Forecasting Daily Stock Movement Using A Hybrid Normalization In Ann

M Abinaya [1], J Ancy John[2]
[1, 2] Dept of CSE
[1, 2] Arunachala College of Engineering for Women

Abstract- *The potential financial benefits of stock market forecasting have drawn a lot of interest. Due to the various interconnected aspects, predicting these markets is a difficult endeavour that necessitates a thorough as well as efficient feature selection procedure to discover the highest useful aspects. Stock price changes are also influenced by previous trading days' movements, which is a time series problem. In stock forecasting, feature selection techniques are commonly used, although most known systems use a single feature selection methodology that probably can neglect some key notions about the regression function that is at the root of the problem relating the variables for input and output. This study employs an artificial neural network (ANN) based generative model to forecast pricing changes in the future by combining features preferred by different feature picking strategies to build an ideal optimal feature group. They begin by calculating an expanded set of 83 technical indicators using day-to-day stock data of six stock indices, and then normalize them using the Hybrid-Normalization (HN) technique. The important features are selected using various types of feature selection techniques and then considering the common features for the stock movement prediction. For stock trend predictions, they used a variety of classifiers such as Support Vector Machine, K Nearest Neighbour and Artificial Neural Network and. The system was given a performance review after simulations were done on 6 stock indices from various portions of the international market. The outcomes show that joining highlighted features got by various feature choice calculations and taking care of them into a profound generative model beats best in class techniques.*

Keywords- Stock market, Hybrid normalization, Artificial neural network

I. INTRODUCTION

Deep learning is a branch of machine learning which is completely based on artificial neural networks, as neural network is going to mimic the human brain so deep learning is also a kind of mimic of human brain. In deep learning, they don't need to explicitly program everything. The concept of deep learning is not new. It has been around for a couple of years now. It's on hype nowadays because earlier they did not have that much processing power and a lot of data.

In human brain approximately 100 billion neurons all together this is a picture of an individual neuron and each neuron is connected through thousands of their neighbors. The question here is how they recreate these neurons in a computer. So, they create an artificial structure called an artificial neural net where they have nodes or neurons. They have some neurons for input value and some for output value and in between, there may be lots of neurons interconnected in the hidden layer.

- Deep Neural Network – It is a neural network with a certain level of complexity (having multiple hidden layers in between input and output layers). They are capable of modeling and processing non-linear relationships.
- Deep Belief Network(DBN) – It is a class of Deep Neural Network. It is multi-layer belief networks.
- Steps for performing DBN :
  - Learn a layer of features from visible units using Contrastive Divergence algorithm.
  - Treat activations of previously trained features as visible units and then learn features of features.
  - Finally, the whole DBN is trained when the learning for the final hidden layer is achieved.
- Recurrent (perform same task for every element of a sequence) Neural Network – Allows for parallel and sequential computation. Similar to the human brain (large feedback network of connected neurons). They are able to remember important things about the input they received and hence enables them to be more precise.

Stock analysis refers to the method that an investor or trader uses to evaluate and investigate a particular trading instrument, investment sector, or the stock market as a whole. Stock analysis is also called equity analysis or market analysis. Investors or traders make buying or selling decisions based on stock analysis information. Stock analysis helps traders to gain an insight into the economy, stock market, or securities. It involves studying the past and present market data and creating a methodology to choose appropriate stocks

for trading. Stock analysis also includes the identification of ways of entry into and exit from the investments.

Stock analysis can be grouped into two broad categories:

Fundamental Analysis

The fundamental stock analysis method involves the evaluation of a business at a basic financial level. Investors use fundamental analysis to determine whether the current price of a company's stock reflects the future value of the company.

Fundamental analysis uses different factors such as the current economic environment and finances of the company to estimate its stock value. Different key ratios are also used to determine the financial health and understand the true value of a company's stock.

Technical Analysis

The technical analysis method involves examining data generated through market activities, such as volume and prices. Analysts following such a type of stock analysis use technical indicators and tools like charts and oscillators to identify patterns that can indicate future price trends or direction.

Technical analysts examine the historical trading data of a security and estimate the future move of the security. It is frequently used for forex and commodities.

