Technologies for Energy, Agriculture, and Healthcare – Shailesh Nikam et al. (eds)

16

Energy Prediction for Efficient Resource Management in IoT-enabled Data Centres

Sarika Mane[1]
Ph. D. Scholar,
K. J. Somaiya College of Engineering,
Somaiya Vidyavihar University,
Vidyavihar, Mumbai
K. J. Somaiya Institute of Technology,
University of Mumbai

Makarand Kulkarni[2], **Sudha Gupta** [3]
K. J. Somaiya College of Engineering,
Somaiya Vidyavihar University,
Vidyavihar, Mumbai

Abstract: Internet of Things (IoT) enabled Data Centres (DC), play a vital role in managing and sustaining modern information-driven infrastructure. Accurate energy prediction is the major requirement of the DC for efficient resource management, management of growing internet infrastructure and increasing demand of digital services. The main challenges in the DC are scalability and the cost effectiveness which are dependent on the accurate energy prediction. Hence there is a need of accurate energy prediction. The work proposes energy prediction model with best agreement between predicted and actual values resulting to approximately zero error and robustness for an IoT enabled DC. The feature normalization concept has been used in energy prediction model to enhance the robustness of the different regression models. Proposed work is validated by comparison with the earlier reported work on energy prediction. Robust Linear Regression-Random Sample Consensus (RLR-RANSAC) and Linear Regression (LR) exhibited remarkable performance with RMSE 6.8176×10^{-17} (KWh),

[1]sarika@somaiya.edu, [2]makarandkulkarni@somaiya.edu, [3]sudhagupta@somaiya.edu

DOI: 10.1201/9781003596707-16

7.9565×10^{-17} (KWh) for hourly time span respectively, as compared with earlier reported work. R-squared (R^2) values approaching 1 indicated a near-perfect fit to the data. The proposed approach demonstrated overall performance improvement and can be applicable in IoT enabled DC environment.

Keywords: Data centre, Energy prediction, Energy efficiency, Energy efficient resource management, Internet of things (IoT), Regression models

1. Introduction

Data Centres (DC) are playing the important role to our nation's energy and information infrastructure. Data centres consume 100x powers compared with large commercial building. 40% powers come from energy needed and cooling requirements. As per (Landrum, 2020), worldwide there are 7,500+ datacetres,7.75M servers are installed per year. On an average one data centre uses the equivalent power of 25,000 homes. Such high consumption of the energy demands the proper resource management to manage growing internet user base and increasing demand of digital services DC. For the efficient and sustainable operations of the data centre facilities, energy management and energy prediction are the closely related aspects. Proactive and strategic approach is required for energy prediction. Known future requirements of the energy consumption can help proper energy management. The work by (Ajayi & Heymann, 2021) reported day ahead energy demand prediction using Artificial Neural Network (ANN). (Hsieh et al., 2020) utilized prediction aware virtual machine consolidation strategy for cloud data centres that use less energy. Neural Network (NN) is used for assessment of cooling power performance of the data centre (Shrivastava et al., 2010).The work (Berezovskaya et al., 2020) proposed modelling toolbox to model the soft datacentre using set of building box. This model is capable to examining the performance and energy saving strategies in dynamic mode. Earliest deadline first based energy aware fault tolerant scheduling using AI driven approaches (Marahatta et al., 2021) for cloud DC was designed. Deep Reinforcement Learning (DRL) was used on real data for the cooling optimization and achieved 15% cooling energy savings and 11% cooling cost reduction (Y. Li et al., 2020). (Gao et al., 2020) used deep learning to develop intelligent solutions that enable a cloud DC while effectively handling the unpredictability of renewable energy. Maximized energy efficiency in cloud data centres for virtual machine (VM) consolidation used strong linear regression prediction(L. Li et al., 2019). To understand the DC electricity need, system dynamic model designed for energy prediction from 2016 till 2030. The work by the (Peoples et al., 2011) is focused on optimization of carbon emission associated with the energy consumption. Power management evaluation strategies (Postema & Haverkort, 2018), system wise utilization reduction (Chisca et al., 2015) are observed in the respective mentioned work. Many researchers worked on various techniques, strategies and Artificial Intelligence (AI) algorithm's implementation. In the existing work it is observed that, emphasis on robustness is missing. This motivated to do work for accurate energy prediction along with robustness integrated AI algorithms.

