cloud computing.

2. Metode Penelitian

2.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan analisis data eksploratif atau *Exploratory Data Analysis* (EDA) untuk mengidentifikasi pola pemanfaatan resource server pada lingkungan *cloud computing*. Metode EDA digunakan untuk mengeksplorasi data secara sistematis melalui analisis statistik deskriptif, visualisasi data, serta identifikasi hubungan antar variabel dalam dataset [10], [11], [12].

Pendekatan ini memungkinkan peneliti untuk memahami karakteristik data monitoring server, seperti distribusi penggunaan resource, tren perubahan dari waktu ke waktu, serta hubungan antara variabel seperti CPU, memori, disk, dan jaringan. Selain itu, EDA juga digunakan untuk mendeteksi adanya anomali atau *outlier* yang dapat memengaruhi hasil analisis.

Dalam konteks cloud computing, analisis terhadap penggunaan resource menjadi penting untuk meningkatkan efisiensi sistem, terutama dalam pengelolaan beban kerja yang dinamis serta optimalisasi penggunaan infrastruktur [15]. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu memberikan gambaran menyeluruh terhadap pola penggunaan resource server.

Untuk meningkatkan kedalaman analisis, penelitian ini juga mengadopsi pendekatan *data-driven* dengan menerapkan teknik clustering menggunakan algoritma K-Means. Metode ini digunakan untuk mengelompokkan data berdasarkan kemiripan pola penggunaan resource server, khususnya pada variabel CPU dan memori [12], [14].

Algoritma K-Means bekerja dengan membagi data ke dalam beberapa cluster berdasarkan kedekatan jarak terhadap centroid, sehingga data dalam satu cluster memiliki karakteristik yang serupa. Dengan metode ini, pola workload server dapat diklasifikasikan ke dalam beberapa kategori, seperti beban rendah (*low workload*), sedang (*medium workload*), dan tinggi (*high workload*).

Dengan demikian, kombinasi antara EDA dan clustering dalam penelitian ini bertujuan untuk menghasilkan analisis yang tidak hanya bersifat deskriptif, tetapi juga mampu memberikan insight yang lebih mendalam terkait pola penggunaan resource serta tingkat efisiensi pemanfaatan resource pada lingkungan cloud computing.

2.2. Sumber Data

Data yang digunakan dalam penelitian ini berasal dari dataset Cloud Datacenter Workload yang tersedia secara publik pada platform Kaggle. Dataset tersebut berisi data monitoring server yang merekam aktivitas penggunaan resource pada lingkungan data center cloud.

Dataset dapat diakses melalui tautan berikut: <https://www.kaggle.com/datasets/ashikhassan007/workloadtrace>. Dataset ini terdiri dari beberapa file log monitoring yang kemudian digabungkan menjadi satu dataset untuk proses analisis. Setelah dilakukan proses penggabungan data, diperoleh dataset dengan ukuran 135.873 baris data dan 11 variabel (135873×11). Data tersebut merepresentasikan aktivitas workload server dalam lingkungan cloud data center.

Variabel utama yang digunakan dalam penelitian ini berkaitan dengan pemanfaatan resource server yang meliputi penggunaan CPU, memori, penyimpanan, serta aktivitas jaringan. Variabel-variabel tersebut digunakan untuk menganalisis pola penggunaan resource pada server dalam lingkungan cloud computing.

2.3. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan kualitas dan konsistensi data sebelum dilakukan analisis lebih lanjut. Proses ini meliputi pembersihan data, penanganan nilai yang hilang (*missing values*), serta penyesuaian format data.

Pada tahap penanganan *missing values*, dilakukan metode data cleaning dengan menghapus baris data yang memiliki nilai kosong (*null values*). Pendekatan ini dipilih untuk menjaga keakuratan analisis serta menghindari bias yang dapat muncul akibat imputasi data, mengingat jumlah data yang tersedia relatif besar sehingga penghapusan sebagian kecil data tidak memengaruhi hasil secara signifikan.

Selain itu, dilakukan proses normalisasi terhadap nilai penggunaan resource, khususnya CPU dan memori, yang memiliki rentang nilai yang sangat besar. Normalisasi dilakukan menggunakan metode Min-Max Scaling, dengan mengubah nilai resource ke dalam skala persentase agar berada dalam rentang yang seragam (0–100%). Secara matematis, proses normalisasi Min-Max dirumuskan sebagai berikut:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \times 100 \quad (1)$$

X = nilai data asli

X_{min} = nilai minimum

X_{max} = nilai maksimum

X' = nilai hasil normalisasi

Proses ini bertujuan untuk menghindari kesalahan interpretasi akibat perbedaan skala data serta

memastikan validitas hasil analisis, terutama pada tahap clustering yang sensitif terhadap perbedaan skala antar variabel.

2.4. Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis untuk memperoleh hasil analisis yang komprehensif terhadap data monitoring server. Secara umum, tahapan penelitian dalam studi ini terdiri dari beberapa langkah sebagai berikut.

