Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 324-336

The Use of Machine Learning for Efficient Energy Management in Big Data-Based Computing Systems

Toni Wijanarko Adi Putra*[1], Nuris Dwi Setiawan[1], Rusito[1]
Email: toni.wijanarko@stekom.ac.id, setyawan_dw@stekom.ac.id, rusito@stekom.ac.id
[1]*Universitas Sains dan Teknologi Komputer, Semarang, Indonesia, 50192*

*Corresponding Author

Abstract

The rapid growth of digital services has intensified the energy demands of data centers, significantly impacting operational costs and global carbon footprints. This study leverages Machine Learning (ML) and Big Data to optimize energy management in data centers, addressing limitations in prior approaches that overlooked real-time variability. A predictive model utilizing the Random Forest algorithm was developed to reduce energy consumption based on dynamic factors, such as workload and environmental conditions like temperature. The research used a six-month dataset consisting of approximately 3 million data points from an operational data center. After preprocessing the data, the model achieved a high predictive accuracy, reflected by an R-squared value of 0.87. The findings demonstrate that the model reduces energy consumption by an average of 11.17% daily, with peak savings of up to 15.56% during off-peak hours. Key metrics, including a Mean Squared Error (MSE) of 0.034 and a Root Mean Squared Error (RMSE) of 0.184, validated the model's effectiveness and reliability. Statistical tests further confirmed its precision within a 95% confidence interval. This study contributes to the academic field by integrating real-time environmental data into predictive modeling, offering a scalable solution for energy-efficient data center operations. The outcomes support sustainability initiatives by mitigating carbon emissions and reducing operational costs. The findings also provide a framework for applying ML in broader industrial contexts requiring efficient energy management. Future research may explore incorporating additional variables, such as user behavior, to further refine predictive capabilities.

Keywords: *Machine Learning, Energy Efficiency, Data Centers, Predictive Analytics*

I. INTRODUCTION

The development of information technology, particularly in Big Data and Machine Learning, has revolutionized various industrial sectors. Big Data refers to large, diverse, and complex datasets that cannot be effectively analyzed using traditional methods (Cao et al., 2022). Meanwhile, Machine Learning (ML), a subset of Artificial Intelligence, enables systems to learn from data, uncover patterns, and make predictions autonomously. The synergy between these technologies has opened new possibilities for optimizing decision-making processes, including energy management in critical infrastructures like data centers (Karthikeyan & Priyakumar, 2022). Data centers play a central role in supporting modern digital services but are among the largest energy consumers globally. It is estimated that data centers contribute to 1% of global electricity use, a figure projected to grow significantly with increasing reliance on cloud computing and digital platforms. This rising energy demand impacts operational costs and exacerbates global carbon footprints (Katal et al., 2023). Addressing this challenge requires innovative approaches that enhance energy efficiency while maintaining the performance and reliability of data center operations.

Received on September 17, 2024; Revised on September 25, 2024; Accepted on October 24, 2024; Published on December 20, 2024
DOI: *10.51903/jtie.v3i3.202*

Current efforts to improve energy efficiency often focus on optimizing individual components of data centers, such as hardware or cooling systems. However, these approaches are limited in scope, as they fail to consider the broader influence of real-time environmental factors, including ambient temperature, humidity, and workload dynamics. Research has demonstrated that energy consumption patterns can shift rapidly in response to these variables, necessitating more adaptive solutions that integrate multiple factors (Guo et al., 2022). Many studies addressing energy efficiency in data centers rely heavily on historical data, which limits their ability to account for variability in real-time operational conditions (Bharany et al., 2022); (He et al., 2022).

Machine Learning has emerged as a promising solution for addressing these limitations. Algorithms such as Random Forest (RF), Neural Networks, and Support Vector Machines (SVM) are widely used in predictive analytics for energy optimization. RF offers a robust approach by constructing multiple decision trees and aggregating their outputs, enabling the algorithm to manage high data variability effectively. Studies have shown that ML-based predictive models can achieve up to 15% improvement in energy efficiency when integrated with real-time dat (Ouatik et al., 2022) a; (Maharana et al., 2022). Despite these advances, most existing studies overlook the interplay between workload dynamics and environmental factors, leaving a critical gap in the literature.

