Electricity Price Forecasting for Cloud Computing Using an Enhanced Machine Learning Model

SALEH ALBAHLI, MUHAMMAD SHIRAZ, (Member, IEEE), AND NASIR AYUB
Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia
Corresponding author: Saleh Albahli (salbahli@qu.edu.sa)

ABSTRACT
Cloud computing is rapidly taking over the information technology industry because it makes computing a lot easier without worries of buying the physical hardware needed for computations, rather, these services are hosted by companies with provide the cloud services. These companies contain a lot of computers and servers whose main source of power is electricity, hence, design and maintenance of these companies is dependent on the availability of steady and cheap electrical power supply. Cloud centers are energy-hungry. With recent spikes in electricity prices, one of the main challenges in designing and maintenance of such centers is to minimize electricity consumption of data centers and save energy. Efficient data placement and node scheduling to offload or move storage are some of the main approaches to solve these problems. In this article, we propose an Extreme Gradient Boosting (XGBoost) model to offload or move storage, predict electricity price, and as a result reduce energy consumption costs in data centers. The performance of this method is evaluated on a real-world dataset provided by the Independent Electricity System Operator (IESO) in Ontario, Canada, to offload data storage in data centers and efficiently decrease energy consumption. The data is split into 70% training and 30% testing. We have trained our proposed model on the data and validate our model on the testing data. The results indicate that our model can predict electricity prices with a mean squared error (MSE) of 15.66 and mean absolute error (MAE) of 3.74% respectively, which can result in 25.32% cut in electricity costs. The accuracy of our proposed technique is 91% while the accuracy of benchmark algorithms RF and SVR is 89% and 88%, respectively.

INDEX TERMS
Data storage, energy saving, electricity price forecasting, XGBoost.

I. INTRODUCTION
Cloud computing is increasingly being used as storage platforms that lowers hardware investments and decreases procurement expenses. Exponential increase in demand for information leads to proportional demand for Data Centers (DCs). DCs consume a lot of power comprising of 2% of the global power utilization. It is expected to rise at the rate of 12% every year [1], [2]. Nearly 39% of power is used for cooling, 45% for running the Information Technology (IT) infrastructure, and 13% for lights [3]. This level of consumption costed the businesses in US 30 billion dollars in 2008 [4].

According to a report by Walker, utilization of distributed computing along with virtualization can improve productivity. However, this approach is still not very common. As indicated by Ericsson, non-virtualized servers use only 6% to 15% of their capacity. This is while virtualized servers can use up to 30% of their capacity [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Eklas Hossain.

In general, DC operators have a few DCs distributed over locations to guarantee reliability by utilizing replication. Being close to the clients will meet the latency requirements. However, distributed DCs in various geographical areas can lead to uncertainty in costs due the dynamic costs of power markets. These power markets can shift to a great extent in cost. Consequently, DC suppliers would build DCs in low temperature locations with low electricity price.

Power markets operate in a deregulated environment and the vendors are free to set the price to attract customers. The volatile nature of the price of power can increase instability in the market and costs can grow ten times within 60 minutes [6]. For instance, the lowest observed cost on the Ontario power market in 2018 was 4.39 CAD/MWh while the highest observed cost was 365.64 CAD/MWh [6]. With the expanded interest of cloud computing and the unstable power cost, research in utilizing volatility in the deregulated power value market is important. It helps to predict value spikes and as a result decrease power utilization during these periods to reduce energy costs.
Content Deliver Networks (CDNs) are used by businesses such as Netflix. They would locate data center nearer to the clients so as to limit the need for long-distance transmission of data and enhance Quality of Service (QoS) [7]. This method could possibly be used to offload capacity from centralized DCs to hubs in the edge of the system so that the businesses can reduce power usage and ultimately reduce power costs.

From past few decades green environment and efficient energy consumption has become a hot topic due to its importance and extreme need. Various researchers have employed state of the art techniques as well as traditional techniques to resolve the issues. For instance, [1] and [4] investigated how power expenses can be marked down in multi-geographical situations. Many researchers suggest that market survey should be carried out on cost of setting up servers at different locations because the cost varies from one geographical location to the other. There additionally exists researches like [8] for the improvement of node scheduling to reduce power cost. Likewise, in [9] the researchers enhanced the route selection for transmission of data. In any case, the referenced papers recommend answers for explicit pieces of an issue instead of providing a competitive singleton solution.

Similarly, many researchers have focused on the diverse effects of machine learning methods on modeling, designing, and forecasting electricity price, particularity in global market. Generally two machine learning techniques are mostly used where the first one is for forecasting electricity price and the later one is for the energy systems. Most of the recent methods use different flavours of deep neural networks such as [10]–[13] as well as the other machine learning techniques methods such as Support Vector Machine (SVM) [14]–[16], Random Forest (RF) [17], Naïve and Decision Tree [18], [19].

