International Journal of Scientific Research and Engineering Development-– Volume X Issue X, Year

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

ALGORITHM PERFORMANCE EVALUATION FOR DATA CENTER ENERGY CONSUMPTION FORECAST

Olawale J. Omotosho*, Chidiebere Enyinnah **, Samson O. Ogunlere***

*(Computer Science Department, Babcock University, Ilishan-Remo, Ogun State, Nigeria)
Email: omotoshoo@babcock.edu.ng

** (Computer Science Department, Babcock University, Ilishan-Remo, Ogun State, Nigeria)
Email: ebereiwobi@gmail.com)

*** (Computer Science Department, Babcock University, Ilishan-Remo, Ogun State, Nigeria)
Email: ogunleres@babcock.edu.ng)

--************************----------------------------------

Abstract

One major problem data centres face is energy management. Energy bill makes the data centre business very expensive to run. Prior knowledge of the energy usage of a data centre is critical in making more informed decision. A comparative analysis of algorithm from different domain will give more insight into the usability of the algorithms in forecasting data centre energy consumption.

Dataset was generated from EnergyPlus simulation platform based on the information from a live data centre. The algorithms considered were ARIMA, SVR and LSTM. These algorithms were compared to determine an optimal algorithm using five (5) performance evaluation metrics and the metrics are MSE, RMSE, MAE, MAPE and Accuracy.

LSTM has the lowest MSE value of 0.016 while SVR and ARIMA had 0.49 and 0.874 respectively. LSTM is the best performing algorithm based on RMSE metric with a value of 0.126 while SVR and ARIMA had 0.220 and 0.935 respectively. For MAE metrics, ARIMA has a score of 0.623, SVR has a score of 0.071 and LSTM has a score of 0.036. Therefore, LSTM has the best performance, followed by SVR and then lastly ARIMA. For MAPE, ARIMA has the value of 0.44, SVR has the value of 0.39 and LSTM has a value of 0.12. Therefore, LSTM has the best performance based on MAPE metrics, followed by SVR and ARIMA. ARIMA has an accuracy of 56%, SVR has accuracy value of 61% and LSTM has an accuracy value of 88%. Therefore, based on accuracy, LSTM is the best algorithm followed by SVR and then ARIMA. The LSTM algorithm was considered to be the best performing algorithm as it had the highest accuracy value and lowest error values.

Keywords — Energy Consumption, Data Centre, Machine Learning, Algorithms, performance evaluation--
************************----------------------------------

1.0 INTRODUCTION

Data centres are powerhouses of most organizations; they are the lifeline that the keeps the world of digitalization moving [7]. Most businesses have gone global, with data being as often as possible used and shared. Due to this, data has gotten to be one of an organization's valuable assets and it must be well protected. Organizations depend greatly on data centres [14].

One major problem data centres encounter is energy management. Energy management is particularly critical for data centres as energy cost is on continuous rise and a good energy management depends on detailed energy usage information [8].

The amount of energy we use keep growing and with that comes greenhouse gas emission [5]. Energy efficiency has become an important concern in the design and management of data centres. Data centre administrators in their daily operations need to understand the power usage patterns to maximize their energy efficiency. Physical power estimation only is not sufficient, as they cannot forecast future power usage. Power consumption forecast plays a vital role as it gives more insight on the factors that causes energy demand and provides the basis for better decision-making in power system planning and operation.

In a data centre, electrical power is used to operate Information and Communications Technology (ICT) equipment and its supporting facilities. [15] noted that about 50% of electrical energy is consumed by IT equipment, which includes servers, storages and networks. Cooling takes about 37% while the remaining power is used by other facilities including Power Distribution System (PDS), Uninterruptible Power Supply (UPS), lights and so on.

These data centres' energy usage and costs are becoming unacceptably high, and placing a great burden on both energy resources and the environment [16].

Gartner estimates that ongoing power costs increase at least 10 percent every year because of cost per kilowatt-hour (kWh) increments in underlying demand, particularly for high-power density servers [10]. Energy represents a large fraction of the operational expenses of a data center, and data center operators are increasingly keen on pushing towards low-power designs [3]. Even with innovative advancements in sustainable energy solutions, the truth is that both big and small data centres use a lot of energy [13].

It is of great importance to achieve the forecast of data centre energy consumption. This study intends to compare the forecasting algorithms to determine the optimal algorithm for building a data centre forecasting model and tool, that will help reduce data centre energy consumption, and in turn, reduce operational cost and environmental impact.