This research addresses this gap by developing an energy optimization model that integrates real-time environmental data, workload fluctuations, and advanced ML techniques. Unlike previous studies that primarily use simpler models, such as Decision Trees, this study employs RF to handle the complexity of dynamic data and achieve higher predictive accuracy. By incorporating six months of real-time data collected from an operational data center, this study evaluates the effectiveness of RF in reducing energy consumption while maintaining system performance (Khan et al., 2022). The objective of this research is to optimize energy consumption in data centers by leveraging the predictive capabilities of ML and the extensive data integration capabilities of Big Data. Specifically, the study aims to demonstrate how real-time environmental and operational data can enhance energy savings, particularly during off-peak hours. This approach not only reduces operational costs but also aligns with global sustainability initiatives aimed at mitigating the environmental impact of high energy consumption (Hassani & Silva, 2023).

This study contributes to the literature by offering a comprehensive solution to energy management in data centers. First, it incorporates real-time variables, such as temperature and workload, into predictive models, addressing a critical gap in existing research. Second, it demonstrates the superior performance of RF in managing complex, high-variability datasets. Finally, it presents a scalable framework for integrating ML and Big Data, providing insights for broader applications across other energy-intensive sectors. These contributions highlight the relevance and novelty of the research, positioning it as a significant advancement in sustainable energy management.

II. LITERATURE REVIEW

A. Energy Management in Data Centers

Energy management in data centers has become a critical issue in recent years, driven by rising global computing demands that impact both energy consumption and carbon emissions (Katal et al., 2023). Modern data centers contribute significantly to the global carbon footprint due to the high energy required to operate servers, cooling systems, and other supporting infrastructure (Buyya et al., 2024). Various efforts have been made to reduce energy consumption in data centers; however, most studies primarily focus on optimizing hardware efficiency or cooling systems without accounting for real-time factors that can dynamically influence energy consumption (Zhang & Liu, 2022). Some studies have attempted to address this issue through predictive algorithm-based approaches that rely on historical data to forecast future energy needs (Shabbir et al., 2022). However, research that depends solely on historical data often fails to capture essential real-time variability, which is critical in data centers where workloads and environmental conditions can shift rapidly (Bharadiya, 2023). Data centers are pivotal to the digital economy but are also significant energy consumers, projected to account for approximately 8% of global electricity consumption by 2030. The rise in demand for cloud services, artificial intelligence workloads, and big data processing has amplified the need for sustainable energy management practices (Buyya et al., 2024). The growing reliance on these facilities has intensified concerns over their environmental impact, particularly in terms of carbon emissions and operational costs.

To address these concerns, sustainability metrics such as Power Usage Effectiveness (PUE) and Carbon Usage Effectiveness (CUE) have been developed to evaluate energy efficiency in data centers. These metrics provide a framework for benchmarking energy usage and identifying inefficiencies in operations (Cao et al., 2022). Optimizing these metrics is crucial for reducing the environmental footprint of data centers while maintaining service reliability and performance. Renewable energy sources, including solar and wind power, are increasingly being integrated into data center energy systems to offset reliance on non-renewable energy. Hybrid models that combine renewable energy with advanced management systems have demonstrated significant potential in reducing energy consumption and enhancing sustainability (He et al., 2022). These efforts highlight the need for innovative approaches that consider the interplay of energy sources, environmental factors, and operational requirements in data centers.

B. Machine Learning for Energy Optimization

Machine Learning has proven to be an effective tool for optimizing various aspects of computing systems, including energy efficiency. Algorithms such as RF, Neural Networks, and Support Vector Machines (SVM) are widely used in studies focused on predicting energy consumption (Zagajewski et al., 2021). Previous studies have employed simpler algorithms like Decision Trees, which, while effective in processing data, often struggle to handle the complexity of changing real-time data (Maharana et al., 2022). In contrast, RF provides greater flexibility and can accommodate more environmental variables, making it well-suited for scenarios where energy consumption is influenced by factors such as temperature, humidity, and workload (Abdel-Razek et al., 2022).

Research by (Awan et al., 2021) found that using Neural Networks to predict energy consumption in data centers achieved up to 10% energy savings. Nevertheless, Neural Networks often face overfitting issues and require large datasets for effective training (Awan et al., 2021). On the other hand, RF can mitigate this issue by constructing multiple decision trees and combining their results to deliver more accurate and stable predictions (Yin et al., 2023). Thus, this study employs RF as the core algorithm in the predictive model to minimize energy consumption due to its capacity to handle real-time data variability and process diverse input types (Fan et al., 2022). Machine Learning (ML) has emerged as a transformative tool for optimizing energy consumption in data centers. Advanced ML techniques, such as Deep Reinforcement Learning (DRL) and Gradient Boosting, have shown promising results in

Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 324-336

identifying complex patterns and making real-time decisions for energy allocation (Yan et al. 2022). DRL has the capability to dynamically adjust computing resources based on current energy demands and workload predictions, ensuring efficient energy usage.

