International Journal of Computer Engineering and Sciences Research

VOL. 04, NO. 02, March-April 2022
Pages 10–17 (ISSN: 2581-8481)
Available online at: www.ijcesr.com

Auto Regressive Integrated Moving Average (ARIMA) Algorithm Application for Mining Operating Cost Forecasting

Teresita L. Toledo
College of Engineering and Information Technology
Surigao State College of Technology
Surigao City, Philippines

Abstract: This paper employed the famous ARIMA(p,d,q) model to forecast the operating cost accounts of selected mining company in Mindanao, Philippines. The operating cost includes but is not limited to personnel cost, mobile cost, material and supplies, contract fees, and other overhead accounts. The simulation result showed that the ARIMA(1,0,1) model appeared to be the statistically appropriate model to forecast the operating cost of the mining company. A forecasted cost of 522,632,000 pesos for the contract fee for the year 2019 was predicted, making it the highest.

Keywords—*arima, cost, forecasting, mining, time series*

I. INTRODUCTION

In economic feasibility studies, most specifically in the budgeting process and of projects, forecasts type analysis is applicable in operating cost analysis. Accordingly, some companies are wanting updates more frequently to their budgeting process because they are not satisfied with the projection of the budget process, wherein, the cost versus benefits are the main problem. Thus, they are looking for a lesser cost forecasting method without sacrificing its quality, which could compromise relevant information [1].

Several decision-makers are considering forecasting accuracy critical. Forecasting methods in various time series exist in combination, separately or both are using linear and nonlinear models. In combining linear and nonlinear models, studies show that to improve forecasting performance, it would be very effective to combine these models. Nonetheless, other assumptions with these existing methods design, their performance in a certain situation might restrict. With this, Auto Regressive Integrated Moving Average (ARIMA)-Artificial Neural Network (ANN) hybrid method is being used that works in the framework in general. The results of this experimental research show that in combining linear and nonlinear models and decomposing the original data all throughout the process of hybridization, the methods of forecasting performance are key factors [2].

Forecasting models [3]-[4] and other data mining techniques [5]–[6] are necessary for the operation analysis of any industry sector where protocols in the projection of budget processes of management decision-making protocols are the basis for essential variables [7].

This paper aims to forecast the operating cost accounts of a selected mining company in Mindanao, Philippines. This operating cost is categorized into personnel cost, mobile cost, material and supplies, contract fees, and other overhead accounts. The famous ARIMA algorithm is used for simulation and a time series analysis model type, in operating cost prediction for the years 2018-2022. Different ARIMA models were observed, and the best model was selected from them.

II. RELATED STUDIES

One of the renowned data mining approaches is prediction [8] which is commonly used in the educational data mining (EDM) [9], crime mining [10], business and finance [11], health [12], and more.

Research in forecasting and prediction is extensive. In response to the problems, various models were developed and utilized in answer to its quest. In this paper [8], forecasting and prediction are categorized into two methods; classical and modern. The inclusion of econometrics-based approaches is said to be classical methods, inferences statistical, and traditional mathematical programming. On the other aspect, if it employs artificial intelligence and algorithms in soft computing that is identified as a typical modern method.

A forecasting approach as an example, where there exists a combination of the strengths of the neural network and multivariate time series models, has been proposed. Basically, in the proposed approach, the exchange rate in the forecasting of Japan, the USA, and the UK by time series was done first, it follows the use of GRNN for the forecasting error correction [13]. Alternatively, the ARIMA model and state-space model performance models were compared for sales prediction in the process of applying both to women's footwear retail sales case study [14]. Along with this, for daily food sales forecasting, an enhanced seasonal ARIMA model was developed. The enhanced method, as a result, revealed, provides deep insights and better prediction for the demand influencing factors effect [15].

According to the literature, the exchange rate in Brazil on its forecasting accuracy using different approaches was examined. Intelligent systems of which they employed like neural networks function on multilayer perceptron and radial basis and the Fore fuzzy system of Takagi-Sugeno against the forecasting traditional methods to mention the linear models of autoregressive moving average (ARMA) and ARMA-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH). As a result, as compared to the traditional ones [16], the intelligent-based methods provided more accurate results.

