International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Enhanced Machine Learning Model for Electricity Price Forecasting for Cloud Computing

K. Niprutha[1], D. Venkata Siva Reddy[2] B.E., MBA, M.Tech, (Ph.D.)

[1]Department Of Computer Science, Besant theosophical College, Madanapalli, Andhra Pradesh, India

[2]Assistant Professor, Head of Department of Computer Science, Besant Theosophical College, Madanapalli, Andhra Pradesh, India

Article Info
Volume 9, Issue 2
Page Number: 99-106

Publication Issue :
May-June-2021

Article History
Accepted : 10 June 2021
Published: 23 June 2021

ABSTRACT

In the IT industry cloud computing is rapidly gaining traction because it eliminates the need for physical computing hardware, which are instead hosted by companies providing cloud services. These firms have a large number of computers and servers whose main power source is electricity. The design and maintenance of these companies therefore depends on the availability of a consistent and cost-effective electricity supply.Energy-hungry are cloud centres. One of the most recent electricity price increases means that maintenance of these centres is to minimise the electricity usage of datacenters and to save energy, as well as to create and efficiently store data placement and to schedule node to download or move storage. A way to solve these issues. In this project, we propose to load and move stocks, predict electricity costs and reduce energy costs in data centres, an Extreme Gradient Boosting (XGBoost) model.Data are divided into 70% training and 30% testing.

Keywords : — Cloud computing, Move stocksExtreme Gradient Boosting (XGBoost).

I. INTRODUCTION

The use of Cloud computing as a storage platform decreases investment in hardware and procurement. Exponential demand for data leads to proportional data center demand (DCs). DCs consume a great deal of power, representing 2% of global power use. It would increase by 12 percent annually [1],[2]. It should rise each year. Almost 39 percent of cooling power, 45 percent for IT infrastructure, and 13 percent for lighting [3] is used. In 2008, businesses cost that level in US$30 billion [4]. This consumption cost companies.

The use of distributed computing and virtualization can increase productivity according to a report by Walker. This approach is not very common, however. Unvirtualized servers use only 6-15% of their capacity, as stated in Ericsson. Virtualized servers are therefore able to use up to 30 percent of their ability [5]. DC operators generally have several DCs distributed over locations, to ensure reliability through replication. The latency requirements are satisfied by being close to customers. Distributed DCs, however, can cause cost uncertainty in different geographical areas due to the dynamic costs of electricity markets. The cost of these power markets may vary greatly. DC suppliers would therefore build DCs at low-temperature, low-electricity locations.

Power markets function in a deregulated environment and vendors can fix the prices attracting customers freely. Within 60 minutes, the volatile nature of the price of power may increase market instability and cost can increase ten times [6]. For example, Ontario's lowest observes in 2018 saw the cost of 4.39 CAD/MWh, with a maximum of 365.64 CAD/MWh [6]. Research into the use of volatility in the deregulated power value market is important with the expanded interest in cloud computing and instable electricity costs. In order to reduce energy cost, it helps predict value increases and consequently decreases power use throughout these periods.

Companies such as Netflix use Content Deliver Networks (CDNs). It would place data centres closer to customers in order to limit the need for long-haul data transmission and improve service quality [7]. This method can possibly be used to discharge capacity in order to reduce energy uses and ultimately to cut electricity costs from central DCs to hubs in the edge of the system.Due mainly to its importance and extreme need, the last few decades have made green environments and efficient energy consumption hot. Different scholars have used state-of-the-art techniques and traditional methods to solve these problems. For example [1] and [4] examined how electricity expenses can be reduced in multi-geography. Many investigators suggest that the cost of setting up servers at various locations should be assessed because the costs vary from one geographic location to another. There are also research projects such as [8] to improve the scheduling of nodes to reduce electricity costs.In [9] the selection of routes for data transmission was also enhanced by the researchers. In any case, the documents referenced recommend answers to the explicit questions rather than provide a competitive one-tone solution.

Likewise, many researchers concentrated on the various effects of machine learning methods on electricity prices, design and forecasting, and on the world market in particular. Generally, the first one is for predicting energy prices, and the second for energy systems, two machine learning techniques are used. Most of the recent methods use different aromas of deep neural networks, such as [10]-[16], Random Forest [17], Naive, and Decision Tree [18] [19], as well as other methods of machine learning, such as Support Vector Machine (SVM) [14]– [16].