Beyond data centers, ML techniques applied in industries like manufacturing and transportation have provided valuable insights into managing dynamic and real-time energy requirements (Santoso et al., 2024). Lessons from these domains indicate the potential for cross-domain applications of ML in improving predictive accuracy and system adaptability in data centers (Strielkowski et al., 2023). Incorporating such strategies could further enhance the capability of ML to address the unique challenges of data center operations. Hybrid ML models that combine Neural Networks and ensemble methods have been particularly effective in energy forecasting. These models leverage the strengths of individual algorithms, enabling them to achieve up to 18% energy savings in high-performance computing environments (Shabbir et al., 2022). Such hybrid approaches offer a promising pathway for integrating diverse data sources and improving energy efficiency in data centers.

C. *Big Data in Energy Management*

Big Data plays a central role in modern computing systems, particularly in the context of energy management. As data volumes continue to grow in data centers, managing and processing large datasets has become a unique challenge (Ouatik et al., 2022). Big Data is characterized by the 3Vs: Volume, Velocity, and Variety. Volume refers to the vast amount of data collected from various sources; Velocity relates to the speed of data collection and processing, while Variety encompasses different data types such as text, images, and sensor signals (González García & Álvarez-Fernández, 2022). Big Data plays a central role in modern energy management systems, particularly in data centers where large-scale operations generate vast amounts of data. The integration of Internet of Things (IoT) devices into data center infrastructures has enabled real-time monitoring of energy consumption, environmental conditions, and system performance. These IoT devices provide critical inputs for ML models, allowing for more responsive and accurate energy management (Panwar et al., 2022).

However, the exponential growth of data volumes in data centers poses scalability challenges. Managing and processing this data requires advanced techniques, such as data partitioning, distributed computing, and cloud-based analytics platforms, to ensure that systems can handle increasing workloads effectively (González García & Álvarez-Fernández, 2022). Addressing these challenges is essential for leveraging the full potential of Big Data in optimizing energy management. In addition to operational optimization, Big Data analytics has been instrumental in predictive maintenance, identifying potential hardware failures before they occur. By minimizing downtime and avoiding over-provisioning of resources, predictive maintenance indirectly contributes to energy efficiency (Bharadiya, 2023). These advancements underscore the importance of integrating Big Data with ML to develop holistic and adaptive energy management solutions.

In energy management, Big Data enables the integration of data from various sources, such as IoT sensors in data centers that record temperature, workload, and energy consumption in real-time (Panwar et al., 2022). Analyzing this data requires substantial computational power and efficient algorithms to detect energy consumption patterns and identify potential savings. (Ouatik et al., 2022) found that integrating Big Data and Machine Learning can improve prediction accuracy in energy management by up to 15%, underscoring the relevance of the approach taken in this study (Ouatik et al., 2022). However, a primary challenge in managing Big

Data is data security and privacy, as large data volumes heighten the risk of information breaches (Rao et al., 2022).

D. Real-Time Data Integration in Predictive Models

The integration of real-time data is critical for the effectiveness of energy management in data centers. Unlike historical data, real-time data provides dynamic insights into fluctuating workloads and environmental conditions, enabling more accurate and responsive predictions (Ouatik et al., 2022). The inclusion of variables such as temperature and humidity in predictive models has proven essential for optimizing cooling systems and reducing energy waste (Katal et al., 2023). Studies have shown that models incorporating real-time data achieve significantly higher accuracy compared to those relying solely on historical datasets (Guo et al., 2022). For example, a study by (Bharadiya, 2023) demonstrated that real-time energy monitoring systems could reduce operational energy consumption by up to 12%. Similarly, dynamic models that continuously adapt to environmental changes have shown substantial gains in energy savings during off-peak hours (He et al., 2022). Despite its advantages, real-time data integration presents challenges, including the need for robust data collection systems and the computational power required to process large volumes of incoming data. Advanced ML algorithms such as Random Forest (RF) offer a solution by efficiently handling diverse and complex datasets while maintaining scalability (Yan et al. 2022).