The notable contributions of the proposed work presented in this paper are as follows:

1. Normalization for enhanced accuracy resulting to approximately zero error and robustness has been achieved in this work
2. Energy prediction for IoT-enabled DC using integrated feature normalization in different regression models such as RLR-RANSAC, LR, and RR.
3. A comprehensive enhancement in overall performance has been presented.

The structure of the paper is organized as follows; proposed methodology is presented in section II, succeeded by the results and discussion in section III, and Conclusions in section IV.

2. Proposed Methodology

Implementation of IoT leverages data acquisition in cloud through gateway. By deploying interconnected sensors and the devices data acquisition can be done for the various parameters such as Power (W), Active Energy (KWh), CPU power consumption (%), GPU power consumption (%), etc. IoT facilitates accurate energy predictions through monitoring these parameters. Figure 16.1 adapted from ("The Impact of IoT on the Data Centre Sector," 2017) has depicted the implementation of an IoT scenario in DC for energy prediction. To carry out the energy predictions we used the data server energy consumption dataset (Asanza, 2021) with 184425 entries (0 to 184424) and Data columns (total 16 features) To understand the data in better manner Exploratory Data Analysis (EDA) is carried out. Through heat map correlation analysis is done. Strong correlation is observed between power factor and power (0.76); also power consumption of CPU (0.68), GPU (0.69) and RAM (0.61) are highly correlated with Active Energy (KWh). Active Energy (KWh) refers to the actual energy consumed is a key metric in energy prediction and efficiency in IoT application.

The proposed methodology is as shown in the Fig. 16.2, using different regression models namely A RLR-RANSAC, LR, RR and Lasso regression. Preprocessing was done to remove unnecessary columns. In the data preprocessing averaging and normalization is used. Averaging is used for the aggregation of the data.

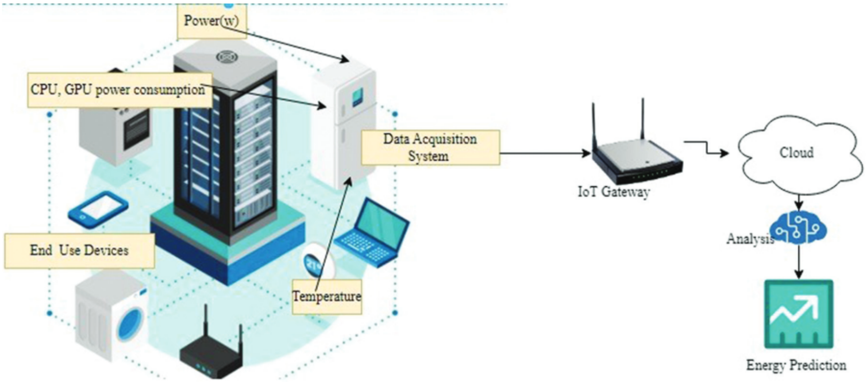

Fig. 16.1 IoT implementation in data centre for energy prediction

Source: Adapted from "The Impact of IoT on the Data Centre Sector." The Irish Advantage, 21 Dec. 2017, www.irishadvantage.com/the-impact-of-iot-on-the-data-centre-sector

