International Journal of Computer Engineering and Sciences Research

VOL. 04, NO. 02, March-April 2022
Pages 18–24 (ISSN: 2581-8481)
Available online at: www.ijcesr.com

Forecasting Salaries and Wages using Auto Regressive Integrated Moving Average (ARIMA) Algorithm

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 costs subdivided into 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. The simulation result showed that the ARIMA(1,0,1) model appeared to be the statistically appropriate model to forecast the operating cost specifically for the salaries and wages prediction of the identified mining company. A forecasted salary and wage cost of 42,387,500 pesos was predicted for the years 2021-2023.

Keywords*—ARIMA; forecasting; mining; salaries; time series*

I. INTRODUCTION

The motivation of employees’ exemplary performances to their respective job descriptions would rely on their salaries and wages scheme. The sustainability of every family’s basic needs relies on the aspects of its income whether employed or self-employed individuals [1]. Enhancing the productivity of an employee through recognition improves organization performance as well [2]. This is deemed appropriate in mining companies having workers receiving monthly salaries.

In the Philippines, the mining industry plays an essential part in the industrial development in the quest for modern advancement and segment's capacity to offer resources of minerals for utilities, manufacturing construction, transportation, and communications sector. Through the exports of mineral ore and other mineral products processed, as a major contributor to foreign currency, it opens employment to localities and other distant places in which mining operations of economic activity are the only basis (Department of Labor and Employment –Industry Career Guide).

According to the International Solidarity Mission on Mining (ISMM), the Philippines visualizes a concrete portrait of the extractive industry for its people and the environment and for the Filipino workers as well. Thus, the country’s mining capital is Caraga Region (Region XIII). Employment in the region ranges only to 23,138 workers which account for 2.6%, mostly unskilled and contractual which ranges from three months to nine months only.

Data mining techniques [3] plays an important function as these mining techniques could generate classification rules in predicting employee performance, having similar attributes to the training dataset [4]. The details can determine unknown cases in the labor market employment [5].

Forgoing studies on salary models and systems prediction were proposed having technique on regression. These models were all functional but had some problems as such: instead for individuals models predicted as a group for salaries aspect, thus, to assimilate fully the result of this in statistics background is extensive.

This paper aims to forecast the Salaries and Wages of selected mining company in Mindanao, Philippines. The renowned ARIMA methodology, a time series analysis type model will be used in this study.

II. RELATED STUDIES

Recently, changes which are subject to vast and sudden that even the most established structures are affecting require all business sectors practical reading into future to be accurate. Thus, forecasts are getting to be very crucial as they are considered to be the language of business in the world and the sign of survival. Estimating the future level of some variables in science is coined into the forecast. Often the variable is in demand and most likely as price or supply. The operation assumption on studied variables of future values is said to be forecasting [6].

Research in forecasting and prediction is huge. According to [7], there are two general categories for forecasting and prediction namely the classical and modern methods. In solving problems, various models were developed and utilized by the researchers which are commonly used in the educational data mining (EDM) [8], crime mining [9], business and finance [10], health [11], and more. Classical methods include econometrics-based approaches, statistical inferences, and traditional mathematical programming while the modern method employs soft computing algorithms and artificial intelligence.

For example, a forecasting approach that combines the strengths of the neural network and multivariate time series models was proposed. In the proposed approach, forecasting the exchange rate of the UK, USA, and Japan was done first by time series, and then GRNN was used to correct the forecasting errors [12]. On the other hand, the performance of state space models and the ARIMA model was compared for predicting sales by applying both to a case study of the women's footwear retail sales [13]. Further, an enhanced seasonal ARIMA model was developed for daily food sales forecasting. The result revealed that the enhanced method provides better prediction and deep insights into the effect of demand influencing factors [14].

In addition, the financial parameters calculations were coined on the company's public annual financial reports. Ensuring accuracy the Altman Z-score model was applied of two mining companies as a sample [15]. Furthermore, an exchange rate in Brazil using different approaches was examined. They employed intelligent systems like multilayer perceptron and radial basis function neural networks and the Takagi–Sugeno fuzzy system versus the traditional methods of forecasting such as autoregressive moving average (ARMA) and ARMA-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) linear models. It was found that the intelligent-based methods provided more accurate results than the traditional ones [17].

Meanwhile, the adaptation of a hybrid methodology combining ARIMA and Deep Neural Network (DNN), which is an ANN model with multiple hidden layers, was considered the optimal model for predicting roll motion compared to the non-hybrid models. It was found that DNN-ARIMA hybrid model showed improved forecast accuracy and was identified to be very effective [18].

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 [19].

Meanwhile, a lot of studies related to salaries and wages prediction using various methods are found in the literature [20]. 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 [21].