E. Random Forest Algorithm in Energy Management

Random Forest (RF) has emerged as a preferred algorithm for energy management due to its ability to handle complex and high-dimensional data. RF operates by constructing multiple decision trees and aggregating their outputs, offering a balance between accuracy and computational efficiency (Zagajewski et al., 2021). Compared to simpler algorithms such as Decision Trees, RF provides greater robustness in managing real-time data variability (Maharana et al., 2022). In energy optimization, RF has been employed to predict energy consumption with high accuracy. For instance, (Awan et al., 2021) achieved a 10% reduction in energy costs using RF-based predictive models integrated with real-time data. The algorithm's flexibility in incorporating diverse variables, such as workload and ambient conditions, makes it particularly suited for energy management in data centers (Fan et al., 2022). However, implementing RF in real-world applications is not without challenges. The algorithm's reliance on a large number of trees increases computational demand, requiring significant processing power. Moreover, parameter tuning, such as selecting the optimal number of trees (n_estimators) or tree depth (max_depth), is crucial to avoid overfitting and ensure model stability (Yin et al., 2023).

F. Challenges in Integrating Machine Learning and Big Data for Energy Optimization

Integrating ML and Big Data for energy management poses several technical and operational challenges. One significant hurdle is ensuring data quality, as noisy or incomplete datasets can lead to inaccurate predictions and reduced model reliability (Munappy et al., 2022). Effective preprocessing techniques, such as normalization and outlier removal, are essential to maintain data integrity (Pfob et al., 2022). Another challenge is the computational infrastructure required for ML-based energy optimization. Processing large datasets and training complex algorithms like RF demand advanced hardware resources, including high-performance computing systems (Li et al., 2023). This requirement often limits the scalability of such solutions in smaller data centers or resource-constrained environments (Strielkowski et al., 2023). Additionally, concerns about data security and privacy add another layer of complexity. Sensitive information collected from data centers must be safeguarded against breaches, which can compromise both operational data and client trust. Implementing encryption protocols and ensuring compliance with data protection regulations are critical to addressing these issues (Rao et al., 2022).

III. RESEARCH METHOD

A. Research Design and Data Collection

This study employs an experimental, quantitative approach aimed at predicting energy consumption in data centers using the RF algorithm. The data utilized in this study was collected from an operational data center over six months and includes energy consumption, system workload, and environmental conditions, such as temperature. The dataset size reaches 1 TB, comprising approximately 3 million data points collected every five minutes. This granularity allows the research to capture variations throughout the day, including changes during peak and off-peak hours.

B. Data Quality and Pre-Processing

Data quality significantly influences the successful implementation of Machine Learning algorithms in energy management (Munappy et al., 2022). To ensure high-quality data, a preprocessing stage was carried out. This involved removing invalid entries, such as missing values and anomalies, and normalizing all features to a scale between 0 and 1 to optimize algorithm performance. Outliers were detected and eliminated using the z-score method, which helped minimize the impact of extreme values on the model. After preprocessing, the dataset was split into 80% training data and 20% testing data to enable robust model training and evaluation while avoiding overfitting.

C. Algorithm Selection and Configuration

The RF algorithm was chosen for its ability to handle large, complex datasets with diverse input variables, such as workload and temperature. RF constructs multiple decision trees and aggregates their outputs to produce stable and accurate predictions. For this study, the algorithm's hyperparameters were carefully tuned:

- Number of Trees (n_estimators): Set to 100, balancing predictive accuracy and processing time.
- Maximum Depth (max_depth): Set to 20 to prevent overfitting while capturing data complexity.
- Splitting Criterion: Gini impurity was selected over entropy for its superior performance in this context.
- Features Considered per Split (max_features): Restricted to half of the total features, enhancing model robustness and reducing overfitting risk.

The RF algorithm's resilience to high data variability and its efficiency in processing large datasets make it particularly well-suited for this study compared to simpler algorithms like Decision Trees or more resource-intensive methods like Neural Networks.

D. Model Validation and Evaluation Metrics

Model validation is essential to ensure the reliability and accuracy of predictions. This study employed three primary evaluation metrics:

- Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values, making it effective for identifying large errors.
- Root Mean Squared Error (RMSE): The square root of MSE, offering an interpretable measure in the same units as the target variable.

- R-squared (R^2): Indicates how well the model explains variability in energy consumption, with values closer to 1 reflecting higher accuracy.

The RF model achieved an MSE of 0.034, an RMSE of 0.184, and an R^2 of 0.87, demonstrating its effectiveness in explaining 87% of the variability in energy consumption data. These results validate the model's capability for accurate energy management in dynamic data center environments.

