International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 2, April 2020, pp. 1524∼1532
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i2.pp1524-1532

Power consumption prediction in cloud data center using machine learning

Deepika T, Prakash P
Department of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India

Article Info

Article history:

Received Jun 6, 2019
Revised Oct 17,2019
Accepted Oct 25, 2019

Keywords:

Cloud computing
Machine Learning
Physical Machine
Power consumption prediction
Virtual Machine

ABSTRACT

The flourishing development of the cloud computing paradigm provides several services in the industrial business world. Power consumption by cloud data centers is one of the crucial issues for service providers in the domain of cloud computing. Pursuant to the rapid technology enhancements in cloud environments and data centers augmentations, power utilization in data centers is expected to grow unabated. A diverse set of numerous connected devices, engaged with the ubiquitous cloud, results in unprecedented power utilization by the data centers, accompanied by increased carbon footprints. Nearly a million physical machines (PM) are running all over the data centers, along with (5 – 6) million virtual machines (VM). In the next five years, the power needs of this domain are expected to spiral up to 5% of global power production. The virtual machine power consumption reduction impacts the diminishing of the PM's power, however further changing in power consumption of data center year by year, to aid the cloud vendors using prediction methods. The sudden fluctuation in power utilization will cause power outage in the cloud data centers. This paper aims to forecast the VM power consumption with the help of regressive predictive analysis, one of the Machine Learning (ML) techniques. The potency of this approach to make better predictions of future value, using Multi-layer Perceptron (MLP) regressor which provides 91% of accuracy during the prediction process.

Corresponding Author:

Deepika T.,
Department of Computer Science and Engineering,
Amrita School of Engineering, Coimbatore,
Amrita Vishwa Vidyapeetham, India.
Email: t_deepika@cb.students.amrita.edu

1. INTRODUCTION

Cloud computing is a technological advancement that furnishing with everything as a service such as storage space to the user, networking, server as well as applications. Infrastructure as a Service(IaaS), Software as a Service(SaaS), and Platform as a Service(PaaS) are the different types of service models, in Cloud computing that can be delivered on demand. Cloud providers offer a pool of virtualized computational resources to customers in the data center, in a pay-as-you-go manner [1]. The virtualized computing services provide IaaS that helps reduce the installation and maintenance cost for computing environments. A cloud data center is associated with a group of connected physical machines (PM) or host used by the organizations for network processing, remote storage and access to enormous data. The data centers are the backbone for the cloud environments. The exponential growth of cloud computing, because of emerging technologies like IIOT (Industrial Internet of Things) applications, big data evolution, and 5G functionality. In 2020,

50 billion connected devices will be in the Internet of Things (IoT) field, the amount of internet traffic as per second is 51974 GB [2]. Consequently, the cloud service providers like AWS, Google cloud and Azure are motivated to extend data centers across the globe to provide on-demand services.

The virtualization technique plays a major role in the data centers - facilitate sharing resources among customers through VMs. Each virtual machine is isolated and used to execute customer applications with the following requirements including its storage capacity, main memory, CPU, I/O capabilities and network bandwidth [3, 4]. Consolidation of physical machines, fault tolerance, and load balancing are some of the key factors that improve cloud computing performance. The PM consolidation occurs through Virtual Machine (VM) migration, when the unavailability of the requested resources by virtual machine from the physical machine, relocation of the virtual machine will take place. The VM is relocated to another physical machine, fulfill the need for VM [5]. The proposed method forecast the power of each VM preliminary to VM migration, based on this prediction and resource availability of PM, then VM migrated to particular PM. The VM power prediction escalates the system availability, minimizes the infrastructure complexity, and reduces the operational cost for cloud providers which helps the customer to pay less amount [6, 7]. There is a need to forecast the VM's power in advance to manipulate the processes fastest and provide more reliable services to customers. The conservation of power can be accomplished through power forecast by applying various machine learning methods. In this work, the regression-based ML strategy is applied to forecast the power consumption of the virtual machine, to enrich the cloud computing infrastructure and to enhance service for IT industries. Moreover, the power utilization of VM's is predicted before the VM allotted to physical machines.