Most of the previous works on electricity price prediction are still in their infancy and lack in terms of accuracy, computational overhead or unable to prove results on real-time data. In this article, we propose a model to measure the effectiveness on forecasting electricity price of the data center of Ontario - Canada, to mitigate energy consumption effects and make considerable cost savings. Our forecasting model can be utilized to analyze the effect of various risk factors on price spike for data-storage and to predict an accurate energy consumption. The model was evaluated by 15 years of historical data gathered from IESO provider to predict electricity price markets. The model used improved XGBoost which indicated that our model is capable of reducing the electricity cost for data-storage up to 25.32% and delivering acceptable future estimation and performance as compared with the random forest and support vector machine methods without employing extra computational overheads. Moreover, the proposed technique is easy to implement and can be utilized in real-time prediction.

A. MOTIVATION

Due to the demand for cloud computing that process large amount of data, the pressure on cloud suppliers to discover better approaches to reduce power expense for data storage never stops. To remain gainful and at the same time meet administration level understandings (SLAs). Because of the instability in the deregulated power cost there is a motivation to explore whether these anomalies can be utilized to bring down power consumption and, thus, power costs. Possibly by offloading storage to nodes likewise to a CDN. This research investigate a specific problem of whether it is valuable or not to use machine learning techniques to leverage a dramatic spike in electricity prices to offload data-storage to lower operational expenses in data centers. In this way, the following research questions are investigated:

1) Is it worthwhile to minimize costs by offloading data storage to nodes before price spikes happen?
2) How accurately can we predict the electricity price?
3) How do machine learning classifiers affect the predication accuracy of the electricity price volatility?

The rest of the paper is organized as follows. Section II will provide the comparative analysis of previous literature and critical evaluation of already presented techniques. In section III, theory components and related works are outlined and how our methods utilizes other works. Section V-A describes the dataset used and machine learning classifiers are evaluated with their metrics. Section V predicts the electricity price with different data analysis and mathematical methods. Finally, we conclude the paper and provide further research areas in section VI.

II. PREVIOUS WORKS

Sustainability is the major concern of today’s life style, there has been a plethora of research catering the issues of electricity consumption and green environment. This section will provide a compact analysis of previous approaches, which are utilized to forecast electricity consumption using various approaches. Moreover, we will highlight the certain issues and shortcomings of the available literature which led us to provide effective and robust solution.

Wang, Huai-zhi, Gang-qiang Li, Gui-bin Wang, Jian-chun Peng, Hui Jiang, and Yi-tao Liu [20] exploited a model based on Multi-Layer Neural Network (MLNN) to estimate the load of electricity and its overall consummation. They also employed the Ensemble technique to discard the diverse errors and cancellation of noise. Although their technique had competitive advantage in terms of accuracy, the technique lacks the robustness because of higher computational time and huge loss rate during testing on real-time data. Similarly, Ping-Huan Kuo and Chiu-jye Huang [1] proposed a hybrid technique for the price forecasting of electricity named as EPNet. The technique was a mesh of LSTM and CNN which provided MAE of 8.84 and MSE of 17.9. Despite of the good results the models provide huge error rates in real time prediction with enormous computational complexity. Moreover, the dataset utilized is highly normalized and the model was unsuccessful to produce similar results on realtime data.

Meanwhile, Umut Ugurlu, Ilkay Oksuz, and Oktay Tas [21] proposed a similar model based on the combination...
of Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM). They concluded the model with MAE of 5.71, but the results are only valid for one-day ahead prediction. Furthermore, the results are not constant and can diverge with changing seasons, which makes this model inefficient in real-time deployment. The models are computationally expensive as well. Similarly, Jesus Lago, Fjode Ridder and BartDe Schutter [22] provided the comparative analysis of DL based approaches which provided solutions on elect city consumption forecasting and green environment. The evaluated and discussed the results of LSTM-DNN, GRU-DNN, CNN and MLP and other 23 benchmark methodologies. They also proposed a DL based algorithm for the prescribed task of electricity price prediction. They stated the results of proposed technique is comparatively better than already preset literature. However, the comparison is performed using a single, highly normalized dataset. The proposed approach is computationally expensive as well as provide false predictions on realtime dataset with huge testing loss.