E. Testing and Results

After training, the model was tested on the 20% test data portion of the total dataset. The testing involved predicting energy consumption based on input variables, such as workload and temperature, and comparing the predictions to actual consumption values. The RF model's evaluation metrics confirmed its reliability and precision, highlighting its applicability for real-time energy management in data centers. This demonstrates the algorithm's robustness and effectiveness in optimizing energy consumption, particularly in complex and variable operational scenarios.

IV. RESULT AND DISCUSSION

This study successfully developed a predictive model using the Random Forest (RF) algorithm to minimize energy consumption in data centers. The results demonstrate the model's ability to significantly reduce energy consumption, particularly during off-peak hours, while maintaining system reliability.

A. Energy Consumption Comparison

The predictive model's effectiveness was evaluated by comparing energy consumption before and after its implementation across different operating times. The experiment was conducted using a dataset of energy consumption and workload data collected from the data center over six months. The predictive model results indicate a significant reduction in energy consumption, especially during off-peak hours. Table 1 provides a detailed comparison of energy consumption and savings achieved.

Table 1. Comparison of data center energy consumption before and after implementing the predictive model

Operating Time	Energy Consumption (kWh)		Energy Savings (kWh)	Savings Percentage (%)
	Before implementing the predictive model	After implementing the predictive model		
00:00 - 06:00	450	380	70	15.56
06:00 - 12:00	600	530	70	11.67
12:00 - 18:00	750	700	50	6.67
18:00 - 00:00	650	580	70	10.77
Total daily	2.450	2.190	260	10.61

The model achieved an average daily energy savings of 10.61%, equivalent to 260 kWh. The highest savings, 15.56%, occurred during off-peak hours (00:00–06:00) when workloads and cooling demands were lower. Conversely, during peak hours (12:00–18:00), the savings dropped to 6.67%, reflecting the challenges of optimizing energy consumption under high workloads.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 3 No. 3 December 2024
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 324-336

B. *Energy Savings Visualization*

The variation in energy savings across different operating times is visualized in Figure 1, emphasizing the model's effectiveness during off-peak hours and its performance challenges during peak periods. Figure 1 highlights the model's adaptability to real-time changes in workload and environmental conditions, making it particularly effective during periods of reduced demand.

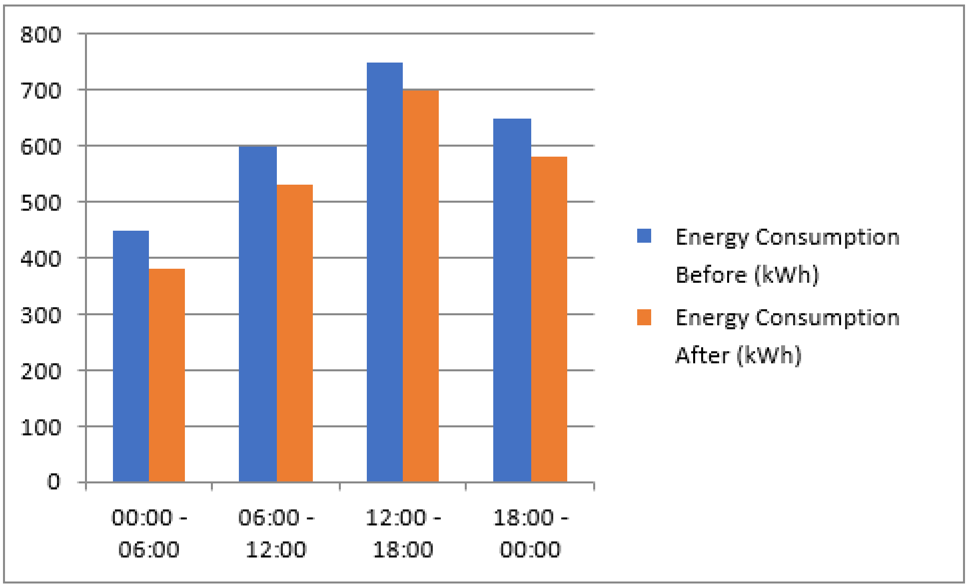

Figure 1. Energy Savings Across Operating Times

C. *Statistical Validation*

To assess the model's reliability, statistical tests were conducted. A 95% confidence interval analysis confirmed that:

- Predicted energy consumption during off-peak hours falls within ±2.5 kWh.
- Predicted energy consumption during peak hours falls within ±5 kWh.

This level of precision underscores the model's robustness and reliability in real-world applications, especially during periods of consistent workloads.