In terms of accuracy, computational overhead or unable to prove results on reliable data, most of the previous works on electricity price prediction are still in its beginnings. In this article, we propose a model to measure the efficiency of the Ontario-Canada Rechenzentrum's prediction of electricity prices, mitigate the impacts of energy consumption and save significant costs. The effect of various risk factors on pricing increases for data storage can be analyzed using our prediction model and accurate energy consumption predictions. The model was evaluated using 15 years of IESO provider historical data to forecast the price of electricity.

The models were improved by XGBoost which showed that the electricity costs for the data storage are reduced by up to 25.32 percent and the future estimate and performance are acceptable in comparison to the forest random and support methods for vector machinery without additional computer overhead. In addition, the technique proposed is easy to apply and in real time prediction can be employed.

II. RELATED WORKS

Sustainability is the biggest concern today. The issues of electricity consumption and the green environment have been investigated in a vast array. A compact analysis of previous approaches used to forecast electricity use using different approaches is provided in this section. We will also highlight certain problems and weaknesses that have led us to provide a robust and effective solution to the available literature. The models used in the Multi-Layer Neural Network (MLNN) model to assess power load and overall consumption were used by Wang, Huai-jiang, Gang-qiang Li, Gui-bin Wang, Jian-chun Peng, Hui Jiang and Yi-tao Liu.[20]

The Ensemble technique was also used to eliminate various noise errors and cancelled noise. Despite the competitive advantage of their technology in terms of precision, the technique lacks strength due to higher computing times and an enormous loss rate when testing on real time data.Likewise, the hybrid technique used to predict electric power, known as EPNet, was proposed for Ping-Huan Kuo and Chiou-Jye Huang [1]. The technique involved an LSTM and CNN mesh, which delivered 8.84 MAEs and 17.9 MSEs. Despite the good results, the models offer huge error rates with enormous computer complexity in real time prediction. In addition, the data set is highly standardized and the model failed to produce similar data results in real time.

Umut Ugurlu, Ilkay Oksuz and Oktay Tas [21] have meanwhile proposed a similar model based on a mix of Gated and Long Short-Term Recurrence Units (GRUs) (LSTM). The model was completed with 5.71 MAE, but results are valid only for one day in advance. Moreover, the results are not consistent and can be different with different seasons, so that in real-time deployment this model is not effective. The models are also computer-cost.Similarly, DL-based analysis provided solutions for elected municipality consumption forecasts as well as a green environment, JesusLago, FjoDe Ridder and BartDe Schutter. [22]

The results of LSTM-DNN, GRU-DNN and CNN and MLP were evaluated and discussed and 23 other benchmarks were discussed. They also suggested a DL-based algorithm for the prescribed prediction task. The results of the technique proposed are comparatively better than the literature already predetermined. The comparison is however made with one highly standardized data set.The approach proposed is computer-computational and provides false forecasts of real time data set with enormous test loss.

By combining Sports Vector Machine (SVM) and Kernel Principle Component analyses Wang, Kun, Chenhan Xu, Yan Zhang, Song Guo, and Albert Y. Zomaya [23] used hybrid methodology to predictor electricity pricing (KPCA). The model was concluded by a 4.6 percent error with the smaller U threshold, while the larger U threshold value was 45.8 percent error. A large overview calculation is introduced, which helped to make the proposed technique more inadequate, due to the use of large data sets including the cost prices for steam, wood, wind, gas and oil.

In addition, prices change during the seasons and may vary from place to place, making the model very statical and dependent on location. The authors also used a net of DNN-based models for the same tasks [24]. Stacked Denoising Autoencoders (SDA). They provided the latest results with an enormous position-based dataset. In addition, [22] authors took advantage of the power of deep learning to improve the accuracy of predictions of electricity costs for the European industry. Functional analysis of the discrepancies in the selection of features and Bayesian optimization were also used.However, the value from the MSE and MAE values is huge and the method cannot handle the problem globally the model was rationally simple with efficient features.

