iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44701

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Electricity Price Forecasting Using Artificial Neural Network

Jainam Gosaliya[1], Amit Andhare[2], Sayali Gulmire[3], Utkarsh Bhalerao[4], Shital Mehta[5]
[1, 2, 3, 4]*Student,*[5]*Prof., Department of Computer Science & Engineering, Dr. Babasaheb Ambedkar Technological University, Loner, India*

Abstract: ***In strength markets there' s no regularity in forecasting of strength rate that is the most vital tasks & foundation for any selection making. In the aggressive strength markets forecasting of the strength rate is essential for purchasers and producers of strength for making plans their operations and additionally to keep the danger of strength. Forecasting additionally performs a very crucial function in financial optimization of strength utilization. Artificial intelligence (AI) alongside with the ELM-Tree Approach has been implemented in rate forecasting that is, the day in advance utilization of the strength and also will expect the month-to-month invoice of the person as in step with their utilization. For complexing the interdependencies among the strength charges, historic load utilization and numerous different factors synthetic neural networks are the excellent technique that can be used. This synthetic neural community technique is used to expect the behaviours of the marketplace primarily based totally on the different factors including historic charges, portions, and different data to forecast the destiny charges and load.***

Keywords: Artificial Neural Network, Electricity Market, Price Forecasting, Decision Tree, Extreme, Machine Learning

I. INTRODUCTION

Prior to 2008, investments in large power-plant and energy infrastructure projects were largely led by big banks in the USA. Since the 2008 market crash, however, investments in these assets reduced significantly, and banks have been dissuaded from investing in long-term and risky energy assets. Nevertheless, energy in the United States has proven to be an innovative and profitable sector in recent years, and the sector started to attract new investors for the capital-intensive energy assets. This trend created a gap in the market between the significant demand for innovative energy projects with great potential for return and investors. The main reason for disconnected investment scene exists, both domestically and abroad, mostly as a result of lack of information and uncertainty about the projected financial performance and the main driving factor for financial return is the price of energy in the long term. The present research aims to get advantage of new powerful data analytics tools and techniques to produce a reliable prediction for the performance of energy particularly the monthly retail price of electricity years in advance. This analysis pipeline would serve as a long-needed tool for every- day investors and analysts to conduct state of the art risk analysis in the energy investment space. In addition to applications in institutional investments, the proposed methodology could aid everyday economists and researchers in making their forecasts and create a noticeable impact on policy makers in energy industry to determine the penetration level for demand response programs and other emerging sustainable energy sources. The energy industry is one that produces ample data thanks to the large-scale implementation of sensors, and the general interest of both the companies and the public. As a rather traditional industry, however, it does not tend to promote the use of cutting-edge data analysis techniques [1]. Internally, analysis is most often conducted by consulting companies who use simple forecasting tools in Excel, combined with their knowledge and intuition about the industry. Better analysis is done by independent data scientists on public data science platforms like Kaggle. However, the energy projects and the required data sets are not available comparing with the large online and offline collection of data in the energy industry. The problem is that there exists a massive gap between the analysis currently conducted and relied upon in the field, and the analysis that could potentially be done given the abundance of online and offline data available in the energy industry. In addition, emergence of simple and powerful machine learning techniques in the last few years initiated a trend in the traditional energy industry to get benefited from the massive pool of data to enhance the 2909 2019 IEEE PES Innovative Smart Grid Technologies Asia offered services and bring new opportunities for investors in the energy sector. The present research can offer analysis that can bridge this gap, provide researchers with long-term insights, and guide investors in viable long-term energy investments. The use of advanced tools can lead to data driven decision making for the investors and to help them to achieve more clarity in the next three to five years. Previous research related to the energy forecast focused on the load forecast.

