Journal of Theoretical and Applied Information Technology

30th September 2023. Vol.101. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMPROVING CASH FORECASTING PERFORMANCE USING A NEW HYBRID-ANN-ARIMA-EM MODEL

ISAM AHMED M. YAQOOB[1], KHAIRUL AZHAR KASMIRAN[2], TEH NORANIS MOHD ARIS[3], NOR AZURA HUSIN[4], MOHD YUNUS SHARUM[5]

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia
Email: [1]gs59654@student.upm.edu.my, [2]k_azhar@upm.edu.my, [3]nuranis@upm.edu.my, [4]n_azura@upm.edu.my, [5]m_yunus@upm.edu.my

ABSTRACT

Budgeting is the technique and science of dividing available and obtainable money between competing needs. Government spending assists programs in tooling up a wide range of services to numerous different slices of the population. Therefore, the demands for further and better services commonly might exceed the government's capability to provide payments for them. Forecasting of cash flow is the most important instrument for any business in general and exclusively in public budgeting. In simple words, forecasting can show if your business will have sufficient and enough cash to execute the business and/or the ability to expand it. The COVID-19 pandemic has caused an economic shock all over the world. The effect of some external factors such as the COVID-19 pandemic is one of the most important drawbacks that make statistical calculations difficult and sometimes intractable because this external factor is unmeasurable. Artificial intelligence and machine learning have been used widely to improve cash forecasting in many circumstances. In this paper, a comparison between ARIMA, ANN, and hybrid models will be done as a first step to highlight their pros and cons. In the second step, the external factor will be measured to assign the weight. Finally, the new proposed model (Hybrid-ARIMA-ANN-EM) will be explained and implemented, thereafter, by using RMSE and MAPE techniques, differentiation will be applied between the latter and other nominated models revealing the variance in forecasting results between them in terms of accuracy.

Keywords: *ARIMA, ANN, Machine Learning, Artificial Neural Network, Cash Forecasting*.

1. INTRODUCTION

Cash flow forecasting is an indispensable and vital issue. However, the most important question is what is the best method of predicting? And what is the opportunity to make a distinct optimization to gain a more accurate prediction? Furthermore, what are the factors or indicators that influence cash flow and should take priority in calculation steps, because there is a wide range of complex and dynamic factors. In addition, it is difficult to build a suitable relation between these indicators (the external or the non-financial factors) and cash flow to observe the effectiveness of the aforementioned factors on cash flow [1-2, 4]. Most cash forecasting research depends on calculating the financial indicators using the previous data in a time series (monthly, yearly, seasonally) to predict the cash flow without giving the real value of considering the impact of non-financial indicators in predicting process, even though the non-financial indicators have considerable influence on the ascending and descending of business more than the financial indicators. Furthermore, calculating the latter is simple because it involves definite numbers and values, whereas calculating the former is more challenging as it lacks precise mathematical values that can be plugged into an equation. As a result, a different conceptual approach and unique equation may need to be devised to make the calculations possible [15].

ARIMA is used widely by many researchers because it has a simple predicting equation, easy to implement with accurate forecasting results. However, the researchers tend to use ARIMA with data that have a stationary behavior, and the most distinct problem with ARIMA is the inability to calculate external factors, and exactly the unmeasurable ones as

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

aforementioned in the abstract [7-9, 30]. S. Bano, N. Nida, and S. A. Irtaza, 2020 [49], used a deep convolutional neural network regression model to predict future prices of crude oil. They made their study because of high crude oil volatility, which has a serious influence on the world economy and the global financial system. The suggested model was tested and verified on the dataset of the crude oil using root mean square error (RMSE) which was represented as a typical method to measure the error of any model in forecasting quantitative data. Whilst other researchers resort to using hybrid models for financial prediction such as designing a model combining ARIMA with one of the ANN models to resolve the problem of stationary and non-stationary data and to get accurate outcomes more than using each model separately. Wibowo, W., Dwijantari, S., & Hartati, A. 2017 [60], the target of their study was to show that "a complex model is not always better than a simpler model". Therefore, they proposed a hybrid ARIMA-ANN, a radial basis function (RBF) using orthogonal least squares. The criteria used to select the simple predicting model are the RMSE and the mean absolute percentage error (MAPE) to predict electricity consumption of the lowest household class in East Java, Indonesia. In Belmahdi, B., Louzazni, M., & Bouardi, A. E., 2020 [64], Büyükşahin, Ü. Ç., & Ertekin, Ş., 2020 [65], this study aimed to increase the accuracy of hybrid ARIMA-ANN models for univariate time series predicting by converting it to a feature-based model. Feature significant scores are defined by gradient boosting trees (GBT). Features with the highest priority/importance score are given as explanatory additional parameters to the hybrid ARIMA-ANN model. The results illustrated high altitude accuracy performance for the suggested model. Nevertheless, some authors prefer to modify the ARIMA model itself to earn highly accurate outcomes whether the data is stationary or not. However, most of these models face the problem of calculating (or setting the weight value of) the external factors. In Y. Peng, K. He, and Q. Yu, 2021 [61], the authors combined two models which are ARIMA and ELM (Extreme learning machine model) to forecast the closing price of stock indexes. Their experiment divided the closing price data into the 80% training set and 20% test set data used to build the ARIMA model, and the trading indicators used as input nodes for the ELM model. After estimating the performance of the integration model, then the number of hidden layer nodes was established. Finally, the test set data was used to forecast the closing price of the index on the designed model, and then the effectiveness of the model was evaluated. The prediction results showed that the combined model is better than using each one separately. J. F. Lorenzato de Oliveira and T. B. Ludermir, 2014 [45] pointed out that statistical linear models such as ARIMA cannot handle nonlinear patterns in time series. Therefore, a support vector regression (SVR) model that represents a nonlinear model was introduced that can map patterns. Thus, a time series can be resolved in linear and nonlinear patterns. The study designed a hybrid system (ARIMA - SVR) model, which was developed by the particle swarm optimization (PSO) algorithm for predictions. The results illustrate that the suggested model achieved promising outcomes for one-step ahead forecasting. The authors L. Yang and H. Yang, 2019 [70] developed an ARIMA model with projection pursuit regression (PPR) to grasp the nonlinear and linear pattern illustrated by a loading series. The combined model was designed by allocating weight coefficients to individual models, where weight coefficients are defined by RMSE. The combined model is implemented for the (5 min) interval load series data of Sichuan Province, China to the mission of half-hour ahead predicting.