The outline of the paper's structure follows. Section 2 reviews past literature work on workload forecast of VM, resource management allocation based on various characteristics of VM. Section 3 deals with the framework for the VM power prediction based on regression-based methods. Section 4 illustrates the ML models and performance evaluation of the proposed approach, through empirical inspection, followed by closing remarks in Section 5 as a conclusion.

2. RELATED WORK

The background research knowledge in the cloud's virtual machine such as forecasting of CPU utilization, resource usage, and management is the effective approaches towards the future in advance. The power supply increases day by day, to run and cool down the utilized devices in the cloud data center and these phenomena increase the operational expenses of cloud service providers. The conservation of power by the data center, the various power aware methodologies were studied. Prediction of power consumption is used to estimate the non-linear future value for better performance of a complex function. Beloglazov and Buyya [8] have applied an Adaptive Threshold algorithm, Local Regression, and Robust Local Regression to evaluate overloaded server, based on CPU utilization in IaaS infrastructure. The threshold is adjusted automatically based on historical analysis of data, manipulate with estimator like Mean absolute deviation, interquartile range. Prevost et al., [9] focused on network load prediction using Autoregressive linear prediction and neural network. The data samples used in this method was less for training to learn the relationship between attributes. Chonglin et al., [10] presented a Tree Regression(TR)-based model to compute the VM power utilization, using cross-validation, based on black box method. The VM and server feature information are gathered based on black box method. They have considered data as linear values for their prediction model. Jingqi et al., [11] presented the Linear Regression method to forecast the workload of cloud services. They also performed the autoscaling process reducing the operational cost of virtual resources through vertical and horizontal scaling. Jitendra et al., [12] proposed a self-adaptive differential evolution algorithm to estimate the workload utilized by the cloud data center using NASA trace and Saskatchewan trace. The authors reviewed fitness function, mutation, and crossover carried out in this method, which was better than other approaches like Particle Swarm Optimization (PSO), Genetic Algorithm (GA) and so on. This method required to minimize the Service Level Agreement (SLA) violations for better service processing.

Flavien et al., [13] explored the challenges, in a cloud environment, to diminish the power consumption of VMs and the operational expenses for cloud vendor. They implemented the ad-hoc framework for VM consolidation; but this approach did not take into account VM requirements like disk space, network bandwidth and time taken by VM to complete a particular task. Hao Xu et al., [14] investigated the power of VM with normalized parameters that satisfy the correlation coefficient of VM's power using Radial Basis Function (RBF) Neural Network. This method used a small number of samples for training and testing data, which

could not get an accurate prediction in the neural network. The estimator used for calculating prediction error was average prediction and maximum prediction error. Minal Patel et al., [15] proposed the Support Vector Regression (SVR) and Autoregressive Integrated Moving Average (ARIMA) method to predict the dirty pages of VM during live migration and determine the migration time of VM depend on time series analysis. The ARIMA model is applied to reduce the dirty pages, network traffic, and memory size based on past statistical data. This approach has less capability to ascertain the built-in features because it formed with single hidden layer as shallow neural network structure. Cortez et al., [16] applied ML algorithms to estimate the resource management of VM, in the cloud platform using the characteristics of Azure workload such as the first party for IaaS and third party for PaaS services. The authors exploited the Fast Fourier Transform to find the category of VM workload and plotted the graph for CPU, memory, CPU core usage per VM and lifetime of VM, using cumulative distribution function. This method used the dynamically linked library(DLL) to accumulate the result after each prediction, while in the next prediction, it checks whether the forecast was valuable using the score of the DLL.

Verma et al., [17] analyzed the workload of VM in order to minimize the power consumption of VM using supervised learning algorithms. They listed the various scheduling approaches to reduce carbon dioxide emissions from data centers. The statistical metrics such as RMSE, R squared and accuracy accomplished with an algorithm to calculate the prediction error. Chang et al., [18] applied the recurrent neural network to forecast and manage the resource allocation to a cloud server. They compared the servers workload prediction results with Time-Delay Neural Network(TDNN) and Regression methods. Witanto et al., [19] proposed the adaptive selector neural network to select the algorithm for reduction of the active VM and compared the results with Linear regression. This method was also focused on Service level agreement(SLA) between customer and cloud service provider but still, SLA is not fulfilled when the customer requirements vary. The above-mentioned literature outline exhibits the potential of machine learning to predict various problems in cloud computing for future evaluation. The aforementioned related works are tabulated below in Table 1.