In the course of the hourly prediction of electricity costs by multivariate models Raviv, Eran, Kees E. Bouwman, and Dick van Dijk [25]. They also conducted size reduction practises to moderate over-fitting effects. However, there are not comparatively better results for MAE and MSE, there are several false previsions. The DNN based model with LSTM mixed for electricity prices were proposed by Mujeeb, Sana, Nadeem Javaid, Mariam Akbar, Rabeiya Khalid, Orooj Nazeer, and Mahnoor Khan

[24].They also worked on the load forecast, but the results of the price estimate tasks are not satisfactory.
The majority of the literature focuses on advanced and deep education techniques, as described above. But some researchers, on the other hand, have also used traditional methods, such as the proposal to use a probabilistic framework to estimate energy per hour through the use of Generalized ELMs (GELM). The model for larger datasets is computationally costly and gives endless results.Likewise, authors focused instead of simplistic training on the selection of functions. They used data acquisition and mutual information (IG) as well as techniques for selecting features. They produced MAE of 4.09 after rigorous testing but only to predict huge datasets offline.
Likewise, the fusion-based technology for estimates and load prediction of electricity was anticipated by Ghasemi, Ali, Hossien Shayeghi, Mohammad Moradzadeh and Mohammad Nooshyar, for example [They have been using LSSVM (Last Square SVM) and Artificial Bee Colony Optimization (ABCO) algorithms.
They used different datasets from the Southern markets of NYISO, the USA and Wales. The above described models are computationally costly and offer huge losses to false predictions that ineffective in actual use. To summaries, the electrical predictions are available to estimate and reduce power consumption in DCs. A large number of literatures are available. Since last few decades, this has been a hot problem. However, current technology does not have the capacity to produce effectively low MAE and MSE results for the global market, most of them computationally costly and unable to operate in real time.

III. THEORY

A. CLOUD COMPUTING

Due to the revolution in cloud computing and how companies use it, the IT industry has changed significantly. Cloud computing on an on-demand computing resource consumes IT resources such as servers, networking, and databases. These resources can be host on the Internet instead of having local servers. Therefore, on-demand payment can minimise investment costs and capital (CAPEX) expenditures on local servers.
Businesses that have traditional data centres must pay a great deal in scale to support services and all requirement with their hardware, software and data storage. They pay less otherwise, but they do not satisfy the requirements of the user. Nevertheless, cloud computing users pay only for services that they require for personal or commercial use. Cloud computing also helps companies to focus on their goals and key business activities by using external technology to house all IT infrastructure to achieve business goals.

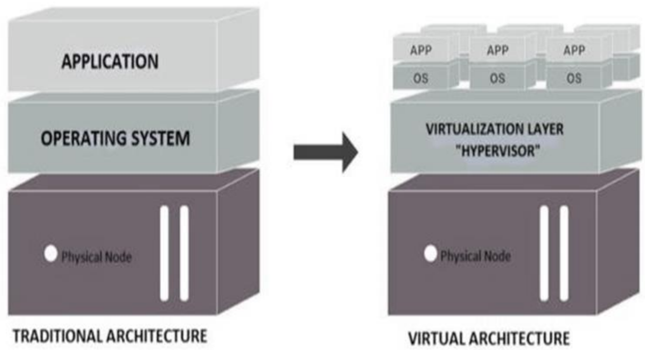

Figure: Traditional vs. virtualized server architecture

Cloud services can ordinarily be divided into three types of services: Service Infrastructures (IaaS), Service Platform (PaaS) and Service-as-Software (SS) (SaaS). IaaS is a key cloud computing service that allows a user to pay asylum. By providing an on-demand environment, PaaS develops, tests, delivers and manages software applications. SaaS is a monthly or annual software delivery to the service over the internet. We rely exclusively on IaaS in our paper.
A public, a private or a hybrid cloud platform may be described. Data centres for third-party providers like, for example, Amazon Web Services (AWS), Microsoft Azure and Google Cloud are available to the public (GCP). In a private environment, data

centres, usually privately hosted and can be used by the owner of the data center, in opposition to the public. Within the hybrid, data centres are combined with private and public cloud platforms which allow the sharing of information between public and private clouds.

An architecture of **virtualization** is an extremely important subject for servers. It helps to run multiple operating system platforms simultaneously on a single server. Data centres can more effectively use IT resources in relation to physical Servers and cost energy by creating different versions of virtual environments on a single machine. Figure 1 describes the concept of virtual architecture.

Figure 1 displays the cloud computing power virtualization architecture. On the same hardware, the virtualization can be mapped to several layers. This allows companies to manage workloads easily and possibly make the multiple layers more scalable than one machine. This helps companies to use IT resources in a data center efficiently. Whilst a single operating system and different applications on the same server can use the traditional architectures.