Wang, Kun, Chenhan Xu, Yan Zhang, Song Guo, and Albert Y. Zomaya [23] exploited a hybrid methodology for electricity price prediction by utilizing combination of Sport Vector Machine (SVM) and Kernel Principle Component Analysis (KPCA). They concluded the model with 4.6% error for the smaller threshold value of U while, 45.8% error for larger threshold value. Due to exploitation of large dataset including the cost price of steam, wood, wind, gas and oil, a huge computation overhead is introduced which contributed towards inefficiency of the proposed technique. Moreover, prices change around seasons and can be variable in different location which makes the model highly static and location dependent. The authors also utilized a mesh of Stacked Denoising Autoencoder (SDA) with DNN based models for the same task [24]. They provided the state-of-the-art results with a huge location dependent dataset. Likewise, the authors of [22] exploited the power of deep learning to enhance the prediction accuracy for electricity cost with respect to market of Europe. Additionally, they employed functional analysis of discrepancy with respect to feature selection and utilized Bayesian optimization. The model was rational simplistic with efficient feature selection technique, however, the MSE and MAE values are huge and the technique is unable to handle the problem globally.

Raviv, Eran, Kees E. Bouwman, and Dick Van Dijk [25] performed hourly based prediction of electricity cost by using multivariate models. They additionally performed dimension reduction practices to moderate the effects over-fitting. However, the results in terms of MAE and MSE are not comparatively better, there are multiple false predictions. Mujeeb, Sana, Nadeem Javaid, Mariam Akbar, Rabiya Khalid, Orooj Nazeer, and Mahnoor Khan [26] suggested a DNN based model with combination of LSTM for the price estimation of electricity. They worked on load prediction as well, but the results are not satisfactory for the tasks of price estimation. As described above, most literature is focused on state-of-the-art deep learning-based techniques. But on the other hand, some researchers worked using traditional methods as well, like [27] proposed a probabilistic framework for the task of hourly price estimation of electricity by utilizing Generalized Extreme Learning Machine (GELM). The model is computationally expensive for larger datasets and provide indefinite results. Similarly, the authors of [28] shifted their focus on feature selection instead of simplistic training. They exploited Information Gain (IG) and Mutual Information (MI), like techniques to perform feature selection. After rigorous testing they provided MAE of 4.09 but the model can only use for offline prediction on huge dataset.

Likewise, Ghasemi, Ali, Hossien Shayeeghi, Mohammad Moradzadeh, and Mohammad Nooshyar [29] anticipated fusion based technique for the price estimation and load prediction of electricity. They utilized Least Square SVM (LSSVM) and employed Artificial Bee Colony Optimization (ABCO) algorithm. Meanwhile, Keles, Dogan, Jonathan Scelle, Florentina Paraschiv, and Wolf Fichtner [30] suggested an ANN based technique or the prescribed task. In the same manner, Fan, Guo-Feng, Li-Ling Peng, and Wei-Chiang Hong [31] came up with the electricity price prediction framework by utilization of a hybrid algorithm of Bi-Square Kernel based model with Phase based Space Reconstruction (PSR-BSK). They employed various datasets of NYISO, USA and Wales markets of South. The models described above are computationally expensive and provide false prediction with huge loss which makes them inefficient in real-time usage. To summarize, there a plenty of literature available on the price prediction of electricity to estimate and reduce the power consummation in DCs. It has been a hot problem since last few decades. However, the present techniques lack the ability to provide efficient results in terms of low MAE and MSE for global market, most of them are computationally expensive and unable to work in real time environment.

III. THEORY AND METHODS
A. CLOUD COMPUTING
IT industry has been changed significantly due to the revolution of cloud computing and how businesses use it. IT resources like servers, networking and databases are consumed heartily by Cloud computing on an on-demand computing resources. Instead of having local servers, these resources can be practically hosted on the Internet. Accordingly, investment costs and capital expenditures (CAPEX) for local servers can be minimized by paying on-demand. Businesses with traditional data centers have to pay much to scale up their hardware, software and data storage to support services and all requirements. Otherwise, they pay less but they will not meet user’s requirements. However, cloud computing helps enterprise companies to focus on their objectives and key business activities to reach business goals by relying on outside technology to host all IT infrastructure.
A content delivery network (CDN) is a distributed network of edge servers in different locations that store large volumes of data with a measurable amount of latency. Nowadays, over half of all traffic is utilizing the CDN services which is the backbone of the Internet to deliver the content and it still grow at a fast pace. Amazon, Dropbox, Facebook, and Netflix are examples serve through CDNs. To reduce the distance between servers and users, Netflix as an example shares data across a geographical area to ensure a user accesses data which is the closest to a server instead of loading the data from an original server and then enhance its QoS [7]. Netflix uses advanced algorithms to predict the target file on the target server at the right time. Therefore, the local server decreases bandwidth costs and adopt to data delivery at scale over a range area. Since the limit in the hubs might be restricted, just the data that accessed repeatedly should be moved to the edge. By using this model, Netflix has expanded its throughput from 8 Gbps from a single server in 2012 to over 90 Gbps from a single server in 2016 [7].