II. LITERATURE SURVEY

The industry of data centers is the fifth largest energy consumer in the world. Distributed green data centers (DGDCs) consume 300 billion kWh per year to provide different types of heterogeneous services to global users. Users around the world bring revenue to DGDC providers according to actual quality of service (QoS) of their tasks. Their tasks are delivered to DGDCs through multiple Internet service providers (ISPs) with different bandwidth capacities and unit bandwidth price. In addition, prices of power grid, wind, and solar energy in different GDCs vary with their geographical locations. Therefore, it is highly challenging to schedule tasks among DGDCs in a high-profit and high-QoS way. A problem is formulated and solved with a simulated-annealing-based bio objective differential evolution (SBDE) algorithm to obtain an approximate Pareto-optimal set. The method of minimum Manhattan distance is adopted to select a knee solution that specifies the Pareto-optimal task service rates and task split among ISPs for DGDCs in each time slot.

Reducing energy consumption in cloud data centers is one of the prime issue in the cloud community. It reduces energy related costs and increases lifespan of high performance computing resources deployed in cloud data centers and it also helps in reducing carbon emissions. Along with energy efficiency, problem of task scheduling is also one of the important problem considered in cloud data centers and it belongs to NP-class problems. With the energy consumption consideration, problem of task scheduling becomes more complex to solve. Metaheuristic algorithms are proven to generate near optimal solutions for task scheduling problem but their scheduling overhead increases vastly as the number of tasks or number of resources increases. The primary objectives in this work are to reduce makespan and computing energy and secondary objectives are to reduce the energy consumed by the resources other than computing resources and reduce execution overhead associated with scheduler. Collectively these objectives guides HIGA for better energy efficiency and performance while reducing the number of required resources (i.e. active racks). It indirectly also reduces cooling energy as they can switch off the rack components (air blower, cooling inlets) once racks become idle. Simulation analysis has been performed over independent task applications as well as real-world scientific applications like CyberShake, Epigenomics and Montage.

High energy consumption in cloud data centers has become one of the main obstacles to green cities, and an urgent problem to be solved. So far, a large number of scheduling algorithms have been developed to reduce energy consumption for executing workflows. However, most existing algorithms have obvious defects in energy and resource efficiency, because they schedule workflow tasks to hosts directly and ignore that the host is so powerful that a single workflow task cannot make full use of its resources.

Data centers are energy intensive buildings that have grown in size and number to meet the increasing demands of a digital economy. This presents a bottom-up model to estimate data center electricity demand in the United States over a 20 year period and examines observed and projected electricity use trends in the context of changing data center operations. Results indicate a rapidly increasing electricity demand at the turn of the century that has significantly subsided to a nearly steady annual electricity use of about 70 billion kWh in recent years. While data center workloads continue to grow exponentially, comparable increases in electricity demand have been avoided through the adoption of key energy efficiency measures and a shift towards large cloud-based service providers. Alternative projections from the model illustrate the wide range in potential electricity that could be consumed to support data centers, with the US data center

workload demand estimated for 2020 requiring a total electricity use that varies by about 135 billion kWh, depending on the adoption rate of efficiency measures during this decade. While recent improvements in data center energy efficiency have been a success, the growth of data center electricity use beyond 2020 is uncertain, as modeled trends indicate that the efficiency measures of the past may not be enough for the data center workloads of the future.

Cloud can be defined as a new computing paradigm that provides scalable, on-demand, and virtualized resources for users. In this style of computing, users can access a shared pool of computing resources which are provisioned with minimal management efforts of users. Yet there are some obstacles and concerns about the use of clouds. Guaranteeing quality of service (QoS) by service providers can be regarded as one of the main concerns for companies tending to use it. Service provisioning in clouds is based on service level agreements representing a contract negotiated between users and providers. According to this contract, if a provider cannot satisfy its agreed application requirements, it should pay penalties as compensation. In this work, they intend to carry out a comprehensive survey on the models proposed in literature with respect to the implementation principles to address the QoS guarantee issue.

Energy efficiency has become one of the major concerns for today's cloud datacenters. Dynamic virtual machine (VM) consolidation is a promising approach for improving the resource utilization and energy efficiency of datacenters. However, the live migration technology that VM consolidation relies on is costly in itself, and this migration cost is usually heterogeneous as well as the datacenter. This work investigates the following bi-objective optimization problem: how to pay limited migration costs to save as much energy as possible via dynamic VM consolidation in a heterogeneous cloud datacenter. To capture these two conflicting objectives, a consolidation score function is designed for an overall evaluation on the basis of a migration cost estimation method and an upper bound estimation method for maximal saved power. To optimize the consolidation score, a greedy heuristic and a swap operation are introduced, and an improved grouping genetic algorithm (IGGA) based on them is proposed. Lastly, empirical studies are performed, and the evaluation results show that IGGA outperforms existing VM consolidation methods.