1.1 Research Contributions

By applying time series analysis to enhance cash forecasting, this study will make a significant contribution to both artificial intelligence and economics. The research will figure out:

- Determine which model is more precise in forecasting (ARIMA, ANN, hybrid models), in terms of accuracy and simplicity.
- Illustrate the importance of calculating the non-financial external variables (such as the COVID-19 pandemic) as a main variable in comparison with the financial variables.
- Measuring the external factor, which is typically unmeasurable. Transform it into a mathematical value (as a weight) that can be used in forecasting calculations.
- Implementing the nominated forecasting models, then the new hybrid proposed model respectively on the selected dataset. Thereafter illustrating the predicted results showing the differences between them with regards to accuracy.

1.2 Background

Numerous investigations have been conducted by scholars in the area of cash projection due to its significant relevance and impact on business operations as a whole. Vanessa Fernandez-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Cortez, David Valle-Cruz, and J. Ramon Gil-Garcia, 2020 [1], made a study about the effects of economic growth (GDP), Social Development Index, level of corruption, and government debt on the public sector (public budget), considering them as non-financial variables. They proposed a mathematical developed model using genetic algorithms. Information was preprocessed and cleaned to prepare the data, and the data were processed using different machine-learning techniques to find patterns and relationships. The mathematical model was used in the optimization process. Finally, different tests were conducted with genetic algorithms to find the optimized budget values. Another study done by José Morales, and José Huanca, 2020 [2], discussed the influence of financial variables such as expenditures in the primary, secondary sector, and tertiary sectors on a government's public budget. They used regression methods (multiple regression) and artificial neural networks (multilayer perceptron) to determine the influence of spending execution on the regional government's public budget. A multiple linear regression model was developed with one dependent variable and three independent variables, a neural network (multilayer perceptron) with an input layer that has three units, a hidden layer with three units, and an output unit for the dependent variable of type scale. Its scale was changed according to the standardized method, which requires the use of the identity activation function for the output layer. Furthermore, distinct research has been written by Yuh-Jen Chen, Jou-AN Lin, Yuh-Min Chen, and Jyun-Han WU, 2019 [3], which talked about the importance of cash forecasting in the private sector (enterprises/companies), however, they made their calculations in terms of using financial indicators which are solvency, operating ability, and profitability and non-financial indicators which are firm size and corporate governance, using multivariate adaptive regression splines (MARS) and the queen genetic algorithm SVR (QGA-SVR) to forecast the comprehensive financial indicators of operating revenue, free cash flow, and net working capital to help enterprises forecast their future financial situation and provide a reference for investment decision making for investors and creditors.

In the private sector and for e-commerce cargo sales Tan Bowen, Zhang Zhe, and Zhang Yulin, 2020 [5], provided a study showing the calculation method of the regular financial indicators to predict the sale of cargo using the ARIMA-BP nonlinear combination model to forecast the sale of merchant's goods. Two separate prediction models, the BP neural network model, and the ARIMA model, were used to predict the sales volume in the next 5 days, and then a mean square error model was established to weigh the fitting and predictive results of the two single predictions. While Xiaoxian Yang, Shunyi Mao, Honghao Gao, Yucong Duan, and Qiming Zou, 2019 [4], introduced their study to calculate the regular financial indicators and their effect on cash forecasting in the public sector (forecasting financial capital flow). Their methodology came with three steps. First, a linear model called YEB_ARIMA is proposed by determining the autocorrelation (ACF) and partial autocorrelation (PACF) parameters, which are optimized by the grid search method. Second, a deep learning model called YEB_LSTM is introduced to strengthen the expressiveness of the model that yields nonlinear transaction features. Finally, a hybrid learning method called YEB_Hybrid has been applied to improve the original weak classifiers.