©IJRASET: All Rights are Reserved

The reason is that the load forecast traditionally drives the investments for the energy infrastructure expansions and also the requirements for building new power plants was driven from the load forecast. In addition, the load forecast indirectly can affect the price of energy in the specific market. Some of the previous research and the applied method for the forecast is summarized in the following: Rahul [2] presented a novel method for long term load forecasting with hourly resolution. The model is fundamentally cantered on Recurrent Neural Network consisting of Long-Short-Term-Memory (LSTM-RNN) cells. The proposed model is found to be highly accurate with a Mean Absolute Percentage Error (MAPE) of 6.54 within a confidence interval of 2.25% which is favourable for offline training to forecast electricity load for a period of five years. Hossein investigated [3] the problem of long-term load forecasting for the case study of New England Network using several commonly used machine learning methods such as feedforward artificial neural network, support vector machine, recurrent neural network, generalized regression neural network, k-nearest neighbours, and Gaussian Process Regression. The results of these methods are compared with mean absolute percentage error (MAPE). Arild presents a framework for price forecasting in hydro- thermal power systems. The presented framework consists of a long-term strategic and a short-term operational model. This research, facilitate more detailed fundamental market modelling to enable realistic multi-market price forecasting. In addition, some of the technical constrained on the price such as cable ramping and reserve capacity are also considered for the price forecast [4]. The impact of load forecast on electricity pricing is considered in the research done by Baifu. In this research, a load forecasting model considering the Costing Correlated Factor (CCF) with deep Long Short-term Memory (LSTM). Also, this paper uses an adaptive Moment Estimation algorithm for network training and the type of neuron is Rectified Linear Unit (ReLU). Baifu concluded that LSTM with CCF can reduce energy cost with acceptable accuracy level [5]. LianLian presented an efficient method for the day-ahead electricity price forecasting (EPF) based on a long-short term memory (LSTM) recurrent neural network model. The applied method is capable of learning features and long-term dependencies of the historical information on the current predictions for sequential data. The proposed method is successfully applied for Australian market at Victoria (VIC) region and Singapore market [6]. Hongqiao proposed a robust predictive model construction and probabilistic forecasting with low-resolution data. In this research, the combination of high- and low-resolution data lead to more effective probabilistic forecast and improve the outcomes [7].

II. THEORY OF SVM, LSSVM AND BFOA

Theories regarding the main predictive tool, namely LSSVM, and BFOA as optimization algorithm are discussed in this section. Electricity price, as illustrated in [25], are highly volatile compared to the load series. Therefore, choosing the right feature affects the efficiency and accuracy of predictions. However, only publicly available data can be used for this purpose. The important input features for the prediction of electricity prices are selected as in [26]. It was found that many existing models of day-ahead electricity price forecast in Ontario were tested on year 2004 [7], [13], [27]– [29]. The analysis of price volatility for each testing period is performed in Section 4.1. Volatility analysis measures the price fluctuation of the test periods. Hence, the price behavior of the test periods should correspond to the volatility index measured. The final day-ahead forecast methodology is presented where LSSVM forecast engine is optimized by BFOA. A. SVM and LSSVM SVM could minimize over-fitting and local minima issues [30], and manage high-dimensional input space well. However, SVM involves high computation. The Least Squares Support Vector Machine (LSSVM) was therefore introduced to decrease the SVM computing load. LSSVM solves a linear equation system over quadratic programming problem (QP) which improves computation speed [31][31][34]. The Karush-Kuhn-Tucker (KKT) linear system is easier than the QP system. LSSVM even retains SVM capabilities that have a great generalization. B. BFOA E.coli bacteria living in the men gut have a unique food search activity. BFOA mimics this mechanism through four main processes; namely chemotaxis, swarming, reproduction, and elimination-dispersal. During chemotaxis, bacteria tumble or swim while avoiding dangerous places. They can also associate with other bacteria by sending attractive signals; or move individually by sending repellent signal. The objective function of each position of bacterium i is depicted by J. During the swarming step, the bacteria that have found the location of the nutrient may trigger nearby bacteria to create a herd. The notation for the objective function during the swarming process is the cell-to-cell attraction and repellent (Jcc). This will be added to the existing objective function, that will ultimately lessening the most recent objective function. Once nutrient is adequate and the environment is at a good temperature, the healthiest bacteria will grow in size and divide in the center to create their own duplication. This will contribute to future generations, while the most unhealthy bacteria will die. This process is referred to as reproduction. BFOA replicates this concept by sorting out the best MAPE in increasing order and retaining half of the total population for reproduction, while the other half will be destroyed.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The final process is elimination dispersal where chemotaxis mechanism will be eliminated and some bacteria will spread to different locations. C. Predicted Error Analyses The accuracy of this developed model is measured by Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), as defined in (1) and (2); respectively. Pactual and Pforecast are the real Hourly Ontario Electricity Price (HOEP) value and it forecast value at hour t, respectively, while N is the number of hours in a week.