Green Computing is a recent trend in computer science, which tries to reduce the energy consumption and carbon footprint produced by computers on distributed platforms such as clusters, grids, and clouds. Traditional scheduling solutions attempt to minimize processing times without taking into account the energetic cost. One of the methods for reducing energy consumption is providing scheduling policies in order to allocate tasks on specific resources that impact over the processing times and energy consumption. In this work, they propose a real-time dynamic scheduling system to execute efficiently task-based applications on distributed computing platforms in order to minimize the energy consumption. Scheduling tasks on multiprocessors is a well know NP-hard problem and optimal solution of these problems is not feasible, they present a polynomial-time algorithm that combines a set of heuristic rules and a resource allocation technique in order to get good solutions on an affordable time scale. The proposed algorithm minimizes a multi-objective function which combines the energy-consumption and execution time according to the energy-performance importance factor provided by the resource provider or user, also taking into account sequence dependent setup times between tasks, setup times and down times for virtual machines (VM) and energy profiles for different architectures.

Green cloud computing has become a major concern in both industry and academia, and efficient scheduling approaches show promising ways to reduce the energy consumption of cloud computing platforms while guaranteeing QoS requirements of tasks. Existing scheduling approaches are inadequate for realtime tasks running in uncertain cloud environments, because those approaches assume that cloud computing environments are deterministic and pre-computed schedule decisions will be statically followed during schedule execution. In this work, they address this issue. They introduce an interval number theory to describe the uncertainty of the computing environment and a scheduling architecture to mitigate the impact of uncertainty on the task scheduling quality for a cloud data center. Based on this architecture, they present a novel scheduling algorithm (PRS1) that dynamically exploits proactive and reactive scheduling methods, for scheduling real-time, aperiodic, independent tasks. To improve energy efficiency, they propose three strategies to scale up and down the system's computing resources according to workload to improve resource utilization and to reduce energy consumption for the cloud data center. They conduct extensive experiments to compare PRS with four typical baseline scheduling algorithms. The experimental results show that PRS performs better than those algorithms, and can effectively improve the performance of a cloud data center.

Increasing demand for acquiring diverse range of services has led to the establishment of huge energy hungry cloud data centers all around the world. Cloud providers face with major concerns to reduce their energy consumption while

ensuring high quality of service based on the Service Level Agreement (SLA). Consolidation is proposed as one of the most effective techniques for online energy saving in cloud environments with dynamic workloads. This work proposes novel proactive online resource management policies to optimize energy, SLA, and number of migrations in cloud data centers. More precisely, this work proposes new prediction algorithm for determination of overloaded hosts as well as novel multi-criteria decision making techniques to select virtual machines. The results of simulations using CloudSim simulator shows up to 98.11 % reduction in the output metric which is representative of energy consumption, SLA violation, and number of migrations, in comparison with state of the art.

The routing problem in energyefficient networks for cloud computing. They propose a robust routing algorithm to reach the higher network energy efficiency, which is based on optimization problem. To attain the highly energy-efficient routing in energy-efficient networks for cloud computing, the link of low utilization is turned into the sleeping state to save the network energy. At the same time, the low link traffic is aggregated to the link with high utilization to enhance the link utilization and to sleep the links as many as possible. They present an optimized link sleeping method to maximize the number of the sleeping links. By targeting the network robustness, a weight adaptive strategy is brought forth to reduce the link congestion and enhance the robustness of the network. Simulation results indicate that our algorithm is effective and feasible to achieve energy-efficient networks for cloud computing.

III. PROPOSED SYSTEM

Stock markets are crucial in the organisation of current commercial frameworks, and anticipating these business sectors is critical for investors, as precise interpretations enable them to manage perils and make informed investment judgements. Stock markets are determined by a variety of aspects, including the economy, industry-specific characteristics, corporate outlook, investor psychology, and the policies of the government, which adds to the difficulty of assessing and forecasting these markets. Fundamental analysis and technical analysis are two types of stock forecasting methodologies. To ascertain a stock's inherent worth, investors turn to fundamental analysis. Technical analysis, on the other hand, assesses equities by looking at statistical trends established by activity in the market, such as previous volumes and prices. Technical analysis is often employed for ephemeral forecasting since the necessary data is released on a regular basis. Relevant information about stocks is retrieved in technical analysis by determining technical indicators from previous prices as well as volumes, and fed into a prediction model.