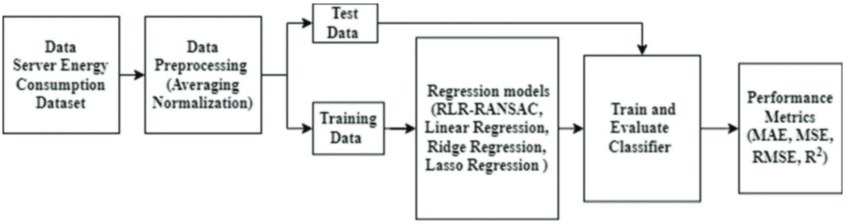

Fig. 16.2 Proposed methodology implementation using various regression models

Source: Author

Normalization reduced biases towards larger-scaled features and promotes numerical stability by ensuring that each feature contributes equally to the training of the regression model. Standard Scalar normalization based on the z score normalization is used. Transformed featuers have mean of 0 and a standard deviation of 1, is given by the equation (1).

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Where, z is the standardized value, x is the original value of the feature, μ is the mean of the feature and σ is the standard deviation of the feature. Standardization helps to address the challenges related to the scale discrepancies, making the models more accurate, robust, and generalizable. The preprocessed data was then split into training and testing sets and further implementation as shown as in Fig. 16.2.

One of the regression models, RLR-RANSAC is used. RANSAC is a technique used to enhance the robustness of the linear models, specifically in the presense of the outliears. It randomly samples a subset of data, fits a model identifies the inliears based on a predefined threshhold. After repetative process the model with the largest inliers is selected as the final model. Compared to classic linear regression, RLR-RANSAC is less sensitive to data outliers. By robustly fitting a linear regression model, RANSAC improves accuracy in energy prediction. RR's L2 regularization term helps to prevent overfitting and stabilizes the model by preventing excessively large coefficients. Incorporation of the L1 regularization, helps in selecting the subset of the most important features. It lead to a simple and more interpretable model. Thus emphasized essential featuers are important in the enhancement of the accuracy.

For the evaluation of regression models different performance metrics are used such as Mean Absolute Error(MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and R-squared (R^2) as per equation (2) and (3) resepectively (Zwingmann, 2022). In reported work RMSE is used for performance measure.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \tag{3}$$

Where, n is number of data points, y_i is actual (observed) value for ith data point, $\hat{y}_i$ is predicted value for ith data point and $\overline{y}_i$ is the mean of the observed values y_i. Thus feature normalization concept implemented in energy prediction to enhance the robustness of the different regression models. The effectiveness of the implementation is evaluated using comprehensive set of metrics. The following subsection provides the detailed analysis of results.

3. Results and Discussion

Table 16.1, presents different model's comparative analysis over a monthly span and based on various performance metrics. The models are evaluated based on their ability to predict energy consumption in terms of KWh. R^2 represents the proportion of variance between active energy and the time span; the values approaching to 1 represented the better fit.

From Table 16.1, it is demonstrated by both RLR-RANSAC and Linear Regression (LR), with minimal MAE, MSE, and RMSE. Their R-squared values of 1 indicated

Table 16.1 Comparison of different models using various performance metrics

Model	Time Span	MAE (KWh)	MSE $(KWh)^2$	RMSE (KWh)	R^2
RLR-RANSAC	Month	3.7977×10^{-14}	2.4095×10^{-27}	4.9087×10^{-14}	1
LR	Month	4.6653×10^{-14}	3.2818×10^{-27}	5.7287×10^{-14}	1
RR	Month	1.0447	1.8360	1.3550	0.9997
Lasso	Month	0.83263	0.93124	0.9650	0.9998

Source: Author

a perfect fit. It is also observed both could able to identify the underlying patterns in the data. These models are well-known for being straightforward and effective and they make strong arguments for precise monthly energy projections. RR and Lasso Regressions introduced regularization to the models, influencing their performance metrics. Although the error rates of both models remain remarkably low, Lasso performs marginally better than RR in terms of MAE, MSE, and RMSE. These models' regularization methods, whose R-squared values are nearly one, help produce predictions that are stable.