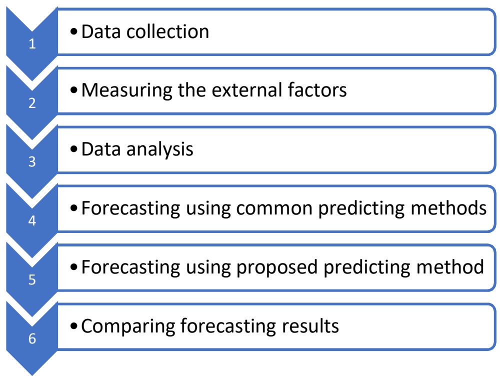
Many researchers did their studies in the forecasting field depending on using the ARIMA equation [5-9], while some of them prefer to use artificial neural network (ANN) models (linear, non-linear, and logistic regression) [14, 35, 61]. A significant study had been done by J. Ying, Y. Wang, C. Chang, C. Chang, Y. Chen, and Y. Liou, 2019 [3] – they believed that most classical machine learning models focus on the estimation and assessment of short-term cash flow. Thus, the lack of traditional ML models would considerably be increased when the duration of the forecasting target fluctuates. They proposed a deep learning framework, Deep Bonds, to construct a forecast model to predict U.S. Treasury Yield with various issue periods. At the same time, the Recurrent Neural Network with Long Short Term Memory (LSTM) architecture was utilized for efficient summarization of U.S. Treasury Yield as distinctive vectors. Relying on the produced distinctive vectors, they predicted future U.S. treasury yield with various issue durations. Their study conducted comprehensive experimental research that relied on a real dataset collected from the website of the resource center of the U.S. Department of the Treasury. The study results showed a remarkable accuracy improvement through the deep learning approach compared with existing works. Other studies done by Y. He, J. Zhang, Z. He, 2019, and M. Ning, Z. He, N. Wang, and R. Liu, 2018 [12] used metaheuristic algorithms

Journal of Theoretical and Applied Information Technology
30th September 2023. Vol.101. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

in predicting cash flow. Other studies in cash forecasting using ANN algorithms such as LSTM and Baison are [22, 51].

2. METHODOLOGY STEPS



3. DATA COLLECTION

The data utilized in this study are authentic and have been sourced from the budget department of the Iraqi Ministry of Finance (MOF) official website. http://mof.gov.iq/pages/ar/FederalBudgetLaw.aspx. Initially, the parameters are divided into two main groups on which the government depends when building the budget. The first group represents the main parameters:

- Revenues: Oil revenues, taxes on income, a wealth tax on goods, services fees, budget share of profits, capital revenue transfers.
- Expenditures: Wages and salaries, services goods, fixed assets maintenance, capital outlays, donations, subsidies and public debt interest, foreign participation, special projects, social protection, investment budget.

The second group is the yearly needs and expenditures of every single spending unit in the government such as the needs and expenditures of each ministry and its departments, and here we will find a large database that contains very different types of needs and expenditures that rely on their discipline.

In this study, fiscal years have been collected from 2015 to 2021. The data have been filtered to shorten the matter and avoid unnecessary duplication in explanation, tables that have been chosen represent two situations: the dataset of one selected year before the COVID-19 pandemic, and the same tables for one selected year when in the pandemic, as shown in tables 1- 4. Note: the value of the numbers in the tables are in Iraqi Dinars.

Table 1: Report revenues by economic classification of the current and investment budget in 2017

Type of revenue	Current Budget	Total Budget
Oil revenues / mineral wealth	65155570.3	65,155,570
Taxes on income / wealth	4533752.74	4,533,764
Commodity taxes / fees output	1764507.15	1,764,507
Fees	790533.491	790,969
Budget share of public sector profits	700874.545	700,875
Revenue capitalism	56945.7629	56,946
Revenue manufacturing	2195243.7	2,202,311
Other income	2083948.89	2,217,231

Table 2: Total revenues and the percentage of each of the total revenue for the current budget in 2017

Total oil revenues	65496777
Total non-oil revenues	11784600
Total revenue	77281377
The ratio of oil revenues to total revenues	85%
The ratio of non-oil revenues to total revenues	15%
Percentage of total revenue	100%

Table 3: Report revenues by economic classification of the current and investment budget in 2020