2.4.1 Pengumpulan Data

Dataset workload cloud datacenter diperoleh dari Kaggle dan digunakan sebagai sumber data utama dalam penelitian.

2.4.2 Data Preprocessing

Tahap ini dilakukan untuk membersihkan dan menyiapkan data sebelum dianalisis. Proses preprocessing meliputi penggabungan dataset, pengecekan data yang hilang (*missing values*), serta penyesuaian format data.

2.4.3 Exploratory Data Analysis (EDA)

Pada tahap ini dilakukan eksplorasi data untuk memahami karakteristik dataset, termasuk distribusi data, nilai minimum dan maksimum, serta hubungan antar variabel.

2.4.4 Visualisasi Data

Data kemudian divisualisasikan menggunakan grafik dan diagram untuk mempermudah interpretasi pola penggunaan resource server.

2.4.5 Analisis Korelasi

Analisis ini dilakukan untuk mengetahui hubungan antara variabel resource server seperti CPU, memori, disk, dan jaringan.

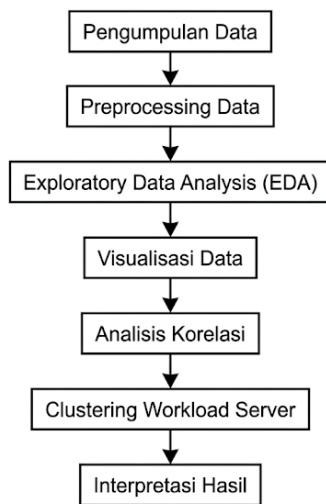
2.4.6 Clustering Workload Server

Metode clustering digunakan untuk mengelompokkan data berdasarkan kemiripan pola penggunaan resource, sehingga dapat mengidentifikasi kategori beban kerja server.

2.4.7 Interpretasi Hasil

Tahap akhir adalah menginterpretasikan hasil analisis untuk memahami pola pemanfaatan resource pada infrastruktur cloud serta memberikan kesimpulan dari hasil penelitian.

Tahapan ini lebih lengkapnya tertuang dalam gambar 1.



Gambar 1. Tahapan Penelitian

2.5. Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini meliputi beberapa metode analisis untuk memahami pola penggunaan resource server pada lingkungan cloud computing.

2.5.1 Analisis Statistik Deskriptif

Analisis ini digunakan untuk menggambarkan karakteristik dasar data seperti nilai rata-rata, median, standar deviasi, nilai minimum, dan maksimum dari setiap variabel resource server.

2.5.2 Visualisasi Data

Visualisasi dilakukan menggunakan grafik seperti histogram, boxplot, dan line chart untuk melihat distribusi data serta pola penggunaan resource dari waktu ke waktu.

2.5.3 Analisis Korelasi

Analisis korelasi digunakan untuk mengidentifikasi hubungan antara variabel resource server seperti CPU usage, memory usage, disk usage, dan network usage. Korelasi antar variabel dianalisis menggunakan matriks korelasi untuk mengetahui tingkat keterkaitan antar resource.

2.5.4 Analisis Tren Penggunaan Resource

Analisis tren dilakukan untuk melihat pola perubahan penggunaan resource server dari waktu ke waktu sehingga dapat memberikan gambaran mengenai dinamika workload pada infrastruktur cloud.

2.5.5 Clustering dengan K-Means

Metode clustering digunakan untuk mengelompokkan data berdasarkan kemiripan pola penggunaan resource server, khususnya CPU dan memori. Algoritma yang digunakan adalah K-Means, yang mengelompokkan data ke dalam beberapa cluster berdasarkan jarak terdekat terhadap centroid. Hasil clustering digunakan untuk mengidentifikasi kategori workload server, yaitu beban rendah, sedang, dan tinggi.

Melalui kombinasi metode analisis tersebut, penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai pola pemanfaatan resource pada lingkungan cloud data center.

2.6. Clustering dengan K-Means

Pada penelitian ini, digunakan metode *clustering* dengan algoritma K-Means untuk mengelompokkan data berdasarkan kemiripan pola penggunaan resource server, khususnya CPU dan memori.

Algoritma K-Means bekerja dengan cara membagi data ke dalam sejumlah cluster berdasarkan jarak terdekat terhadap centroid. Setiap data akan dikelompokkan ke dalam cluster yang memiliki jarak paling dekat dengan pusat cluster (centroid), sehingga data dalam satu cluster memiliki karakteristik yang serupa.

Sebelum dilakukan proses clustering, data terlebih dahulu melalui tahap normalisasi untuk memastikan bahwa setiap variabel memiliki skala yang seragam, sehingga hasil pengelompokan tidak bias terhadap variabel dengan rentang nilai yang lebih besar.

