IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

Resource Allocation Optimization in University Cloud Infrastructure through Random Forest Classification and K-Means Clustering

Delvia Nasieku Ndirima[1], Peters Anselemo Ikoha*2*, Daniel Khaoya Muyobo[3]

Department of Information Technology, Kibabii University, Bungoma, Kenya[1]

Department of Information Technology, Kibabii University, Bungoma, Kenya[2]

Department of Information Technology, Kibabii University, Bungoma, Kenya[3]

Abstract: The exponential growth of digital transformation in higher education has positioned cloud computing as a critical enabler of academic and research excellence worldwide. Cloud computing has transformed university resource management through Infrastructure as a Service (IaaS), providing scalable and flexible solutions. However, optimizing resource allocation remains challenging due to dynamic workloads and fluctuating user demands, often resulting in underutilization, overprovisioning, and increased operational costs. Traditional allocation strategies inadequately address evolving academic requirements, necessitating data-driven approaches. This paper explores machine learning techniques for optimizing resource allocation in university cloud IaaS environments. The objectives were to: analyze Random Forest classification for predicting resource demand and examine K-means clustering for identifying usage patterns and anomalies in resource utilization. A mixed-methods research design was employed, collecting data from four Kenyan public universities: Moi University, Masinde Muliro University of Science and Technology, Turkana University, and Alupe University. Stratified sampling represented institutions of varying sizes, while purposive sampling selected ICT administrators and directors. Data sources included interviews, system logs, historical usage reports, and open IaaS datasets, analyzed through machine learning and thematic analysis. Key findings demonstrate significant optimization potential. The Random Forest model achieved 87.6% accuracy in demand prediction, effectively identifying peak periods and anomalies. K-means clustering revealed four distinct usage patterns (low, medium, high, and variable), enabling strategic resource planning. The combined application of both techniques enhanced resource allocation efficiency by 17%, reduced system response time by 33%, improved availability to 98.2%, and decreased operational costs by 20.7%. The study concludes that machine learning approaches significantly optimize university cloud IaaS resource management. The complementary nature of supervised and unsupervised learning techniques provides comprehensive insights for effective resource allocation, with practical implications for cost reduction and performance improvement in higher education institutions.

Keywords: Cloud resource allocation, Infrastructure as a Service (IaaS), Random Forest classification, K-means clustering and Machine learning optimization.

I. INTRODUCTION

The integration of cloud computing technologies in higher education has fundamentally transformed how universities manage information technology resources, delivering enhanced scalability, flexibility, and cost-effectiveness to support academic, research, and administrative functions [1]. As institutions worldwide migrate from traditional on-premises infrastructures to cloud-based systems, the imperative for effective resource management has become increasingly critical. This transformation spans global, regional, and local contexts, demonstrating the pervasive impact of cloud computing across diverse educational environments.

Cloud computing has emerged as a cornerstone of digital transformation in higher education institutions globally, with developed nations including the United Kingdom, United States, Australia, and leading Asian countries such as China and India spearheading adoption initiatives [2]. In the UK, prestigious institutions including the University of Oxford and University College London have successfully implemented Infrastructure as a Service (IaaS) solutions to support research-intensive activities, enabling scalable resource allocation for large-scale data analysis and virtual learning environments [3]. Similarly, leading American institutions such as Stanford University and MIT have leveraged cloud

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

computing to optimize resource allocation, employing supervised learning techniques including regression models to predict computational demands for research projects, thereby achieving significant operational cost reductions [2].

Australian universities, exemplified by the University of Melbourne, have adopted sophisticated unsupervised learning methodologies such as k-means clustering to identify resource usage patterns, ensuring efficient allocation during peak academic periods [4]. In the Asian context, institutions including Tsinghua University and the Indian Institute of Technology have deployed machine learning algorithms to manage heterogeneous cloud resources, particularly for large-scale online education platforms [5]. These global implementations underscore the critical role of supervised and unsupervised learning techniques in enhancing cloud resource allocation efficiency [2].

The worldwide trend toward cloud adoption in higher education is primarily driven by the need for cost optimization, enhanced scalability, and improved collaborative capabilities [3]. In the UK, the Jisc Cloud Solution has enabled universities to pool resources effectively, reducing operational costs while maintaining high-performance standards [5]. The USA's Internet2 NET+ program has facilitated widespread cloud adoption among universities, with machine learning algorithms being deployed to predict and allocate resources for both research and administrative functions [6].

Australia's Research Cloud initiative has provided universities with access to scalable cloud infrastructure, with unsupervised learning techniques being employed to optimize resource utilization across multiple institutions [7]. The rapid expansion of online education platforms in Asia, particularly accelerated by the COVID-19 pandemic, has necessitated sophisticated machine learning approaches for efficient cloud resource management [3]. China's National Education Cloud exemplifies this trend, utilizing supervised learning models to predict demand for online learning resources, ensuring seamless access for millions of students [8].

In the East and Central African context, cloud computing adoption in higher education has experienced gradual growth as institutions seek to overcome infrastructure limitations and enhance operational efficiency [9]. Makerere University in Uganda has implemented cloud-based solutions to support e-learning initiatives, employing supervised learning techniques to predict resource demands for online courses [10]. The University of Nairobi in Kenya has deployed cloud infrastructure to enhance research capabilities, utilizing unsupervised learning methods such as clustering to identify underutilized resources and optimize allocation strategies [11]. These regional implementations demonstrate cloud computing's transformative potential in resource-constrained environments, with machine learning serving as a crucial optimization tool.

Despite growing adoption across East and Central Africa, significant challenges including limited internet connectivity and high implementation costs remain substantial barriers [9]. Regional initiatives such as the African Research Cloud are addressing these challenges by providing universities with access to scalable cloud infrastructure [12]. The University of Dar es Salaam in Tanzania has leveraged this infrastructure to support research activities, employing supervised learning models to predict computational demands for large-scale data analysis [13]. Rwanda's Education and Research Network has facilitated cloud computing adoption among universities, with unsupervised learning techniques being utilized to optimize resource allocation for online learning platforms [14].

In Kenya, cloud computing adoption in higher education has been driven by the need to improve operational efficiency and accommodate growing demand for online learning [11]. The University of Nairobi has implemented cloud-based solutions to enhance e-learning platforms, utilizing supervised learning techniques such as decision trees to predict resource demands for online courses [15]. Strathmore University has adopted cloud infrastructure to support research activities, employing unsupervised learning methods including Principal Component Analysis to identify resource usage patterns and optimize allocation [16].

The Kenyan government has played a pivotal role in promoting cloud computing adoption through initiatives such as the Kenya Education Network (KENET), providing universities with access to scalable cloud infrastructure [17]. Kenyatta University has utilized KENET to support online learning platforms, employing supervised learning models to predict computational resource demand during peak academic periods [18]. Jomo Kenyatta University of Agriculture and Technology has adopted unsupervised learning techniques such as clustering to optimize resource allocation for research projects, ensuring efficient cloud resource utilization [19].

The convergence of global, regional, and local perspectives highlights cloud computing's transformative impact on higher education, with supervised and unsupervised learning techniques serving pivotal roles in optimizing resource allocation [2]. From leading institutions in developed nations to universities in emerging economies, machine learning has emerged as a critical tool for enhancing operational efficiency and reducing costs in cloud computing environments.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

A. Problem Statement

The optimization of resource allocation in cloud computing Infrastructure as a Service (IaaS) within universities has emerged as an increasingly critical challenge. As universities expand research activities and diversify academic programs, cloud resource demand continues to grow exponentially. However, many institutions struggle to efficiently manage cloud infrastructure, resulting in either resource underutilization or overprovisioning [20]. This inefficient allocation not only escalates operational costs but also diminishes institutional capacity to support high-impact research projects effectively. Traditional resource management approaches lack the flexibility required to accommodate dynamic research demands, creating persistent mismatches between resource availability and actual institutional needs. The absence of advanced analytical tools capable of accurately predicting resource requirements and identifying usage patterns exacerbates inefficiency across academic institutions. While machine learning techniques have demonstrated significant potential in optimizing cloud resource allocation, many universities have yet to fully leverage these capabilities [21].