Recently, a model for predicting OTOP’s products was developed using K-Nearest Neighbor (KNN) in 5, 10, 15, 20, and 25 k-fold cross-validations and K-NN with k values assigned with 3, 5, and 7. The model with 5 folds cross-validation and K-NN with k=3 yields the best prediction with 87.73% accuracy. Moreover, the authors suggested to compare the reliability of the results using Naïve Bayes, C4.5, and Rule-based algorithms in order to search for the optimal model for the prediction [17].

Meanwhile, a lot of studies related to electric consumption prediction using various methods are found in the literature [18]. For example, the use of a univariate dataset was observed to feed in ARIMA(p,d,q) model in predicting electric energy usage in Eastern Saudi Arabia. Findings showed that the ARIMA(1,1,0) model is the optimum model to be used in forecasting their electric usage in the monthly data [19].

In addition, to measure the growth of electrical demand in Chile, the statistic linear parametric model was used, specifically for the residential use [20]. Meanwhile, a short-term electricity demand forecasting in Queensland, Australia, was done using Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), and Multivariate Adaptive Regression Spline (MARS), model. The accuracy of the forecast was evaluated using statistical metrics such as RMSE and MAE [21].

On the other hand, the estimation of the electricity demand in Turkey was done using ARIMA. The result revealed that their current official projection of electric consumption is way too high than the forecasted result generated using ARIMA. The overestimation in the forecast of electric consumption affects energy policy and the electricity market [22]. Meanwhile, forecasting Turkey’s electric consumption was done using the artificial neural network (ANN) and regression models. The variables used were the GDP, population rate, historical data on electricity consumed, and other demographic variables and datasets in Turkey. The result revealed that prediction made using ANN was more effective than those of the multiple linear regression and power regression models [23].

Moreover, the ARIMA algorithm was employed to forecast electric consumption in a healthcare institution. The datasets used were obtained from Apollo Hospital in India for the year 2005 to 2016. Various forecasting validation tools and statistical metrics such as AIC, SBC, root mean squared error (RSME), and mean percentage error (MPE) were considered in selecting the appropriate model for forecasting electric consumption in monthly, bimonthly, and quarterly periods. The result showed that ARIMA(2,1,3) was the most appropriate model to be used in forecasting monthly series, while the ARIMA(2,1,1) model is used for both bi-monthly and quarterly series. The data was analyzed using the three series, and it was found that the best method to be used in forecasting the electric consumption of the said healthcare institution is the monthly forecasting method. The ARIMA(2,1,3) has the lowest RMSE and MPE value among the three models [24].

III. METHODOLOGY

A. Dataset

The datasets used in this paper are the operating cost of a mining company in the Philippines categorized as personnel cost, mobile cost, material and supplies, contract fees, and other overhead accounts for the years 2014 to 2018.

B. Autoregressive Integrated Moving Average

The ARIMA model is considered one of the most widely used methodologies in time series forecasting that aims to describe the autocorrelations in the data and use the ARIMA(p,d,q) notation. *p* denotes the order of the auto regression process (AR), *d* refers to the degree of differentiation involved (I), and *q* refers to the (MA), which is the order of the Moving Average. The mathematical expression of the model is:

$$Y_t=\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+\ldots+\phi_p Y_{t-p}+e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\ldots-\theta_q e_{t-q} \quad (1)$$

where Y_t is the variable value at time *t*, while$\emptyset$ and θ are the model parameters of (AR) and (MA) and e_t is the residual term representing random disturbances that cannot be predicted. Figure 1 shows the algorithm flowchart of ARIMA.