From Table 16.2, RLR-RANSAC algorithm, showed improved forecast accuracy in energy usage. The reported work showed RMSE of 0.0018(KWh) for hourly forecasts and 0.01350(KWh) for daily predictions for LR-Robust linear. Robust Linear -RANSAC performed well, showed RMSE of 6.8176 x10-17 (KWh) for hourly forecasts and 1.6362×10^{-15} (KWh) for daily predictions. Linear Regression also performed well, showed RMSE of 7.9565×10^{-17} (KWh) for hourly forecasts and 1.9095×10^{-15} (KWh) for daily predictions. In comparison with the reported work all error values are less and approximately zero, have resulted enhanced accuracy.

Table 16.2 Comparison of RMSE performance with reported work

Model	Time Span	RMSE (KWh)
LR-Robust linear (Estrada et al., 2022)	Hour	0.0018
	Day	0.01350
RLR-RANSAC (Proposed Approach)	Hour	6.8176×10^{-17}
	Day	1.6362×10^{-15}
	Month	4.9087×10^{-14}
LR (Proposed Approach)	Hour	7.9565×10^{-17}
	Day	1.9095×10^{-15}
	Month	5.7287×10^{-14}

Source: Author

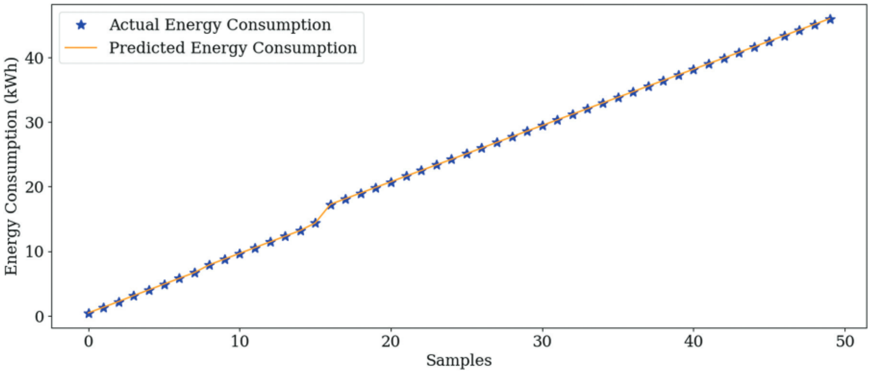

Fig. 16.3 Energy prediction using RLR-RANSAC

Source: Author

Energy Prediction using RLR-RANSAC is as shown in Fig. 16.3. Figure 16.3 showed a nearly perfect fit. Error is approximately 0 so the actual energy consumption is superimposed by predicted energy consumption. But it may also raises concerns over potential over fitting in real time implementation, evaluating the model on new unseen data.

4. Conclusions

The proposed approach with normalization integration implemented with regression models for robustness and accuracy enhancement. The AI plays significant role in addressing energy prediction related aspects. RLR-RANSAC and LR exhibited remarkable performance with RMSE 6.8176×10^{-17} (KWh)' 7.9565×10^{-17} (KWh) for hourly time span respectively, as compared with earlier reported work. In comparison with other models performance RR (1.3550 (KWh)) and Lasso (0.9650 (KWh)) performed well for monthly time span respectively. RMSE values close to the 0 indicated that the models predictions are very accurate and almost indistinguishable from the actual values. RMSE values are consistently extremely small suggested that the models are robust and reliable for predicting energy consumption over the specified time with the errors that are practically negligible. R-squared values approaching 1 indicated a near-perfect fit, able to capture the relations between the variables leading to accurate predictions for all models. Thus integration of normalization with regression models can provide the transformative solution to achieve greater efficiency. It can result in cost effectiveness and scalability along with resource management. In future scope, real time implementation over fitting issue needs to be resolved. Additionally, performance under dynamic, real world condition needed to be resolved using by employing regularization, cross validation, advanced hyperparameter tuning methods.