Jumlah cluster (K) yang digunakan dalam penelitian ini adalah sebanyak 3 cluster, yang merepresentasikan kategori workload server, yaitu beban rendah (low workload), beban sedang (medium workload), dan beban tinggi (high workload).

Penentuan jumlah cluster dilakukan menggunakan metode Elbow Method, yaitu dengan mengamati grafik hubungan antara jumlah cluster (K) dengan nilai *inertia*. Titik siku (*elbow*) pada grafik menunjukkan jumlah cluster optimal yang digunakan dalam proses clustering.

Secara matematis, algoritma K-Means bertujuan untuk meminimalkan jumlah kuadrat jarak antara data dengan centroid, yang dirumuskan sebagai berikut:

$$J = \sum_{i=1}^k \sum_{j=1}^n \|x_j - c_i\|^2 \tag{2}$$

di mana:

x_j = data ke-j

c_i = centroid cluster ke-i

k = jumlah cluster

Implementasi algoritma K-Means dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan bantuan library *Scikit-learn*, yaitu melalui fungsi `KMeans()` untuk proses pengelompokan data.

2.7. Tools dan Software

Penelitian ini menggunakan bahasa pemrograman Python sebagai alat utama dalam proses pengolahan dan analisis data. Seluruh proses analisis dilakukan menggunakan platform Google Colab, yang merupakan lingkungan komputasi berbasis cloud sehingga memungkinkan eksekusi kode secara efisien tanpa memerlukan instalasi lokal.

Beberapa library Python yang digunakan dalam penelitian ini antara lain: (1). Pandas digunakan untuk proses manipulasi dan pengolahan data; (2). NumPy digunakan untuk komputasi numerik; (3). Matplotlib dan Seaborn digunakan untuk visualisasi data; (4). Scikit-learn (sklearn) digunakan untuk implementasi algoritma K-Means Clustering

Penggunaan Google Colab dalam penelitian ini memberikan kemudahan dalam proses eksplorasi data, visualisasi, serta implementasi algoritma machine learning secara interaktif dan terintegrasi.

3. Hasil dan Pembahasan

3.1. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan dataset *Cloud Datacenter Workload* yang diperoleh dari platform Kaggle. Dataset tersebut terdiri dari beberapa file log monitoring server yang kemudian digabungkan menjadi satu dataset untuk proses analisis.

Setelah dilakukan proses penggabungan dan preprocessing, diperoleh dataset dengan jumlah 135.873 data observasi dan 11 variabel yang merepresentasikan penggunaan resource server, seperti CPU usage, memory usage, disk usage, dan network usage.

Data dalam dataset ini mencerminkan aktivitas workload server pada interval waktu tertentu, sehingga memungkinkan dilakukan analisis terhadap pola pemanfaatan resource pada lingkungan cloud computing.

Tabel 1. Tabel Variabel Utama Dataset

Variabel	Keterangan
CPU usage	Tingkat penggunaan CPU pada server
Memory usage	Tingkat penggunaan RAM pada server
Disk usage	Aktivitas penggunaan penyimpanan
Network usage	Aktivitas lalu lintas jaringan
Timestamp	Waktu pencatatan aktivitas server

Berdasarkan pengamatan awal terhadap dataset pada tabel 1, terlihat bahwa nilai penggunaan resource server memiliki variasi yang cukup beragam. Variasi ini menunjukkan adanya perbedaan beban kerja yang diproses oleh server pada setiap waktu pengamatan. Karakteristik data tersebut menjadi dasar dalam analisis lebih lanjut terhadap pola penggunaan resource pada sistem cloud.

3.2. Hasil Analisis Data

3.2.1 Statistik Resource Server

Analisis statistik deskriptif dilakukan untuk mengetahui karakteristik dasar dari penggunaan resource server. Parameter yang dianalisis meliputi nilai rata-rata, minimum, dan maksimum dari setiap variabel resource.

Tabel 2. Statistik Deskriptif Resource Server

Resource	Rata-rata	Min	Median	Max
CPU usage (%)	2.17	0.0	1.00	87.70
Memory usage (KB)	262,226.99	0.0	218,10 0.80	2,929,720. 00
Disk read (KB/s)	17.47	0.0	0.07	45,404.87
Disk write (KB/s)	32.36	0.0	7.73	6,259.47
Network received (KB/s)	13.73	0.0	0.07	7,214.80
Network transmitted (KB/s)	3.42	0.0	0.47	2,291.60

Berdasarkan hasil analisis statistik deskriptif pada tabel 2, penggunaan CPU memiliki nilai rata-rata sebesar 2,17% dengan median sebesar 1,00%. Hal ini menunjukkan bahwa sebagian besar aktivitas server berada pada tingkat penggunaan yang rendah. Nilai maksimum CPU sebesar 87,7% mengindikasikan bahwa terdapat kondisi tertentu dimana terjadi lonjakan beban kerja yang cukup tinggi, meskipun tidak terjadi secara sering.