A number of approaches, as reported in literatures, have been applied in forecasting data centre energy consumption. With the ongoing advancements in the use of machine learning for forecasting, a number of studies have considered using the machine learning approach in predicting the power consumption of data centres. However, the studies from literature have used either deep learning, neural network or other machine learning approaches and none to the best of our knowledge evaluates the performance of algorithms for data centre energy consumption from different domain. Therefore, comparing a statistical algorithm like Autoregressive Integrated Moving Average (ARIMA) with machine learning algorithm like Support Vector Regression (SVR) and Deep learning Long Short Term Memory (LSTM) sequence algorithm will offer more insights into the usability of these algorithms.

Methodology Overview

In order to accomplish the stated objective, the following methods were employed:

Data centre simulation was built using EnergyPlus simulation software. Information was gathered from one of the leading data centre located in Lagos, Nigeria. The information from the data centre informed the simulation.

The selected algorithms are ARIMA time series algorithm, Support Vector Regression forecasting algorithm and Deep learning Long-short term memory (LSTM) sequence algorithm. ARIMA was considered because this study is using Time Series dataset and ARIMA is a time series algorithm to predict future trends. It is very popular with forecasting problems. Support Vector Regression is a good machine learning algorithm with good reports from literatures while LSTM is a good algorithm that can be used as forecasting algorithm when talking about deep learning because of its recurrent architecture and ability to solve the long term dependency problem.

A comparative analysis of the three algorithms specified above was carried out to determine the optimal algorithm. The performance of the algorithms was evaluated based on five performance metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) Mean Absolute Percentage Error (MAPE) and Accuracy.

2.0 REVIEW OF LITERATURE

[2] carried out a survey of techniques for Data centres' energy consumption modelling and energy demand forecasting. They discovered that IT equipment and facility infrastructure majorly consume energy. They also classified known existing machine learning techniques for forecasting data centre energy usage. They concluded that most of the studies focused mostly on energy effectiveness of lower hardware levels of data centres and less on higher aggregated levels which is needed for demand response programs. [21] focused on using control knobs that modulate the power consumption of I.T equipment in a data centre. The study proposed a hierarchical optimization framework and a suite of algorithms for the upper layer of the hierarchical framework. Both Empirical and experimental approaches were adopted for the study. Employing control knobs for data centre utility bill optimization is complex and remains difficult. Moreso, [24] came up with a power optimization strategy (PowerNets) that leverages workload correlation analysis to jointly minimize the total power consumption of servers and data centre networks. The network topology used in this experiment was fat-tree topology and this makes it difficult to generalize. [1] identified that there was a problem with high energy consumption which has led to high operational cost and low efficiency of data centres. The objectives were to evaluate the benefit of green energy prediction for increasing the data centre job throughput while not sacrificing service jobs response time constraints and also, to reduce the peak power level of data centres to decrease the utility bill. Event-based simulation was used and the study proposed approaches to energy efficiency problem of data centres from multiple dimensions that are complementary to each other. [18] proposed an optimization algorithm called Data Centre-wide Energy-Efficient Resource Scheduling (DCEERS) for efficient resources scheduling in data centres. The DCEERS considers the current workload in the data centre before performing scheduling. The whole idea of the algorithm is to be sure that the amount of resources usage at a particular time is the minimum number of resources required. The

challenge is to find the minimum amount of resources required at a particular point referred to as the Minimum Cost Multi Commodity Flow (MCMCF). The study therefore, applied Benders Algorithm the estimate MCMCF. [23] proposed a Hierarchical Energy Optimization (HERO) model aimed at lowering the algorithm complexity and verifying several heuristics algorithms to achieve reduced energy usage. Given a DCN topology and a traffic matrix, the possibility of turning off some network elements (i.e., routers, switches, and links) hierarchically was evaluated without violating the network connectivity and QoS constraints. Putting some idle components of data center off is good for saving energy and cost but it also comes with its challenges. [6] applied Ant Colony System (ACS) algorithm to tackle the problem of unnecessary energy consumption by idle virtual machines which in some cases accounts for 50%-70% of the total server energy.

[17] proposed Holt-Winter forecasting algorithm to tackle the issue of forecasting the energy usage of future data centres. The Holt-Winter algorithm that is based on time series uses seasonality to make predictions. The proposed approach was compared with the traditional approach being used currently in literature in order to give the optimal forecasts. The result shows that the proposed approach outperforms similar algorithms proposed in literature, with highly variable workloads which is particularly of interest for future data centres. Nevertheless, a full comparison against well established forecasting algorithms will determine the generalization of this method.