B. MACHINE LEARNING

Machine learning is a method of analyzing data with the help of its algorithms to look for patterns then predict unseen data. It enables to use resources more effectively by learning from previous processed data. Without being explicitly programmed, a machine learning algorithm receives data in order to build its own logic and adaptively enhance its performance. In essence, the main machine learning problems are characterized as supervised or unsupervised. In supervised learning, the dataset and result are given to build a model and predict future outputs. Whereas unsupervised learning feeds by input but without labeled or categorized data. The aim of unsupervised learning is to figure out patterns and create different market segments in data to derive meaning [36].

Cloud computing paradigm may intelligently get benefit of the functionalities offered by machine learning. As considered in related works, machine learning methods can effectively forecast energy costs. Hence, machine learning techniques have been applied for energy management and could be applicable to estimate future electricity prices [1], [26]. Thus, the electricity market have been largely influenced by electricity price forecasting. In this article, however, we create a model to predict electricity prices by comparing the overall performance of XGBoost, Random Forest and Support Vector Regression (SVR) classifiers.

IV. METHODS

This work is divided into four different stages. First we gather the data from different sources and arranged for analysis. Second, data has been explored in detail to understand various data characteristics and discover more information. Third, the data is predicted with different machine learning classifiers to generate electricity price forecasts with the tuned model which will help further in the fourth step. However, this section will follow the clarified structure.

A. DATA COLLECTION AND PREPARATION

Data from Ontario - Canada from the provider IESO was used in this article [6].
B. PRICE FORECASTING

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 extend from 0 to ∞. They are adversely situated scores: Lower estimates are better.

Following evaluation metrics given the data \((x_1, x_2, \ldots, x_n)\) and predictions of the data-set \((y_1, y_2, \ldots, y_n)\) we can formulate MSE and MAE as follows, where x is actual and y are predicted values. \([37]\)

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| \quad (1)
\]

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2} \quad (2)
\]

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \quad (3)
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| \quad (4)
\]

To further evaluate spike prediction we take in consideration of the confusion matrix in which we use True positive (TP) and False negative (FN) which are defined as:

\[
TP = \frac{TP}{TP + FP} \quad (5)
\]

where FP is false positive. If TP value is closer to 1 it is much better.

\[
FN = \frac{FN}{FN + TP} \quad (6)
\]

where value of FN closer to 0 is considered better.

Equation (5) suggests that at the point when an information provided is delegated a positive sample (let’s say class 1) and it is really a positive model, similarly Equation (6) at the point when an information provided is delegated a negative example (let’s say class 0) yet it is really a positive precedent. In our analysis using XGBoost which gave us the best result it was observed that with learning rate 0.1, max depth as 3 and number of estimators as 25, the probability of true positive was 0.31 and the probability of false negative was 0.68 which further suggests that there is good amount of difference in actual and predicted values.

C. OPTIMIZATION

With a single data center system and different distance of nodes, M were considered. For every hour, the power cost was surveyed to explore whether it was advantageous to offload capacity to nodes. It was constantly less expensive, as far as cost to offload to nodes. The data in the model is expected to be updated regularly. For instance, Facebook, WhatsApp and Telegram messages hit 1 billion users, so that they are stored as much data flow as they need in the data center. Hence, the data was possibly moved if a value spike happens. With this case, a cell phone has been represented as a node so that it stores data at the point avoiding neither charging the node nor connecting to an electricity provider depending on the node’s owner uses to charge the node.

Thus, Figure 2 visually shows the set-up on a map. It shows a single data center 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 center’s power and cooling typically is more than the cost of the IT equipment inside it”, she came to this conclusion after

**FIGURE 2.** An illustration of the data center interconnect map. The cloud data centers represent as servers, and each server connect with multiple nodes. For each hour, the power cost was surveyed to explore whether it was advantageous to offload capacity to nodes.
she found out most cloud companies use methods with are very expensive when setting up their companies.

Notations used in our optimization model are introduced in Table 1.

The problem can be modeled in a set of formulas called P0 which varies with respect to values of of \( x_a \) and \( x_b \) as seen below:

\[
P_0: \quad \text{min } w = \sum_{a=1}^{TN} x_a \cdot X + \sum_{b=1}^{TM} x_b \cdot c_b
\]

subject to

\[
x_a \leq x_i, \quad \forall a = 1, \ldots, TN
\]

\[
x_b \leq j_i, \quad \forall b = 1, \ldots, TM
\]

\[
\sum_{a=1}^{TN} x_a + \sum_{b=1}^{TM} x_b = \sum_{i=1}^{X} x_i
\]

\[
x_a, x_b \geq 0, \quad \forall a = 1, \ldots, TN, \forall b = 1, \ldots, TM
\]

where the variable \( x_a \) is set up as an integer which denotes the massive troves of data to store at server \( a \) and \( x_b \) is the volume of data to offload to node \( b \). \( X \) is the power cost, \( c_b \) represents the expense to store data at node \( j \), \( a \) the limit of server \( i \) and \( b \) the limit of node \( b \). \( c_j \) is a symbolic cost attribute specifies a threshold level of value spike to avoid storing the whole data in the edge at once, but instead to data center. \( w \) is a function \( f(a, b) \) which is minimized to give the optimal values of \( a \) and \( b \). In the cost estimation, the cost is set to zero for all instances to continue pay for the electricity by node’s owner.

