IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, May 2024

Enhanced Machine Learning for Cloud Computing Electricity Cost Prediction

Ganesh Kumar U[1], Keerthana G M[2], Sahana Y[3], Yalageri Akshay[4], Mr. Usman K[5]

Students, Department of Computer Science[1,2,3,4]

Assistant Professor, Department of Computer Science[5]

Ballari Institute of Technology and Management, Bellary, Karnataka, India

Abstract: *Cloud computing is becoming popular in the tech industry as it offers convenient computing without the need to buy physical hardware. Instead, companies provide these services using their own computers and servers powered by electricity. However, running these data centers consumes a lot of energy, and with rising electricity prices, minimizing energy usage has become a big challenge. One way to tackle this is by efficiently managing data storage and scheduling tasks. In this article, we suggest using an advanced machine learning modelcalled Extreme Gradient Boosting (XGBoost) to predict electricity prices and optimize data storage, helping to reduce energy costs in data centers.*

Keywords: Electricity Price Forecasting for Cloud Computing

I. INTRODUCTION

The increasing adoption of cloud computing as storage platforms has led to reduced hardware investments and procurement expenses. This trend aligns with the exponential growth in demand for information, resulting in a proportional increase in the need for Data Centers (DCs). However, DCs consume a substantial amount of power, accounting for 2% of global power utilization and projected to rise by 12% annually. Cooling accounts for nearly 39% of this power consumption, while 45% is used for running the Information Technology (IT) infrastructure, and 13% for lighting. This level of energy consumption incurred a significant cost for businesses in the US, amounting to $30 billion in 2008. In response to these challenges, numerous researchers have explored the diverse impacts of machine learning methods on modelling, designing, and forecasting electricity prices, particularly in the global market. Typically, two machine learning techniques are employed: one for forecasting electricity pricesand the other for energy systems. Recent approaches often leverage various flavours of deep neural networks, as well as other machine learning methods such as Support Vector Machine (SVM), Random Forest (RF), Naive Bayes, and Decision Tree.

II. LITERATURE SURVEY

In the literature review, Wang et al. investigated electricity price forecasting using a hybrid structured deep neural network model, which involved authors such as Huai-zhi Wang, Gang-qiang Li, Gui-bin Wang, Jian-chun Peng, Hui Jiang, and Yitao. Their methodology employed a multi-layer neural network with Mean Absolute Error (MAE) and Root Mean Square Error (MSE) algorithms. The outcomes of their study revealed a substantial error rate in MAE of 8.84 and MSE of 17.9, indicating promising results in electricity price forecasting. Ugurlu et al. explored proactive power and thermal aware optimizations for energy-efficient cloud computing. Authored by Umut Ugurlu, Ilkay Oksuz, and Oktay Tas, their methodology utilized Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM). Their study resulted in a better MAE of 5.21, surpassing the performance of previous models in energy-efficient cloud computing optimization. Wang et al. conducted an evaluation of data center energy and cost-saving strategies, with authors including Kun Wang, Chenhan Xu, Yan Zhang, Song Guo, and Albert Y. Zomaya. They employed Stacked Denoising Autoencoder (SDAE) with Deep Neural Networks (DNN) in their methodology. While the model yielded a 4.6% error rate for small datasets, it faced significant challenges with larger datasets. Zahid et al. investigated electricity price and load forecasting in smart grids usingenhanced convolutional neural network and enhanced support vector regression techniques. Authored by M. Zahid, F. Ahmed, N. Javaid, R. Abbasi, H. Z. Kazmi, A. Javaid, M. Bilal, M. Akbar, and M. Ilahi, their methodology incorporated Generalized Extreme Learning Machine (GELM) with deep

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 3, May 2024

learning techniques. However, the outcomes did not significantly improve the MAE and MSE values in electricity price forecasting.

Albahli et al. introduced an enhanced machine learning model for electricity price forecasting in cloud computing, authored by Saleh Albahli, Muhammad Shiraz, and Nasir Ayubm. Their methodology involved employing the XGBoost Model with node scheduling and data placement. The outcomes demonstrated an accuracy of 91%, surpassing the benchmark algorithm's accuracy of 81% in electricity price forecasting for cloud computing.

III. PROPOSED SYSTEM

Numerous researchers have examined how machine learning methods can impact modelling, design, andforecasting of electricity prices, especially in the global market. Typically, two main machine learning techniquesare employed: one for predicting electricity prices and the other for managing energy systems. Recent methods often utilize various types of deep neural networks. In our model, we incorporate machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Naive Bayes, and Decision Tree.