Type of revenue	Current Budget	Total Budget
Oil revenues and mineral wealth	51312620.95	52231452.85
Taxes on income and wealth	1618623.569	16400679.33
Commodity taxes and fees output	396358.2469	1068312.694
Fees	607073.7873	2544339.027
Budget share of public sector profits	1045339.747	1299908.609
Revenue capitalism	83046.13017	83046.13017
Revenue manufacturing	2152253.799	2152253.799

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Other income	9175211.356	9175211.356
Total	66390527.59	84955203.8

Table 4: Total revenues and the percentage of each of the total revenue for the current budget in 2020

Total oil revenues	21512812.21
Total non-oil revenues	1057987.09
Total revenue	22570799.3
The ratio of oil revenues to total revenues	95 %
The ratio of non-oil revenues to total revenues	5 %
Percentage of total revenue	100 %

4. MEASURING THE EXTERNAL FACTORS

In most fields when the researcher tries to start determining the factors that may affect the predictions, the external factors will appear to have the greatest effect on the final results more than the standard variables (measurable/mathematical variables) which are typically used to construct the forecasting equation. The external factors are different as domains vary, they may be environmental conditions, customer demands, customer comments (string data, voice messages), local unemployment rates, governmental regulations, or pandemic disasters like the COVID-19 pandemic that has been mentioned in this study. These external factors can have the direct and utmost impact on the final forecasting results. In general, the mathematical normal variables can affect a slight gradient in changing values through its movements across time series in periods. While the external variables can make the values of the data have a sharp incline or decline, which does not happen normally when the predicting equation eliminates the external factors.

Technically, the variable that represents the external factor should be multiplied by the main variables of the forecasting equation, because it will have an extreme effect on the other values in the time series. Therefore, multiplication will show the exact influence of the external factor on the whole data values in a general sense, it is commonly observed that external factors tend to exert a deleterious impact on the data. However, this phenomenon may not hold universally in all instances. Sometimes they can have a positive influence on the data and make the data slope rise. In general, there are three primary conditions when we try to calculate the effect of the external factor, look at figure 1:

- A positive value is when the external factor already exists and has an affirmative impact.
- Negative value when the external factor already exists and has an unfavourable impact.
- The external factor value =1 if it does not exist.

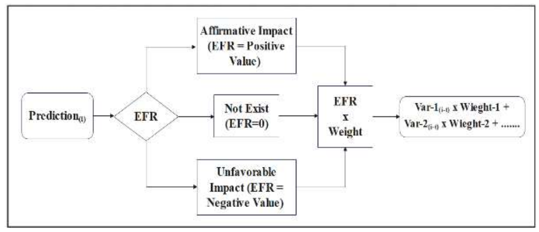


Figure 1: Measuring the external factor

5. EFFECTIVE STRATEGIES FOR SELECTING HYBRID MODELS

ARIMA is the simplest method to get a reasonable accuracy of forecasting in terms of time series. It has a simple traditional equation with minimum inputs, and the final forecasting results can be calculated with a few simple steps. ARIMA calculations for time series are straightforward, however, accurate results cannot be guaranteed in all cases. ARIMA is not equipped to handle certain data problems, such as non-stationarity, which can be a major limitation in its applicability. Despite the non-stationarity limitation of ARIMA, authors commonly choose to integrate it with an ANN model as a means to enhance the accuracy of the outputs PROPHET, which was innovated by Facebook utilizes this strategy [25, 53]. Furthermore, ARIMA has another limitation, it doesn't directly resolve the problem of integrating the non-financial factors that may have a remarkable effect on forecasting. Therefore, the latter should be extended further to overcome these obstacles [8, 33]. Other prediction models, such as linear regression, perceptrons, SMOReg, and random forest, are intended to grip specific prediction scenarios, like anticipating significant dates or events and estimating trends in people's behavior in particular fields, but they may not be suitable for all prediction situations. Certain techniques have restrictions when applied to large datasets, while others like Multivariate Adaptive Regression Splines (MARS) pose limitations such as susceptibility to overfitting, complexity, and difficulty in interpretation. Additionally, MARS may not perform well in dealing with missing data [3]. As for random forest, it generates a huge number

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of trees which causes a delay in executing the algorithm and makes it slow and ineffective for real-time forecasting. Technically, these models are fast to train, but extremely slow in creating predictions once they are trained. Furthermore, in the case of neighborhood-based models, the algorithm will not give accurate forecasts for time frames outside of the training data [46, 57].

5.1 Starting Data Analysis

This section aims to provide an exposition of the process of revenue forecasting. To that end, it shall showcase the utilization of the ARIMA model, and also eight distinct ANN forecasting models, to illustrate the range of accuracy of revenue prediction using the latter models.

5.2 Forecasting Using Common Predicting Methods

By using Python, Weka, and Excel software, common predicting methods have been implemented on the total and detailed revenues for the years 2015-2021. Look at tables 5-8 and figures 2-8.