Furthermore, the data representation method has a significant influence on the efficiency of a learning model. As a result, altering unprocessed data before integrating it into a model often increases the learning model's performance. Thus, this phase will examine the prediction performance of six different stock indices using various feature selection strategies. They use a novel Hybrid Normalization technique and a new feature selection approach with the goal of improving the reliability of stock movement forecasts. K Nearest Neighbor, Artificial Neural Network and Support Vector Machine classification scores are also compared.

The realization of the methodology is discussed in this section. Primarily, it becomes vital to understand what the related models are, as well as their learning or performing capabilities.

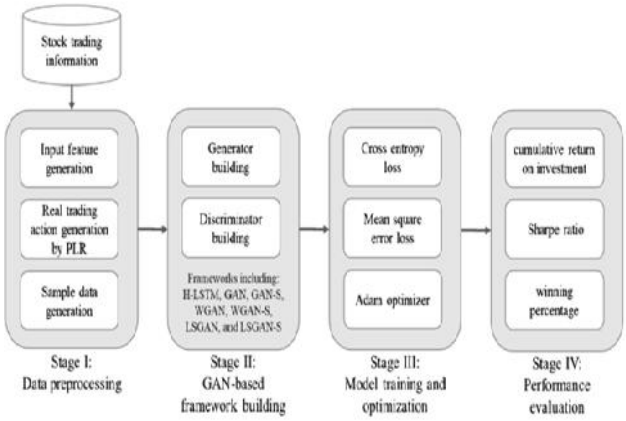


Fig 1 Block Diagram of Proposed System

Elucidate all the details of the listed deep learning models (LSTM, Bi-LSTM, ARIMA, GRU) and schematically represent them in subsequent sections, evaluating all the results, and interpreting the models sequentially as the deep learning methods are pivotal in this section. Then, explaining all the datasets which we used along with the proposed method workflow of this different algorithm will be presented and investigated. To get better estimation of the results, it is highly required to differentiate datasets for training and testing purposes. For which study about different epochs, batch sizes are carried out.

IV. MODULES DESCRIPTION

Data Collection

Primarily the foremost step is the collection or acquisition of a suitable dataset in many machine learning

predictive analytics projects. In the implementation of this project, stock market prices of the listed eight well-known tech firms or companies that seizes a large amount of market percentage, enlisted in the Indian stock market, required data was acquired from the Yahoo finance, beginning from January 2012 till including December 2022.

The attributes listed in the datasets of the stock rate were available as opening rate, last rate, high and low rates, and adjacent close rate of listed companies' stocks. The eight companies of which datasets have been used are listed as: (ITC) Indian Tobacco Company, (HCL) Hindustan Computers Limited, WIPRO, Redington, CYIENT, (L&T) Larsen and Toubro, (TCS) Tata Consultancy Services, (TECHM) Tech Mahindra.

For prevention of demolition of the present model of price shifting range, we selected stocks having total trading days more than 90% for the project. For the closing index, to fill in missing values we utilized the closing rate of trading day earlier.

Data Pre-processing

Before commencing with the model training, to prevent errors and outliers required data will be trained as per the rules of uniformity and consistency. As data collected will contain numerical and categorical variables, hence the raw data needed to be pre-processed for the removal of missing values and unstructured data. Data pre-processing especially for substantial data-based projects becomes a crucially important step for the elimination of unnecessary points and normalizing the data for everyday usage, data transformation improves quality and makes the data more usable. In this kind of projects data quality has given higher priority than huge quantity of data as accuracy of data results in more impactful output with its precision and accuracy. Involvement of data segregation, data cleaning or organization, data standardization, data scaling, etc., i.e., data standardization and normalization as well as encoding categorical data. Moreover, operations of data transformation also took place with machine learning and deep learning models.