IV. METHODOLOGY

This work comprises four stages. Firstly, we collect data from various sources and prepare it for analysis. Secondly, we thoroughly explore the data to understand its characteristics and uncover additional insights. Thirdly, we use different machine learning classifiers to predict the data, generating electricity price forecasts using a tuned model. These forecasts serve as a basis for the fourth step, where further analysis and actions are undertaken basedon the predictions.

Data Collection and Preparation

Data from Ontario - Canada from the provider IESO was used in this article.

Data Exploration

We have used 15 years of historical data from 2003-2018 which merged into a single csv. To get an overview of the entire data set the data was plotted as a time series in Figure 3. According to the key statistics of data-set, themin value, max value, mean value and standard deviation value is -138.79,1891.14, 35.30 and 33.56, respectively. Looking at the entire data set it is rather clear the price fluctuates significantly and suffer from severe price spikes. This is also reflected in the key statistics where we can see that the standard deviation is as large as the mean value. Moreover, the maximum price reaches above 1800 CAD. Even if that is just a single measurement, several large price points appear in Figure 1.

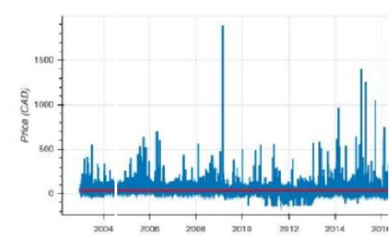

Figure 1. Historical data set as a times series (2003-2018).

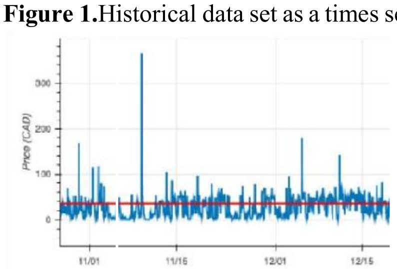

Figure 2. Energy price forecasting data between November and December

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, May 2024

Figure 2 shows the prices in a shorter time period. In fact, the data that was used in the price forecast. It is easierto see in Figure 2 how to price behaves in a closer look. We can see the price fluctuates around the mean but suffers from price spikes. Consequently, this Figure indicates there is opportunities for offloading storage.

Prediction

Our model was developed using three different machine learning algorithms, specifically, XGBoost, Random Forest, and Support vector machine in order to improve the prediction of electricity prices:

- XGBoost
- Random Forest
- Support Vector Machine

All classifiers used the same separation of training and test data to ensure a fair comparison between the methods. We utilized the train_test_split function so as to make the split. The test size = 0.3 inside the capacity shows the level of the information that ought to be held over for testing. It's for the most part around 70/30 or 80/20. To avoid over-fitting and under-fitting we have applied K-cross validation technique such that we ensure that the comparison between the models is fair. During the evaluation it was find that K = 3 was most suited as more folds will take up more memory and since we are taking lower value of K, error will be less due to variance. To understand how much should be used an XGBoost model with the default setting given by sci-kit learn was run as a baseline model on different amount of data.

To evaluate these machine learning models, we have used Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics. The MAE and the RMSE can be utilized together to analyze the variety in the errors in a lot of estimates. The RMSE will dependably be bigger or equivalent to the MAE; the more noteworthy contrast between them, the more prominent the variation in the individual errors in the example data. On the off chance that the RMSE will be equal to MAE, at that point every one of the error are of a similar extent. Both the MAE and RMSE can extendfrom 0 to ∞. They are adversely situated scores: Lower esteems are better.

Figure 3 shows the MSE and MAE with different amount of data on a XGBoost model with the default set parameters. We can depict from the Figure that the MSE increases significantly with a larger data set and MAE decreases. Since the MSE increases multiple times compared to MAE and we are forecasting a data set sufferingfrom large spikes we choose to use a minimal data set since the MSE increases significantly with a larger data set for the used model.

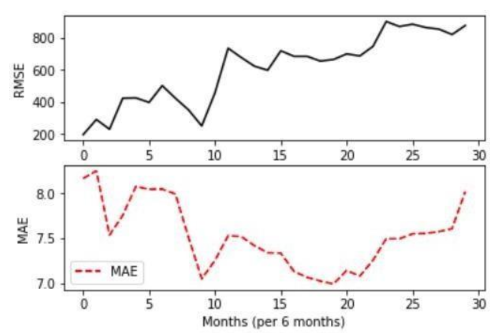

Figure 3. Forecast results with different amount of data on the XGBoost model, with increasing of RMSE value and more decreasing of MAE.

Figure 4 shows a time series of the test set as well as the predicted prices in a graph. The orange line represents the true price whereas the blue line represents the forecast price. The forecast values are near to the actual values, which means that our proposed model outperforms in the context of predicting the electricity price. We can see the blue line imitates the orange line quite well. Indicating the low MAE. However, it is hard the depict from the

Figure whether the blue line is hitting the spikes or is one step behind. We can also see that many times, the blueline does not resemble the extreme values. Indicating the higher MSE and poor P(t/p).