Table 5: Forecasting using Linear Regression and Perceptron

Years	Linear Regression	Perceptron
2015	66,390,527.59	66,390,527.59
2016	54,327,966.40	54,327,966.40
2017	77,281,376.61	77,281,376.61
2018	96,995,931.31	96,995,931.31
2019	27,935,886.24	27,935,886.24
2020	22,570,799.30	22,570,799.30
2021	24,994,967.39	24,994,967.39
2022	19,064,421.44	79,598,431.51
2023	10,598,474.84	62,183,560.98
2024	2,132,528.24	17,339,622.11
2025	6,333,418.35	12,008,888.74

Table 6: Forecasting using SMOreg and Random Forest

Years	SMOreg	Random Forest
2015	66,390,527.59	66,390,527.59
2016	54,327,966.40	54,327,966.40
2017	77,281,376.61	77,281,376.61
2018	96,995,931.31	96,995,931.31
2019	27,935,886.24	27,935,886.24
2020	22,570,799.30	22,570,799.30
2021	24,994,967.39	24,994,967.39
2022	29,988,128.99	26,424,900.47
2023	21,401,577.54	25,915,825.17
2024	13,047,402.83	25,915,825.17
2025	6,836,465.24	25,915,825.17

Table 7: Forecasting using Decision Stump and ZeroR

Years	Decision Stump	ZeroR
2015	66,390,527.59	66,390,527.59
2016	54,327,966.40	54,327,966.40
2017	77,281,376.61	77,281,376.61
2018	96,995,931.31	96,995,931.31
2019	27,935,886.24	27,935,886.24
2020	22,570,799.30	22,570,799.30
2021	24,994,967.39	24,994,967.39
2022	25,167,217.65	52,928,207.84
2023	25,167,217.65	52,928,207.84
2024	25,167,217.65	52,928,207.84

Table 8: Forecasting using Random Tree and REP Tree

Years	Random Tree	REP Tree
2015	66,390,527.59	66,390,527.59
2016	54,327,966.40	54,327,966.40
2017	77,281,376.61	77,281,376.61
2018	96,995,931.31	96,995,931.31
2019	27,935,886.24	27,935,886.24
2020	22,570,799.30	22,570,799.30
2021	24,994,967.39	24,994,967.39
2022	22,570,799.30	52,928,207.84
2023	22,570,799.30	52,928,207.84
2024	22,570,799.30	52,928,207.84
2025	22,570,799.30	52,928,207.84

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

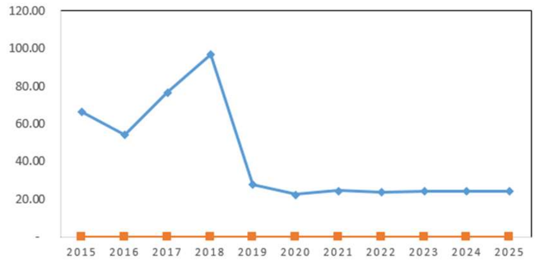


Figure 2: Forecasting using Linear Regression

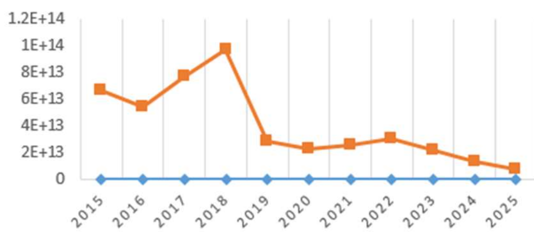


Figure 3: Forecasting using Perceptron

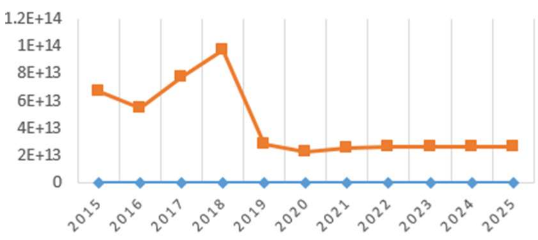


Figure. 4: Forecasting using SMOreg

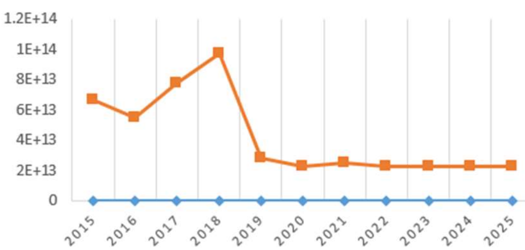


Figure 5: Forecasting using Random Forest

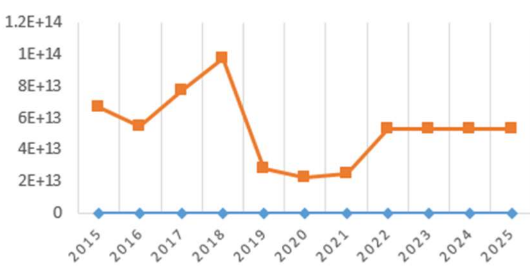


Figure 6: Forecasting using Random Tree

Figure 7: Forecasting using REP Tree

Figure 8: Forecasting using ARIMA

Table 9: Forecasting using ARIMA

Years	Revenues/Actual values	Prediction
2015	54327966.4	55327966.4
2016	77281376.61	60359246.99
2017	96995931.31	65804671.51
2018	27935886.24	87138653.96
2019	22570799.3	62465908.78
2020	24994967.39	25253342.77
2021	23782883.35	23782883.35
2022	24388975.37	24388925.37
2023	24085929.36	24085929.36
2024	24237452.36	24237452.36
2025	54327966.4	55327966.4