Penggunaan memori menunjukkan rata-rata sebesar 262.226,99 KB dengan nilai maksimum mencapai 2.929.720 KB. Hal ini menunjukkan bahwa penggunaan memori cenderung lebih stabil dibandingkan CPU, namun tetap mengalami peningkatan pada kondisi tertentu.

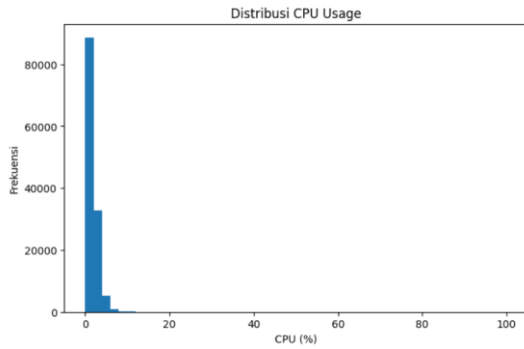
Aktivitas disk read dan disk write menunjukkan nilai rata-rata yang relatif rendah, dengan sebagian besar data berada di sekitar nilai mendekati nol. Namun demikian, nilai maksimum yang cukup tinggi menunjukkan adanya aktivitas baca/tulis yang intens pada kondisi tertentu.

Sementara itu, penggunaan jaringan baik pada sisi penerimaan (*network received*) maupun pengiriman (*network transmitted*) juga menunjukkan pola yang serupa, yaitu dominasi nilai rendah dengan beberapa lonjakan pada kondisi tertentu.

Secara keseluruhan, hasil ini menunjukkan bahwa pemanfaatan resource server cenderung tidak merata dan didominasi oleh kondisi beban rendah, dengan sesekali terjadi lonjakan penggunaan resource.

3.2.2 Distribusi Penggunaan CPU

Distribusi penggunaan CPU dianalisis menggunakan histogram untuk melihat sebaran nilai CPU usage dalam dataset. Histogram CPU usage ini dapat dilihat pada gambar 2.



Gambar 2. Histogram Distribusi CPU Usage

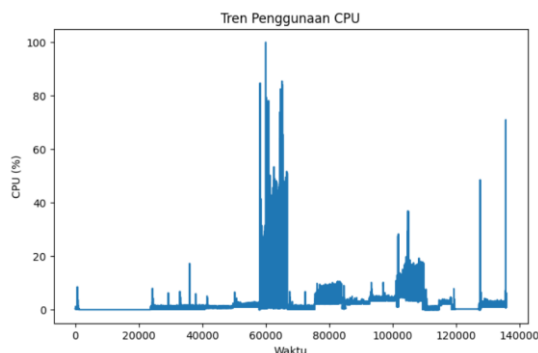
Berdasarkan hasil visualisasi, terlihat bahwa sebagian besar nilai penggunaan CPU terkonsentrasi pada rentang rendah, yaitu di bawah 10%. Hal ini menunjukkan bahwa server lebih sering berada dalam kondisi beban ringan.

Selain itu, distribusi data menunjukkan pola yang condong ke kanan (*right-skewed*), dimana terdapat sejumlah kecil data dengan nilai CPU yang tinggi hingga mendekati 90%. Nilai-nilai tersebut merepresentasikan kondisi lonjakan beban kerja yang terjadi pada waktu tertentu.

Pola distribusi ini mengindikasikan bahwa penggunaan CPU tidak merata dan cenderung bersifat sporadis, dengan dominasi kondisi penggunaan rendah serta sesekali terjadi peningkatan beban yang signifikan.

3.2.3 Tren Penggunaan Resource Server

Analisis tren dilakukan untuk melihat pola perubahan penggunaan resource server dari waktu ke waktu.



Gambar 3. Tren Penggunaan CPU Server

Berdasarkan hasil visualisasi pada gambar 3, penggunaan CPU menunjukkan pola fluktuatif dengan beberapa lonjakan pada titik waktu tertentu. Hal ini mengindikasikan adanya variasi beban kerja yang dinamis pada sistem.

Secara umum, sebagian besar nilai penggunaan CPU berada pada tingkat rendah, namun terdapat beberapa periode dimana terjadi peningkatan penggunaan yang cukup signifikan. Pola ini

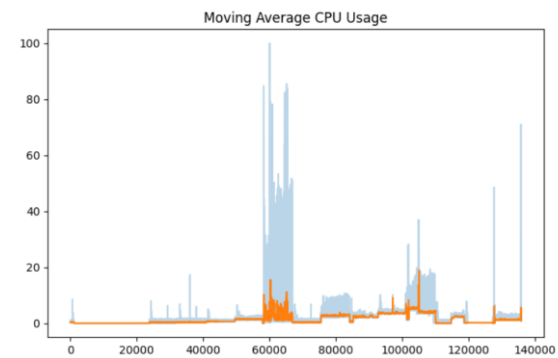
menunjukkan bahwa aktivitas komputasi pada server tidak berlangsung secara konstan, melainkan bergantung pada beban kerja yang dijalankan.