References

1. Ajayi, O., & Heymann, R. (2021). Data centre day-ahead energy demand prediction and energy dispatch with solar PV integration. *Energy Reports*, *7*, 3760–3774. https://doi.org/10.1016/j.egyr.2021.06.062
2. Asanza, V. (2021). *Data Server Energy Consumption Dataset* [dataset]. IEEE. https://ieee-dataport.org/open-access/data-server-energy-consumption-dataset
3. Berezovskaya, Y., Yang, C.-W., Mousavi, A., Vyatkin, V., & Minde, T. B. (2020). Modular Model of a Data Centre as a Tool for Improving Its Energy Efficiency. *IEEE Access*, *8*, 46559–46573. https://doi.org/10.1109/ACCESS.2020.2978065
4. Chisca, D. S., Castineiras, I., Mehta, D., & O'Sullivan, B. (2015). On Energy- and Cooling-Aware Data Centre Workload Management. *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 1111–1114. https://doi.org/10.1109/CCGrid.2015.141
5. Estrada, R., Asanza, víctor, Torres, D., Bazurto, A., & Valeriano, I. (2022). Learning-based Energy Consumption Prediction. *Procedia Computer Science*, *203*, 272–279. https://doi.org/10.1016/j.procs.2022.07.035
6. Gao, J., Wang, H., & Shen, H. (2020). Smartly Handling Renewable Energy Instability in Supporting A Cloud Datacenter. *2020 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 769–778. https://doi.org/10.1109/IPDPS47924.2020.00084
7. Hsieh, S.-Y., Liu, C.-S., Buyya, R., & Zomaya, A. Y. (2020). Utilization-prediction-aware virtual machine consolidation approach for energy-efficient cloud data centers. *Journal of Parallel and Distributed Computing*, *139*, 99–109. https://doi.org/10.1016/j.jpdc.2019.12.014
8. Landrum, J. (2020, July 17). Did You Know These Data Center Facts? *Enabled Energy*. https://enabledenergy.net/insights/did-you-know-these-data-center-facts/
9. Li, L., Dong, J., Zuo, D., & Wu, J. (2019). SLA-Aware and Energy-Efficient VM Consolidation in Cloud Data Centers Using Robust Linear Regression Prediction Model. *IEEE Access*, *7*, 9490–9500. https://doi.org/10.1109/ACCESS.2019.2891567
10. Li, Y., Wen, Y., Tao, D., & Guan, K. (2020). Transforming Cooling Optimization for Green Data Center via Deep Reinforcement Learning. *IEEE Transactions on Cybernetics*, *50*(5), 2002–2013. https://doi.org/10.1109/TCYB.2019.2927410
11. Marahatta, A., Xin, Q., Chi, C., Zhang, F., & Liu, Z. (2021). PEFS: AI-Driven Prediction Based Energy-Aware Fault-Tolerant Scheduling Scheme for Cloud Data Center. *IEEE Transactions on Sustainable Computing*, *6*(4), 655–666. https://doi.org/10.1109/TSUSC.2020.3015559
12. Peoples, C., Parr, G., & McClean, S. (2011). Energy-aware data centre management. *2011 National Conference on Communications (NCC)*, 1–5. https://doi.org/10.1109/NCC.2011.5734700

13. Postema, B. F., & Haverkort, B. R. (2018). Evaluation of Advanced Data Centre Power Management Strategies. *Electronic Notes in Theoretical Computer Science*, *337*, 173–191. https://doi.org/10.1016/j.entcs.2018.03.040
14. Shrivastava, S. K., VanGilder, J. W., & Sammakia, B. G. (2010). *Data Center Cooling Prediction Using Artificial Neural Network*. 765–771. https://doi.org/10.1115/IPACK2007-33432
15. The Impact of IoT on the Data Centre Sector. (2017, December 21). *The Irish Advantage*. https://irishadvantage.com/the-impact-of-iot-on-the-data-centre-sector/
16. Zwingmann, T. (2022). *AI-Powered Business Intelligence*. A book by O'Reilly Media, Inc. ISBN: 9781098111472,45-67