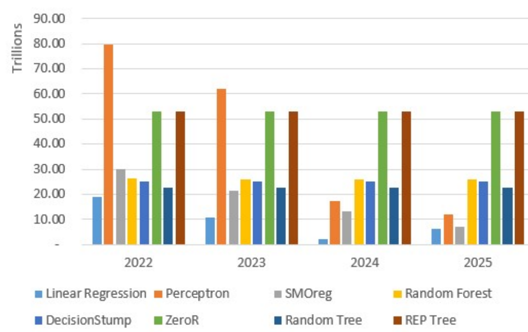


Figure 9: Cash forecasting of the revenues for the predicted years 2022 – 2025

Finally, figure 9 shows the cash forecasting of the revenues for the predicted years of 2022 – 2025, done by coding in Python.

Based on the analysis of the tables and graphs provided above, it can be inferred that the utilization of conventional forecasting techniques such as Linear Regression, Perceptron, SMOreg, Random Forest, Decision Stump, ZeroR, Random Tree, REP Tree, ARIMA, without any enhancement to their prediction equations, could potentially result in a less precise forecast outcome. For example, if we look at Table-7 and Figure 10 which shows the use of linear regression to forecast the years 2022-2025, it can be seen clearly that the slope and the numbers register a plunge from 2022 - 2025 to reach the lowest value which is (6,333,418,354,026.00). Technically, the aforementioned results are not accurate and do not reflect the economic situation of the selected period, because, essentially, there are no expectations for any external non-financial factors such as COVID-19 that may cause an immerse in the oil revenues for the selected years (2022-2025). Therefore, the expected results of forecasting should be near the real values of the oil revenues for the period 2017-2019. The main reason behind these results is that most forecasting methods essentially depend on taking the behaviour of the data in a previous period and calculating the average range to get their predicted results.

6. FORECASTING BY USING HYBRID-ANN-ARIMA-EM

The proposed model has been applied in Equation 1 to obtain predictions for the years 2017-2021, then a comparison has been done with the original data set for the same period as illustrated in table 10, figure 10.

$$\text{Hybrid} - \text{ANN} - \text{ARIMA} - \text{EM} = \text{AVG}\,(\text{AF}\mathbf{i} + \phi\mathbf{i} + \text{Ơi}) \times \text{Wi} \quad (1)$$

AF = ARIMA forecasting of the forecasted period
Φ = the value of the same period
Ơ = the value of the next period
W = weight ratio

The weight ratio is computed through Equations 2 and 3 as shown below:

$$W = 1 + \frac{1}{1e^{-\alpha}} \quad (2)$$

$$\alpha = \text{AVG(TS)} \quad (3)$$

Where:
TS = the whole period time of the selected time series
AVG = average function

Table 10: Forecasting using Hybrid-ANN-ARIMA-EM

Years	Revenues	Prediction
2017	77281376.61	65804671.51
2018	96995931.31	87138653.96
2019	27935886.24	28465908.78
2020	22570799.3	25253342.77
2021	24994967.39	23782883.35

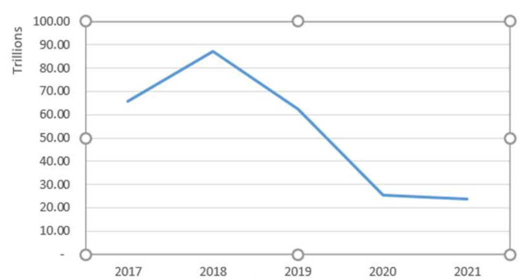


Figure 10: Forecasting by using the proposed Hybrid-ANN-ARIMA-EM model

7. COMPARING FORECASTING RESULTS

By calculating RMSE in equation 4, and MAPE in equation 5 for each forecasting model, an accuracy comparison has been done as shown in table 11,12 shows the RMSE and MAPE of the best model in terms of accuracy.