Table 1. Comparison of algorithms for VM resource requirements prediction

Author(s)	Method	Goal	Weakness	Performance better than
John J. Prevost et al.,(2011)	Auto Regressive Linear prediction	Network load prediction	Need for extension to multiple resources	-
Beloglazov et al., (2012)	Adaptive Threshold algorithm Local Regression, Robust Local Regression	Predict overloaded server based on CPU utilization	Multiple migration not discussed	Heuristic algorithm
Jingqi Yang et al.,(2014)	Linear Regression	Workload prediction of service cloud	Network load is not considered	Hidden Markov process
Chang et al., (2014)	Neural network	Resource allocation	SLA violation	TDNN and Regression method
Chonglin et al., (2015)	TR-based method	Compute VM power	Examine only Linear values	Linear Regression Regression tree
Hao Xu et al., (2016)	RBF Neural Network	VM power prediction	Considered less VM samples	Short term prediction models
Minal Patel et al.,(2016)	ARIMA	Dirty pages prediction in VM	Slow execution	Support vector regression
Verma et al.,(2017)	Supervised learning methods	Forecast the VM's workload	Operational time of VM and CPU usage not taken into account	Gaussian process, Ridge Regression and so on
Cortez et al., (2017)	Gradient boosting tree, Random Forest	Resource management	Resource exhaustion	-
Jitendra et al., (2018)	Self-adaptive differential evolution algorithm	Workload prediction	Improve the SLA for better prediction	Particle Swarm Optimization, Genetic algorithm
Witanto et al.,(2018)	Adaptive selector Neural Network	Resource management	SLA varies with different QOS requirements	Local Regression

3. SYSTEM MODEL

The aspects of the proposed method to forecast the power utilization of VM in a proactive manner. Figure 1 shows an overall framework for the proposed system which focuses on the prediction of VM power utilization. The proposed framework is comprised of different components, which includes cloud information service module, resource provisioner module, machine learning module, and decision-making module. A multiple VM request from the customer is registered in cloud information service module to deploy their system and application. The resource provisioner module allocates the resources to the virtual machines based on the decision of cloud manager, whenever needed. This module is responsible for satisfying the service request for customers according to the service level agreement(SLA). The ML module inspects the repository of VMs historical data and then selects the data for training and testing phase. These are retrieved by the decision-making module for power prediction. The cloud management monitors the other modules and takes the decision in the appropriate situation. The cloud data center consists of connected hosts in which each host is allocated with multiple VMs. The virtual machine monitor (VMM) is a layer which controls each virtual machine located in the physical machines. The VMM receives the result from the cloud manager and allocates the VM to the preferable PM. The unexpected creation of virtual machine instance in the physical host or assignment of a task to existing VM, ensue in changes of VM attributes; consequent fluctuations in power consumption occur in the corresponding physical host. In this scenario, the power anomalies can be regulated, through prediction, at any point from the historical data, before the change in power consumption.