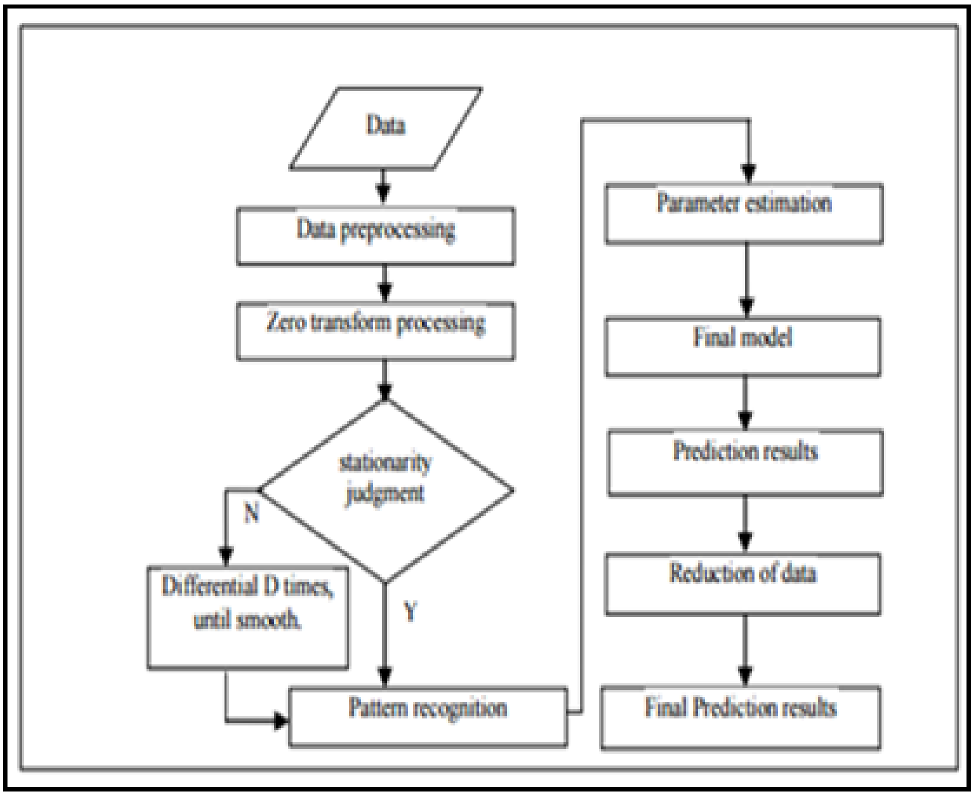

Figure 1. Algorithm flowchart of ARIMA

IV. Results and Simulation

The data that were used in this study are the indexed datasets of operating cost accounts of a mining company in Mindanao, Philippines to wit: personnel cost, mobile cost, material and supplies, contract fees and other overhead accounts from years 2014-2018. This paper used ARIMA(1,0,1) model in determining the operating costs for the next five years. The simulation was done using GRETL software application

A. Graphical and Statistical Method

Figure2 displayed the time series plot and forecasted 5-year personnel cost for years 2019-2023. A stable trend is evident in the graph which denotes the steady personnel cost to wit: salaries, wages & allowances, 13th-month pay and bonuses, other fringe benefits, leave commutation, sss, medicare employer's contributions, hdmf employer's contribution, Philhealth employer's contribution, mess supplies and foodstuff, hospital and medical expenses, training and seminars, retirement/pension expenses, and sports and recreational expenses. Meanwhile, the specific values for the predicted personnel cost in the next five years, having a 95% confidence interval are shown in Table I.

TABLE I. FORECASTED PERSONNEL COST WITH 95% INTERVAL

Year	Forecast	Lo 95	Hi 95
2019	42,023,200	37,735,900	46,310,500
2020	42,382,300	36,362,400	48,402,200
2021	42,387,500	36,367,300	48,407,600
2022	42,387,500	36,367,400	48,407,700
2023	42,387,500	36,367,400	48,407,700