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}}{\text{N}} \quad (4)$$

Where:
N = number of non-missing data points
x = actual time series
y = estimated time series

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (5)$$

n = number of non-missing data points
A_t = actual value
F_t = forecast value

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 11: RMSE and MAPE of tested models

Model	RMSE	MAPE
Linear Regression	1.15	0.78
Perceptron	1.16	0.01
SMOreg	1.06	0.60
Random Forest	8.26	0.40
DecisionStump	8.47	0.42
ZeroR	2.96	0.05
Random Tree	9.06	0.48
REP Tree	2.79	0.05
ARIMA	1.20	0.59
Hybrid-ANN-ARIMA-EM	1.04	0.009

Table 12: RMSE and MAPE for the best model

Min RMSE	Model	Min MAPE	Model
1.04	Hybrid-ANN-ARIMA-EM	0.009	Hybrid-ANN-ARIMA-EM

The results in table 12 show that the proposed hybrid model Hybrid-ANN-ARIMA-EM has the lowest RMSE (1.04) and MAPE (0.009) compared to other selected models, which means that the new proposed model has made a distinct improvement in forecasting.

8. CONCLUSION

Cash forecasting is an unavoidable and decisive part of economic and financial analysis. Technically, predictions drive decision-making. Furthermore, accurate decisions come from accurate forecasting, and using sophisticated hybrid statistical-AI models appears to be the best choice to forecast accurate results. The study illustrated that the new proposed Hybrid-ANN-ARIMA-EM model has the best accuracy in terms of time series forecasting for the given situation, compared to other common forecasting models. In the future, the model could be optimized so that can be able to handle the variety and variation of possible external factors with accurate results and minimize the prediction error.

REFERENCES

[1] V. Fernandez-Cortez, D. Valle-Cruz and J. R. Gil-Garcia, "Can Artificial Intelligence Help Optimize the Public Budgeting Process? Lessons about Smartness and Public Value from the Mexican Federal Government," 2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG), Buenos Aires, Argentina, 2020, pp. 312-315, doi: 10.1109/ICEDEG48599.2020.9096745.

[2] Morales, J., & Huanca, J. (2020). Regression model and neural network applied to the public spending execution. International Journal of Advanced Computer Science and Applications, 11(4), 195-200. doi:10.14569/IJACSA.2020.0110426.

[3] Y. -J. Chen, J. -A. Lin, Y. -M. Chen and J. -H. Wu, "Financial Forecasting With Multivariate Adaptive Regression Splines and Queen Genetic Algorithm-Support Vector Regression," in IEEE Access, vol. 7, pp. 112931-112938, 2019, doi: 10.1109/ACCESS.2019.2927277.

[4] X. Yang, S. Mao, H. Gao, Y. Duan and Q. Zou, "Novel Financial Capital Flow Forecast Framework Using Time Series Theory and Deep Learning: A Case Study Analysis of Yu'e Bao Transaction Data," in IEEE Access, vol. 7, pp. 70662-70672, 2019, doi: 10.1109/ACCESS.2019.2919189.

[5] T. Bowen, Z. Zhe and Z. Yulin, "Forecasting method of e-commerce cargo sales based on ARIMA-BP model*," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 133-136, doi: 10.1109/ICAICA50127.2020.9181926.

[6] Li M , Ji S , Liu G . Forecasting of Chinese E-Commerce Sales: An Empirical Comparison of ARIMA, Nonlinear Autoregressive Neural Network, and a Combined ARIMA-NARNN Model[J]. Mathematical Problems in Engineering, 2018, 2018:1-12

[7] Zhang Fang, Zhou Zongfang. Research on Sales Portfolio Forecast Based on ARIMA Model and BP Neural Network [J]. Management Scientist: Academic Edition (04):38-43+80.

[8] Y. Du, "Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, 2018, pp. 2854-2857.

[9] N. Chen-xu and W. Jie-sheng, "Auto regressive moving average (ARMA) prediction method of bank cash flow time series," 2015 34th Chinese Control Conference (CCC), 2015, pp. 4928-4933, doi: 10.1109/ChiCC.2015.7260405.

[10] K. C. Varghese and A. M. Perdon, "Cashflow forecasting with linear models," 2019 International Conference on Control, Automation and Diagnosis (ICCAD), Grenoble, France, 2019, pp. 1-6, doi: 10.1109/ICCAD46983.2019.9037919.