Untuk memperjelas kecenderungan pola tersebut, digunakan pendekatan *moving average* yang menunjukkan tren penggunaan CPU secara lebih halus. Hasilnya memperlihatkan bahwa meskipun terdapat fluktuasi jangka pendek, tren umum penggunaan CPU tetap berada pada tingkat rendah.

Sementara itu, penggunaan memori cenderung lebih stabil dibandingkan CPU, dengan variasi yang tidak terlalu signifikan. Hal ini menunjukkan bahwa perubahan beban kerja lebih berpengaruh terhadap penggunaan CPU dibandingkan memori.

Secara keseluruhan, hasil analisis tren ini menunjukkan bahwa pola penggunaan resource server bersifat dinamis dan tidak merata, dengan dominasi kondisi beban rendah serta adanya lonjakan penggunaan pada waktu tertentu.

Untuk memperjelas pola tren penggunaan CPU, digunakan metode *moving average* yang berfungsi untuk menghaluskan fluktuasi data jangka pendek.



Gambar 4. Moving Average CPU Usage

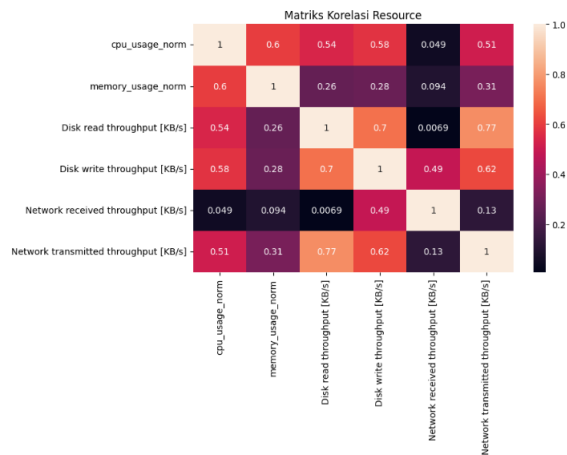
Berdasarkan hasil visualisasi pada gambar 6, garis *moving average* menunjukkan pola yang lebih stabil dibandingkan data asli, sehingga kecenderungan umum penggunaan CPU dapat diamati dengan lebih jelas.

Hasil analisis menunjukkan bahwa meskipun terdapat lonjakan penggunaan CPU pada beberapa titik waktu, tren rata-rata penggunaan CPU tetap berada pada tingkat rendah. Hal ini mengindikasikan bahwa lonjakan yang terjadi bersifat sementara dan tidak berlangsung dalam jangka waktu yang lama.

Dengan demikian, penggunaan *moving average* membantu dalam mengidentifikasi pola utama dari data serta mengurangi pengaruh noise, sehingga interpretasi terhadap tren penggunaan resource menjadi lebih akurat.

3.2.4 Analisis Korelasi Resource Server

Analisis korelasi dilakukan untuk mengidentifikasi hubungan antar variabel resource server, seperti CPU, memori, disk, dan jaringan.



Gambar 5. Matriks Korelasi Resource Server

Berdasarkan hasil matriks korelasi pada gambar 5, terlihat bahwa hubungan antara CPU usage dan memory usage menunjukkan korelasi positif dengan tingkat kekuatan sedang. Hal ini mengindikasikan bahwa peningkatan aktivitas CPU cenderung diikuti oleh peningkatan penggunaan memori.

Sementara itu, korelasi antara CPU dengan resource lainnya seperti disk dan jaringan relatif lebih rendah. Hal ini menunjukkan bahwa aktivitas pada disk dan jaringan tidak selalu berbanding lurus dengan penggunaan CPU, sehingga masing-masing resource dapat dipengaruhi oleh jenis workload yang berbeda.

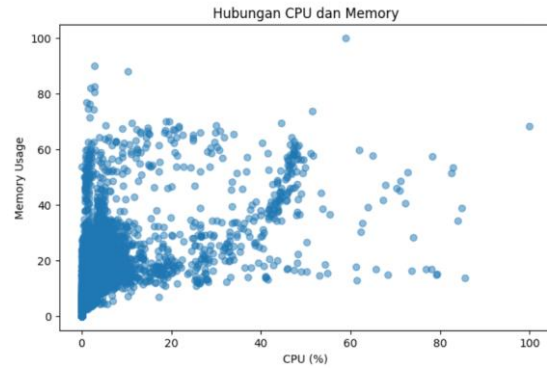
Secara keseluruhan, hasil analisis ini menunjukkan bahwa tidak semua resource memiliki keterkaitan yang kuat, sehingga analisis multi-resource diperlukan untuk memahami pola penggunaan server secara lebih komprehensif.

3.2.5 Hubungan CPU dan Memory Usage

Hubungan antara CPU usage dan memory usage dianalisis menggunakan scatter plot untuk melihat pola keterkaitan antara kedua variabel.