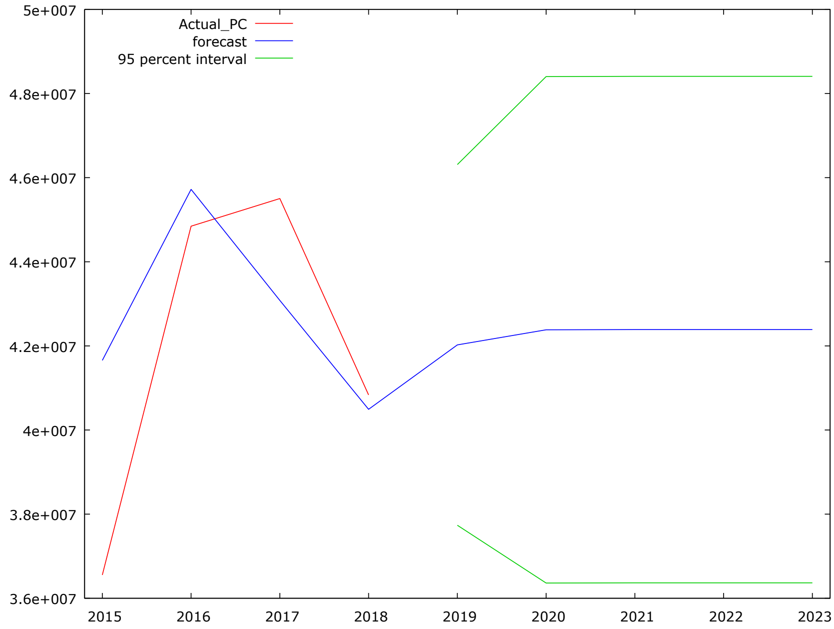

Figure 2. Forecasted personnel cost for 2019-2023

Figure3 shows the time series plot and forecasted 5-year mobile cost for years 2019-2023. An increasing trend from the year 2016-2023 is evident in the graph, which denotes an increase in mobile costs related to tires, batteries, accessories, spare parts supplies, fuel, oil, and lubricant supplies used. Meanwhile, the specific values for the predicted mobile cost in the next five years, having a 95% confidence interval are shown in Table II.

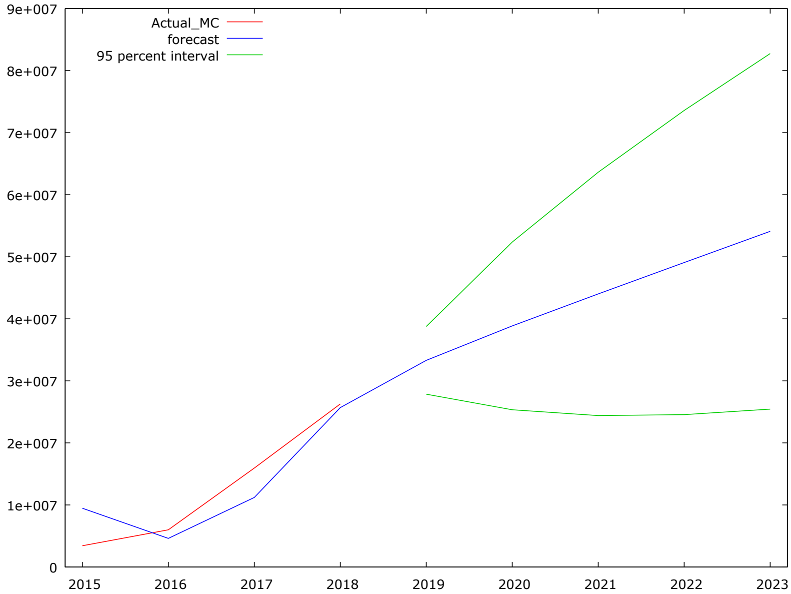

Figure 3. Forecasted mobile cost for 2019-2023

TABLE II. FORECASTED MOBILE COST WITH 95% INTERVAL

Year	Forecast	Lo 95	Hi 95
2019	33,302,800	27,848,100	38,757,600
2020	38,855,500	25,339,800	52,371,300
2021	44,014,800	24,402,300	63,627,300
2022	49,069,100	24,556,300	73,581,800
2023	54,095,200	25,442,300	82,748,100

Figure4 shows the time series plot and forecasted 5-year materials and supplies cost for years 2019-2023. An increasing trend from the year 2018-2019 is evident in the graph, which denotes an increase in cost-related construction, electrical and hardware supplies, household supplies, safety and environmental control supplies, medicines and medical supplies, laboratory supplies, expendable tools supplies, and stationeries and office supplies used. Meanwhile, the specific values for the predicted materials and supplies cost in the next five years, having a 95% confidence interval are shown in Table III.