III. MODEL DEVELOPMENT

The model developed is tested in the Ontario Electricity Market (OEM), that is managed by the Independent Electricity System Operator (IESO). Ontario is known for being one of the most volatile markets for electricity prices due to the adoption of a single settlement of real-time energy, therefore poses a great challenge to the price forecaster. Hence, the forecast model developed is compared with some of the current models for the same test data and power market [7], [13], [27]–[29]. OEM data is publicly accessible at http://www.ieso.ca/. The test data comprise of six weeks representing the three prominent seasons in year 2004 as shown in [26]. Volatility Analysis The following step is performing to measure the price fluctuation of each test periods. Volatility refers to the standard deviation of the arithmetic or logarithmic result of the T period. The value of volatility index is lies between 0 to1, where the price series is highly volatile when volatility index close to 1. However, most studies of volatility analysis adopted the logarithmic return [32], [33]. The logarithmic return, over the time h, is formulated in (3): where; pt = price at time t. Hence, the volatility index for a day (24 hour), σh,24(d) is defined as (4): where d is the index for the day,c൘ച൘ ക̄ · ൘൘ is the average of the logarithmic return for each day, and h = 1, to represent hourly logarithmic return. Thus, σ1,24 is denoted as a price changeover the day from one hour to the next.cB. Input Feature Selection The next step is choosing the significant input features to train the LSSVM. Correlation analysis was carried out to seethe correlation among price and some other input features as shown in [34]. HOEP and demand for the preceding 15 days care used as predictive inputs to represent daily and weekly impacts and also to inhibit losing some significant input [6],c[13], [35]–[40]. Therefore, the number of input used is [(15cdays x 24 hours price) + (15 days x 24 hours demand) = 720]c[34]. The optimization process of BFOA is initiated with random positions for each bacterium in the 722 dimensions, which representing 720 input features, gamma, and sigma's. LSSVM-BFOA Flow of Optimization BFOA has four main processes or loops; representing chemotaxis, swarming, reproduction, and elimination-dispersal. The brief flow of optimization process is illustrated as in Fig. 1[41]. The outermost loop is the process of elimination and dispersal, while the process of chemotaxis takes place within the innermost loop. During the chemotaxis process, the optimized LSSVM parameter values and the number of inputs selected by the BFOA will be trained in the LSSVM to generate fitness values (J) or MAPEs. During chemotaxis, bacteria tumble or swim. When tumbling, bacteria rotate clockwise, while during swimming, bacteria rotates counterclockwise. Random direction is created during tumbling, which represented as a vector Δ. By associating the Δ and the step height of C, new position of the bacterium is created. Each time the bacteria changed their positions, the new J will be calculated after having LSSVM training. Similarly, the new Jcc will be calculated to be added with the current J which would ultimately reduce the real MAPE. The other way of bacteria's movement is swimming, which helps the bacteria move faster. Similarly, by associating Δ with the step height of C, each bacterium changes position. Therefore, the optimized number of inputs and values of the LSSVM parameters will be trained in the LSSVM to produce J. Thereafter, Jcc is calculated and then added to the current J to reduce the final error. Swimming is stopped when the current MAPE is higher than the previous MAPE, or the total number of swims is approached. The chemotaxis process continues for other bacteria, and repeated for a predefined number of chemotaxis.

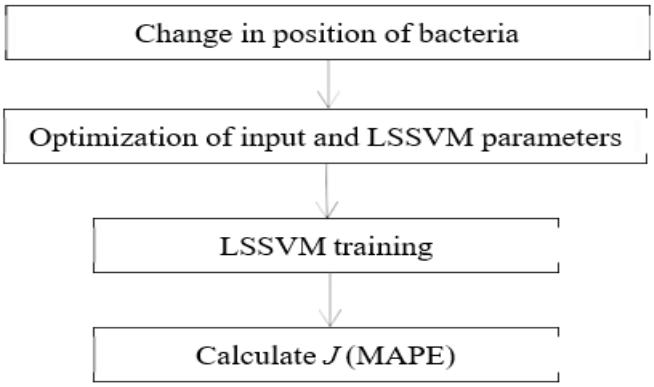


Fig 1**:** Flow chart of LSSVM-BFOA optimization

IV. CASE STUDY

A. Data Collection and Cleaning

Despite the ample availability of public energy data on the web, obtaining consistent and reliable energy price data for a controlled geographic region was a major challenge. Data was gathered from tens of sources and merged, then compared with that published by each hub belonging to energy companies in California. Inconsistencies were found in almost every data table obtained from merging data from various sources. Finally, reliable energy data set is prepared after the comprehensive data cleaning in the Energy Information Administration's OpenData database.