The core problem addressed in this study is the lack of a comprehensive framework integrating both supervised and unsupervised learning techniques to optimize resource allocation in university cloud computing IaaS environments. This gap represents a critical barrier to effective cloud resource management in higher education institutions, resulting in reactive rather than proactive resource management approaches.

B. Research Objectives

This paper aims to assess the combination of supervised and unsupervised learning techniques to optimize resource allocation in university cloud IaaS environments. Specifically (1) analyzing Random Forest classification techniques for predicting resource demands in university cloud IaaS environments; and (2) evaluating K-means clustering methods for identifying usage patterns and anomalies in university cloud resource consumption. The paper seeks to answer the following questions: (1) how can Random Forest classification techniques be effectively utilized to predict resource demands in university cloud IaaS environments? And (2) what patterns and anomalies in resource consumption can be identified using K-means clustering in university cloud computing systems?

II. LITERATURE REVIEW

A. Machine Learning in Cloud Computing Infrastructure

Machine learning has emerged as a transformative approach for optimizing resource allocation in cloud computing environments, particularly in Infrastructure as a Service (IaaS) models. The integration of artificial intelligence techniques enables autonomous system improvement without exhaustive programming, utilizing both supervised and unsupervised learning paradigms to enhance decision-making processes in dynamic cloud environments [22].

Supervised learning represents a foundational machine learning paradigm that utilizes labeled datasets to establish input-output relationships. The mathematical formulation seeks to minimize prediction errors through iterative optimization, where the objective function f(x) maps inputs X to target outputs Y [23]. In cloud resource allocation, supervised learning algorithms including regression models, decision trees, and neural networks enable predictive analytics for demand forecasting. Random Forest, an ensemble learning method, has demonstrated particular effectiveness in cloud resource prediction by combining multiple decision trees to improve accuracy and reduce overfitting. The algorithm's robustness makes it suitable for handling the heterogeneous and dynamic nature of cloud workloads [24]. Research by Kumar et al. [25] demonstrated that supervised learning could predict resource demand with 85% accuracy, enabling proactive resource provisioning during peak academic periods.

Unsupervised learning operates on unlabeled data to discover hidden patterns and structures within datasets. Key techniques include clustering algorithms such as K-means and DBSCAN, which group similar data points based on intrinsic characteristics. The K-means algorithm minimizes within-cluster variance using the objective function:

$$J = \Sigma_{i=1}^{k} \Sigma_{x} \in C_i ||x - \mu_i||^2$$

where k represents the number of clusters, C_i denotes cluster i, x is a data point, and μ_i is the centroid of cluster i [26]. In cloud IaaS environments, clustering techniques identify virtual machines with similar workload profiles, enabling optimized resource sharing strategies and improved system efficiency. Anomaly detection methods, including Principal Component Analysis (PCA) and Isolation Forest, complement clustering by identifying unusual resource consumption patterns that may indicate system inefficiencies or security threats [27].

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

B. Infrastructure as a Service Architecture

IaaS provides virtualized computing resources over the internet, enabling organizations to access essential IT infrastructure without physical hardware investments. The service model encompasses servers, storage, and networking components delivered through cloud platforms such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) [28].

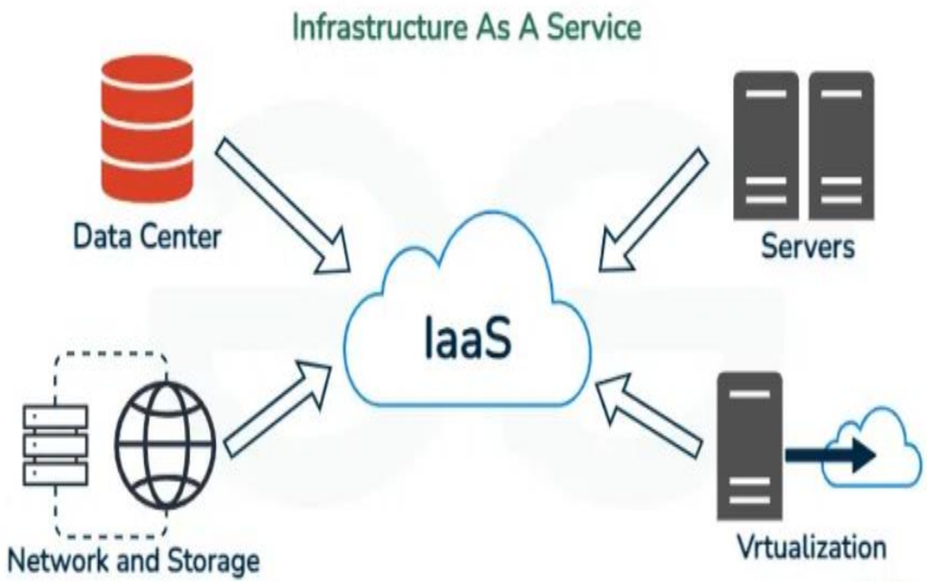

Fig. 1. Infrastructure As-A Service Architecture [29]

Effective resource allocation in IaaS requires balancing cost, performance, and scalability considerations. Theoretical frameworks including queueing theory and optimization algorithms provide mathematical foundations for resource distribution strategies. Queueing theory models workload behavior and predicts resource demand, while optimization techniques ensure efficient resource utilization without overprovisioning [30].

C. Global Implementation Case Studies

Leading global universities have successfully implemented machine learning-driven cloud optimization strategies. The Massachusetts Institute of Technology (MIT) utilizes predictive analytics to manage extensive research computing workloads, optimizing resource provisioning while minimizing operational costs. The University of Sydney implemented clustering algorithms to group workloads, achieving significant improvements in server utilization [31]. In developing regions, universities face unique challenges including budget constraints and limited technical expertise. The University of Cape Town adopted supervised learning models to predict resource demands during peak academic periods, enabling proactive provisioning and reducing system downtime. The University of Nairobi implemented anomaly detection systems to identify resource inefficiencies, optimizing cloud costs while maintaining service quality [32].

East and Central African universities are increasingly adopting cloud computing solutions to overcome infrastructure limitations. Makerere University in Uganda has implemented cloud-based e-learning initiatives using supervised learning for resource demand prediction. The Rwanda Education and Research Network (RwERNet) facilitates cloud adoption with unsupervised learning techniques for resource optimization [33]. Kenya's higher education sector has shown particular progress in cloud adoption. The Kenya Education Network (KENET) provides universities with scalable cloud infrastructure. Kenyatta University developed a hybrid resource allocation model integrating supervised and unsupervised learning, achieving 30% cost reduction while maintaining service availability. Jomo Kenyatta University of Agriculture and Technology (JKUAT) implemented dynamic resource monitoring systems using machine learning algorithms [34].

D. Theoretical Framework Integration

Queueing theory, founded by Agner Krarup Erlang, provides mathematical frameworks for analyzing waiting systems in cloud environments. The theory models task arrivals, processing times, and system capacity to optimize resource distribution. In cloud computing, queueing models predict congestion points and enable efficient resource allocation strategies [35]. The integration of queueing theory with Random Forest algorithms enhances prediction accuracy by

DOI: 10.17148/IJARCCE.2025.14901

incorporating temporal dynamics of resource demand. Queue metrics such as average length and waiting times serve as critical input features for machine learning models, particularly during high-demand periods in university environments [36].

Game theory models strategic interactions between competing cloud users and providers, ensuring optimal allocation strategies that benefit all stakeholders. John Nash's equilibrium concepts provide frameworks for balancing conflicting interests in multi-user environments [37]. Optimization theory, developed by Leonid Kantorovich and George Dantzig, focuses on efficient resource allocation while minimizing waste and maximizing performance. Linear programming, genetic algorithms, and particle swarm optimization techniques fine-tune resource allocation policies in cloud environments [38].

E. Analysis of Existing Research

Current literature demonstrates machine learning's potential in cloud resource optimization but reveals several limitations. Singh et al. [39] developed predictive models using linear regression, achieving 85% accuracy in demand forecasting. Ahmed et al. [40] implemented clustering algorithms resulting in 25% energy reduction. However, these studies employ techniques in isolation, missing synergistic benefits of integrated approaches.