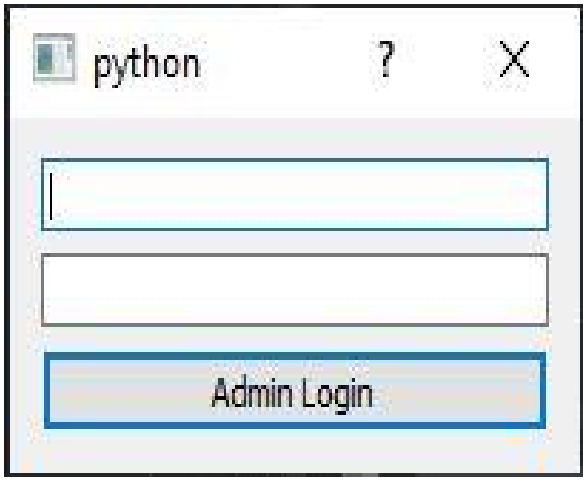
Training of Models

For better analysis of the expected long-term time series, one of the following methods will be applied before the training algorithm, in which the data is divided into various intervals for training. As discussed earlier, mentioned deep learning models which we are going to predict stock prices are- GRU, ARIMA, LSTM, and Bi-LSTM, and, in order to proceed before the data pre-processing step, it becomes mandatory to split the dataset into the training and testing datasets separately. In the process of this project, the dataset will split into the ratio of 80:20 training and testing sets. This ratio was chosen in such a way as to satisfy all the requirements of the model to avoid the volatility of data as it is a predictive analytics project. Since it involves time series prediction, hence it becomes necessary to train the model on most of the notable data owing correlation of data. For the cross-validation part, the data was also set aside. Splitting of training and testing datasets in the specific ratio is essential as it is more impactful in the performance of the models. Having large differences within dataset values can lead to under fitting while similar training data and test data can lead to the over fitting of models. Consequently, to get a real picture of the performance of various models it is vital to have an appropriate ratio of the train-to-test dataset which in turn would render reliable results.

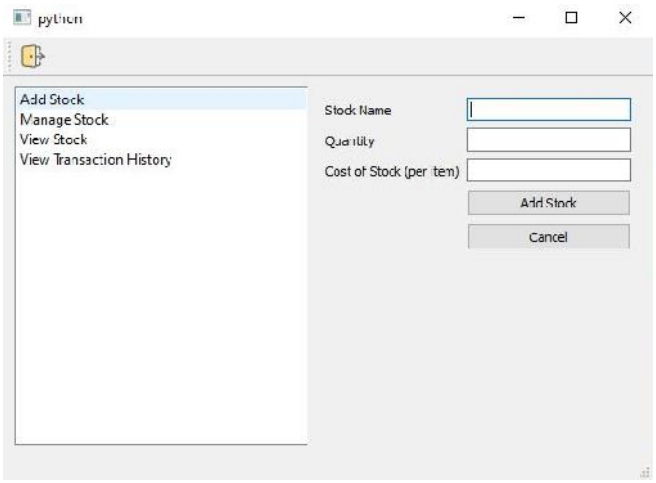
Testing of Models

In the last stage of the proposed work, which can be significant through the name itself that evaluation of the models' performance on trained dataset based on some efficient and useful performance parameters or metrics is being discussed in this section. For making tests of models as close to real-time as possible all the four models are tested with all the possible combinations by using 20% of total dataset for the models ARIMA, LSTM, GRU & Bi-LSTM. The evaluation parameters or metrics to check performance were Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). A testing parameter based on percentage errors or relative errors which is well-known and efficient is termed MAPEAPE. In MAPE, the accuracy of a given model be considered as highly predictive, if it is having a lower value. RMSE is defined as the square root of the mean of squared differences between actual and predicted values hence, stating model can be considered better if it is having lower RMSE value. MSE is used to check how close estimates or forecasts are to actual values hence, the lower the MSE, the closer is forecast to the actual.