B. Feature Correlation Analysis

The following factors were extracted for use as input features in the analysis. They were chosen due to their high correlation with electricity price. Numbers indicate correlation of featurewith Average Monthly Retail Electricity Price:

- Natural Gas Consumed, Electric Power Sector (.85)
- Electricity Generated by Coke (-0.66)
- Net Electricity Generation (.81)
- Net Electricity Imports (.75)
- Natural Gas Consumed by Industrial Sector (.74)
- GDP (.7)
- Renewables (.77)
- Solar (.71)
- Geothermal (.65)
- Wind (.62)
- Biofuels (.72)

The correlation of unrelated factors that were highly correlated with the average price of electricity were surprising. The interesting findings was related to the high correlation of the consumption and production of renewable energy features to the retail price of electricity including solar, geothermal, wind, and biofuel and basically overall renewable generation were heavily correlated with the retail price of electricity. Our findings seems to agree that there exists a relationship explaining the influence of the price of electricity on the adoption of renewables. As detailed in Figure 1, the correlation plots produced by Natural Gas Consumption and GDP (both highly correlated with electricity prices) are scatter plots with a clear positive correlation (a positive trend is shown in the plots). However, the correlation plot produced by Solar Consumption and Production, Wind Consumption, and the production and consumption of Renewables as a whole (Figure 2) indicate a curve wherein the increasing price of electricity seems to cause a reactionary increase in the adoption of renewable energy.

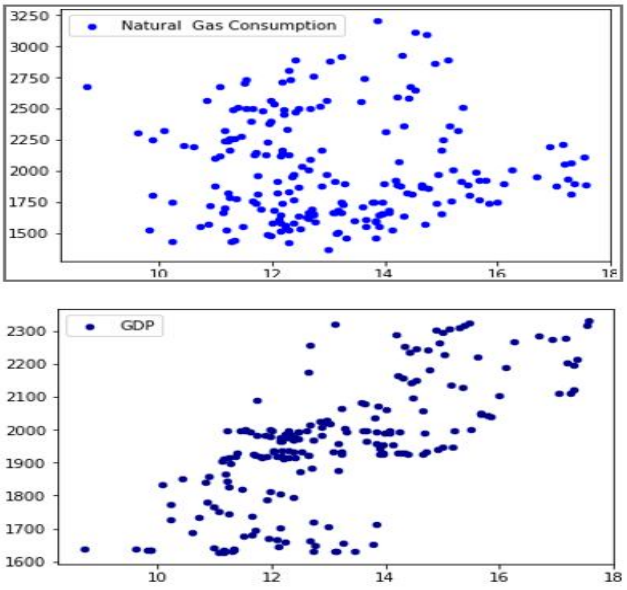


Fig 2: Natural Gas Consumption and GDP vs. Electricity Price

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Unlike the well-established fossil fuels industry, where a number of non-quantitative factors (political climate, foreign trade, etc) unpredictably influence the price of electricity, the main driver behind trends in production and consumption of renewable energy is demand; when electricity is cheap, there is little incentive for people to incur the heavy overhead costs of buying renewable solutions such as solar panels. And, as electricity gets more expensive and renewables become more affordable, people feel increasingly comfortable switching due to more promising savings. Due to its relative renewable- friendliness, California's energy prices would likely be much more predictable. Conclusively, more than a third of our most highly-correlated features connected to renewables. Our analysis indicates that the monthly price of electricity, when controlled for region, has a high seasonality with a relatively predictable increase of average prices occurring each year for the recent years. Our 3-year ARIMA time series predictor was able to capture complex trends to a great degree and performed extremely well, with an average monthly error of less than half a cent on the 2014-2017 test set, and an accuracy within a cent of the true price in ~90% of the months (Figure 3). Our 5-year ARIMA predictor, however, overfittedto trends coming up to 2012 and projected prices to beconsistently higher than the actual.