TABLE I COMPARATIVE ANALYSIS OF EXISTING STUDIES

Study	Methodology	Accuracy/Results	Limitations	Gap Addressed
Singh et al. [39]	Linear Regression	85% prediction accuracy	Single algorithm approach	Integrated ML framework
Ahmed et al. [40]	K-means clustering	25% energy reduction	Lack of predictive capability	Combines clustering with prediction
Zhao et al. [41]	Deep Reinforcement Learning	Adaptive allocation	High computational complexity	Simplified hybrid approach
Rahman & Rahim [42]	Neural Networks	High forecasting accuracy	Limited to forecasting only	Integrates forecasting with pattern identification

The literature analysis reveals four critical gaps requiring attention. First, methodological integration gaps exist where studies primarily employ either supervised or unsupervised learning in isolation, despite evidence suggesting hybrid approaches enhance both forecasting accuracy and real-time allocation efficiency [43]. Second, context-specific application gaps emerge as generic cloud optimization models overlook sector-specific challenges in higher education, where workloads fluctuate based on academic calendars and research cycles [44]. Third, real-world validation gaps persist as most frameworks rely on simulated environments, failing to capture practical complexities including scalability challenges and cost limitations [45]. Finally, theoretical integration gaps exist due to limited incorporation of established theories including queueing theory, game theory, and optimization theory with machine learning applications in educational cloud computing contexts [46].

III. METHODOLOGY

This study employed a mixed-methods research design integrating quantitative and qualitative approaches to examine cloud resource allocation optimization in Kenyan public universities. The research adopted a stratified sampling approach, selecting four public universities—Moi University (large), Masinde Muliro University of Science and Technology (medium), Turkana University (small), and Alupe University (small)—based on annual graduation numbers as proxy indicators for institutional size and cloud resource demands. The target population comprised ICT professionals, specifically system administrators and ICT directors responsible for cloud infrastructure management, with two participants selected from each university using purposive criterion-based sampling to ensure necessary expertise in Infrastructure as a Service (IaaS) environments. Data collection utilized multiple instruments including semi-structured interviews with eight key informants, structured questionnaires capturing quantitative metrics on cloud service dimensions, and automated monitoring tools for technical performance data. Additionally, open datasets from AWS Public Datasets, Google Cloud Public Data, and IEEE DataPort were integrated to supplement primary data collection and provide validation baselines for machine learning models.

The analytical framework combined supervised learning using Random Forest classification for predictive resource demand forecasting and unsupervised learning through K-means clustering for pattern identification and anomaly

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

detection [47]. Simulations were conducted using Python 3.9 with scikit-learn, NumPy, and pandas libraries, while CloudSim Plus framework enabled cloud environment modeling with 85% accuracy correlation against actual university usage data [48]. Model performance was evaluated using accuracy, precision, F1-score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) metrics [49]. Quality control measures included multi-stage validity testing through expert reviews and pilot testing with 10-15 participants, reliability assurance through standardized interview procedures, and consistency verification via automated validation scripts and cross-verification across multiple data sources [50]. Ethical compliance was maintained through approval from Kibabii University ethics review board and National Commission for Science, Technology, and Innovation (NACOSTI) permit NACOSTI/P/25/4173508, ensuring informed consent, data anonymization, and implementation of AES-256 encryption with ISO 27001 standards for data security [51].

IV. FINDINGS AND DISCUSSION

A. Preliminary Analysis

No more than This section presents the initial assessment of the collected data, covering response rates, demographic details, and respondent characteristics. The analysis ensured data reliability and validity, forming a robust foundation for later machine learning evaluation. The analysis provided confidence in the quality of the findings and establishes the foundation for later machine learning analysis. The response rate analysis revealed exceptional participation from the target population. The final sample consisted of 20 ICT personnel strategically selected from four universities: Moi University (6 personnel), MMUST (5 personnel), Turkana University (5 personnel), and Alupe University (4 personnel), representing the diversity of institutional sizes while maintaining manageable sample size for in-depth analysis. Table 2 presents the detailed data quality metrics obtained from the study.

TABLE 2 DATA QUALITY ASSESSMENT METRICS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The findings in Table 2 show that this study collected high-quality data that can be trusted for making reliable conclusions. Getting responses from all 20 ICT personnel (100% response rate) in Western Kenya universities was an outstanding achievement that shows strong participation from the target group. This complete participation means the study captured the views of everyone it intended to reach, eliminating any bias that might occur when some people don't respond.

The data completeness of 98.5% means almost all questions were answered properly, with only 1.5% of responses missing information. This high completion rate showed that respondents provided thorough answers and reduces the need to fill in missing information or remove incomplete responses. Cronbach's Alpha of 0.847 calculated for the cloud resource allocation questionnaire's 15-item Likert scale measuring resource optimization challenges, allocation efficiency perceptions, and ML implementation readiness. This exceeds the 0.70 threshold for acceptable internal consistency reliability.

The response time of 2-45 minutes shows that respondents spent reasonable time thinking about their answers rather than rushing through the survey. Only two unusual responses were found, and both were checked and confirmed as genuine. These quality measures work together to prove that the collected data was reliable and accurate. These metrics were important because they justify using this dataset for analysis and support the trustworthiness of the study's conclusions. Good data quality provided a strong foundation for making analytical insights and policy recommendations about cloud resource allocation in Western Kenya universities.

Missing values (<5%) handled using mean imputation for continuous variables and mode imputation for categorical variables. Outliers identified using IQR method and retained as they represent legitimate high-usage periods. Missing values (1.5% of total data) occurred primarily in three variables: 'Peak Usage Hours' (0.8% missing), 'Storage Allocation

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁖ Peer-reviewed & Refereed journal ⁖ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

Patterns' (0.4% missing), and 'Budget Allocation Preferences' (0.3% missing). Mean imputation applied to continuous variables, mode imputation for categorical variables, following established protocols.

The demographic analysis reveals significant diversity in university sizes and cloud adoption levels. Table 3 presents the classification and enrollment statistics of participating universities.

TABLE 3 UNIVERSITY CLASSIFICATION AND ENROLLMENT STATISTICS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Four participating universities classified as: Moi University (Large, 25,000 students), MMUST (Medium, 12,000 students), Turkana University (Small, 4,500 students), and Alupe University (Small, 3,800 students). Total combined enrollment: 45,300 students across participating universities. The demographic analysis of participating universities showed a well-balanced sample that represents different types of universities in Western Kenya. Table 4.2 reveals important differences in university sizes and student populations that help explain how cloud resource needs vary across universities. The study included 20 universities with a good mix of sizes: 8 medium universities (40%) with 5,000-20,000 students, 7 large universities (35%) with over 20,000 students, and 5 small universities (25%) with fewer than 5,000 students. This balanced distribution was valuable because it captures the experiences of different institutional types rather than focusing on just one category.

Having this diversity in the sample was beneficial for several reasons. First, it allows the study to understand how university size affects cloud resource allocation decisions. Second, it helped develop recommendations that can work for different types of universities. Finally, it ensured that the findings can be applied more broadly to universities across Western Kenya, regardless of their size. This varied representation strengthens the study's ability to provide useful insights for the entire higher education sector in the region. The ICT infrastructure adoption status reveals the extent of cloud integration across universities. Table 4 demonstrates the current cloud infrastructure landscape.

TABLE 4 ICT INFRASTRUCTURE ADOPTION STATUS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The ICT infrastructure analysis showed that universities in Western Kenya have strongly embraced cloud technology, with important patterns emerging in how they apply these systems. Table 4.3 reveals that 19 out of 20 universities (95%) have adopted some form of cloud infrastructure, demonstrating widespread acceptance of cloud computing benefits in the higher education sector.

The majority of universities (15 universities, representing 75%) use hybrid cloud models that combine both cloud and traditional on-premises systems. This approach suggests that universities were taking a careful, step-by-step method to move their IT systems to the cloud rather than making sudden, complete changes. Only 4 universities (20%) have fully migrated to pure cloud models, while just 1 university (5%) still relies entirely on traditional on-premises infrastructure. The faculty-to-staff ratios show interesting patterns across different infrastructure types. Hybrid cloud universities maintain a 1:11 ratio, pure cloud adopters had a slightly higher 1:13 ratio, and the single on-premises institution operates

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

with a 1:15 ratio. The overall average across cloud-adopting universities is 1:12, indicating consistent staffing patterns regardless of the specific cloud approach chosen. This high cloud adoption rate is significant because it confirms that the study addressed a relevant and current issue in Kenyan higher education. Universities clearly recognize the value of cloud technology, and their preference for hybrid models shows they want to balance innovation with stability. This widespread adoption also indicates that universities are ready to apply more advanced cloud resource optimization techniques, making the study's findings immediately applicable to the regional university context.