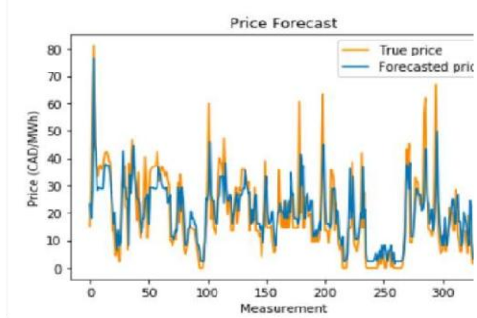

Figure 4. Price forecasting with tuned model as a time series for the test data

Optimixation

In this stage, we analyzed the cost-effectiveness of offloading capacity to different nodes within a single data centre system, considering various distances. Through hourly surveys of power costs, it was consistently more economical to offload capacity to nodes. The model assumes regular updates to the data, such as when messaging platforms like Facebook, WhatsApp, and Telegram reach 1 billion users, leading to increased data storage needsin the data centre. In such instances, data may be moved if there is a spike in value. For example, a cell phone canact as a node, storing data locally to avoid additional charging or connection to an electricity provider, depending on the owner's preferences for charging the node.

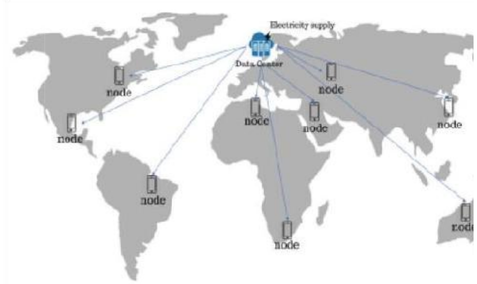

Figure 5. An illustration of the data centre interconnect map. The cloud data centers represent as servers, and eachserver connect with multiple nodes. For each hour, the power cost was surveyed to explore whether it was advantageous to offload capacity to nodes.

Thus, Figure 5 visually shows the set-up on a map. It shows a single data centre representing a server which provides cloud computing services to M connected nodes. The lighting symbol represents electricity supply which means the server is powered by electricity. The capacity of offloading the storage can be represented as arrows, starts from the data center and ends at the target nodes. According to Carolyn Duffy Marsan of the Network World, ‘‘The cost of a data centre’s power and cooling typically is more than the cost of the IT equipment inside it’’, shecame to this conclusion after she found out most cloud companies use methods which are very expensive when setting up their companies.

V. CONCLUSION

The main goal of this research is to explore whether using machine learning techniques can help minimize energy consumption in cloud data centers during dramatic spikes in electricity prices. We analyze daily spot electricity price data from 2003 to 2018 to predict Ontario electricity returns, addressing challenges posed by price spikes and volatility in the Ontario electricity market. Our study evaluates the performance of a cost savings model across different standard deviation (std) values, demonstrating approximately 50% cost savings as std increases. Ultimately, our model achieves an 85.66 accuracy for price forecasting, with a Mean Squared Error (MSE) of 6.66. With these forecasts, our optimized model successfully reduces electricity costs by up to 25.32%. Furthermore, results from a small testing platform indicate significant potential savings in electricity costs, suggesting even greater savings on a larger scale

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 3, May 2024

REFERENCES

[1]. P.-H. Kuo and C.-J. Huang, ''An electricity price forecasting model by hybrid structured deep neural networks,'' Sustainability, vol. 10, no. 4, p. 1280, Apr. 2018.

[2]. P. A. Garcia, ''Proactive power and thermal aware optimizations for energy-efficient cloud computing,''Ph.D. Dissertation, Dept. Electron. Eng., Tech. Univ. Madrid, Madrid, Spain, 2017.

[3]. Z. Song, X. Zhang, and C. Eriksson, ''Data center energy and cost saving evaluation,'' Energy Procedia,vol. 75, no. 1, pp. 1255–1260, Aug. 2015.

[4]. M. Zahid, F. Ahmed, N. Javaid, R. Abbasi, H. Z. Kazmi, A. Javaid, M. Bilal, M.Akbar, and M. Ilahi, ''Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids,'' Electronics, vol. 8, no. 2, p. 122, Jan. 2019.

[5]. SALEH ALBAHLI, MUHAMMAD SHIRAZ, (Member, IEEE), AND NASIR AYUBM, "Electricity Price Forecasting for Cloud Computing Using an Enhanced Machine Learning Model" Extreme Gradient Boosting (XGBoost) vol.11, no. 1, p. 143, Oct 2020