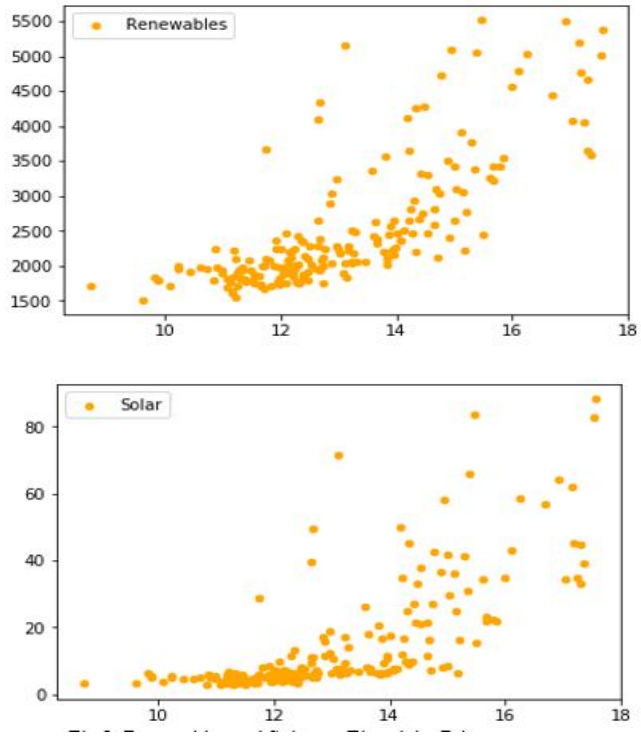


Fig 3: Renewables and Solar vs. Electricity Price

We believe this error to be the result of a diminished train set of 2001-2012, and the the fact that 2013-2014 were particularly trendsetting years in renewable and electricity pricing; with no data during those years, the algorithm produced a less reliable prediction.

A. Feature-based Model Prediction:

After preprocessing, S-ARIMA was used to synthetically augment the data from the 11 selected features from 2001- 2014 in order to create projected values to compare with the 2014-2017 test set. The plots below present monthly projected values in 2014-2017 overlayed with the true values of 8 of 11 (most reliable) features, in addition to further projection into the future. Upon performing linear regression on the augmented dataset, we obtained MSE = 0.015, MAE = 0.096 and RMSE = 0.125, training accuracy as 84.9% and test accuracy as 64% which suggested some degree of overfitting. Narrowing down to the 6 most seasonal features and training several other models yielded drastically improved results. Results from the 4 best performing algorithms are presented below.

1) Logistic Regression: MAE = 11.68, MSE = 267.68,RMSE = 16.36.
2) SVM: MAE = 16.58, MSE =433.33, RMSE = 20.81
3) KNN: MAE = 9.68, MSE = 136, RMSE = 11.66
4) Rand Forest: MAE = 10.86, MSE = 201.38, RMSE =14.19