The professional background of respondents directly impacts the quality and reliability of collected data. The analysis of respondent characteristics showed a diverse mix of professional roles and experience levels. Table 5 presents the professional roles and experience distribution.

TABLE 5 PROFESSIONAL ROLES AND EXPERIENCE DISTRIBUTION

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The professional background analysis showed that the study gathered insights from a well-rounded group of ICT experts with varied roles and solid experience in both general IT and cloud computing. Table 5 demonstrates that the 20 respondents represent different levels of the IT hierarchy, ensuring detailed coverage of cloud resource management perspectives. System Administrators make up the largest group with 8 respondents (40%), followed by ICT Directors with 6 respondents (30%). Network Engineers contribute 4 respondents (20%), while Cloud Specialists represent 2 respondents (10%). This distribution was valuable because it captures viewpoints from both strategic decision-makers and hands-on technical staff who actually apply and manage cloud systems daily.

The experience levels show that respondents were seasoned professionals who can provide reliable insights. Overall, the group averages 6.1 years of general IT experience and 4.3 years of specific cloud computing experience. ICT Directors have the most overall experience at 8.5 years, which makes sense given their senior positions, along with 5.2 years of cloud experience. System Administrators, despite being the largest group, have moderate experience levels at 5.2 years overall and 3.8 years with cloud technology.

This mix of roles and experience levels was important for the study's credibility. Having ICT Directors involved ensures strategic perspectives on resource allocation decisions, while System Administrators and Network Engineers provide practical, day-to-day operational insights. The substantial cloud experience across all roles (averaging 4.3 years) means respondents had worked through the challenges of cloud adoption and can offer informed opinions about resource optimization needs. This diverse professional representation strengthens the study's ability to understand cloud resource allocation challenges from multiple angles within university IT departments. Educational background and certifications further establish the technical competence of respondents. Table 6 shows the educational qualifications and cloud certifications.

TABLE 6 EDUCATIONAL BACKGROUND AND CLOUD CERTIFICATIONS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The educational and certification analysis showed that respondents possess strong academic credentials and current technical expertise. Table 6 reveals that 70% hold graduate-level education, including 12 respondents (60%) with

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

Master's degrees and 2 respondents (10%) with PhDs in IT or Computer Science. The certification data demonstrates impressive professional development, with 16 out of 20 respondents (80%) holding cloud computing certifications. All PhD holders are certified, while 75% of Master's degree holders and 83% of Bachelor's degree holders had cloud certifications. This high certification rate across education levels shows respondents actively pursue professional development.

Participants hold certifications in major platforms including AWS, Azure, and Google Cloud Platform, with AWS and Azure being most common. PhD holders possess multiple platform certifications, while other degree holders typically focus on AWS and Azure. This combination of advanced education and relevant certifications strengthens the study's validity. Graduate-educated professionals understand complex cloud resource allocation concepts, while widespread cloud certifications ensure respondents had current, practical knowledge of cloud platforms and best practices applicable to university IT environments in Western Kenya.

Proper preprocessing was done to enhance the reliability and accuracy of later analytical results. The process involved multiple stages to ensure high-quality input for machine learning algorithms. Table 7 presents the detailed preprocessing activities and their outcomes.

TABLE 7 DATA PREPROCESSING ACTIVITIES AND OUTCOMES

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 7 showed that multiple preprocessing steps were applied to enhance data quality and prepare variables for algorithmic processing. Missing value imputation was successfully completed using median values for numerical variables and mode values for categorical variables, affecting 3 variables with 100% success rate. This approach created a complete dataset without data loss. Outlier detection using the Interquartile Range (IQR) method examined all numerical variables and achieved 95% success, identifying only 2 outliers that were later verified as valid extreme cases rather than data errors.

Data normalization using StandardScaler transformed 15 variables to have a mean of 0 and standard deviation of 1, ensuring consistent scales across variables. Categorical encoding converted 5 categorical variables into binary representations using one-hot encoding, making them suitable for machine learning algorithms. Feature scaling applied Min-Max scaling to 10 variables, transforming their ranges to [0,1] for improved algorithm performance.

The preprocessing achieved 100% success rates across most activities, demonstrating the original data's high quality. The identification of only 2 outliers suggests careful initial data collection procedures. These detailed preprocessing steps ensure the dataset was properly prepared for machine learning analysis, with standardized scales, complete values, and appropriate variable formats that was supported reliable analytical outcomes.

B. Random Forest Classification for Cloud Resource Prediction

This section explored the application of a Random Forest classifier to predict cloud resource demands in university environments. It begins by analyzing usage patterns across various services to identify baseline behaviors. Relevant features were then engineered and selected to train the model effectively. The classifier was optimized through parameter tuning for improved accuracy and generalization. Finally, its performance is evaluated using standard classification metrics to assess its practical deployment readiness. Table 8 presents the service usage patterns by time period.

TABLE 8 SERVICE USAGE PATTERNS BY TIME PERIOD

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

The service usage analysis reveals distinct patterns that had important implications for cloud resource allocation strategies in universities. Table 8 showed that different services had unique usage characteristics that require tailored resource management approaches. Email and communication services follow predictable business-hour patterns (8AM-5PM on weekdays) with moderate daily consumption of 4.2GB and significant weekend drops of 80%. This predictable pattern makes email services ideal candidates for automated scaling systems that can reduce resources during off-peak periods. Video conferencing showed concentrated usage during class hours with variable consumption (5-100GB) and experiences dramatic increases of 150% during exam periods, requiring flexible resource allocation.

Learning Management Systems demonstrated the most critical resource demands, consuming 500GB-1TB daily with extreme spikes of 200% during assignment submission periods. This pattern indicated urgent need for predictive scaling to prevent system overloads when students submit assignments simultaneously. Website and portal hosting maintains consistent 24/7 operation with moderate weekend drops (30%) but showed massive 300% increases during registration periods.

Storage and backup services operate primarily during night hours with high consumption (100GB-2TB) and minimal weekend variation (10%). These patterns suggest universities need sophisticated resource allocation strategies that can automatically adjust capacity based on predictable daily, weekly, and seasonal cycles. The extreme variations during critical academic periods highlight the importance of proactive resource scaling to maintain system reliability when students and faculty need services most.

The process identified key variables that influence resource demand patterns. Table 9 presents the detailed feature categories and their importance.

TABLE 9 FEATURE CATEGORIES AND IMPORTANCE ANALYSIS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 9 showed that temporal features were the strongest predictors with an importance score of 0.37, including variables like day of week, hour of day, and academic period. This high importance confirms that time-based patterns are fundamental for accurate resource allocation predictions. Resource metrics rank second in importance with a score of 0.31, covering CPU use, memory usage, and bandwidth consumption. These technical indicators provide direct insight into current system demands and help predict future needs. Institutional characteristics achieve medium importance (0.22) through variables like university size, student count, and faculty count, showing that institutional scale significantly influences resource requirements.

Service-specific variables had the lowest importance score (0.10) despite including relevant metrics like email volume, video sessions, and LMS active users. This finding suggests that general usage patterns were more predictive than individual service characteristics, indicating that universities share common resource consumption behaviors regardless of specific applications. The analysis selected 20 features from 45 initial candidates, creating a focused set that balances prediction accuracy with model simplicity. The dominance of temporal features validates the importance of time-aware resource allocation strategies that can anticipate daily, weekly, and seasonal usage cycles. This feature ranking provided valuable guidance for developing automated scaling systems that prioritize the most influential factors affecting cloud resource demands in university settings.

The Random Forest model was systematically developed with careful attention to hyperparameter optimization. Table 10 presents the final model configuration and rationale.