The prediction prices influence the storage cost at each hour that can be decreased by the objective \( w \). Constraint (8) and Constraint (9) ensures data is dynamically allocated to the server capacity and for each node respectively. Constraint (10) guarantees that the data allocated to storage has been distributed on multiple nodes as well as servers in data centers. Finally, Constraint (11) makes sure that only integer values can be chosen on \( x_a \) and \( y_b \) whereas the negative storage is neglected.

The algorithmic of \( P_0 \) is clarified in the following steps:

- Methods that described in section IV can be used to predict electricity prices.
- For every hour, predicted electricity prices can be solved with \( P_0 \).

Based on the solution provided by \( P_0 \), we can estimate the cost with the help of electricity prices.

The optimization is based on prediction of price whether or not the prediction is correct because the actual expense will be determined by the actual price. Hence the model of electricity price can be decided by the predicted price so that the cost determined from the correct prices. Thus, the hourly cost can be written as follows:

\[
cost = \sum_{a=1}^{TN} x_a \cdot Y \cdot E \cdot 10^{-6}
\]

where \( E \) in Equation (10) is the utilization of energy in W per server each hour and \( p_2 \) is the corrected price of electricity.

To have a similar unit as \( E \), the hourly cost of electricity price in CAD/MWh changed to be CAD/Wh.

V. RESULTS AND ANALYSIS

The analysis of predicting the daily spot prices is implemented in three different steps as follows:

1) Data investigation
2) Price Forecasting
3) Optimization

A. 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, the min 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 3.

Figure 4 shows the prices in a shorter time period. In fact, the data that was used in the price forecast.

It is easier to see in Figure 4 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.
FIGURE 4. Energy price forecasting data between November and December.

FIGURE 5. Auto-correlation and partial auto-correlation functions for 100 lags at Ontario.

FIGURE 6. Correlation heatmap between different variables to show the effect on the energy price. The most related variables are colored by the bright.

FIGURE 7. Forecast results with different amount of data on the XGBoost model, with increasing of RMSE value and more decreasing of MAE.

To understand how past data impacts the current price, the auto-correlation and partial auto-correlation functions were plotted in Figure 5.

Figure 5 clearly shows that the data is seasonal - meaning recurring dependencies exist. Therefore, these lags may add relevant information to the model. Furthermore, we see that the older the data gets, the lower the correlation becomes. Figure 6 shows the correlation between all the different selected features in a heat-map. The brighter the color, the higher the correlation.

From this Figure, we can extract the most correlated features as a starting point for the model. We see that the closest lags, the forecasts as well as the price point 24 hour back provide rather high correlation in comparison to the other features. Among the date related features, Hour is the only feature that seem to provide any relevant information to the model.

Not surprisingly we see the correlation become smaller from lag $k = 1$ to lag $k = 5$. This makes sense, as the older the value, the smaller the effect on the current value it should have. Comparing this with the features for the date, we can see a significant difference. Equally as for the lag we can see that the correlation decreases as we go from micro perspective to macro perspective (hour to year) i.e. knowing what year plays a smaller role comparing to knowing which hour a datapoint took place.

Since only the hour seem to give some valuable information regarding the price, the remaining date features will be dropped.

B. PRICE FORECASTING

Three different electricity price forecasting techniques used: XGBoost, Random Forest and Support Vector Regression (SVR).

Figure 7 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 suffering from large spikes we choose to use a minimal data set since the $MSE$ increases significantly with a larger data set for the used model.

To find the optimal set of features for each method, features were subsequently added to the model starting with the feature acknowledging the highest correlation according to Figure 6.

Table 2 shows the different feature sets and Table 3 shows the results from the feature selection.