[11] approached energy efficiency and infrastructural performance of data centres based on the accuracy of efficiently mapping requests with the physical infrastructure using virtual machine algorithm. The algorithm combined greedy strategy with limited search so as to assign virtual machines to computers nodes and physical data storage systems to storage servers and map virtual machines and virtual data storage physical resources obtained.

[9] followed three stages in addressing data center effectiveness: Stackelberg bargaining approach, adoption of Modified Thomson Bargaining Solution and conduction of simulation analysis. The Studied employed RKSBS (Raiffa Kalai Smorodinsky Bargaining Solution) and MTBS. These methods were used to determining the best dynamic energy pricing strategy and simulation analysis to verify the superiority of the first two phases whose results proved effective and feasible.

3.0 PERFORMANCE EVALUATION OF ALGORITHMS

Data center simulation was built with EnergyPlus simulation software. Information like the temperature maintained (for different weather conditions), list of equipment, type and number of cooling systems, the total energy consumed monthly etc. were gotten from a real data center and this informed the simulation. The energy consumption log file generated from the data center simulation was split based on the 80:20 hold out rule.

Performance evaluation of the forecasting algorithms was carried out to determine the optimal algorithm for data centre energy consumption forecast. The algorithms are Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR) and Long Short Term Memory (LSTM). The choice of these algorithms were based on type of data available for this study and how well they perform. ARIMA was considered because the dataset for this study is a time Series dataset and ARIMA is a traditional statistical time series algorithm used to predict future trends. It is very popular with forecasting problems. Support Vector Regression is one of the machine learning algorithm used to fit data for forecasting and also has good reports from literatures. Deep learning is highly valuable in the energy sector as the algorithms is suited for large dataset and LSTM is a recurrent deep learning model, which means it has the ability to remember what has happened previously and therefore makes it a usable algorithm for energy forecast.

A comparative analysis of the algorithms was carried out to determine the optimal algorithm. The performance of the algorithms was calculated based on five performance metrics:

1. Mean Squared Error (MSE)
2. Mean Absolute Error (MAE)
3. Root Mean Squared Error (RMSE)
4. Mean Absolute Percentage Error (MAPE)
5. Accuracy.

MAE is calculated as:

$$E = \frac{1}{n}\sum|x - \hat{x}| \qquad \text{Equation (1)}$$

Where x is the value from the testing data, $\hat{x}$ is the value derived from the model and n is the total count of validation data.

RMSE on the other hand is calculated as below:

$$E = \sqrt{\frac{1}{n}\sum(x - \hat{x})^2} \qquad \text{Equation (2)}$$

MSE is calculated as :

$$E = \frac{1}{n}\sum(x - \hat{x})^2 \qquad \text{Equation (3)}$$

MAPE is calculated as :

$$E = \frac{1}{n}\sum|x - \hat{x}|/x \qquad \text{Equation (4)}$$

A lower value of MAE, MSE, MAPE and RMSE will equate a better model.

Accuracy on the other hand is calculated as

$$\text{Accuracy} = (1 - \text{MAPE}) \times 100 \qquad \text{Equation (5)}$$

Unlike other metrics, a higher accuracy value indicate a better model.

Algorithm comparison using MSE

MSE estimates the error between the actual test value and the predicted value. The lower the value of MSE the better the performance of the algorithm. Table 1 shows that LSTM has the lowest MSE value of 0.016. Therefore, based on MSE metrics LSTM performs better than the other two algorithms and ARIMA has the worst performance out of the three algorithms.

TABLE 1: Comparison Table for MSE Metrics

Algorithm	MSE score
ARIMA	0.874
SVR	0.049
LSTM	0.016

Algorithm comparison using RMSE

RMSE is the square root of MSE value, and it tries to remove the effect of the square operation on the MSE value. Like the case of MSE, the lower the RMSE value the better the performance of the algorithm. Table 2 shows the RMSE score for each of the algorithms. From the table, LSTM has the lowest RMSE value; therefore, LSTM is the best performing algorithm based on RMSE metric with a value of 0.126. The second best algorithm is SVR with a value 0f 0.220 while the worst performing algorithm is ARIMA with a score of 0.935.