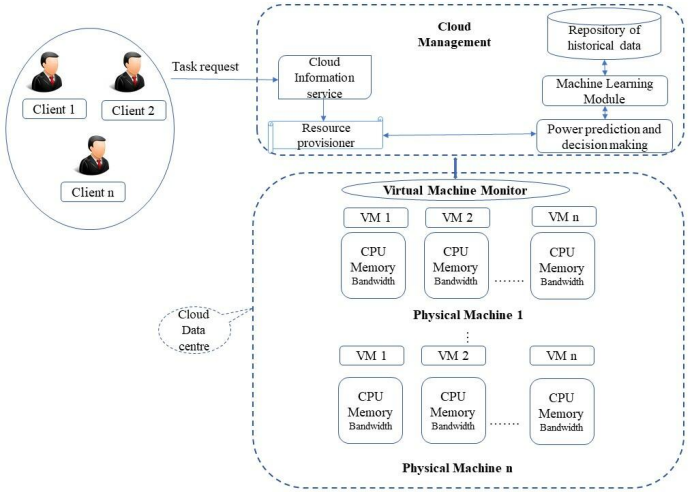

Figure 1. Framework of the proposed system

4. POWER PREDICTION OF VM BY APPLYING MACHINE LEARNING

4.1. Forecasting methods

The machine learning models are used to learn the features of the dataset in a flawless manner, to forecast the VM metrics like CPU, memory, and power. The complex correlation between the input variables can be handled by effective learning algorithms among the massive amount of traced data contains the numerous VM. The dataset can be handled with normalization, feature selection and find the relationship among features, through correlation method. The supervised machine learning algorithms will predict the target variables based on input and output variables [20, 21]. The raw dataset contains the target variable as a continuous value; so, it comes under the category of Regression predictive model. The Regression model is used to predict the response variable from analyzing the relationship between multiple independent variables and one dependent variable [22]. The regression model can be assessed through the root mean square error using the formula noted below

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \widehat{y_j})^2} \quad (1)$$

where 'n' is the number of observations in the raw dataset, 'y_j' is the forecasted value and '$\widehat{y_j}$' is the actual value of the observation. The various regression methods have been trained and tested on the dataset, to generate the RMSE value, using the prediction method between the test data and regression models. For both training and testing data based on the metrics, provide the output as a score of the model, prediction error, running time, performance consistency and so on. The best model is selected based on the score for prediction of attributes such as VM's CPU(MHZ) usage, Memory(GBs) usage, and Power consumption. The following regression methods were used in the prediction process.

4.2. Regression types

The shrinkage algorithms like Least Absolute Shrinkage and selection operator regression, Ridge regression are effective for multicollinearity problem. The variables in the dataset are highly correlated with each other that results in poor prediction, can be overcome by these algorithms. The Elastic Net is the hybrid of Lasso and Rigid methods. The aforementioned algorithms are regularisation techniques to avoid the overfitting of data [23]. The result of the Lasso, Ridge, and Elastic Net Regression are compared with other regression methods.

4.3. Random forest regressor

Random Forest is one of the ensemble ML algorithms used for regression analysis. It uses the bagging technique and selects the features for best node splitting and to construct the multiple decision trees subsequently averaging the value of all decision tree to predict the accuracy [24]. This approach will learn how to predict the future value with the help of currently observed value. The RMSE metric is used to calculate the difference between the real observed value and forecasted value by the regressor.

4.4. K nearest neighbor (KNN) regression

The K nearest neighbor forecasts the power utilized by each VM, and based on the feature similarity, it collects the average of the training test. The distance metric, Euclidean distance defines the distance between the new value and training value, using the formula

$$Euclidean \quad distance = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \tag{2}$$

For tuning the hyperparameter 'k', KNN uses the k-fold cross-validation to choose the right value of k, and sum up all losses of each 'k' value, to estimate the score of the algorithm. The cost function of 'k' drops in some period of time, and again increase it further whilst find the 'k' value using the elbow method. Figure 2 depicts the value of RMSE decreases while increasing the 'k' value. The optimum value of 'k' is determined through parameter tuning to achieve a better score of the KNN regressor algorithm.

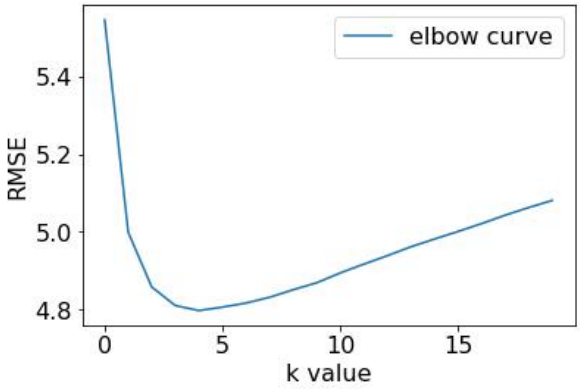

Figure 2. Optimum value of k

4.5. Multi-layer perceptron (MLP) regressor

The MLP is the precursor for an artificial neural network. An MLP Regressor uses back propagation to train the data based on the perceptron, which consist of an input layer, hidden layer, and the output layer. The neurons composed, in each of the layer and hidden layer with activation function to produce output for a given input node or neuron; in this model, it uses the Relu activation function within the hidden layer. The performance of MLP Regressor improved high while compared to other models.