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

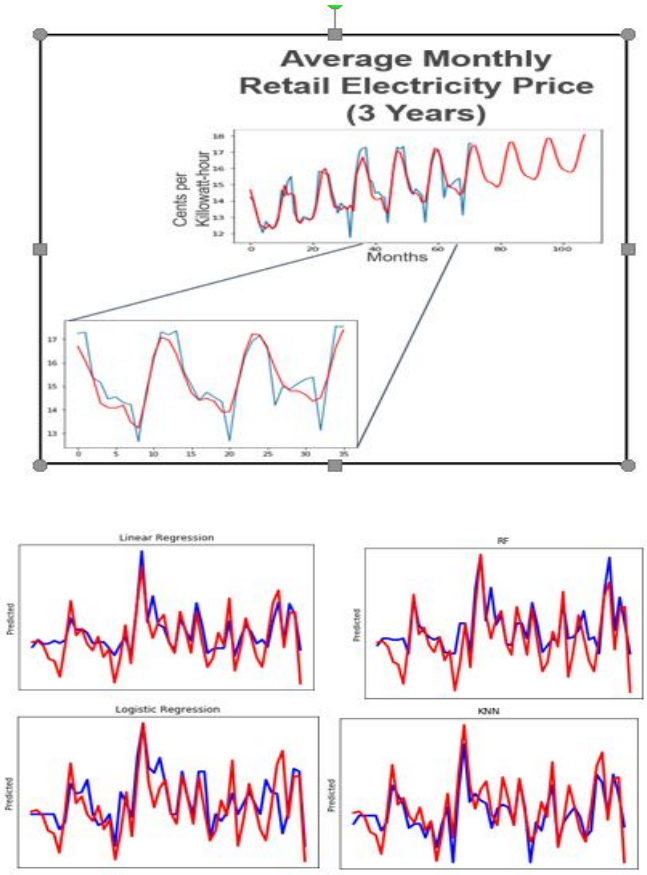


Fig 4: Performance Comparison of Machine Learning Algorithms

The graphs above illustrate the predicted plot in blue and actual plot in red. When the number of features were reduced to 6 by eliminating the 5 less reliable datasets, the Training accuracy fell to 82% and the test accuracy increased to 69%. Reducing the number of features in this case reduced overfitting and yielded in better test accuracy. The best performing algorithm, logistic regression using the 6 most reliable features, was able to outperform the 3 year and 5 year S-ARIMA time series predictors, with an MSE (Mean Squared Error) of only 0.22. The achived results outperformsthe reviewed literature and the industry standard.

V. CONCLUSIONS

The availability of energy data and the emergence of simple yet powerful machine learning algorithms in recent years, energy price forecasting has become easier and more accurate than ever before. Our findings suggest that state of the art analysis in this space can be conducted using widely available open data sources, feature discovery techniques and machine learning. A difficult aspect of energy forecasting, and a fundamental limitation of this technique, is that there are massively influential factors that simply cannot be predicted; the political/economic climate, weather patterns, OPEC and international policy changes all drive the price of electricity on a daily basis. One solution we would like to explore in the future, given more time and resources, is to look at a way to provide several perturbance-flexible estimates to be tuned with domain knowledge (in Economics or Politics, for instance). Using the time series data augmentation method described above, we hope to develop a dynamic model capable of changing weights and adding or adjusting the influence of features over a particular time bracket in the future, and providing several possible outcomes based on qualitative parameters like the probability of beneficial or adversarial political or economic events occurring at a given time. Conceivably, this model would utilize online learning from databases updated in real time providing constant feedback into the pipeline, and would be able to translate real world economic changes to a corresponding change in the feature set and weights used in the model. This would allow the user to consider different scenarios by changing the contributing factors of the prediction. Logical next steps in this direction would be to develop a consistent method for quantifying perturbations, and to investigate how to incorporate and translate domain knowledge into changes in the prediction algorithm's parameters.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