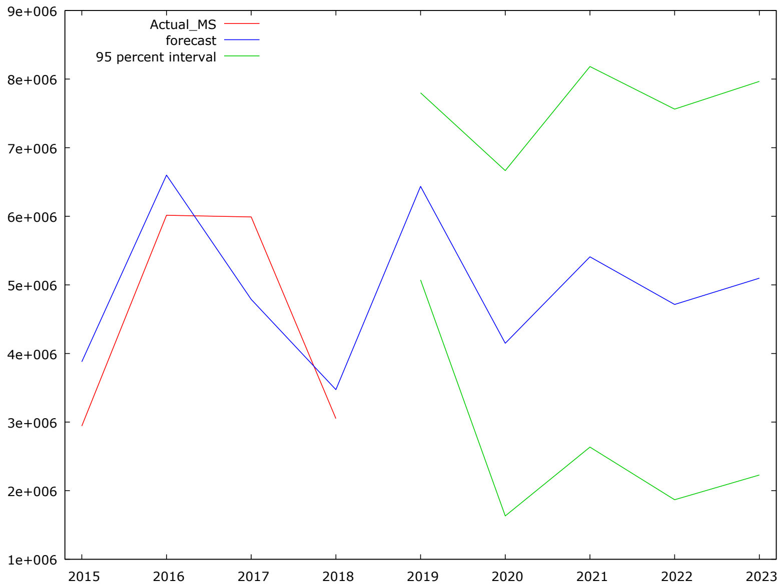

Figure 4. Forecasted materials and supplies cost for 2019-2023

TABLE III. FORECASTED MATERIALS AND SUPPLIES COST WITH 95% INTERVAL

Year	Forecast	Lo 95	Hi 95
2019	6,436,004.44	5,072,158.31	7,799,850.56
2020	4,149,181.50	1,632,296.96	6,666,066.04
2021	5,409,232.93	2,635,564.55	8,182,901.30
2022	4,714,937.99	1,867,887.81	7,561,988.16
2023	5,097,498.14	2,228,540.16	7,966,456.11

Figure 5 shows the time series plot and forecasted 5-year materials and supplies cost for years 2019-2023. An increasing trend from the year 2018-2019 and a decreasing cost a year after are evident in the graph, which denotes an unsteady in cost contract fees. Meanwhile, the specific values for the predicted contract fees cost in the next five years, with a 95% confidence interval are shown in Table IV.

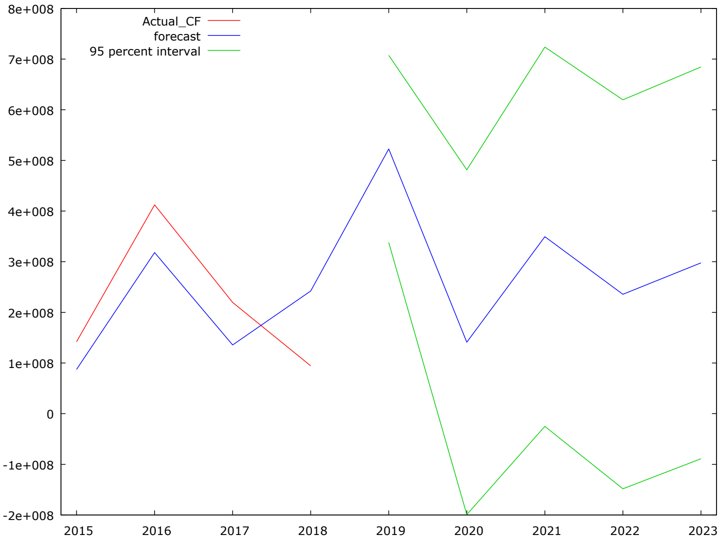

Figure 5. Forecasted contract fees cost for 2019-2023

TABLE IV. FORECASTED CONTRACT FEES COST WITH 95% INTERVAL

Year	Forecast	Lo 95	Hi 95
2019	522,632,000	337,789,000	707,476,000
2020	141,061,000	98,988,400	481,387,000
2021	349,367,000	101,554,300	723,743,000
2022	235,649,000	99,561,000	619,589,000
2023	297,730,000	102,446,000	684,474,000

Figure 6 shows the time series plot and forecasted 5-year overhead fees for the years 2019-2023. Rise and fall of cost are evident in the graph which denotes an increase and decrease of overhead costs to wit: transportation, traveling, representation, and industrial relations fees, communications, dues, and subscriptions/periodicals, freight and handling fees, repairs and maintenance fees, equipment, office, house, lot, and space rentals, light and water fees, taxes and licenses fees, commission expense, royalty fee – claim owner, the royalty fee for the government, excise tax, insurance and bond premium, professional fee, legal documentation, and other fees and services, mobilization and demobilization cost, pump boat loss, plant crop damages, other expense such that in SDMP, bank charges, depreciation expense, depletion expenses, facilitation fee, liability for mine rehabilitation, gain and loss on disposal of equipment, and miscellaneous expense among others. Meanwhile, the specific values for the predicted overhead fees in the next five years, having a 95% confidence interval are shown in Table V.