Negative lag is the amount of time that a successor activity can start before the completion of predecessor activity. Observing the results with different features it is clear that simply adding a feature does not necessarily mean it improves the model.
TABLE 2. Feature sets.

| Notation | Feature set |
|----------|-------------|
| A        | lag k = -1  |
| B        | lag k = -1 & lag k = -2 |
| C        | lag k = -1, lag k = 1 & lag k = -2 |
| D        | lag k = -1, lag k = 1, lag k = 2 & lag k = -2 |
| E        | lag k = -1, lag k = 1, lag k = 2, lag k = 3 & lag k = -3 |
| F        | lag k = -1, lag k = 1, lag k = 2, lag k = 3, lag k = 4 & lag k = -4 |
| G        | lag k = -1, lag k = 1, lag k = 2, lag k = 3, lag k = 4, lag k = 5 & lag k = -5 |
| H        | lag k = -1, lag k = 1, lag k = 2, lag k = 3, lag k = 4, lag k = 5, lag k = 6 & lag k = -6 |

TABLE 3. Feature selection with forecast results for all different models.

| Model            | Feature Set | MSE  | MAE  |
|------------------|-------------|------|------|
| XGBoost          | A           | 121.63 | 7.62 |
| XGBoost          | B           | 108.67 | 7.63 |
| XGBoost          | C           | 110.31 | 7.59 |
| XGBoost          | D           | 110.31 | 7.58 |
| XGBoost          | E           | 104.67 | 7.49 |
| XGBoost          | F           | 108.33 | 7.38 |
| XGBoost          | G           | 108.33 | 7.39 |
| XGBoost          | H           | 94.92  | 6.94 |
| Random Forest    | A           | 195.4  | 10.26 |
| Random Forest    | B           | 200.91 | 10.32 |
| Random Forest    | C           | 155.79 | 10.32 |
| Random Forest    | D           | 159.58 | 9.28 |
| Random Forest    | E           | 157.01 | 8.93 |
| Random Forest    | F           | 124.85 | 7.59 |
| Random Forest    | G           | 122.52 | 7.75 |
| Random Forest    | H           | 106.63 | 7.23 |
| SVR              | A           | 111.28 | 7.78 |
| SVR              | B           | 111.28 | 7.79 |
| SVR              | C           | 192.05 | 11.38 |
| SVR              | D           | 194.54 | 11.42 |
| SVR              | E           | 220.36 | 11.52 |
| SVR              | F           | 249.36 | 13.75 |
| SVR              | G           | 250.08 | 13.76 |
| SVR              | H           | 216.36 | 12.77 |

E.g. for XGBoost we can see that the result is improved in terms of MSE when adding an additional feature from feature set A to B. On the other hand, when adding even more additional features the MSE increases. However we can see that the MAE constantly decreases when adding more features, although marginally. Nevertheless we find the best solution of feature sets for feature set H, which includes all the relevant lags as well as the hour. This is the case for both XGBoost and Random Forest.

Looking at the results for the SVR, it behaves rather different when adding additional features compared to Random Forest and XGBoost. Rather than yielding higher accuracy, the accuracy decreases rather drastically when adding more features. With exception between feature set G and feature set H. However, the best result is yet given with the minimum feature set - A.

Since XGBoost with feature set H gives the best result, that is the model used for further tuning. Following parameters are iteratively tuned for XGBoost: number of estimators, max depth, learning rate and gamma. The optimal solution was found when number of estimators = 25, max depth = 3, learning rate = 0.1 and gamma = 0. Table 4 shows the performance error of our proposed model with the state of the art algorithms.

The performance error of our proposed model is less than other techniques as in Table 4.

We can now see that the forecast has improved in terms of MAPE, RMSE, MSE and MAE. It can be clearly seen that MAPE, RMSE, MAE and MSE of proposed technique is better than state of the art techniques in Figure 8. Although it couldn’t cross the DL based techniques but it is computationally inexpensive and can be deployed in real time environment. In addition, to the MSE and MAE, \( P(t/p) = 0.32 \) and \( P(f/n) = 0.68 \) represents the accuracy of how well the model predicts the spikes.

Figure 9 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 blue line does not resemble the extreme values. Indicating the higher MSE and poor \( P(t/p) \).
C. OPTIMIZATION

Note: ∈ means is a member of The optimization result of Cost Saving (CAD) is 1915.77 and Cost Saving is 25.32% with a randomized in servers and nodes capacities Given a normal distribution with variance mean and standard distribution, the storage sizes with regards to both the servers and the nodes were randomized. In our model, we consider TN = 5 servers with capacities \( x_i \in TN(1000, 20) \). To make energy-aware, efficient and zero consumption, we utilize a one-and-off plan to shut servers down in case of offloading. In addition, homogeneous servers are being considered to make energy consumption more balancing when active. Following [38] all active servers are set to 240 W/h. Additionally, we implement TM = 1000 nodes with capacities \( j_i \in TN(5, 2) \) GB. We suppose in our model that storage capacity of nodes achieve the maximum energy optimization for offloading in terms of avoiding energy station and then the node’s free-storage space and cost. Additionally, the execute and transfer cost between the data center and nodes were neglected. Now, it is possible to define the logic with the parameters TN \( \gg TM \) and \( x_i \gg j_i \), \( \forall a = 1, \ldots, TN \) and \( \forall b = 1, \ldots, TM \). Likewise, \( j_i \) ought to be a realistic number that could apply for a cell phone or some other portable devices with a smaller-capacity.