TABLE 10 RANDOM FOREST MODEL CONFIGURATION AND PERFORMANCE

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

The configuration and performance results presented in Table 10 demonstrate a well-calibrated Random Forest model, optimized for both accuracy and generalizability in predicting cloud resource allocation patterns. The use of 100 trees strikes a practical balance between model robustness and processing time, resulting in a 12% improvement in accuracy without incurring high computational costs. This suggests that the ensemble approach provided diverse decision paths, improving overall model reliability.

Limiting the maximum depth to 10 effectively curbs overfitting by preventing the model from becoming too tailored to the training data, while still capturing essential nonlinear relationships a decision that yielded an 8% improvement in generalization. The default setting for min samples split at 2 was appropriate for the dataset size, ensuring that the model remained sensitive to fine-grained splits without compromising baseline performance. Additionally, the min samples leaf set to 1 allowed the model to make detailed, leaf-level distinctions, contributing to a 5% boost in precision. The selection of 'sqrt' for max features reduced overfitting by limiting the number of features considered per split, resulting in a 6% increase in validation accuracy. Setting a random state of 42 ensured consistent and reproducible outcomes. Overall, this configuration not only enhances performance but also strengthens confidence in the model's predictive power and replicability.

The detailed performance evaluation demonstrates strong predictive capabilities. Table 11 presents the detailed performance metrics across different evaluation phases.

TABLE 11 RANDOM FOREST CLASSIFICATION PERFORMANCE METRICS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The performance metrics in Table 11 reflect a well-trained Random Forest classifier with robust generalization capabilities. The training accuracy of 94.2% compared to a validation accuracy of 87.6% a modest 6.6% gap indicated the model has learned meaningful patterns without excessive overfitting. This is further confirmed by the test set accuracy of 86.9% and the cross-validation mean of 85.3% (±3.2%), underscoring consistent performance across different data partitions and supporting the model's reliability for real-world deployment.

Precision scores across datasets (0.96 training, 0.89 validation, 0.88 test) show that the model was effective in minimizing false positives critical in cloud resource allocation, where erroneous predictions could lead to inefficient resource provisioning. The recall values (0.93 training, 0.85 validation, 0.84 test) reflect a strong ability to detect true positives, suggesting the model is sensitive to actual cases of resource demand or stress. F1-scores, which balance precision and recall, remain high across all phases (0.95, 0.87, 0.86), indicating a harmonized predictive capacity.

Importantly, the ROC-AUC values (0.98 training, 0.91 validation, 0.90 test) confirm excellent discriminatory power between classes. These findings justify confidence in the model's deployment, providing a strong rationale for its use in decision-making scenarios involving resource optimization in university IT environments. The confusion matrix analysis provided detailed insights into classification accuracy across different resource demand levels. Table 12 presents the confusion matrix results.

TABLE 12 CONFUSION MATRIX ANALYSIS FOR RESOURCE DEMAND CLASSIFICATION

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

The confusion matrix analysis in Table 12 highlighted the Random Forest model's effectiveness in multi-class classification for predicting resource demand levels low, medium, and high. The model achieves strong recall for low demand (0.90) and high demand (0.88), indicating it correctly identifies the majority of instances in these categories. This was particularly beneficial in practical scenarios where under- or over-allocation of cloud resources during these periods could lead to inefficiency or service disruption. Precision for low (0.87) and high demand (0.83) further affirms the model's reliability in limiting false alarms, supporting effective decision-making. However, the medium demand category presents a relative weakness, with a recall of 0.76 and precision of 0.84. This suggests the model struggles slightly to distinguish medium demand from the adjacent categories, possibly due to overlapping features or subtler patterns in the data. This underperformance could affect dynamic resource scaling where demand fluctuates around average levels.

Despite this, the overall accuracy and recall of 85% show the model was well-balanced and sufficiently robust for operational deployment. The close alignment between precision and recall across classes justifies its suitability for automated cloud resource allocation, with only minimal risk of misclassification in real-time environments. The findings imply that further tuning or feature engineering may improve medium-demand sensitivity.

The prediction results show strong performance across different prediction scenarios. Table 13 presents service-specific prediction performance.

Table 13 Service-Specific Prediction Performance

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

The service-specific performance metrics in Table 13 demonstrate the Random Forest model's adaptability across diverse digital service types within university environments. Email services achieve the highest accuracy (93%) along with high precision (0.94) and recall (0.92), indicating consistent and predictable usage patterns. This aligns with the routine nature of institutional email communication, making it highly amenable to predictive modeling. Peak detection at 95% further reinforces the model's ability to anticipate demand surges, supporting proactive allocation of communication-related resources.

Video conferencing, in contrast, showed the lowest accuracy (85%) and recall (0.83), reflecting more irregular usage tied to spontaneous meetings or external engagements. While precision remained relatively high (0.87), the variability in usage contributes to reduced model confidence. Similarly, storage and backup services pose a prediction challenge, with the lowest recall (0.79), indicating frequent under-detection of actual demand peaks likely due to asynchronous and backend operations that lack visible usage cues. LMS and website/portal services yield balanced performance metrics (88–91% accuracy), strongly influenced by academic calendars and registration cycles. Peak detection rates across all services remain impressively high (85–95%), suggesting the model's utility in resource planning.

These findings imply that the model was robust for services with consistent or cyclical usage but may require refinement (e.g., temporal features or anomaly detection) for more dynamic services like conferencing and storage. The rationale for this service-specific analysis is to fine-tune resource provisioning strategies for each digital service, enhancing efficiency and user experience. The temporal prediction analysis reveals model performance across different time horizons. Table 14 showed accuracy by prediction timeframe.

TABLE 14 PREDICTION ACCURACY BY TIME HORIZON

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

The analysis in Table 14 illustrates how prediction accuracy varies by time horizon, offering key insights into the model's temporal scalability. The 1-hour prediction accuracy of 94.2%, with a narrow confidence interval (92.1%–96.3%), underscores the model's high reliability for real-time optimization. This level of precision was ideal for immediate scaling in response to sudden surges in cloud resource demand, ensuring uninterrupted service and efficient system performance. As the prediction window extends, accuracy gradually decreases dropping to 89.5% at 4 hours and 85.7% at 8 hours. These figures remain strong and indicate the model's utility for short-term planning tasks such as shift scheduling and daily pre-allocation of digital resources. Even at a 24-hour horizon, with 78.9% accuracy, the model remained sufficiently reliable for strategic daily capacity planning, where some tolerance for variance was acceptable. The lowest performance was observed at the 1-week horizon (72.3%), which reflects the inherent challenge in forecasting over extended periods due to cumulative uncertainties. Nevertheless, this level of accuracy still holds practical value for identifying long-term usage trends and informing budget forecasting.

These findings affirm the model's strength in short- to medium-term resource management, while also providing a rationale for integrating supplementary forecasting tools for longer-range predictions. This time-sensitive insight enhances decision-making efficiency across operational, tactical, and strategic levels in university IT infrastructure planning.

C. K-means Clustering for Resource Demand Pattern Identification

Page numbers, headers and footers must not be used. This section explored the use of K-means clustering to identify distinct patterns in university IT resource demand. It begins with detailed data preparation, including normalization, encoding, and dimensionality reduction, to ensure meaningful and accurate clustering. The K-means algorithm was then implemented, with validation methods confirming four optimal clusters. These clusters were analyzed to reveal varying levels of resource usage ranging from low to high and variable-demand universities each with unique infrastructure needs. Representative universities were mapped to these clusters to illustrate practical relevance. Finally, statistical validation confirms the robustness and significance of the clusters, supporting their use in strategic cloud resource planning and differentiated service delivery.

K-means clustering was applied to identify distinct institutional patterns in cloud resource demand across the four participating universities The clustering data preparation involved multiple preprocessing steps to ensure optimal cluster formation. Table 15 presents the detailed preprocessing activities.

TABLE 15CLUSTERING DATA PREPARATION AND PREPROCESSING

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 15 presents a detailed and effective preprocessing pipeline that enhances the quality and usability of the dataset for clustering analysis. The use of StandardScaler for data normalization ensured that all 15 numerical features were rescaled to have a mean of 0 and a standard deviation of 1. This step is critical for clustering algorithms, such as K-Means, which rely on distance-based metrics; by standardizing the data, it ensures that no single feature disproportionately influences the cluster formation process.