Hasil visualisasi yang dapat dilihat pada gambar 6, menunjukkan adanya kecenderungan hubungan positif, dimana peningkatan penggunaan CPU diikuti oleh peningkatan penggunaan memori.

Namun demikian, sebaran data yang cukup luas menunjukkan bahwa hubungan tersebut tidak bersifat linear sempurna. Terdapat variasi nilai memori pada tingkat CPU yang sama, yang mengindikasikan bahwa penggunaan memori tidak hanya dipengaruhi oleh CPU, tetapi juga oleh faktor lain seperti jenis proses atau aplikasi yang berjalan pada server.

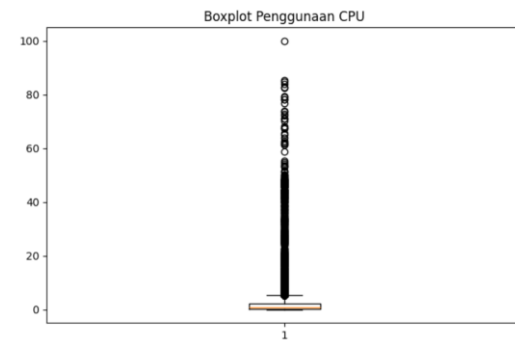


Gambar 6. Scatter Plot CPU dan Memory Usage

Hal ini menunjukkan bahwa meskipun terdapat hubungan antara CPU dan memori, keduanya tidak selalu bergantung secara langsung dalam setiap kondisi workload.

3.2.6 Analisis Variasi Penggunaan CPU

Variasi penggunaan CPU dianalisis menggunakan boxplot untuk melihat distribusi data serta keberadaan nilai pencilon (*outlier*).



Gambar 7. Boxplot Penggunaan CPU Server

Hasil menunjukkan bahwa sebagian besar nilai CPU berada pada tingkat rendah, dengan median yang mendekati nilai minimum.

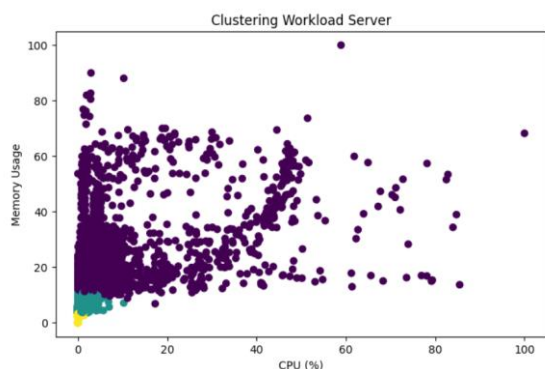
Selain itu, terdapat sejumlah nilai pencilon yang menunjukkan adanya lonjakan penggunaan CPU pada kondisi tertentu. Nilai-nilai tersebut merepresentasikan aktivitas workload yang lebih intens dibandingkan kondisi normal.

Keberadaan outlier ini menunjukkan bahwa penggunaan CPU tidak bersifat stabil, melainkan dipengaruhi oleh beban kerja yang bersifat dinamis. Dengan demikian, meskipun mayoritas kondisi server berada pada beban rendah, terdapat periode tertentu yang mengalami peningkatan penggunaan resource secara signifikan.

3.2.7 Clustering Workload Server

Untuk memperoleh pemahaman yang lebih mendalam mengenai pola penggunaan resource, dilakukan analisis clustering menggunakan metode K-Means. Proses clustering dilakukan terhadap data

yang telah dinormalisasi menggunakan metode Min-Max Scaling, sehingga setiap variabel memiliki skala yang seragam. Clustering digunakan untuk mengelompokkan data berdasarkan kemiripan pola penggunaan CPU dan memori. Penentuan jumlah cluster (K) dilakukan menggunakan metode Elbow, yang menunjukkan bahwa jumlah cluster optimal pada penelitian ini adalah sebanyak 3.



Gambar 8. Clustering Workload Server

Hasil clustering yang terlihat pada gambar 8, menunjukkan bahwa data dapat dikelompokkan ke dalam tiga kategori workload, yaitu beban rendah (*low workload*), beban sedang (*medium workload*), dan beban tinggi (*high workload*).

Sebagian besar data termasuk dalam kategori beban rendah, yang menunjukkan bahwa server lebih sering berada dalam kondisi penggunaan resource yang minimal. Sementara itu, cluster dengan beban sedang dan tinggi muncul pada kondisi tertentu yang mengindikasikan adanya peningkatan aktivitas komputasi.

Hasil ini memperkuat temuan sebelumnya bahwa distribusi penggunaan resource server tidak merata dan cenderung didominasi oleh kondisi beban rendah dengan sesekali lonjakan penggunaan.

Dengan adanya pengelompokan ini, pola workload server dapat dipahami dengan lebih jelas, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan terkait optimasi resource pada sistem cloud.