4.6. EXPERIMENTS AND RESULTS

4.6.1. VM power

The utilization of the power in a virtual machine can be computed with the power consumption VM's CPU, VM's memory, VM's IO, and so on [25]. The equation for VM power calculation is defined as below,

$$P_{VM} = P_{CPUUtilize} + P_{Memory} + P_{IO} \tag{3}$$

where P_{VM} is the amount of power consumed by VM, $P_{CPUUtilize}$,P_{Memory}, P_{IO} are the power consumption of VM requirements such as CPU, memory and IO respectively.

4.6.2. Data model

The raw dataset, collected from Azure VM workload, contains VM requirement details like CPU utilization for minimum and maximum usage, memory space, CPU core, and VM lifetime and so on, available in Github [16]. Over ten lakhs of VM's are monitored, and collected data from each VM, for 24 hours per day, for four months continuously. Every VM's detail relates to five-minute VM CPU utilization readings and other features. The task/data-driven model uses this dataset to forecast the power consumption of VM, with the help of error estimators.

4.6.3. Performance evaluation

The input to the different regression algorithms is raw dataset divided as 80% for training and 20% for the testing set. The performance of each method is validated. The evaluation will be how far into the future prediction value. The proposed method depicts the performance of machine learning models along with the accuracy of testing and training data.

4.6.4. Result and analysis

The dataset with a collection of different services for VM types, based on their CPU, and memory usage, the better forecasting method was chosen, based on the score value of each machine learning algorithm.

Figure 3,4 depicts the amount of CPU and memory utilization per service. Figure 5 illustrates the plotting of learning curve to exhibit the predict performance through validated error and training error for random forest regressor analysis. The prediction error and score of the results are used investigating the overall performance of the algorithms. Figure 6,7,8,and 9 represents the effect of different machine learning models on the actual value, and predicted the value, of the random VM's power. Figure 10 shows the better performance of the MLP regressor model correctly predicts the actual value for each record of VM. This profound approach observes the past VM data, and evaluates the upcoming service, to achieve the goal of the prediction process.