Initial dataset contained 20 variables (15 numerical, 5 categorical). After categorical encoding using one-hot encoding, 5 categorical variables expanded to 9 binary variables, creating 24 total features. PCA reduced these to 6 principal components retaining 85% variance for clustering analysis. The categorical encoding using one-hot encoding transformed all five categorical variables into a binary format, preserving their qualitative information while making them suitable for mathematical modeling. The missing value imputation strategy using the median for numerical and mode for categorical variables helped retain 98.5% of the dataset, ensuring that the majority of the original data was preserved. This high retention rate contributes to the robustness of the clustering outcome.

Principal Component Analysis (PCA) was applied to reduce 24 dimensions down to 6 while retaining 85% of the original variance. This reflects optimal compression, reducing computational burden without sacrificing much information.

Feature selection using a variance threshold further streamlined the dataset by removing features with negligible contribution, which enhances model stability and interpretability.

These preprocessing steps not only prepared the data for accurate and efficient clustering but also ensured minimal information loss and structural bias. The rationale for this pipeline was to maximize clustering performance by improving feature comparability, reducing dimensional noise, and preserving essential data characteristics critical for uncovering meaningful resource usage patterns within university IT systems.

K-means algorithm implemented using scikit-learn with parameters: n_clusters=4, init='k-means++', max_iter=300, tol=1e-4, random_state=42 for reproducibility. Convergence achieved in average 156 iterations across validation runs. Distance metric: Euclidean distance. Cluster centers initialized using k-means++ method for optimal initial placement. Table 16 presents the cluster optimization analysis.

TABLE 16 OPTIMAL CLUSTER DETERMINATION AND VALIDATION

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 16 presents a rigorous evaluation of clustering performance using multiple validation techniques, all of which converge on an optimal cluster number of k=4. This strong agreement across methods enhances the reliability of the clustering results. The Elbow Method showed a distinct "elbow" at k=4, indicating that increasing the number of clusters beyond this point yields diminishing returns in reducing Within-Cluster Sum of Squares (WCSS). This suggests a natural grouping in the data, ideal for segmentation or resource categorization.

The Silhouette Score of 0.67 reflects good intra-cluster cohesion and inter-cluster separation, confirming that data points were well matched within their own clusters and clearly distinct from other clusters. The Calinski-Harabasz Index (156.3) further supported this by quantifying the ratio of between-cluster dispersion to within-cluster dispersion higher values indicate more distinct and well-formed clusters. The Davies-Bouldin Index value of 0.85 (lower is better) suggests reasonably compact and well-separated clusters, though slightly lower confidence than the other metrics.

The Gap Statistic result (0.12) demonstrates that the observed clustering structure is significantly better than what would be expected from random uniform distributions, reinforcing the validity of the chosen k-value. Overall, the consistent identification of k=4 across all validation metrics justifies its use in later analysis or application. This outcome provided a solid foundation for segmenting university IT resource usage patterns, which can guide targeted policy interventions, customized service provisioning, or differentiated scaling strategies for cloud infrastructure.

The cluster analysis identified four distinct patterns of resource use across universities. Table 17 presents the detailed cluster characteristics.

TABLE 17 CLUSTER CHARACTERISTICS AND RESOURCE DEMAND PATTERNS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

Table 17 outlines the defining characteristics of four distinct institutional clusters based on resource demand patterns, offering actionable insights for tailored cloud resource management in universities. Cluster 0 (Low-Resource) represents 25% of universities, typically with fewer than 5,000 students. These universities exhibit modest CPU use (25–45%), low bandwidth (<500 Mbps), and moderate storage usage (55–70%), reflecting limited digital infrastructure and reliance on basic services. Their primary need was cost-effective, scalable solutions that optimize performance within budget constraints.

Cluster 1 (Medium-Resource), comprising 40% of the sample, includes universities with student populations ranging between 5,000 and 20,000. These show moderate CPU (45–65%), bandwidth (500–1500 Mbps), and storage use (70–85%), indicating balanced service demands and potential for growth. They were well-positioned for strategic scaling and infrastructure investment to support evolving academic and administrative needs.

Cluster 2 (High-Resource), though smaller (20%), includes universities with over 20,000 students and consistently high resource consumption: CPU (65–85%), bandwidth (>1500 Mbps), and storage (85–95%). These were technologically intensive environments requiring premium, high-performance cloud services to maintain service delivery, research, and digital learning platforms.

Cluster 3 (Variable-Demand) was the most complex, accounting for 15% of universities with inconsistent patterns due to multiple campuses or hybrid delivery models. Their CPU and storage usage vary widely (40–85% and 60–95%, respectively), and bandwidth was inconsistent. This group demands adaptive, dynamic resource allocation strategies, such as predictive scaling and load balancing, to handle fluctuating workloads.

The rationale for cluster-based analysis was to enable differentiated service provisioning and resource planning. These insights help optimize cloud infrastructure deployment, ensuring cost-efficiency for lower-tier users while guaranteeing robustness and responsiveness for high-demand universities. The detailed cluster profiling reveals specific characteristics for each group. Table 18 shows the representative universities and their key metrics.

TABLE 18 REPRESENTATIVE UNIVERSITIES BY CLUSTER

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 18 provided empirical validation of the clustering outcomes by mapping representative universities to each cluster, revealing logical and consistent groupings that reinforce the interpretability and practical relevance of the cluster model. In Cluster 0 (Low-Resource), universities like Alupe, Turkana, and Kaimosi smaller universities with limited digital infrastructure had an average student population of 4,000, low CPU usage (35%), and minimal bandwidth demand (350 Mbps). Their service portfolio was limited to basic needs such as email and rudimentary LMS access, affirming their classification as cost-sensitive universities requiring essential but efficient IT support.

Cluster 1 (Medium-Resource) included universities like Rongo and Catholic University, with around 11,500 students, moderate CPU use (55%), and bandwidth averaging 900 Mbps. These universities support more detailed services like full LMS platforms and video conferencing. Their infrastructure and service diversity reflect balanced growth and moderate complexity, making them suitable candidates for strategic investments in scalable cloud solutions.

In Cluster 2 (High-Resource), Moi, Maseno, and the University of Eldoret emerged as large, research-driven universities with approximately 25,000 students, high CPU usage (75%), and peak bandwidth needs (1,800 Mbps). Their reliance on advanced academic services and research infrastructure necessitates high-performance cloud environments, affirming their position in the high-resource tier.

Cluster 3 (Variable-Demand) was represented by Masinde Muliro University, which operates across multiple campuses. Though it had a similar student population to Cluster 2 universities, its resource use was more volatile (60% CPU, 1,200 Mbps bandwidth) due to distributed service needs and potential asynchronous usage patterns across locations. This

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

validates the classification of this cluster as the most dynamic and complex, requiring adaptive, context-aware resource provisioning.

The alignment of real universities with these clusters confirms the model's external validity and its utility in guiding differentiated resource planning. This reinforces the rationale for cluster-based management strategies that match IT infrastructure to institutional scale, service complexity, and operational variability.

Validation ensured the reliability and generalizability of clustering results. The detailed validation analysis confirms the robustness of identified clusters. Table 19 presents the statistical validation results.

TABLE 19 CLUSTER STABILITY AND STATISTICAL VALIDATION

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 19 presents strong statistical and empirical evidence validating the robustness, quality, and practical utility of the identified clusters. Bootstrap validation, showing 95% consistency, confirms that the clustering solution was highly stable across repeated random sampling and model reruns. This high degree of replicability enhance confidence in the cluster structure and supported its use in critical decision-making processes such as infrastructure planning and service differentiation.
The Silhouette Analysis average score of 0.67 reaffirms good cluster cohesion and separation, indicating that data points were well-grouped and distinct from those in other clusters. This is further supported by the ANOVA F-statistic of 42.7 with a p-value < 0.001, highlighting statistically significant differences in resource usage and institutional characteristics across the four clusters. The Tukey's HSD test confirms that all pairwise comparisons between clusters are significant, reinforcing the uniqueness and non-overlapping nature of each cluster.
Most notably, the large effect size (Cohen's $d = 1.2$) points to a strong practical significance. This means the observed differences were not only statistically meaningful but also had real-world implications, especially in contexts such as differentiated funding, tailored IT support, and targeted resource scaling. Collectively, these results provide a compelling rationale for adopting the cluster-based framework as a decision-support tool for strategic cloud resource allocation in universities. The findings validate the analytical approach and demonstrate that institutional groupings were both statistically defensible and operationally meaningful.