D. *Model Evaluation Metrics*

The model's performance was measured using key evaluation metrics, as presented in Table 2. An R^2 value of 0.87 indicates that the model can explain 87% of the variability in energy consumption based on input variables such as workload and temperature. The Adjusted R^2 value of 0.85 confirms the model's robustness, even when accounting for complexity, while the MAE of 0.15 kWh highlights minimal prediction error.

Table 2. Evaluation Metrics of the Predictive Model

Metric	Value
Mean Squared Error (MSE)	0.034
Root Mean Squared Error (RMSE)	0.184
R-squared (R^2)	0.87
Adjusted R^2	0.85
Mean Absolute Error (MAE)	0.15 kWh
Explained Variance Score (EVS)	0.87

E. Observations on Workload Influence

The model demonstrated varying levels of effectiveness based on workload conditions. Off-Peak Hours (00:00–06:00): The highest savings were achieved due to reduced workload variability and lower cooling requirements. Peak Hours (12:00–18:00): Despite accurate predictions, energy savings decreased due to increased cooling and computational demands.

Discussion

This study demonstrates that applying Machine Learning in data center energy management can result in substantial energy savings. The predictive model developed using the Random Forest (RF) algorithm achieved an average daily energy savings of 10.61%, with peak savings reaching 15.56% during off-peak hours. These findings align with prior research emphasizing the potential of predictive models to optimize energy consumption. For instance, (Awan et al., 2021) reported up to 10% energy savings through the integration of Machine Learning and real-time data in data centers. This study surpasses that benchmark, particularly during periods of reduced workloads when the data center operated below optimal capacity.

Previous studies, such as those by (Maharana et al., 2022), utilized simpler models like Decision Trees, which often struggled to handle dynamic real-time variables, including workload and temperature fluctuations. In contrast, this study employed RF, which has proven superior in managing complex, real-time data due to its ensemble approach and robustness against overfitting. The advantages of RF over algorithms such as Neural Networks and Support Vector Machines (SVM) have been well-documented, particularly regarding its ability to process large, diverse datasets efficiently (Zagajewski et al., 2021).

Additionally, the energy savings achieved in this study are comparable to findings by (Ouatik et al., 2022), who reported 15% improvement in predictive accuracy through the integration of Big Data and Machine Learning in energy management. However, unlike studies that primarily rely on historical data (Bharadiya, 2023), this research incorporates real-time environmental variables, such as temperature and workload. These variables enhanced the model's responsiveness and accuracy, enabling higher energy savings during off-peak hours when the system demand was lower. This real-time integration underscores the importance of adapting to dynamic environmental and operational conditions in data centers.

Despite its promising results, the implementation of this predictive model faces several challenges. The model's reliance on high-quality data means that incomplete or noisy datasets can undermine its performance. Effective data preprocessing techniques, such as normalization and outlier removal, are crucial to ensure reliable predictions. Furthermore, large-scale implementation demands significant computational resources, including advanced storage systems and processing power, which could pose barriers to adoption in smaller or resource-constrained data centers (Pramanik et al., 2022).

The novelty of this research lies in its real-time, data-driven approach to optimizing energy consumption with higher accuracy than previous studies. By leveraging real-time environmental and operational variables, such as temperature and workload, the RF-based model adapts effectively to dynamic conditions, offering a responsive solution to the challenges of energy management. Unlike earlier studies that focused solely on historical data or employed simpler algorithms like Decision Trees, this research combines Machine Learning with Big Data to deliver a predictive model that is both accurate and stable when handling large-scale data. These advancements contribute significantly to the literature on sustainable energy management and offer a framework for broader applications in other energy-intensive industries. This study highlights the potential for further research in enhancing predictive accuracy by incorporating additional real-time variables, such as humidity, server activity patterns, or renewable energy integration. It also underscores the importance of developing cost-effective computational infrastructure to make such solutions accessible to a wider range of data centers. These steps are critical for achieving more efficient and sustainable energy management practices on a global scale.

CONCLUSION AND RECOMMENDATION

Conclusion

This study successfully demonstrates that integrating Machine Learning with Big Data can significantly enhance energy efficiency in data centers. Using the RF algorithm, the developed predictive model managed to reduce energy consumption by up to 15.56% during specific hours without compromising system performance. This approach offers a practical solution for addressing the energy management challenges faced by modern data centers, especially considering rising operational costs and environmental impacts due to high energy consumption. A unique contribution of this study lies in its use of real-time data efficiently processed by Machine Learning algorithms, resulting in highly accurate energy consumption predictions. Consequently, this research makes a valuable contribution to the literature on data-driven energy optimization, while also supporting global initiatives to reduce carbon emissions. Furthermore, it enriches the discourse on integrating Big Data and Machine Learning in energy management and paves the way for broader applications in various industries reliant on large-scale computing systems. Despite the significant findings, the study faced limitations, such as dependence on the quality of collected data and the requirement for robust computing infrastructure. These limitations should be considered when applying this model on a larger scale or in different industries.