Further, the statistic linear parametric model was used to measure the growth of electrical demand in Chile, particularly in residential use [22]. 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 [23].

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 [24]. 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 [25].

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 [26], [27].

III. METHODOLOGY

A. Dataset

The datasets used in this paper are the historical data of a mining company's salaries and wages 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. Fig. 1 shows the algorithm flowchart of ARIMA.

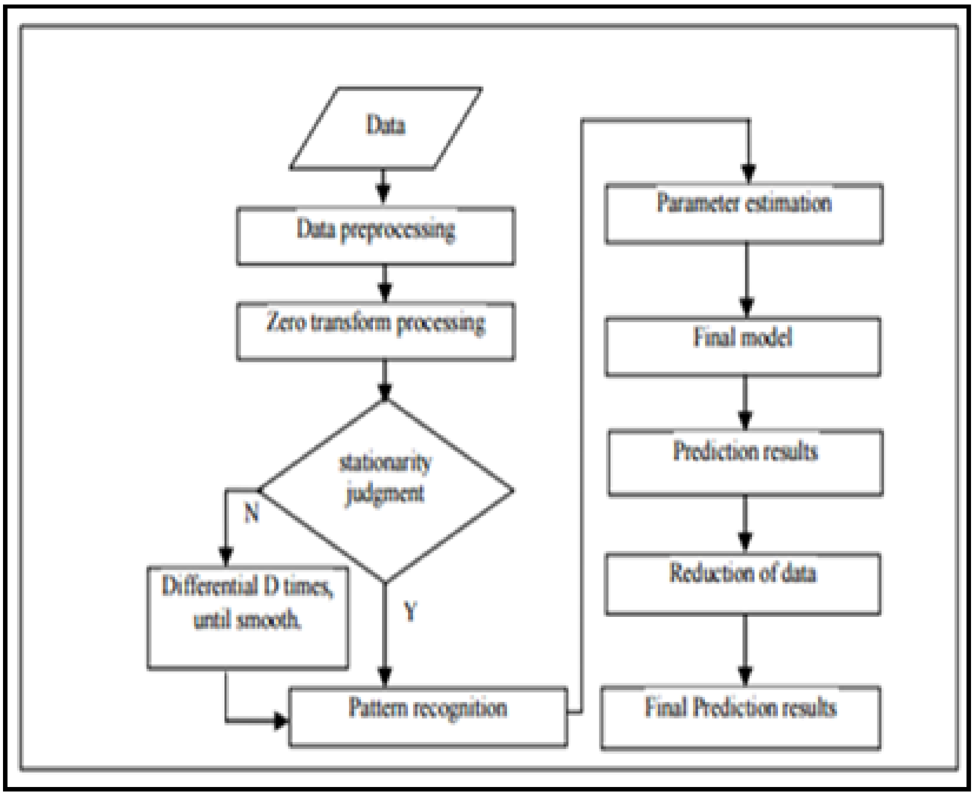

Figure 1. Algorithm flowchart of ARIMA

IV. Results and Simulation

The data that were used in this study are the indexed datasets of operating cost accounts of selected mining company situated 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

Figure 2 and Figure 3 shows the time series plot and forecasted 5-year personnel cost for years 2019-2023, respectively.

Figure 2. Time series plot for salaries and wages cost from 2004-2018

In Figure 3, a stable trend is evident in the graph which denotes the steady cost for salaries and wages cost specifically covering 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.