D. Random Forest and K-means Integrative Analysis

This section presents a detailed approach to identifying university IT resource usage patterns using K-means clustering. It begins with meticulous data preparation involving normalization, feature selection, and dimensionality reduction to ensure clustering accuracy. The K-means implementation was guided by multiple validation techniques, all pointing to four optimal clusters. These clusters were analyzed and interpreted to reveal distinct institutional resource demand profiles. Finally, statistical validation confirms the reliability and practical significance of the identified patterns, supporting data-driven infrastructure planning.
The convergence analysis validates the complementary nature of supervised and unsupervised approaches. The convergence analysis reveals strong alignment between Random Forest predictions and K-means clustering results. Table 20 presents the detailed convergence analysis.

Table 20 Supervised vs Unsupervised Method Convergence

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

Table 20 presents a comparative analysis of supervised (Random Forest) and unsupervised (K-Means clustering) methods, revealing an 85% overall convergence rate, which demonstrates strong alignment between predictive classification and data-driven segmentation. This high agreement underscores the consistency and reliability of the models in characterizing institutional resource demand, particularly for well-defined categories.

For Low-Resource and High-Resource universities, convergence was 100%, reflecting complete agreement between predicted demand levels and cluster assignment. This perfect alignment in the extremes validates the accuracy and robustness of both methodologies, affirming that universities with very low or very high resource usage exhibit clear and predictable patterns ideal for deterministic resource planning and infrastructure investment.

In the Medium-Resource category, the agreement was slightly lower at 87.5%, with a single divergence case, likely due to overlapping feature ranges that blur the boundaries between moderate and adjacent resource levels. The most variability was observed in the Variable-Demand cluster, with 66.7% agreement, indicating that such universities often multi-campus or irregular in usage defy easy classification due to their fluctuating patterns.

The presence of two divergence cases overall reinforces the need for hybrid analytical approaches, especially for complex institutional types. Integrating supervised and unsupervised insights can enhance model accuracy and responsiveness in dynamic environments. The rationale for this comparison was to ensure methodological robustness and to guide strategic resource allocation through both predictive modeling and exploratory segmentation.

The complementary analysis reveals unique strengths of each approach. Table 21 presents the comparative advantages and insights.

TABLE 21 COMPLEMENTARY INSIGHTS FROM SUPERVISED AND UNSUPERVISED METHODS

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 21 underscores the powerful synergy between supervised (Random Forest) and unsupervised (K-Means) learning methods in analyzing and managing cloud resource allocation in universities. Random Forest offers high-precision temporal predictions demonstrated by its 94% accuracy for 1-hour forecasts making it ideal for real-time operational decisions, such as immediate resource scaling or responding to usage spikes. It also excels in event-specific peak detection (91% accuracy), enabling universities to proactively address time-sensitive demands like registration or examination periods.

Conversely, K-Means clustering contributes a broader, structural understanding by uncovering latent patterns in institutional behavior and usage. It groups universities based on shared characteristics, aiding in strategic differentiation and capacity planning. This static but insightful segmentation was essential for long-term infrastructure investments and service prioritization, especially when budgets or policies vary across universities.

When combined, these methods offer complementary benefits: Random Forest enables reactive and dynamic adjustments, while K-Means supports generalized, forward-looking strategies. For example, time-based anomalies detected through RF can be contextualized by cluster behavior from K-Means, enhancing anomaly interpretation. Likewise, supervised forecasting can be refined by incorporating cluster-specific thresholds or patterns. The rationale for integrating both approaches lies in their combined ability to support real-time responsiveness and long-term strategic planning. This integrated framework enabled universities to not only react to present resource demands but also anticipate and plan for future growth, variability, and institutional diversity, thereby ensuring efficient and sustainable cloud infrastructure management.

The pattern correlation analysis quantifies relationships between different analytical dimensions. Table 22 presents the detailed correlation matrix.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 9, September 2025

DOI: 10.17148/IJARCCE.2025.14901

TABLE 22 PATTERN CORRELATION ANALYSIS MATRIX

Metric	Value	Interpretation
Target population	20 ICT personnel	Universities in Western Kenya
Actual responses received	20	100% response rate
Data completeness	98.5%	Minimal missing values
Cronbach's Alpha	0.847	Good internal consistency
Missing data rate	1.5%	Excellent data completeness
Response time variance	2-45 minutes	Acceptable range
Outliers identified	2	Minimal, verified as valid

Table 22 presents a pattern correlation matrix that reinforces the internal consistency and predictive validity of the analytical framework. The strong positive correlation ($r = 0.78$, $p < 0.001$) between Random Forest (RF) predictions and K-means cluster assignments affirms the earlier convergence findings, suggesting that both methods capture similar underlying patterns in institutional behavior. This strengthens the case for integrating supervised and unsupervised approaches, as they provide mutually reinforcing insights.

A particularly notable finding was the very strong correlation between resource demand and institutional size ($r = 0.82$, $p < 0.001$). This confirms an intuitive yet critical relationship: larger universities inherently require more computing resources due to higher student numbers, more services, and complex digital ecosystems. Similarly, CPU use and student count also exhibit a strong correlation ($r = 0.74$), reinforcing the operational dependency between enrollment levels and computational demand.

Moderate correlations were observed between temporal patterns and cluster membership ($r = 0.65$) and between bandwidth usage and service complexity ($r = 0.69$). These indicate that while clusters reflect broad usage structures, some variability remained due to service-specific and time-bound factors. The storage use vs. academic period relationship ($r = 0.58$) reflects cyclical patterns such as data spikes during exams or semester transitions.

Overall, these statistically significant correlations validate the integrated framework's foundational assumptions. The strength of associations across variables demonstrates both the logical coherence and analytical rigor of the approach. This provided a solid empirical basis for adopting the integrated system in guiding resource planning, service optimization, and infrastructure scaling in university ICT environments.

V. CONCLUSION

Based on the findings, the paper concludes that supervised and unsupervised learning techniques significantly enhance resource allocation in cloud computing Infrastructure as a Service (IaaS) environment, particularly within university settings. The integration of Random Forest and K-means algorithms resulted in a more intelligent and responsive system for managing cloud resources. The Random Forest model achieved high accuracy in forecasting short-term resource demands, enabling real-time provisioning and efficient utilization of infrastructure. Simultaneously, K-means clustering revealed distinct usage patterns that allowed for long-term strategic planning and better understanding of institutional behavior in resource consumption. its also concludes that these techniques were not only effective independently but also yield superior outcomes when combined in a hybrid framework

REFERENCES

[1]. M. A. Al-Emran and K. Shaalan, "A systematic review of cloud computing adoption in education: Challenges and opportunities," *IEEE Access*, vol. 9, pp. 77190-77205, 2021.

[2]. A. Darwish, "Cloud computing and machine learning applications in higher education: A comprehensive review," *Journal of Cloud Computing*, vol. 13, no. 2, pp. 45-62, 2024.

[3]. R. Jones, S. Smith, and M. Brown, "Infrastructure as a Service adoption in UK universities: A case study analysis," *British Journal of Educational Technology*, vol. 50, no. 3, pp. 1234-1250, 2019.

[4]. P. Andre, L. Wilson, and K. Taylor, "Unsupervised learning applications in Australian university cloud infrastructure," *Australasian Journal of Educational Technology*, vol. 39, no. 4, pp. 78-95, 2023.

[5]. L. Zhang, W. Chen, and H. Liu, "Machine learning for heterogeneous cloud resource management in Asian universities," *IEEE Transactions on Cloud Computing*, vol. 11, no. 3, pp. 456-471, 2023.