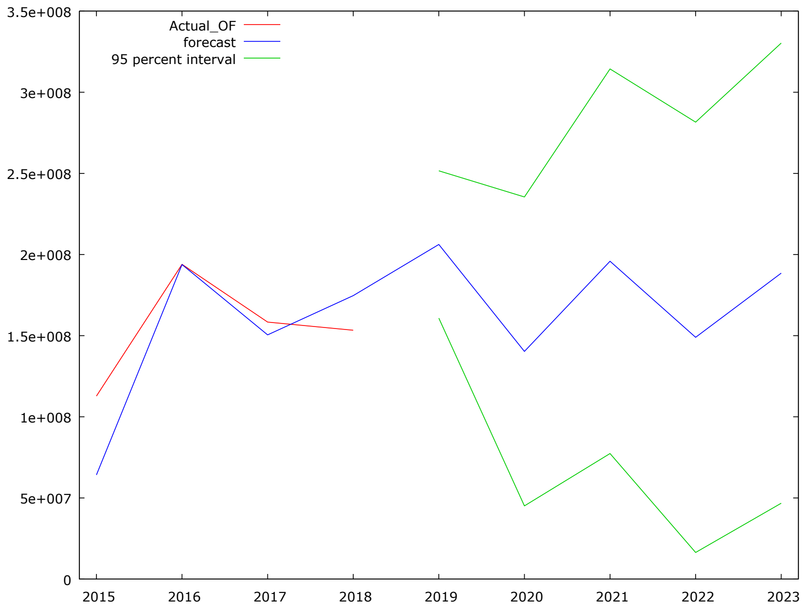

Figure 6. Forecasted overhead fees cost for 2019-2023

TABLE V. FORECASTED OVERHEAD FEES COST WITH 95% INTERVAL

Year	Forecast	Lo 95	Hi 95
2019	206,194,000	160,802,000	251,586,000
2020	140,287,000	45,095,500	235,479,000
2021	195,868,000	77,374,900	314,361,000
2022	148,996,000	16,400,200	281,592,000
2023	188,524,000	46,749,500	330,298,000

V. CONCLUSION

In this paper, the ARIMA(1,0,1) model was used to forecast the operation costs as to personnel cost, mobile cost, material and supplies, contract fees, and other overhead accounts of a mining company in the Philippines for the year 2019-2023. The forecast showed an increasing trend for mobile cost in the years 2019-2023, and a steady trend for personnel cost throughout the next five years. Meanwhile, there is a zigzag pattern denoting an increase and decrease of cost forecasted for the materials and supplies cost, contract fees, and overhead fees throughout the five-year forecast.

REFERENCES

[1] J. Rodriques da Silva, G. Fortunato, S. Bastos, "Operating cost budgeting methods: quantitative methods to improve the process," *Producao*, vol. 26, no. 4, 2016

[2] Buyuksahin, Ertekin, "Improving forecasting accuracy of time series data using ARIMA-ANN hybrid method and empirical mode decomposition," *Neurocomputing*, vol. 361, pp. 151-163, 2019.

[3] A. J. P. Delima, “Applying Data Mining Techniques in Predicting Index and non-Index Crimes,” *Int. J. Mach. Learn. Comput.*, vol. 9, no. 4, pp. 533–538, 2019.

[4] A. J. P. Delima, A. M. Sison, and R. P. Medina, “A Modified Genetic Algorithm with a New Crossover Mating Scheme,” *Indones. J. Electr. Eng. Informatics*, vol. 7, no. 2, pp. 165–181, 2019.

[5] U. O. Cagas, A. J. P. Delima, and T. L. Toledo, “PreFIC : Predictability of Faculty Instructional Performance through Hybrid Prediction Model,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7, pp. 22–25, 2019.

[6] A. J. P. Delima, “Predicting Scholarship Grants Using Data Mining Techniques,” *Int. J. Mach. Learn. Comput.*, vol. 9, no. 4, pp. 513–519, 2019.

[7] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, “Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines,” *Int. J. Electr. Power Energy Syst.*, vol. 67, pp. 431–438, 2015.

[8] M. J. Rezaee, M. Jozmaleki, and M. Valipour, "Integrating dynamic fuzzy C-means , data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange," *Physica A*, vol. 489, pp. 78–93, 2018.

[9] R. Ahuja, A. Jha, R. Maurya, and R. Srivastava, "Analysis of Educational Data Mining," *Adv. Intell. Syst. Comput.*, pp. 897–907, 2018.

[10] P. Vrushali, M. Trupti, G. Pratiksha, and G. Arti, "Crime Rate Prediction using KNN," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 1, pp. 124–127, 2018.

[11] P. Carmona, F. Climent, and A. Momparler, "Predicting failure in the U.S. banking sector: An extreme gradient boosting approach," *Int. Rev. Econ. Financ.*, pp. 1–54, 2018.