When looking at the results it is important to remember that the potential saving is highly related to the number of predicted spikes. The standard deviation and mean cost are also another element influence the price saving, for instance a higher for both standard deviation and mean cost will smooth up prices and bring several chances for more effective price saving.

After going through the optimization result of Cost Saving (CAD) and Cost Saving, obviously that conceivable price savings happen. As expected, experimental result between cost savings show that even for a small test dataset, the outcomes are better than the reality that likely sustain over other constraints which may influence the price costs, such as communications costs. Hence, those adjustments lead to better prices and thus cost savings. One could see that in Ontario which helped to reduce costs up to 25.32% constitutes for more than 1900 CAD during about 2 weeks for a single Data Center with just 5 servers with a size of 1000 GB for each server. The cost savings will be increased with a large Data Center reach for example to 5000 servers, the savings would translate to 49,4 million CAD/year. Consequently, saving energy consumption in reality can be done based on some factors and further constraints which are not discussed here, but in our model, predicted spikes can lead to drop drastically in energy consumption comparing to other constraints in real world.

The motivation around moving to a small dataset with 5 servers was a result of computational restrictions. Simultaneously, expanding the server size with a large data center will significantly influences the calculation time multi hours for each simulation as shown in our experimental results. The reason behind re-scaling of various capacities such as: nodes and servers were not implemented here was because these rescalings do not affect the result. Since, savings on electricity and cost are associated to offload storage of data, the total capacity has to be identical on all servers and nodes to enable offloading data in the data center and efficiently decrease energy consumption for data storage. In either case, reducing the nodes count but keep the capacity in a large form does not affect the cost’s estimations and savings. The cost saving (CAD) value is 1159.58 and cost saving is 25.32%, where TN = 3 and \( x_i \in TN(1000, 20) \) and TM = 1000 with \( j_i \in TN(3, 2) \).

Table 5 shows some results from the experiment conducted. We implement several optimization of node storage costs to conduct a patchwork of analyses on changes in use cases to show the efficiency of offloading storage to some nodes other than other. The reason behind that is that we have considered the distance between physical nodes or connection points for data transmission and data center. In Table 5, we applied our optimization model by taking three extra different node storage costs by selecting multiple standard deviation (std) values using different distributions.

It is our contention that the cost of these extra node storage to be more suitable to offload to the cloud may yet offset the cost through the added features. Hence, the optimization results with a randomized in the servers and nodes capacities, and also a different mean and standard distribution show the effective added features on the storage space using our reliable testing model that ultimately help allocate server resources to the nodes.

It is not surprising then, that our model has been designed to be as cost savings as possible, so that whenever increasing the standard deviation (std), our results offer significant cost savings. However, adding a new std value efficiency leads to reducing in the cost. For instance, saving resources have been reduced by more than half when the std increments from 0.1 to 0.5. A possible reason for this could be that many spikes are just above the threshold and fall under the threshold when the limit for offloading is constrained. The remaining spikes are significantly above the threshold and consequently the effect on increases the std is not as significant.

VI. CONCLUSION AND FUTURE STUDIES

In this article, we propose a model for Ontario - Canada electricity price prediction. The main objective of this research is to investigate a specific problem of whether it is valuable or not to use machine learning techniques to leverage a dramatic spike in electricity prices to offload data storage to minimize the energy consumption in cloud data centers. Moreover, we analyze the forecasting result of daily spot electricity price, during 2003-2018, to predict Ontario electricity returns. Electricity prices are challenging the industries to address price spikes or volatility of prices of Ontario.

| std | Cost Saving (CAD) | Cost Saving (%) |
|-----|------------------|-----------------|
| 0.1 | 1354.19         | 20.24          |
| 0.3 | 726.16          | 9.42           |
| 1   | 555.24          | 7.03           |

TABLE 5. Optimization results with different node costs.
electricity market. We studied the performance of our cost savings model on different standard deviation (std) values. The results show the efficiency of our model by saving cost storage with approximately 50% when the std increased. Ultimately it was possible to forecast the price with an accuracy of 85.66 and 6.66 for MSE and MAE respectively. Considering these forecasts, our optimized model to offload storage of data in data centers has successfully reduced electricity costs up to 25.32%. More importantly, aforementioned data for a small testing platform shows that a significant electricity cost savings are possible which indicate that taking a larger testing platform expect to be reduced, potentially saving significant sums.