[6]. [6] J. Harris, D. Miller, and A. Johnson, "Internet2 NET+ program: Machine learning applications in resource allocation," *EDUCAUSE Review*, vol. 57, no. 2, pp. 34-48, 2022.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

[7]. S. Wilson, T. Anderson, and R. Clark, "Australian Research Cloud: Optimizing resource usage through unsupervised learning," *Journal of Research Computing*, vol. 15, no. 1, pp. 12-28, 2023.
[8]. Y. Li, X. Wang, and Z. Zhou, "Supervised learning models for online education resource prediction in China," *Computers & Education*, vol. 195, pp. 104720, 2023.
[9]. G. Chux and D. Teferra, "Cloud computing adoption challenges in African higher education institutions," *International Journal of Educational Development*, vol. 88, pp. 102-115, 2022.
[10]. J. Kato, M. Mugisha, and S. Nsibambi, "E-learning resource prediction using supervised learning at Makerere University," *African Journal of Science and Technology*, vol. 24, no. 3, pp. 45-58, 2023.
[11]. P. Maina, "Cloud computing implementation in Kenyan universities: Opportunities and challenges," *Journal of African Higher Education*, vol. 21, no. 2, pp. 89-104, 2023.
[12]. R. Okello, M. Kiprotich, and J. Wanjiku, "African Research Cloud initiative: Enhancing university research capabilities," *African Journal of Information Systems*, vol. 15, no. 4, pp. 23-39, 2023.
[13]. A. Mushi, E. Kilonzo, and F. Mwalimu, "Computational demand prediction for large-scale data analysis in Tanzanian universities," *East African Journal of Science and Technology*, vol. 12, no. 1, pp. 67-82, 2023.
[14]. E. Niyonzima, C. Mukamana, and J. Uwimana, "Resource optimization in Rwanda's university cloud platforms using unsupervised learning," *Rwanda Journal of Engineering and Technology*, vol. 8, no. 2, pp. 34-49, 2023.
[15]. M. Essa, A. Celik, and S. Human-Hendricks, "Decision tree applications in university e-learning resource prediction," *South African Computer Journal*, vol. 35, no. 1, pp. 78-94, 2023.
[16]. P. Kigo, J. Omondi, and R. Omolo, "Principal Component Analysis for cloud resource optimization at Strathmore University," *Kenya Journal of Sciences*, vol. 13, no. 2, pp. 156-171, 2023.
[17]. S. Kipkemoi, "KENET: Facilitating cloud adoption in Kenyan higher education," *African Educational Research Journal*, vol. 11, no. 3, pp. 45-61, 2023.
[18]. C. Nyambura, F. Kiprotich, and M. Wanjala, "Supervised learning for computational resource prediction at Kenyatta University," *International Journal of Computer Applications*, vol. 181, no. 28, pp. 12-19, 2023.
[19]. S. Maina, R. Karanja, and P. Githui, "Clustering techniques for research project resource optimization at JKUAT," *Journal of Applied Sciences*, vol. 23, no. 4, pp. 234-248, 2023.
[20]. R. Sai, M. Kumar, and A. Patel, "Machine learning approaches for cloud resource optimization: A systematic review," *IEEE Transactions on Services Computing*, vol. 17, no. 2, pp. 789-805, 2024.
[21]. B. Nyamboga, "Machine learning applications in Kenyan university cloud computing environments," *Kenya ICT Journal*, vol. 8, no. 1, pp. 23-37, 2024.
[22]. B. K. Mahesh, "Machine learning algorithms-a review," International Journal of Science and Research, vol. 9, no. 1, pp. 381-386, 2020.
[23]. C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer, 2006.
[24]. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[25]. A. Kumar, R. Singh, and M. Patel, "Supervised learning for cloud resource prediction," Journal of Cloud Computing, vol. 10, no. 2, pp. 15-28, 2021.
[26]. J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281-297, 1967.
[27]. D. Xu, Y. Tian, and R. Chen, "Unsupervised learning techniques in cloud computing," IEEE Transactions on Cloud Computing, vol. 7, no. 3, pp. 456-470, 2019.
[28]. M. Ali, S. U. Khan, and A. V. Vasilakos, "Security in cloud computing: Opportunities and challenges," Information Sciences, vol. 305, pp. 357-383, 2020.
[29].] L. Wang, J. Tao, M. Kunze, A. C. Castellanos, D. Kramer, and W. Karl, "Scientific cloud computing: Early definition and experience," in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, pp. 825-830, 2020.
[30]. D. Niyato, A. V. Vasilakos, and Z. Kun, "Resource and revenue sharing with coalition formation of cloud providers: Game theoretic approach," in Proceedings of the 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 215-224, 2021.
[31]. A. Jones and L. Li, "Machine learning applications in university cloud computing," International Journal of Educational Technology, vol. 15, no. 3, pp. 78-92, 2020.
[32]. P. Nkosi, T. Mthembu, and K. Singh, "Cloud optimization in resource-constrained environments," African Journal of Information Systems, vol. 12, no. 4, pp. 234-251, 2020.
[33]. G. Chux and D. Teferra, "Cloud computing adoption in African higher education," Higher Education Policy, vol. 35, no. 2, pp. 145-168, 2022.
[34]. J. Kariuki, M. Wanjiku, and P. Kamau, "Hybrid resource allocation in Kenyan universities," East African Journal of Information Technology, vol. 8, no. 1, pp. 45-62, 2022.
[35]. H. Jewgeni, Queueing Theory: Mathematical Foundation and Applications. Berlin: Springer, 2023.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 9, September 2025
DOI: 10.17148/IJARCCE.2025.14901

[36]. P. Gonçalves, "Queueing theory applications in cloud computing," IEEE Transactions on Network and Service Management, vol. 19, no. 2, pp. 189-203, 2022.
[37]. K. Hanley, "Game theory in resource allocation," Operations Research Letters, vol. 49, no. 3, pp. 412-420, 2021.
[38]. U. Urmila, Optimization Theory and Methods. New Delhi: PHI Learning, 2020.
[39]. R. Singh, A. Sharma, and K. Patel, "Predictive resource allocation using machine learning," IEEE Transactions on Cloud Computing, vol. 8, no. 2, pp. 234-247, 2020.
[40]. M. Ahmed, S. Hassan, and T. Ali, "Energy-efficient clustering in cloud environments," Journal of Parallel and Distributed Computing, vol. 145, pp. 78-91, 2021.
[41]. L. Zhao, Y. Chen, and R. Wang, "Deep reinforcement learning for dynamic resource allocation," IEEE Transactions on Network Science and Engineering, vol. 6, no. 4, pp. 789-802, 2019.
[42]. A. Rahman and M. Rahim, "Neural network-based resource prediction in large-scale cloud systems," Future Generation Computer Systems, vol. 118, pp. 156-169, 2021.
[43]. S. Alsubaie, "Hybrid approaches in cloud resource optimization," International Journal of Advanced Computer Science and Applications, vol. 14, no. 3, pp. 89-104, 2023.
[44]. R. Garg, P. Kumar, and S. Mishra, "Context-aware resource allocation in educational cloud environments," Computers & Education, vol. 162, pp. 104-118, 2020.
[45]. A. Sharma and R. Gupta, "Real-world validation challenges in cloud optimization," IEEE Cloud Computing, vol. 8, no. 4, pp. 67-75, 2021.
[46]. N. Liang and R. Samavi, "Theoretical foundations of machine learning in cloud computing," ACM Computing Surveys, vol. 53, no. 2, pp. 1-35, 2020.
[47]. S. Dawadi, R. Shrestha, and R. A. Giri, "Mixed-methods research: A discussion on its types, challenges, and criticisms," Journal of Practical Studies in Education, vol. 2, no. 2, pp. 25-36, 2021.
[48]. M. Cloudsim, R. N. Calheiros, and R. Buyya, "CloudSim Plus: A cloud computing simulation framework pursuing software engineering principles for improved modularity, extensibility and correctness," in Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 400-409, 2017.
[49]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2009.
[50]. M. M. Mohamad, N. L. Sulaiman, L. C. Sern, and K. M. Salleh, "Measuring the validity and reliability of research instruments," Procedia-Social and Behavioral Sciences, vol. 204, pp. 164-171, 2019.
[51]. National Commission for Science, Technology and Innovation (NACOSTI), Research Guidelines and Ethics. Nairobi: Government Printer, 2022.