TABLE 2 Comparison Table for RMSE Metrics

Algorithm	RMSE Score
ARIMA	0.935
SVR	0.220
LSTM	0.126

Algorithm comparison using MAE

MAE metric measures the absolute error value between the predicted value and the actual value. A low value indicate a minimal error between the predicted and the actual value and therefore indicate a better performance. ARIMA has a score of 0.623, SVR has a score of 0.071 and LSTM has a score of 0.036. Therefore, LSTM has the best performance, followed by SVR and then lastly ARIMA.

TABLE 3 Comparison Table for MAE Metrics

Algorithm	MAE Score
ARIMA	0.623
SVR	0.071
LSTM	0.036

Algorithm comparison using MAPE

MAPE measures the absolute ratio of error between predicted and actual value where error is the difference between the predicted and actual value. Like all the previously mentioned metrics, a low value indicate a better algorithm. ARIMA has the value of 0.44, SVR has the value of 0.39 and LSTM has a value of 0.12. Therefore, LSTM has the best performance based on MAPE metrics, followed by SVR and ARIMA.

TABLE 4 Comparison Table for MAPE Metrics

Algorithm	MAPE Score
ARIMA	0.44
SVR	0.39
LSTM	0.12

Algorithm comparison using ACCURACY

Accuracy measures how well the algorithm can predict the value. However since this is not a classification problem. Accuracy is calculated as (1-MAPE) *100 and unlike other metrics, a higher value indicate a good algorithm. ARIMA has an accuracy of 56%, SVR has accuracy value of 61% and LSTM has an accuracy value of 88%. Therefore, based on accuracy, LSTM is the best algorithm followed by SVR and then ARIMA.

TABLE 5 Comparison Table for Accuracy

Algorithm	Accuracy (%)
ARIMA	56%
SVR	**61%**
LSTM	88%

Algorithm comparison summary

Table 6 shows the summary of the algorithms based on the five metrics. LSTM performed better based on the five metrics. Therefore, LSTM was selected as the best performing algorithm.

Table 6 Metrics value for the three Algorithms

	ARIMA	SVR	LSTM
MSE	0.874	0.049	0.016
RMSE	0.935	0.220	0.126
MAE	0.623	0.071	0.036
MAPE	**0.44**	0.39	0.12
ACCURACY (1 - MAPE)	**56%**	61%	88%

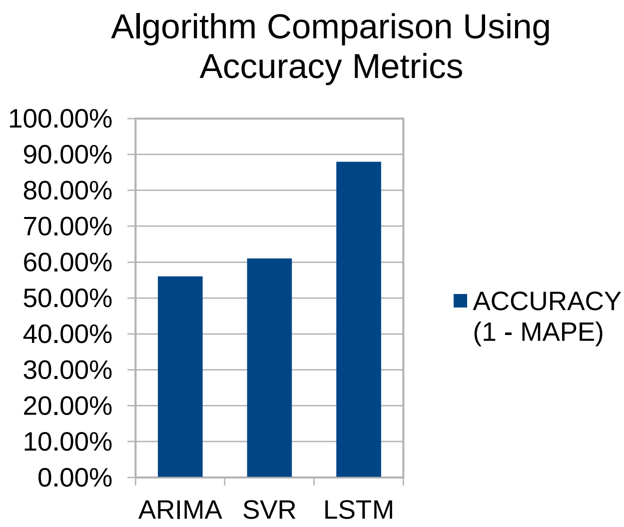

Fig. 1: Graph showing the performance of the three algorithm.

4. 0 FUTURE WORK

Future work will include developing a model based on the optimal algorithm which in this case, is LSTM algorithm. The model that will be developed will further be utilized to produce a data center energy forecast tool. The forecast model/tool will be evaluated with existing models/tools.

5.0 CONCLUSION

To determine the optimal algorithm, the test data was used. Some performance metrics were used to compare the algorithms. The algorithms considered were ARIMA, SVR and LSTM. The performance metrics used for the evaluation are MSE, RMSE, MAE, MAPE and Accuracy. LSTM recorded the highest accuracy and lowest error values followed by SVR and ARIMA respectively. The result from this study gives more insight into the usability of the selected algorithms for energy forecast in a data center.

REFERENCES

[1] Aksanli, B. (2015). *Energy and cost-efficient data centers.* (Doctoral dissertation, University of California, USA). Available online at http://engineering.sdsu.edu/~aksanli/papers/Thesis_Aksanli.pdf accessed on 5th, June 2020

[2] Dayarathna, M., Wen, Y., Member, S., & Fan, R. (2016). Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, *18*(1), 732–794.

[3] Ganesh, L., Weatherspoon, H., Marian, T., & Birman, K. (2013). Integrated approach to data center power management. *IEEE Transactions on Computers, 62(6)*, 1086-1096.