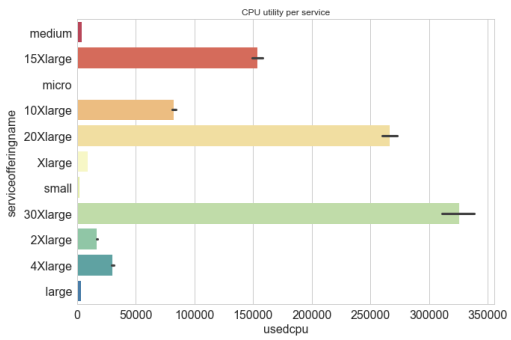

Figure 3. CPU utilization

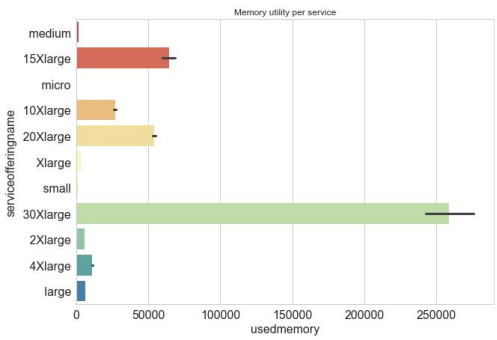

Figure 4. Memory usage

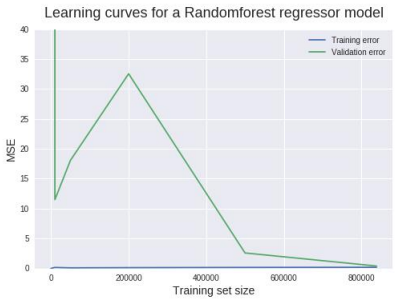

Figure 5. Random forest learning curve

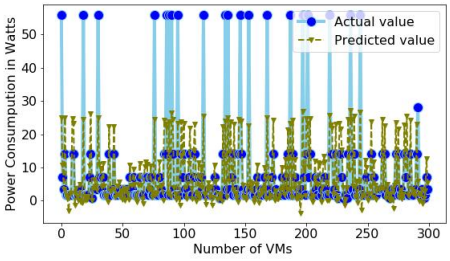

Figure 6. Lasso, rigid and elastic net regression

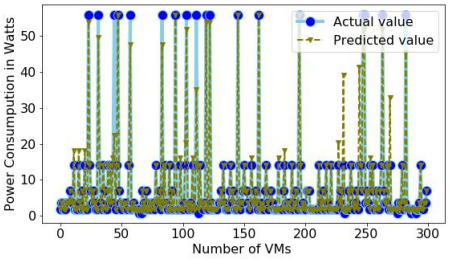

Figure 7. KNN regressor

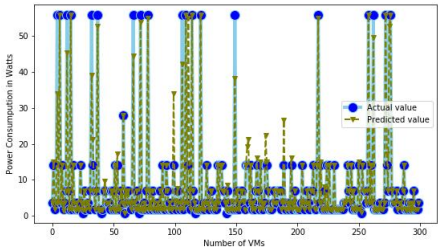

Figure 8. Random forest regression

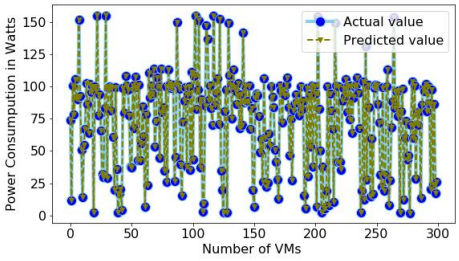

Figure 9. MLP regressor

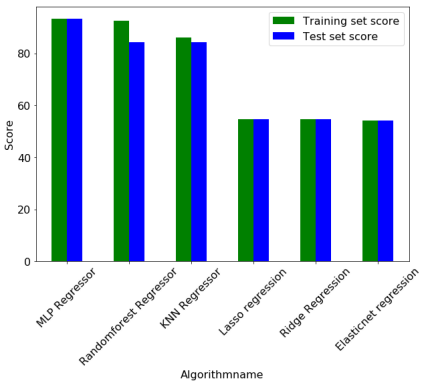

Figure 10. Comparison of all algorithms

5. CONCLUSION

In this paper, proactive methods can forecast the sudden fluctuation in power consumption, due to the changes in VM attributes, ahead of time, happens through historical performance data of a large number of VM's. The Regression based different machine learning algorithms were tested in the historical dataset to predict the VM power consumption. The MLP regressor model estimates the actual power value of VM to overcome the future uncertainty in power consumption of VM and the proposed framework has a potency to proceed the cloud manager as proactive to forecast the future power consumption of VM for efficient power management in the cloud data center. This scenario in cloud computing technic implies to provide a reliable environment to customers. The future enhancement will allow the data center to understand the characteristics of VM in advance with better prediction and VM migration model which leads to power consumption.

REFERENCES

[1] Ismaeel, R. Karim, and A. Miri, "Proactive dynamic virtual-machine consolidation for energy conservation in cloud data centres,"*Journal of Cloud Computing*, vol. 7, no. 1, p. 10, 2018.

[2] K. Mason, M. Duggan, E. Barrett, J. Duggan, and E. Howley, "Predicting host cpu utilization in the cloud using evolutionary neural networks," *Future Generation Computer Systems*, vol. 86, pp. 162-173, 2018.

[3] P. Prakash, G. Kousalya, S. K. Vasudevan, and K. K. Rangaraju, "Distributive power migration and management algorithm for cloud environment," *Journal of Computer Science*, vol. 10, no. 3, p. 484, 2014.

[4] F. Zhang, G. Liu, X. Fu, and R. Yahyapour, "A survey on virtual machine migration: Challenges, techniques, and open issues," *IEEE Communications Surveys and Tutorials*, vol. 20, no. 2, pp. 1206-1243, 2018.

[5] N. Janani, R. S. Jegan, and P. Prakash, "Optimization of virtual machine placement in cloud environment using genetic algorithm," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 10, no. 3, pp. 274-287, 2015.

[6] Z. Usmani and S. Singh, "A survey of virtual machine placement techniques in a cloud data center," *Procedia Computer Science*, vol. 78, pp. 491-498, 2016.

[7] H. Zhao, J. Wang, F. Liu, Q. Wang, W. Zhang, and Q. Zheng, "Power-aware and performance-guaranteed virtual machine placement in the cloud,"*IEEE Transactions on Parallel and Distributed Systems*, vol. 29, no. 6, pp. 1385-1400, 2018.