A **CDN** (Content Delivery Network) is a distributed network of edge servers at various locations which store high volumes of latency data. CDN services, the backbone of the Internet, deliver content today and continue to grow rapidly, use over half of all traffic. Examples of CDN servers are Amazon, Dropbox, Facebook and Netflix. As an example, Netflix shares information across the geographic area to reduce the distance between servers and users, ensuring the users access the data closest to a server instead of loading the data from their original servers, thereby improving its QoS [7].

In order to predict the target file on the target server in time, Netflix uses advanced algorithms. The local server therefore reduces bandwidth costs and allows data to be transmitted on a scale across the range. Since the hubs limit could be limited, the data repeatedly accessed should only be shifted to the edge. Netflix's output from 8 Gbps on a single server in 2012 to more than 90 Gbps on a single Server in 2016 [7] has now been extended by using this model.

B. MACHINE LEARNING

Machine learning is a method used by its algorithms to analyze data, to find models and to predict unseen data. By learning from previously processed data, it enables more efficient use of resources. A learning algorithm receives data without explicit programming, with the aim of building its own logic and improving its efficiency accordingly.

Machine learning is a means by its algorithms for analyzing data to search for patterns and then predict invisible data. By learning from previously processed data it enables resources to be used more effectively. An algorithm of machine learning receives data without explicit programming to create its own logic and to improve its performance. Essentially, the main problems with the machine learning are defined as being supervised or uncontrolled. Data set and output are provided for building a model and predicting future outputs in supervised learning. Whereas unattended learning provides information through input, but without labelled or classified data.

The paradigm of cloud computing can intelligently benefit from the functionality of machine learning. Machine learning techniques can effectively predict energy costs, as they are taken into account in related work. Therefore, energy management machine teaching techniques were applied to the estimation of future electricity prices [1]. This could be applicable. The electricity market has thus been influenced largely by predictions of electricity prices. However, in this article we create a model to predict energy prices, comparing the overall performance of SVR classifiers and XGBoost.

IV. METHODOLOGY

This work is split into four stages. First, we collect and analyze the data from different sources. Secondly, data has been explored in detail in order to understand and learn about various data characteristics. Third, with various machine learning classifiers, the data are predicted to create forecasts for the electricity price with a tuned system, helping in the fourth step further. But the clarified structure will be followed.

Ontario data - Canada was used in this by IESO provider.In order to improve the prediction of electricity prices, our model was created using 3 different machinery learning algorithms, especially XGBoost, Random Forest and Support vector machines:

- XGBoost
- Random Forest
- Support Vector Machine

In order to achieve a fair comparison between methods all classificators used the same training and tested data separation. We used the function train test split to divide up. The test size = 0.3 within capacities shows the level of data that should be kept for testing. Most of it is approximately 70/30 or 80/20. We have applied K- Cross validation technology to prevent over-riding and under-riding to make sure the comparison between models is fair.

It was observed during the assessment that K = 3 were most suitable because more folds will take more memory and since we take less than K, the variance is less likely to cause error. To understand how much a standard setting of the XGBoost model provided by sci-kit learning should be used, a basic model with various quantities of data was executed.

We used Root Square Error, Mean Absolute Percentage Error (MAPE), MSF and MSF as assessment metrics to evaluate these machine models. In order for us to evaluate these models we have used the Root Mean Square Error (RMSE). The MAE and RMSE can be used together in many estimates to analyse the difference in errors. The RMSE depends more or less on the MAE; the more significant the difference of the individual error in the example data is. The more important is the contrast between them. If the RMSE is unlikely to be the same as MAE, every error will be similar at this point. MAE and RMSE may range from 0 to 0, respectively. They are unfavorably positioned: Better are the lower estimates.

After measurements of data (x1, x2,. ., xn) data and data-set(y1, y2,.,., yn) forecast data we will be able to form MSE and MAE, where x is current and y values are forecast as follows.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right|$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_j - y_j)^2}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

We take account of the confusion matrix with True Positive (TP) (FN) defined as a Spike Prediction in order to further evaluate Spike Prediction.

$$TP = \frac{TP}{TP + FP}$$

If FP is false positive it is much better if the TP value is close to 1.The above equation shows that the positive sample (say Class 1) is delegated to the point at which the information provided.

$$FN = \frac{FN}{FN + TP}$$

where FN value is considered better nearer to 0.At the time a negative example (let's say class 0) is delegated to the information provided but this is really a positive precedent. In our XGBoost analysis, we found that with a learning rate of 0.1, max depth of 3 and number of estimates of 25, we were able to use this best result. The probability for true good was 0.31 and false negative was 0.68 which also indicates that the actual values and predicted values differ greatly.