[4] Ghribi, C., Hadji, M., & Zeghlache, D. (2013). Energy efficient VM scheduling for cloud data centers: Exact allocation and migration algorithms. *13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, 671-678.*

[5] Guillard, E. (2020). Forecasting energy consumption using machine learning and AI. Available online at https://www.dexma.com/blog-en/forecasting-energy-consumption-using-machine-learning-and-ai/ accessed on 17th, January 2022.

[6] Hossain S.K.M, Ema .S.A & Sohn H, (2020) Rule-Based Classification Based on Ant Colony Optimization: A Comprehensive Review. *Applied Computational Intelligence and Soft Computing. https://doi.org/10.1155/2022/2232000*

[7] Jie Yee, O. (2020). What are Data Centers and why are they important? Available online at https://www.makeuseof.com/what-are-data-centers-and-why-are-they-important/

[8] Kelley, M. (2019). Energy Management for Data Centers. Available online at https://www.setra.com/blog/datacenters#:~:text=Energy%20management%20is%20especially%20crucial,energy%20consumption%20and%20daily%20usage.

[9] Kim S. (2020) Adaptive Data Center Management Algorithm Based On The Cooperative Game Approach. *IEEE Access*, 3461-3470. DOI: 10.1109/access.2020.3047887

[10] Kleyman, B. (2018). Be aware of these 5 data center trends in 2018. Available online at https://www.datacenterknowledge.com/manage/be-aware-these-5-data-center-trends-2018,

[11] Kostenko V. & Chupakhin A. (2020) Live Migration Schemes in Data Centers. Programming and Computing Software, 46(5), 312-315. https://doi.org/10.1134/S0361768820050035

[12] Li, Y., Wen, Y., Guan, K., & Tao, D. (2018). Transforming cooling optimization for green data center via deep reinforcement learning. Available online at https://arxiv.org/pdf/1709.05077.pdf,

[13] Marashi, A. (2020). How to improve data center power consumption & energy efficiency. Available online at https://www.vxchnge.com/blog/power-hungry-the-growing-energy-demands-of-data-centers

[14] Mishra, R. (2021) Importance of Data Centers. Available online at https://medium.com/@htshosting/importance-of-data-centers-827fd46ffc5d

[15] Mukaffia, A., Ariefa, R., Hendradjita, W., & Romadhona, R. (2017). Optimization of cooling system for data center. *Procedia Engineering 170*, 552-557.

[16] Rahman, A., Liu, X., & Kong, F. (2014). A survey on geographic load balancing-based data center power management in the smart grid environment. *IEEE Communications Surveys & Tutorials*, *16*(1), 214-233.

[17] Rossi, M., & Brunelli, D. (2015). Forecasting data centers power consumption with the Holt-Winters method. *2015 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS) Proceedings*, 210-214. doi:10.1109/eesms.2015.7175879

[18] Shuja, J., Bilal, K., Madani, S. A., & Khan, S. U. (2014). Data center energy efficient resource scheduling. *Cluster Computing*, *17*(4), 1265–1277.

[19] Shuja, J., Bilal, K., Madani, S. A., Othman, M., Ranjan, R., Balaji, P., & Khan, S. U. (2016). Survey of techniques and architectures for designing energy-efficient data centers. *IEEE Systems Journal*, *10*(2), 507–519.

[20] Swetha, P (2018) A comparison of energy efficient adaptation algorithms in cloud data centres. *Blekinge Institute of Technology, Faculty of Computing, Department of Computer Science and Engineering*

[21] Wang C., Urgaonkar B., Wang Q. & Kesidis G. (2013). A hierarchical demand response framework for data center power cost optimization under real-world electricity pricing. *Proceedings - IEEE Computer Society's Annual International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunications Systems, MASCOTS*, 305-314.

[22] Wiboonrat, M. (2020). Energy Management in Data Centers from Design to Operations and Maintenance. *IEEE International Conference and Utility Exhibition on Energy, Environment and Climate Change (ICUE)*

[23] Zhang, Y., Member, S., & Ansari, N. (2015). HERO : Hierarchical Energy Optimization for Data Center Networks. *IEEE Systems Journal, 9(2),* 406–415.

[24] Zheng, K., Zheng, W., Li, L., & Wang, X. (2016). PowerNetS: Coordinating data center network with servers and cooling for power optimization. *IEEE Transactions on Network and Service Management, 14*(3), 661-67