[8] A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers," *Concurrency and Computation:Practice and Experience*, vol. 24, no. 13, pp. 1397-1420, 2012.

[9] J. J. Prevost, K. Nagothu, B. Kelley, and M. Jamshidi, "Prediction of cloud data center networks loads using stochastic and neural models," in *2011 6th International Conference on System of Systems Engineering*, pp. 276-281, IEEE, 2011.

[10] C. Gu, P. Shi, S. Shi, H. Huang, and X. Jia, "A tree regression-based approach for vm power metering," *IEEE Access*, vol. 3, pp. 610-621, 2015.

[11] J. Yang, C. Liu, Y. Shang, B. Cheng, Z. Mao, C. Liu, L. Niu, and J. Chen, "A cost-aware auto-scaling approach using the workload prediction in service clouds," *Information Systems Frontiers*, vol. 16, no. 1, pp. 7-18, 2014.

[12] J. Kumar and A. K. Singh, "Workload prediction in cloud using artificial neural network and adaptive differential evolution," *Future Generation Computer Systems*, vol. 81, pp. 41-52, 2018.

[13] F. Quesnel, H. K. Mehta, and J.-M. Menaud, "Estimating the power consumption of an idle virtual machine," in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pp. 268-275, IEEE, 2013.

[14] H. Xu, X. Zuo, C. Liu, and X. Zhao, "Predicting virtual machine's power via a rbf neural network," in *International Conference on Swarm Intelligence*, pp. 370-381, Springer, 2016.

[15] M. Patel, S. Chaudhary, and S. Garg, "Machine learning based statistical prediction model for improving performance of live virtual machine migration," *Journal of Engineering*, vol. 2016, 2016.

[16] E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura, and R. Bianchini, "Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms," in *Proceedings of the 26th Symposium on Operating Systems Principles*, pp. 153-167, ACM, 2017.

[17] N. Verma and A. Sharma, "Workload prediction model based on supervised learning for energy efficiency in cloud," in *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, pp. 66-71, IEEE, 2017.

[18] Y.-C. Chang, R.-S. Chang, and F.-W. Chuang, "A predictive method for workload forecasting in the cloud environment," in *Advanced Technologies, Embedded and Multimedia for Human-Centric Computing*, pp. 577-585, Springer, 2014.

[19] J. N. Witanto, H. Lim, and M. Atiquzzaman, "Adaptive selection of dynamic vm consolidation algorithm using neural network for cloud resource management," *Future Generation Computer Systems*, vol. 87, pp. 35-42, 2018.

[20] C. Sammut and G. I. Webb, *Encyclopedia of machine learning and data mining*, Springer, 2017.

[21] H. Brink, J. Richards, and M. Fetherolf, *Real-world machine learning*, Manning Publications Co., 2016.

[22] L. Breiman, *Classification and regression trees. Routledge*, 2017.

[23] G. James, D. Witten, T. Hastie, and R. Tibshirani, "Linear model selection and regularization," in *An*

introduction to statistical learning, pp. 203-264, Springer, 2013.
[24] J. Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, and K. Li, "A parallel random forest algorithm for big data in a spark cloud computing environment," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 4, pp. 919-933,2016.
[25] Z. Jiang, C. Lu, Y. Cai, Z. Jiang, and C. Ma, "Vpower: Metering power consumption of vm," in *2013 IEEE 4th International Conference on Software Engineering and Service Science*, pp. 483-486, IEEE, 2013.

BIOGRAPHY OF AUTHORS

Deepika T received the B.Tech and M.E degrees from Anna University, in 2010 and 2012, respectively, where she is currently pursuing the Ph.D. degree in Computer Science and Engineering, Amrita School of Engineering, Coimbatore. His research interests include Cloud Computing, Machine Learning and Image Processing.

Dr. Prakash P received the Ph.D. degree in Information and Communication Engineering from Anna University, in 2016. He is currently serving as Assistant Professor at department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore. His research interests include Cloud Computing, Big data analytics, Automata Theory and Analysis of Algorithms. He is also exploring the integration and data analysis of Internet of Things (IoT) with cloud computing.