3.3. Pembahasan

Berdasarkan hasil analisis yang telah dilakukan, terlihat bahwa pemanfaatan resource server pada lingkungan *cloud computing* menunjukkan pola yang tidak merata dan cenderung didominasi oleh penggunaan pada tingkat rendah. Hal ini ditunjukkan oleh nilai rata-rata penggunaan CPU sebesar 2,17% serta distribusi data yang condong ke nilai rendah (*right-skewed*).

Hasil analisis tren juga menunjukkan bahwa penggunaan CPU bersifat fluktuatif dengan beberapa lonjakan pada waktu tertentu. Namun demikian, berdasarkan pendekatan *moving average*, diketahui

bahwa tren umum penggunaan CPU tetap berada pada tingkat rendah. Hal ini mengindikasikan bahwa lonjakan penggunaan resource yang terjadi bersifat sementara dan tidak berlangsung secara kontinu.

Analisis korelasi menunjukkan adanya hubungan positif dengan tingkat sedang antara CPU dan memori, yang mengindikasikan bahwa peningkatan aktivitas komputasi cenderung diikuti oleh peningkatan penggunaan memori. Namun, hubungan antara CPU dengan resource lain seperti disk dan jaringan relatif lemah, sehingga menunjukkan bahwa masing-masing resource dipengaruhi oleh karakteristik workload yang berbeda.

Selain itu, hasil analisis clustering menunjukkan bahwa sebagian besar data berada pada kategori beban rendah (*low workload*), dengan hanya sebagian kecil yang termasuk dalam kategori beban sedang dan tinggi. Hal ini memperkuat temuan bahwa penggunaan resource server tidak merata dan cenderung didominasi oleh kondisi penggunaan yang rendah.

Kondisi tersebut mengindikasikan adanya potensi *underutilization*, dimana kapasitas resource server yang tersedia belum dimanfaatkan secara optimal. Dalam konteks *cloud computing*, kondisi ini dapat berdampak pada inefisiensi penggunaan resource, terutama pada model layanan berbasis biaya penggunaan (*pay-as-you-go*).

Oleh karena itu, diperlukan strategi pengelolaan resource yang lebih adaptif, seperti penerapan *autoscaling* atau *dynamic resource allocation*, untuk menyesuaikan kapasitas resource dengan kebutuhan sistem secara real-time. Dengan demikian, efisiensi penggunaan resource pada infrastruktur cloud dapat ditingkatkan secara optimal.

4. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan terhadap dataset Cloud Datacenter Workload, dapat disimpulkan bahwa pemanfaatan resource server pada lingkungan *cloud computing* menunjukkan pola yang dinamis dan tidak merata.

Hasil analisis statistik deskriptif menunjukkan bahwa penggunaan CPU memiliki rata-rata sebesar 2,17% dengan mayoritas nilai berada pada tingkat rendah. Hal ini mengindikasikan bahwa sebagian besar server berada dalam kondisi beban ringan.

Distribusi penggunaan CPU yang bersifat *right-skewed* serta hasil visualisasi histogram menunjukkan dominasi penggunaan rendah dengan adanya beberapa lonjakan pada kondisi tertentu. Pola ini diperkuat oleh analisis tren yang menunjukkan fluktuasi penggunaan CPU yang tidak konstan, namun secara umum tetap berada pada tingkat rendah.

Analisis korelasi menunjukkan bahwa terdapat hubungan positif dengan tingkat sedang antara penggunaan CPU dan memori, sedangkan hubungan dengan resource lain seperti disk dan jaringan relatif lebih lemah. Hal ini menunjukkan bahwa tidak semua resource memiliki keterkaitan yang kuat dalam setiap kondisi workload.

Hasil clustering menggunakan metode K-Means menunjukkan bahwa pola workload server dapat dikelompokkan menjadi tiga kategori, yaitu beban rendah, sedang, dan tinggi. Mayoritas data berada pada kategori beban rendah, yang mengindikasikan bahwa pemanfaatan resource server belum optimal.

Secara keseluruhan, hasil penelitian ini menunjukkan adanya potensi *underutilization* pada resource server, dimana kapasitas yang tersedia belum dimanfaatkan secara maksimal dalam sebagian besar waktu operasional sistem.

Penelitian ini memberikan kontribusi dalam bentuk analisis berbasis data terhadap pola pemanfaatan resource server menggunakan pendekatan Exploratory Data Analysis (EDA) dan metode clustering K-Means. Hasil penelitian ini dapat menjadi dasar dalam memahami pola workload server serta membantu pengambilan keputusan dalam pengelolaan resource pada lingkungan cloud computing.

Penelitian ini memiliki beberapa keterbatasan, yaitu penggunaan dataset yang bersifat sekunder serta tidak mempertimbangkan faktor eksternal seperti jenis aplikasi, konfigurasi sistem, maupun karakteristik beban kerja secara spesifik yang dapat memengaruhi penggunaan resource server. Selain itu, analisis yang dilakukan masih bersifat deskriptif dan belum mencakup prediksi penggunaan resource di masa mendatang.