Moving forward, we expect that further research could improve our model by investigating more parameters tuning and constraints. Moreover, our model can be applied to many practical forecasting scenarios, for example, various geographical area can be investigated using different types of data instead of one type. Energy, load forecasts, weather data are examples of other types of data.

However, our model shows the power of using machine learning in predicting energy forecasting of the data center that can further result in reducing price spike risks for data storage. More specifically, XGBoost shows significant electricity cost savings and the highest accuracy average over other classifiers (Random Forest and Support Vector Regression). Therefore, studying different classifiers such as: neural networks will be other dimensions of extension. Additionally, clustering may help to identify a lower error of spikes and then Figure our missed/undesired spike error on the forecast.

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. 1290, 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] S. Walker. (Aug. 2016). *Three Ways to Boost Datacenter Utilization*. [Online]. Available: https://cloudblog.e ricsson.com/digital-services/3-ways-to-boost-datacenter-utilization

[6] (2018). I E SO. [Online]. Available: https://www.i eso.ca

[7] K. Florence. (Mar. 2016). *How Netflix Works With ISPs Around the Globe to Deliver a Great Viewing Experience*. [Online]. Available: https://media.netflix.com/en/company-blog/how-netflix-works-with-isps-around-the-globe-to-deliver-a-great-viewing-experience

[8] S. Rahaman, A. Gupta, M. Tornatore, and B. Mukherjee, “Dynamic work-load migration over backbone network to minimize data center electricity cost,” *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 570–579, Jun. 2018.

[9] C. Natalino, L. Chiaravaglia, F. Idzikowski, L. Woinski, and P. Monti, “Joint optimization of failure management costs, electricity costs, and operator revenue in optical core networks,” *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 291–304, Mar. 2018.

[10] P.-H. Kuo and C.-J. Huang, “A high precision artificial neural networks model for short-term energy load forecasting,” *Energies*, vol. 11, no. 1, p. 213, Jan. 2018.

[11] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, “Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach,” *Energy*, vol. 118, pp. 999–1017, Jan. 2017.

[12] F. Ahmed, M. Zahid, N. Javaid, A. B. M. Khan, Z. A. Khan, and Z. Murtaza, “A deep learning approach towards price forecasting using enhanced convolutional neural network in smart grid,” in *Proc. Int. Conf. Emerg. Internet Working, Data Web Technol.* Cham, Switzerland: Springer, 2019, pp. 271–283.

[13] K. Amarasinghe, D. L. Marino, and M. Manic, “Deep neural networks for energy load forecasting,” in *Proc. IEEE 26th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2017, pp. 1483–1488.

[14] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, “A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization,” *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.

[15] A. W. Services. (2018) *How Does Cloud Computing Work?*. [Online]. Available: https://aws.amazon.com/what-is-cloud-computing/
SALEH ALBAHLI received the B.Sc. degree from Qassim University, Saudi Arabia, in 2005, the MIT degree from The University of Newcastle, Australia, in 2010, and the Ph.D. degree from Kent State University, USA, in 2016, all in computer science. He is currently an Associate Professor with the College of Computer, Qassim University. He has attended various conferences, symposiums, workshops, training, and presented many seminars. He has more than 15 years of experience in both the IT industry and academia in Saudi Arabia, Australia, and USA. His research interests include big data analytics, data science, machine learning, and optimization strategies.

MUHAMMAD SHIRAZ (Member, IEEE) received the master’s degree in computer science from Allama Iqbal Open University at Islamabad, Pakistan, in 2007. He is currently pursuing the Ph.D. degree, under the Commonwealth Scholarship Program, with the Faculty of Computer and Information Technology, University of Malaya, Malaysia. He is also an Assistant Professor with the Department of Computer Science, Federal Urdu University of Arts, Sciences and Technology at Islamabad, Islamabad, Pakistan. He is also an Active Researcher with the Mobile Cloud Computing Research Group, FCSIT. His current research interests include distributed applications design for ubiquitous networks, distributed systems, lightweight applications, smart client applications and optimization strategies, and mobile cloud computing. He is a Reviewer of the Journal of Network and Computer Applications.

NASIR AYUB received the M.S. degree in software engineering from COMSATS University Islamabad, Islamabad, in 2019, and the degree from the Federal Urdu University of Arts, Science and Technology at Islamabad. He was an Active Researcher with the Communication over Sensors Research Lab, COMSATS University Islamabad. He is currently a Lecturer with the Department of Computer Science, Federal Urdu University of Arts, Science and Technology at Islamabad, Islamabad, Pakistan. He is also working on the electricity theft detection in the Smart Grid. His current research interests include energy management, big data, machine learning, and optimization strategies.

* * *