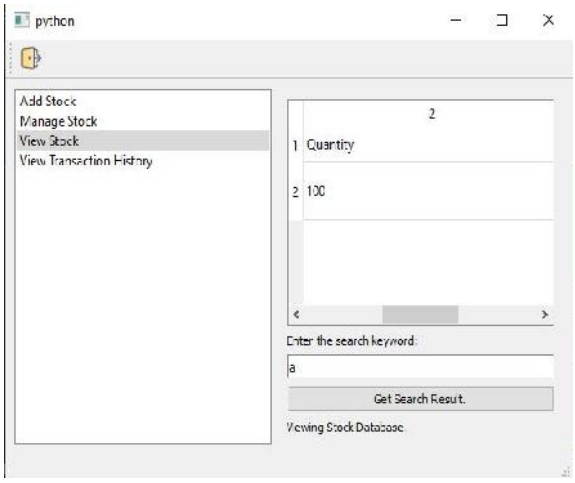
V. SCREEN SHOTS



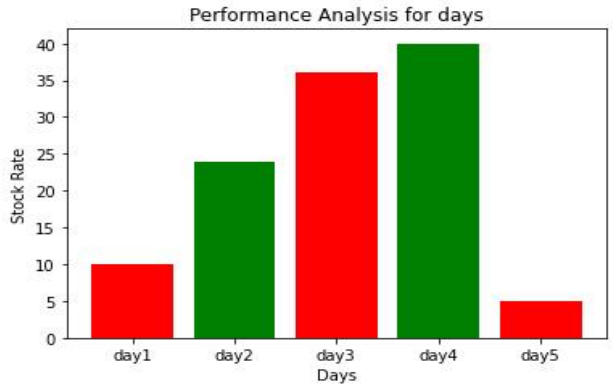
Login Page



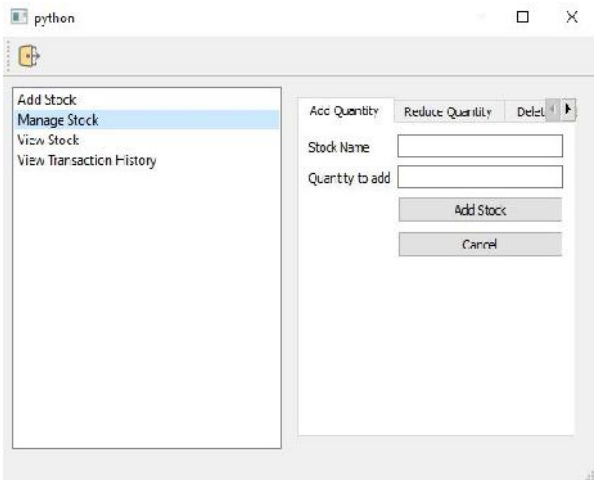
Add Stock Structure



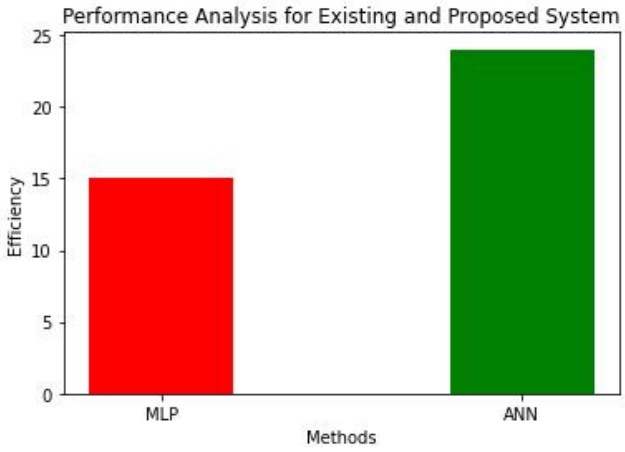
View Stock Structure



Performance Analysis of days



Manage Stock Structure



Comparison Graph for existing and proposed

VI. CONCLUSION

Stock forecasting is an important study issue that has gotten a lot of attention because of the possible financial rewards. The choosing of representative features and the development of a suitable prediction model are crucial for accurate stock price prediction. The machine learning

processes' accuracy outcomes are substantially impacted by the normalisation technique used for the input data normalisation. This work is a part of the pipeline we proposed for stock market prediction. Here the main focus is an ensembled approach for the feature selection by applying intersection operation on the optimal sets selected by different feature selection algorithms pertaining to three different categories namely filter, wrapper and embedded. Pipeline uses our published normalization technique that is Hybrid Normalization method. The efficiency of various feature selection procedures is therefore considered in this research. Thus, this work takes into account the goodness of various feature selection techniques. Feature selection, like other machine learning problems, is an eminent aspect of stock forecasting, but most studies use a single feature selection approach that may not take into account all notions about the elemental regression function that links the input variables to the output variables. As a result, merging features from several disjoint approaches will result in a more optimal feature set and improved prediction model performance. The number of features chosen for the optimal set is the minimum by the proposed Intersection Feature Selection method as it takes the goodness of different categories of feature selection techniques. The number of features chosen for the optimal feature set using IFS was significantly reduced to within a range of 24 – 30 features which was less in comparison to IG, FFS and LASSO. Thus it is found that Intersection Feature Selection gives better stock prediction accuracy when combined with the Hybrid Normalization technique. The study also found that ANN generated models outperformed SVM and KNN models for stock predictions. A variety of alternative feature selection strategies can be discovered in the literature that can be mixed to build a more optimal feature collection.