Recommendation

Based on the study's findings, several recommendations are suggested for further development. First, future research should focus on developing a more adaptive predictive model that can better accommodate the variability in data center workloads. Improved predictive accuracy could be achieved by integrating additional variables, such as user data usage patterns or more comprehensive environmental conditions. Additionally, future studies could test this model's application in data centers of varying sizes and characteristics to explore its flexibility and adaptability. Second, further research is needed on the development of interactive animated stickers that can be used on other digital platforms, such as educational games or health applications. More interactive stickers could enhance the effectiveness of conveyed messages, especially in the context of educational or health campaigns. This type of research is expected to support various public awareness programs on important issues through more engaging and communicative visual media.

REFERENCES

Abdel-Razek, S. A., Marie, H. S., Alshehri, A., & Elzeki, O. M. (2022). Energy Efficiency through the Implementation of an AI Model to Predict Room Occupancy Based on Thermal Comfort Parameters. *Sustainability*, *14*(13), 7734. https://doi.org/10.3390/su14137734

Awan, U., Shamim, S., Khan, Z., Zia, N. U., Shariq, S. M., & Khan, M. N. (2021). Big Data Analytics Capability and Decision-Making: The Role of Data-Driven Insight on Circular Economy Performance. *Technological Forecasting and Social Change*, *168*, 120766. https://doi.org/10.1016/j.techfore.2021.120766

Bharadiya, J. P. (2023). The Role of Machine Learning in Transforming Business Intelligence. *International Journal of Computing and Artificial Intelligence*, *4*(1), 16–24. https://doi.org/10.33545/27076571.2023.v4.i1a.60

Bharany, S., Sharma, S., Khalaf, O. I., Abdulsahib, G. M., Al Humaimeedy, A. S., Aldhyani, T. H. H., Maashi, M., & Alkahtani, H. (2022). A Systematic Survey on Energy-Efficient Techniques in Sustainable Cloud Computing. *Sustainability*, *14*(10), 6256. https://doi.org/10.3390/su14106256

Buyya, R., Ilager, S., & Arroba, P. (2024). Energy-Efficiency and Sustainability in New Generation Cloud Computing: A Vision and Directions for Integrated Management of Data Centre Resources and Workloads. *Software: Practice and Experience*, *54*(1), 24–38. https://doi.org/10.1002/spe.3248

Cao, Z., Zhou, X., Hu, H., Wang, Z., & Wen, Y. (2022). Toward a Systematic Survey for Carbon Neutral Data Centers. *IEEE Communications Surveys and Tutorials*, *24*(2), 895–936. https://doi.org/10.1109/comst.2022.3161275

Fan, G. F., Zhang, L. Z., Yu, M., Hong, W. C., & Dong, S. Q. (2022). Applications of Random Forest in Multivariable Response Surface for Short-Term Load Forecasting. *International Journal of Electrical Power and Energy Systems*, *139*, 108703. https://doi.org/10.1016/j.ijepes.2022.108073

González García, C., & Álvarez-Fernández, E. (2022). What Is (Not) Big Data Based on Its 7Vs Challenges: A Survey. *Big Data and Cognitive Computing*, *6*(4), 158. https://doi.org/10.3390/bdcc6040158

Guo, B., Yu, J., Yang, D., Leng, H., & Liao, B. (2022). Energy-Efficient Database Systems: A Systematic Survey. *ACM Computing Surveys*, *55*(6), 1–53. https://doi.org/10.1145/3538225

Hassani, H., & Silva, E. S. (2023). The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing*, *7*(2), 62. https://doi.org/10.3390/bdcc7020062

He, H., Shen, H., Hao, Q., & Tian, H. (2022). Online Delay-Guaranteed Workload Scheduling to Minimize Power Cost in Cloud Data Centers Using Renewable Energy. *Journal of Parallel and Distributed Computing*, *159*, 51–64. https://doi.org/10.1016/j.jpdc.2021.09.002

Karthikeyan, A., & Priyakumar, U. D. (2022). Artificial Intelligence: Machine Learning for Chemical Sciences. *Journal of Chemical Sciences*, *134*(1), 1–20. https://doi.org/10.1007/s12039-021-01995-2