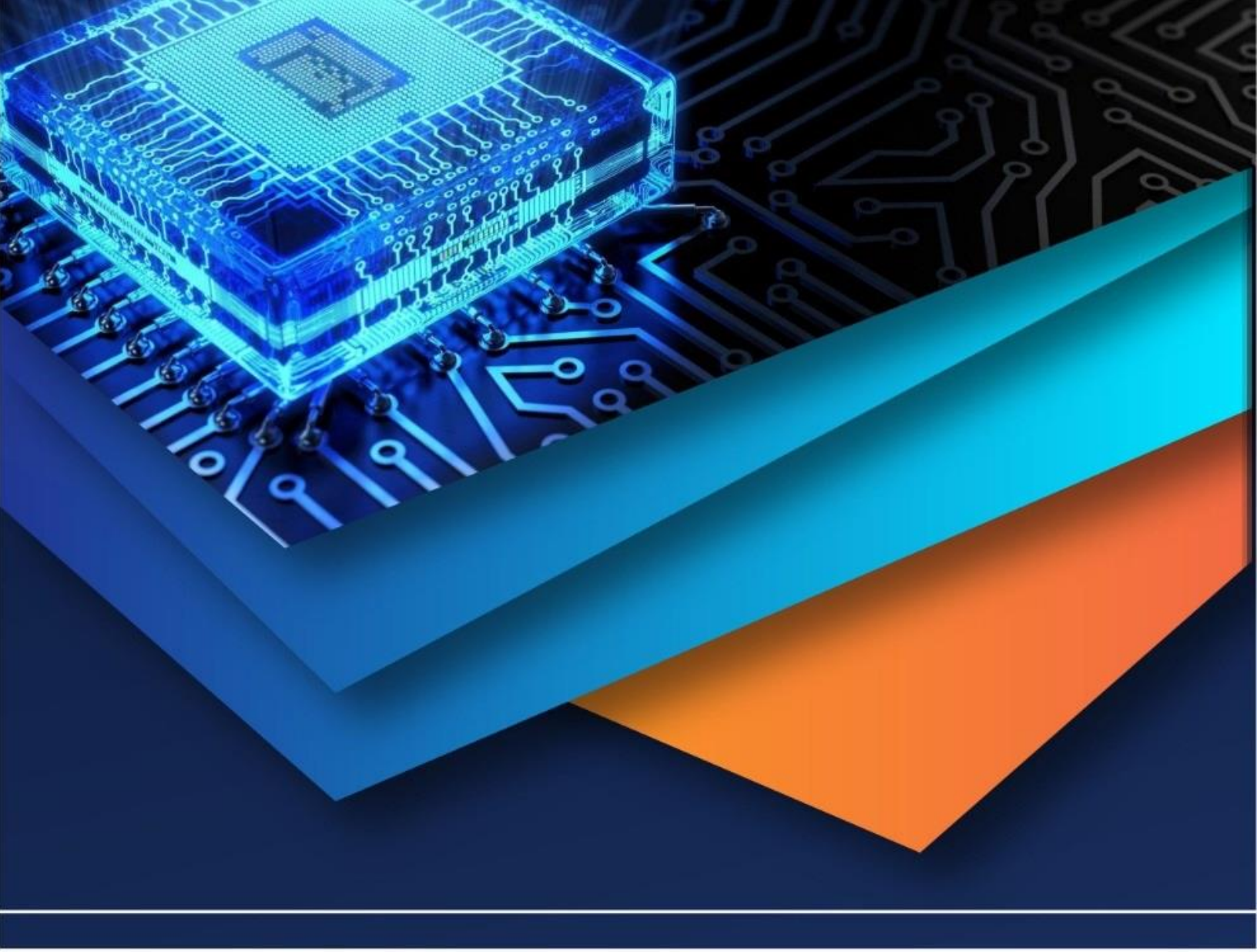
VI. ACKNOWLEDGMENT

The research is funded, in part, by the FUNDAMENTAL RESEARCH GRANT SCHEME (FRGS) (FRGS/1/2017/TK04/FKE-CERIA/F00331) given by Malaysia's Ministry of Higher Education. We also would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for technical and moral supports in the course of this study.

REFERENCES

[1] V. Gundu and S. P. Simon, "PSO–LSTM for short term forecast of heterogeneous time series electricity price signals," J. Ambient Intell. Humaniz. Comput., Jul. 2020.
[2] J. Olamaee, M. Mohammadi, A. Noruzi, and S. M. H. Hosseini, "Day- ahead price forecasting based on hybrid prediction model," Complexity, vol. 21, no. S2, pp. 156–164, Nov. 2016.
[3] G. Mestre, J. Portela, A. Muñoz San Roque, and E. Alonso, "Forecasting hourly supply curves in the Italian Day-Ahead electricity market with a double-seasonal SARMAHX model," Int. J. Electr. Power Energy Syst., vol. 121, p. 106083, Oct. 2020.
[4] T. Pinto, T. M. Sousa, I. Praça, Z. Vale, and H. Morais, "Support Vector Machines for decision support in electricity markets strategic bidding," Neurocomputing, vol. 172, pp. 438–445, Jan. 2016.
[5] X. Qiu, P. N. Suganthan, and G. A. J. Amaratunga, "Short-term Electricity Price Forecasting with Empirical Mode Decomposition based Ensemble Kernel Machines," Procedia Comput. Sci., vol. 108, pp. 1308–1317, 2017.
[6] H. Shayeghi and A. Ghasemi, "Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme," Energy Convers. Manag., vol. 74, pp. 482–491, Oct. 2013.
[7] V. Sharma and D. Srinivasan, "A hybrid intelligent model based on recurrent neural networks and excitable dynamics for price prediction in deregulated electricity market," Eng. Appl. Artif. Intell., vol. 26, no.5–6, pp. 1562–1574, 2013.
[8] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique," Energy Convers. Manag., vol. 50, no. 12, pp. 2976–2982, 2009.
[9] N. Amjady and A. Daraeepour, "Design of input vector for day-ahead price forecasting of electricity markets," Expert Syst. Appl., vol. 36, no. 10, pp. 12281–12294, Dec. 2009.
[10] F. Keynia, "A new feature selection algorithm and composite neural network for electricity price forecasting," Eng. Appl. Artif. Intell., vol. 25, no. 8, pp. 1687–1697, 2012.
[11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
[12] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)