Penelitian selanjutnya disarankan untuk menggunakan data real-time serta mengintegrasikan metode analisis yang lebih kompleks, seperti machine learning prediktif, untuk memperkirakan kebutuhan resource secara lebih akurat. Selain itu, penerapan teknik optimasi seperti autoscaling dan dynamic resource allocation dapat dikaji lebih lanjut untuk meningkatkan efisiensi penggunaan resource pada sistem cloud.

Daftar Rujukan

- [1] Saputra, D., & Pratama, A. 2022. Analisis performa cloud computing menggunakan metode monitoring resource server. *Jurnal Teknologi Informasi dan Ilmu Komputer*. 9(2): 123–130. <https://doi.org/10.25126/jtiik.202292123>
- [2] Nugroho, A., & Setiawan, H. 2023. Analisis penggunaan resource pada cloud computing menggunakan pendekatan

data mining. *Jurnal Informatika*. 10(1): 45–53. <https://doi.org/10.31311/ji.v10i1.15230>

- [3] Ramadhan, M., & Kurniawan, B. 2021. Implementasi monitoring server berbasis Grafana dan Prometheus pada cloud infrastructure. *Jurnal RESTI*. 5(4): 678–685. <https://doi.org/10.29207/resti.v5i4.3200>
- [4] Widyaningsih, I., Haq, A., & Anggoro, T. 2024. Monitoring kinerja virtual machine pada lingkungan Google Cloud Platform menggunakan Prometheus dan Grafana. *Indonesian Journal of Artificial Intelligence and Data Mining*. 7(1): 45–52. <https://doi.org/10.20961/ijai.v10i1.109143>
- [5] Yulvianda, R., & Ismail, M. 2023. Desain dan implementasi sistem monitoring sumber daya server menggunakan Zabbix dan Grafana. *Jurnal Ilmiah Mahasiswa Sistem Informasi dan Komputer*. 8(2): 110–118. <https://doi.org/10.33998/jakom.2023.3.1.712>
- [6] Adil, S.B., & Beeh, Y.R. 2024. Implementasi monitoring sistem perusahaan on-premises dan cloud menggunakan Jenkins. *Jurnal Ilmiah Manajemen Informatika dan Komputer*. 9(2): 87–95. <https://doi.org/10.35870/jimik.v5i2.832>
- [7] Huda, N., Munthe, I.R., & Juledi, A. 2024. Optimisasi manajemen sumber daya pada sistem operasi untuk lingkungan cloud computing. *Jurnal Ilmu Komputer dan Informatika*. 6(1): 25–33. <https://doi.org/10.55338/jikomsi.v7i1.2721>
- [8] Chauhan, N., Singh, R., & Sharma, P. 2024. A systematic literature review on task allocation and performance management techniques in cloud data center. *International Journal of Cloud Computing*. 13(2): 155–170. <https://doi.org/10.1504/IJCC.2024.10057492>
- [9] Rivangga, S., Yanuarti, R., & Nilogiri, A. 2024. Implementasi cloud computing pada prediksi kualitas udara berbasis Internet of Things. *JUSTINDO*. 9(1): 65–73. <https://doi.org/10.32528/justindo.v9i2.1574>
- [10] Prasetyo, E., & Wibowo, A. 2022. Analisis data eksploratif untuk pemanfaatan resource server berbasis cloud. *Jurnal Sistem Informasi*. 8(2): 90–98. <https://doi.org/10.21609/jsi.v8i2.1450>
- [11] Syaifudin, A., dkk. 2024. Implementasi Exploratory Data Analysis untuk Analisis Data Lemak Tubuh. *IC-Tech*. 20(1): 1–6. <https://doi.org/10.47775/icttech.v20i1.328>
- [12] Radhi, M., dkk. 2021. Analisis Big Data dengan Metode EDA dan Visualisasi Menggunakan Jupyter Notebook. *JUSIKOM*. 6(2): 102–109. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475>
- [13] Fitriani, L., & Hidayat, T. 2021. Penerapan data mining untuk analisis pola penggunaan resource server. *Jurnal Teknologi dan Sistem Komputer*. 9(3): 155–162. <https://doi.org/10.14710/jtsiskom.9.3.2021.155-162>
- [14] Santoso, B., & Nugraha, R. 2023. Implementasi K-Means clustering untuk analisis workload server. *Jurnal Informatika dan Sistem Informasi*. 9(1): 33–41. <https://doi.org/10.28932/jutisi.v9i1.5678>
- [15] Saxena, D., & Singh, A.K. 2022. A proactive autoscaling and energy-efficient VM allocation framework for cloud data centers. *Journal of Cloud Computing*. 11(4): 1–15. <https://doi.org/10.1186/s13677-022-00289-4>