Katal, A., Dahiya, S., & Choudhury, T. (2023). Energy Efficiency in Cloud Computing Data Centers: A Survey on Software Technologies. *Cluster Computing*, *26*(3), 1845–1875. https://doi.org/10.1007/s10586-022-03713-0

Khan, T., Tian, W., Ilager, S., & Buyya, R. (2022). Workload Forecasting and Energy State Estimation in Cloud Data Centres: ML-Centric Approach. *Future Generation Computer Systems*, *128*, 320–332. https://doi.org/10.1016/j.future.2021.10.019

Li, J., Herdem, M. S., Nathwani, J., & Wen, J. Z. (2023). Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management. *Energy and AI*, *11*, 100208. https://doi.org/10.1016/j.egyai.2022.100208

Maharana, K., Mondal, S., & Nemade, B. (2022). A Review: Data Pre-Processing and Data Augmentation Techniques. *Global Transitions Proceedings*, *3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Munappy, A. R., Bosch, J., Olsson, H. H., Arpteg, A., & Brinne, B. (2022). Data Management for Production Quality Deep Learning Models: Challenges and Solutions. *Journal of Systems and Software*, *191*, 111359. https://doi.org/10.1016/j.jss.2022.111359

Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting Student Success Using Big Data and Machine Learning Algorithms. *International Journal of Emerging Technologies in Learning*, *17*(12), 236–251. https://doi.org/10.3991/ijet.v17i12.30259

Panwar, S. S., Rauthan, M. M. S., & Barthwal, V. (2022). A Systematic Review on Effective Energy Utilization Management Strategies in Cloud Data Centers. *Journal of Cloud Computing*, *11*(1), 95. https://doi.org/10.1186/s13677-022-00368-5

Pfob, A., Lu, S.-C., & Sidey-Gibbons, C. (2022). Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Medical Research Methodology*, *22*(1), 282. https://doi.org/10.1186/s12874-022-01758-8

Pramanik, P. K. D., Pal, S., & Mukhopadhyay, M. (2022). Healthcare Big Data. In *Research Anthology on Big Data Analytics, Architectures, and Applications* (pp. 119–147). IGI Global. https://doi.org/10.4018/978-1-6684-3662-2.ch006

Rao, N. T., Neal Joshua, E. S., & Bhattacharyya, D. (2022). An Extensive Discussion on Utilization of Data Security and Big Data Models for Resolving Healthcare Problems. *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems*, *19*, 311–324. https://doi.org/10.1016/b978-0-323-90032-4.00001-8

Santoso, J. T., Wibowo, M. C., & Raharjo, B. (2024). Optimizing Deep Learning Efficiency through Algorithm-Hardware Co-design. *Journal of Advances in Information Technology*, *15*(10), 1163–1173. https://doi.org/10.12720/jait.15.10.1163-1173

Shabbir, N., Kütt, L., Jawad, M., Husev, O., Rehman, A. U., Gardezi, A. A., Shafiq, M., & Choi, J. G. (2022). Short-Term Wind Energy Forecasting Using Deep Learning-Based Predictive

Analytics. *Computers, Materials and Continua*, *72*(1), 1017–1033. https://doi.org/10.32604/cmc.2022.024576

Strielkowski, W., Vlasov, A., Selivanov, K., Muraviev, K., & Shakhnov, V. (2023). Prospects and Challenges of the Machine Learning and Data-Driven Methods for the Predictive Analysis of Power Systems: A Review. *Energies*, *16*(10), 4025. https://doi.org/10.3390/en16104025

Yin, L., Li, B., Li, P., & Zhang, R. (2023). Research on Stock Trend Prediction Method Based on Optimized Random Forest. *CAAI Transactions on Intelligence Technology*, *8*(1), 274–284. https://doi.org/10.1049/cit2.12067

Zagajewski, B., Kluczek, M., Raczko, E., Njegovec, A., Dabija, A., & Kycko, M. (2021). Comparison of Random Forest, Support Vector Machines, and Neural Networks for Post-Disaster Forest Species Mapping of the Krkonoše/Karkonosze Transboundary Biosphere Reserve. *Remote Sensing*, *13*(13), 2581. https://doi.org/10.3390/rs13132581

Zhang, Y., & Liu, J. (2022). Prediction of Overall Energy Consumption of Data Centers in Different Locations. *Sensors*, *22*(10), 3704. https://doi.org/10.3390/s22103704