Figure: An image of the interconnection map of the data centre. Cloud data centres are servers, with multiple nodes connected to each server. The energy cost was studied for every hour to see if the ability to download nodes was beneficial.

V. CONCLUSION

We propose a model for Ontario in this article: Canada's prediction on electricity prices. The principal objective of this research is to examine whether or not machine learning technology can be valuable to exploit a dramatic rise in electricity prices, to discharge data storage, in order to minimise cloud data centre energy consumption. Furthermore, we analyse the forecast result of the price of electricity in Ontario in 2003–2018.

However, the power of machine learning for predicting the data center's energy forecasting, which can further reduce price increases for data storage, is shown in our model. In particular, XGBoost shows significant savings in electricity costs and highest precision over other classifications (Random Forest and Support Vector Regression). There are thus other dimensions of extension to study various clusters, for example: neural networks. Furthermore, clustering can help detect a lower spike error and can then show our missed/unwanted spike error in the forecast.

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

[6]. (2018). IESO. [Online]. Available: https://www.ieso.ca

[7]. K. Florance. (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-ispsaround-the-globe-to-deliver-a-great-viewing-experience

[8]. S. Rahman, A. Gupta, M. Tornatore, and B. Mukherjee, "Dynamic workload 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. Chiaraviglio, F. Idzikowski, L. Wosinska, 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].N. Ayub, N. Javaid, S. Mujeeb, M. Zahid, W. Z. Khan, and M. U. Khattak, “Electricity load forecasting in smart grids using support vector machine,” in Proc. Int. Conf. Adv. Inf. Netw. Appl. Cham, Switzerland: Springer, 2019, pp. 1–13.

[16].J.-P. Liu and C.-L. Li, “The short-term power load forecasting based on sperm whale algorithm and wavelet least square support vector machine with DWT-IR for feature selection,” Sustainability, vol. 9, no. 7, p. 1188, Jul. 2017.

[17].Q. Huang, Y. Li, S. Liu, and P. Liu, “Short term load forecasting based on wavelet decomposition and random forest,” in Proc. Workshop Smart Internet Things (SmartIoT), 2017, p. 2.

[18].C. C. Yang, C. S. Soh, and V. V. Yap, “A non-intrusive appliance load monitoring for efficient energy consumption based on naive Bayes classifier,” Sustain. Comput., Informat. Syst., vol. 14, pp. 34–42, Jun. 2017.

[19].S. Ø. Ottesen, A. Tomasgard, and S.-E. Fleten, “Prosumer bidding and scheduling in electricity markets,” Energy, vol. 94, pp. 828–843, Jan. 2016.

[20].H.-Z. Wang, G.-Q. Li, G.-B. Wang, J.-C. Peng, H. Jiang, and Y.-T. Liu, “Deep learning-based ensemble approach for probabilistic wind power forecasting,” Appl. Energy, vol. 188, pp. 56–70, Feb. 2017.

[21].U. Ugurlu, I. Oksuz, and O. Tas, “Electricity price forecasting using recurrent neural networks,” Energies, vol. 11, no. 5, p. 1255, May 2018.

[22].J. Lago, F. De Ridder, and B. De Schutter, “Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms,” Appl. Energy, vol. 221, pp. 386–405, Jul. 2018.

[23].K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya, “Robust big data analytics for electricity price forecasting in the smart grid,” IEEE Trans. Big Data, vol. 5, no. 1, pp. 34–45, Mar. 2019.

[24].L. Wang, Z. Zhang, and J. Chen, “Short-term electricity price forecasting with stacked denoising autoencoders,” IEEE Trans. Power Syst., vol. 32, no. 4, pp. 2673–2681, Jul. 2017.

[25].E. Raviv, K. E. Bouwman, and D. van Dijk, “Forecasting day-ahead electricity prices: Utilizing hourly prices,” Energy Econ., vol. 50, pp. 227–239, Jul. 2015.

[26].D.Venkata shiva Reddy, "Enhancing Data Privacy and Efficiency for Secure Cloud Storage Model", Journal of Emerging Technologies and Innovative Research (JETIR),vol.6 (4), 2019.