[12] K. Lakshmi, D. I. Ahmed, and G. Siva Kumar, "A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques," *2018 Ijsrset*, vol. 4, no. 1, 2018.

[13] A. Chen and M. T. Leung, "Regression neural network for error correction in foreign exchange forecasting and trading," *Comput. Oper. Res.*, vol. 31, pp. 1049–1068, 2004.

[14] P. Ramos, N. Santos, and R. Rebelo, "Performance of state space and ARIMA models for consumer retail sales forecasting," *Robot. Comput. Integr. Manuf.*, vol. 34, pp. 151–163, 2015.

[15] N. S. Arunraj and D. Ahrens, "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," *Int. J. Prod. Econ.*, vol. 170, pp. 321–335, 2015.

[16] A. A. P. Santos, N. C. A. da Costa, and L. dos S. Coelho, "Computational intelligence approaches and linear models in case studies of forecasting exchange rates," *Expert Syst. Appl.*, vol. 33, pp. 816–823, 2007.

[17] N. Suhermi, Suhartono, D. D. Prastyo, and B. Ali, "Roll motion prediction using a hybrid deep learning and ARIMA model," *Procedia Comput. Sci.*, vol. 144, pp. 251–258, 2018.

[18] J. Tarapitakwong, B. Chartrungruang, and N. Tantranont, "A Classification Model for Predicting Standard Levels of OTOP ' s Wood Handicraft Products by Using the K-Nearest Neighbor," *Int. J. Comput. Internet Manag.*, vol. 25, no. 2, pp. 135–141, 2017.

[19] Y. T. Chen, "The factors affecting electricity consumption and the consumption characteristics in the residential sector—a case example of Taiwan," *Sustain.*, vol. 9, no. 8, 2017.

[20] R. E. Abdel-Aal and A. Z. Al-Garni, "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis," *Energy*, vol. 22, no. 11, pp. 1059–1169, 1997.

[21] H. Verdejo, A. Awerkin, C. Becker, and G. Olguin, "Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile," *Renew. Sustain. Energy Rev.*, vol. 74, no. January, pp. 512–521, 2017.

[22] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Adv. Eng. Informatics*, vol. 35, no. April 2017, pp. 1–16, 2018.

[23] E. Erdogdu, "Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey," *Energy Policy*, vol. 35, no. 2, pp. 1129–1146, 2007.

[24] M. Kankal, A. Akpinar, M. I. Kömürcü, and T. Ş. Özşahin, "Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables," *Appl. Energy*, vol. 88, no. 5, pp. 1927–1939, 2011.

[25] H. Kaur and S. Ahuja, "Time Series Analysis and Prediction of Electricity Consumption of Health Care Institutions Using ARIMA Model," in *Sixth International Conference on Soft Computing for Problem Solving*, 2017, vol. 547.

[26] N. A. C. Atienza, J. Renzo, A. T. Jao, J. A. D. S. Angeles, E. Lance, and T. S. Jr, "Prediction and Visualization of Electricity Consumption in the Philippines Using Artificial Neural Networks , Particle Swarm Optimization , and Autoregressive Integrated Moving Average," *2018 3rd Int. Conf. Comput. Commun. Syst.*, pp. 135–138, 2018.

[27] J. D. Urrutia, F. L. T. Mingo, and C. N. M. Balmaceda, "Forecasting Income Tax Revenue of the Philippines Using Autoregressive Integrated Moving Average (Arima) Modeling : a Time Series Analysis," *Am. Res. Thoughts*, vol. 1, no. 9, pp. 1938–1992, 2015.

[28] K. Molebatsi and M. Raboloko, "Time Series Modelling of Inflation in Botswana Using Monthly Consumer Price Indices," *Int. J. Econ. Financ.*, vol. 8, no. 3, pp. 15–22, 2016.

[29] D. A. Kuhe and R. C. Egemba, "Modeling and Forecasting CPI Inflation in Nigeria: Application of Autoregressive Integrated Moving Average Homoskedastic Model," *J. Sci. Eng. Res.*, vol. 3, no. 2, pp. 57–66, 2016.

[30] S. Makirdakis, E. Spiliotis, and V. Assimakopoulos, "The accuracy of machine learning (ML) forecasting methods versus statistical ones : Extending the results of the M3-Competition," *Neural Networks*, 2017.