Future work, failure effects on the amounts of consumed power will be studied. Deep learning techniques will be used in order to highly predicate the utilization of servers and learn different related parameters to the scheduling and consolidation. Additionally, plan to enhance the scheduling algorithm with a load balancing technique.

REFERENCES

[1] Yuan, Haitao, Jing Bi, MengChu Zhou, Qing Liu, and Ahmed Chiheb Ammari. "Biobjective task scheduling for distributed green data centers." IEEE Transactions on Automation Science and Engineering (2020).

[2] Sharma, Mohan, and Ritu Garg. "HIGA: Harmony-inspired genetic algorithm for rack-aware energy-efficient task scheduling in cloud data centers." Engineering Science and Technology, an International Journal 23, no. 1 (2020): 211-224.

[3] Li, Chen, Lisu Huo, and Huangke Chen. "Real-time workflows oriented hybrid scheduling approach with balancing host weighted square frequencies in clouds." IEEE Access 8 (2019): 40828-40837.

[4] Shehabi, Arman, Sarah J. Smith, Eric Masanet, and Jonathan Koomey. "Data center growth in the United States: decoupling the demand for services from electricity use." Environmental Research Letters 13, no. 12 (2020): 124030.

[5] Ghahramani, Mohammad Hossein, MengChu Zhou, and Chi Tin Hon. "Toward cloud computing QoS architecture: Analysis of cloud systems and cloud services." IEEE/CAA Journal of Automatica Sinica 4, no. 1 (2019): 6-18.

[6] Wu, Quanwang, Fuyuki Ishikawa, Qingsheng Zhu, and Yunni Xia. "Energy and migration cost-aware dynamic virtual machine consolidation in heterogeneous cloud datacenters." IEEE transactions on Services Computing 12, no. 4 (2021): 550-563.

[7] Juarez, Fredy, Jorge Ejarque, and Rosa M. Badia. "Dynamic energy-aware scheduling for parallel task-based application in cloud computing." Future Generation Computer Systems 78 (2018): 257-271.

[8] Chen, Huangke, Xiaomin Zhu, Hui Guo, Jianghan Zhu, Xiao Qin, and Jianhong Wu. "Towards energy-efficient scheduling for real-time tasks under uncertain cloud computing environment." Journal of Systems and Software 99 (2020): 20-35.

[9] Arianyan, Ehsan, Hassan Taheri, and Saeed Sharifian. "Novel heuristics for consolidation of virtual machines in cloud data centers using multi-criteria resource management solutions." The Journal of Supercomputing 72, no. 2 (2019): 688-717.

[10] Jiang, Dingde, Zhengzheng Xu, Jindi Liu, and Wenhui Zhao. "An optimization-based robust routing algorithm to energy-efficient networks for cloud computing." Telecommunication Systems 63, no. 1 (2021): 89-98.

[11] Areekul, P., Senjyu, T., Toyama, H., Yona, A., 2020. A hybrid arima and neural network model for short-term price forecasting in deregulated market. IEEE Transactions on Power Systems Pwrs.

[12] Box, G.E.P., Jenkins, G.M., 2021. Time series analysis: Forecasting and control. Journal of Time 31, 238–242.

[13] Chandar, S.K., Sumathi, M., Sivanandam, S.N., 2020. Prediction of stock market price using hybrid of wavelet transform and artificial neural network. Indian Journal of Science & Technology 9.

[14] [Ding, X., Zhang, Y., Liu, T., Duan, J., 2020. Deep learning for event-driven stock prediction, in: Proceedings of the Twenty-Fourth International Joint Conference on

Artificial Intelligence, IJCAI 2020, Buenos Aires, Argentina, July 25-31, 2015, pp. 2327–2333.

[15] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y., 2019. Generative adversar-ial nets, in: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2019, December 8-13 2019, Montreal, Quebec, Canada, pp. 2672–2680.