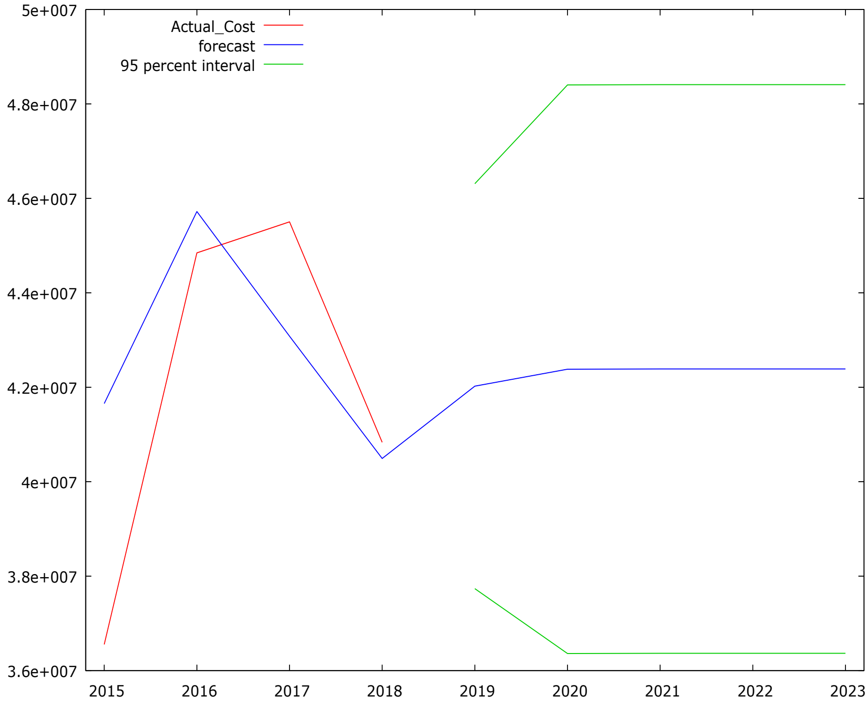

Figure 3. Forecasted salaries and wages cost for 2019-2023

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

V. CONCLUSION

In this paper, the ARIMA(1,0,1) model was used to forecast the operation costs as to salaries and wages of a mining company in the Philippines for the year 2019-2023. The forecast showed an increasing trend for salaries and wage costs in the years 2018-2021, and a steady trend throughout the years 2022-2023. The forecasted highest cost for salaries and wages is expected in the years 2021-2023 amounting to 42,387,500 pesos.

REFERENCES

[1] O. E. Sule, S. I. Amuni, K. A. Obasan, and H. A. Banjo, "Wages and salaries as a motivational tool for enhancing organizational performance. A survey of selected Nigerian workplace," *EuroEconomica*, vol. 34, no. 1, 2015.

[2] B. Nagaraju, "Impact of Salary on Employee Performance Emperical Evidence From Public and Private Sector Banks," *Int. J. Mark. Hum. Resour. Manag.*, vol. 8, no. 4, pp. 43–51, 2017.

[3] Q. A. Al-Radaideh and E. Al Nagi, "Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 2, pp. 144–151, 2012.

[4] P. Khongchai and P. Songmuang, "Implement of salary prediction system to improve student motivation using data mining technique," *Proc. - 11th 2016 Int. Conf. Knowledge, Inf. Creat. Support Syst. KICSS 2016*, pp. 0–5, 2017, doi: 10.1109/KICSS.2016.7951419.

[5] Y. A. Alsultanny, "Labor market forecasting by using data mining," *Procedia Comput. Sci.*, vol. 18, pp. 1700–1709, 2013, doi: 10.1016/j.procs.2013.05.338.

[6] J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, and A. Lachhab, "Forecasting of demand using ARIMA model," *Int. J. Eng. Bus. Manag.*, vol. 10, pp. 1–9, 2018, doi: 10.1177/1847979018808673.

[7] 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, doi:

10.1016/j.physa.2017.07.017.

[8] R. Ahuja, A. Jha, R. Maurya, and R. Srivastava, "Analysis of Educational Data Mining," *Adv. Intell. Syst. Comput.*, pp. 897–907, 2018, doi: 10.1007/978-981-13-0761-4.

[9] 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.

[10] 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, doi: 10.1016/j.iref.2018.03.008.

[11] 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.

[12] 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, doi: 10.1016/S0305-0548(03)00064-9.

[13] 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, doi: 10.1016/j.rcim.2014.12.015.

[14] 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, doi: 10.1016/j.ijpe.2015.09.039.

[15] D. Zlatanovic, M. Bugarin, V. Milisavljevic, and V. Zlatanovic, "Forecasting the financial distress of mining companies: Tool for testing the key performance indicators," *Min. Metall. Eng. Bor*, vol. 8836, no. 1, pp. 73–80, 2016, doi: 10.5937/mmeb1601073z.

[16] J. F. Mgaya, "Application of ARIMA models in forecasting livestock products consumption in Tanzania," *Cogent Food Agric.*, vol. 5, no. 1, 2019, doi: 10.1080/23311932.2019.1607430.

[17] 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, doi: 10.1016/j.eswa.2006.07.008.

[18] 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, doi: 10.1016/j.procs.2018.10.526.

[19] 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.

[20] 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, doi: 10.3390/su9081484.

[21] 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, doi: 10.1016/S0360-5442(97)00032-7.

[22] 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, doi: 10.1016/j.rser.2017.01.110.

[23] 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, doi: 10.1016/j.aei.2017.11.002.

[24] 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, doi: 10.1016/j.enpol.2006.02.013.

[25] 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, doi: 10.1016/j.apenergy.2010.12.005.

[26] 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, doi: 10.1007/978-981-10-3325-4.

[27] 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, doi: 10.1109/ccoms.2018.8463351.

[28] 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, doi: 10.6084/M9.FIGSHARE.1469539.

[29] 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, doi: 10.5539/ijef.v8n3p15.

[30] 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.

[31] 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