[11] L. Lokmic and K. A. Smith, "Cash flow forecasting using supervised and unsupervised neural networks," Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, Como, Italy, 2000, pp. 343-347 vol.6, doi: 10.1109/IJCNN.2000.859419.
[12] J. Wang, C. Ning and W. Cui, "Time series prediction of bank cash flow based on grey neural network algorithm," 2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF), 2015, pp. 272-277, doi: 10.1109/ICEDIF.2015.7280205.
[13] X. Li, H. Dong and S. Han, "Multiple Linear Regression with Kalman Filter for Predicting End Prices of Online Auctions," 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2020, pp. 182-191, doi: 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00042.
[14] K. C. Varghese and A. M. Perdon, "Cashflow forecasting with linear models," 2019 International Conference on Control, Automation and Diagnosis (ICCAD), Grenoble, France, 2019, pp. 1-6, doi: 10.1109/ICCAD46983.2019.9037919.
[15] T. Q. Sun and R. Medaglia, ³Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare,´ Gov. Inf. Q., vol. 36, no. 2, pp. 368±383, Apr. 2019.
[16] Y. He, J. Zhang and Z. He, "Metaheuristic Algorithms for Multimode Multiproject Scheduling With the Objective of Positive Cash Flow Balance," in IEEE Access, vol. 7, pp. 157427-157436, 2019, doi: 10.1109/ACCESS.2019.2944746.
[17] M. Ning, Z. He, N. Wang and R. Liu, "Metaheuristic Algorithms for Proactive and Reactive Project Scheduling to Minimize Contractor’s Cash Flow Gap under Random Activity Duration," in IEEE Access, vol. 6, pp. 30547-30558, 2018, doi: 10.1109/ACCESS.2018.2828037.
[18] D. Valle-Cruz, ³Public value of e-government services through emerging technologies,´ Int. J. Public Sect. Manag., vol. 32, no. 5, pp. 473±488, Jul. 2019.
[19] Li Jie, Wang Yuxia, Zhao Xudong. Prediction Methods for Commodity Sales of E-commerce Enterprises [J]. Statistics and Decision, 2018 (22).
[20] B. Kim, "Integrating Risk Assessment and Actual Performance for Probabilistic Project Cost Forecasting: A Second Moment Bayesian Model," in IEEE Transactions on Engineering Management, vol. 62, no. 2, pp. 158-170, May 2015, doi: 10.1109/TEM.2015.2404935.
[21] X. Guo, "The Research of Forecasting Cash Inflow and Outflow Based on Time Series Analysis," 2015 8th International Symposium on Computational Intelligence and Design (ISCID), 2015, pp. 68-71, doi: 10.1109/ISCID.2015.289.
[22] C. Chou, J. Park and E. Chou, "Predicting Stock Closing Price After COVID-19 Based on Sentiment Analysis and LSTM," 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021, pp. 2752-2756, doi: 10.1109/IAEAC50856.2021.9390845.
[23] J. Ying, Y. Wang, C. Chang, C. Chang, Y. Chen and Y. Liou, "DeepBonds: A Deep Learning Approach to Predicting United States Treasury Yield," 2019 Twelfth International Conference on Ubi-Media Computing (Ubi-Media), 2019, pp. 245-250, doi: 10.1109/Ubi-Media.2019.00055.
[24] D. La Torre, D. Liuzzi and S. Marsiglio, "A Note on Stochastic Logistic Shocks and Economic Growth," 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), 2019, pp. 1-5, doi: 10.1109/ICMSAO.2019.8880319.
[25] N. Agarwal and S. S. Bhadouria, "Crime detection in rural areas using enhanced prophet routing algorithm in DTN," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016, pp. 1-5, doi: 10.1109/CDAN.2016.7570898.
[26] A. Hui, Z. Qing and L. Xuan, "The research of East Asian monetary cooperation based on symmetry of economic shocks," 2011 International Conference on Management Science & Engineering 18th Annual Conference Proceedings, 2011, pp. 841-846, doi: 10.1109/ICMSE.2011.6070058.

[27] G. Mengov and I. Nenov, "Towards an Intelligent Neural System for Economic and Business Forecasting," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 340-344, doi: 10.1109/IS48319.2020.9200124.

[28] Ayodele Ariyo Adebiyi, Aderemi Oluyinka Adewumi, Charles Korede Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction", Journal of Applied Mathematics, vol. 2014, Article.

[29] J. J. Wang, J. Z. Wang, Z. G. Zhang, and S. P. Guo, "Stock index forecasting based on a hybrid model," Omega, vol. 40, no. 6, pp. 758–766, 2012.

[30] A. Meyler, G. Kenny, and T. Quinn, "Forecasting Irish Inflation Using ARIMA Models," Technical Paper 3/RT/1998, Central Bank of Ireland Research Department, 1998.

[31] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," Expert Systems with Applications, vol. 37, no. 1, pp. 479–489, 2010.

[32] H. D. Chacón, E. Kesici and P. Najafirad, "Improving Financial Time Series Prediction Accuracy Using Ensemble Empirical Mode Decomposition and Recurrent Neural Networks," in IEEE Access, vol. 8, pp. 117133-117145, 2020, doi: 10.1109/ACCESS.2020.2996981.M. Khashei, M. Bijari, and G. A. R. Ardali, "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)," Neurocomputing, vol. 72, no. 4–6, pp. 956–967, 2009.

[33] E. Khan, "Neural fuzzy based intelligent systems and applications," in Fusion of Neural Networks, Fuzzy Systems, and Genetic Algorithms Industrial Application, C. J. Lakhmi and N. M. Martin, Eds., The CRC Press International Series on Computational Intelligence, pp. 107–139, CRC Press, New York, NY, USA, 2000.

[34] Y. Chen, B. Yang, J. Dong, and A. Abraham, "Time-series forecasting using flexible neural tree model," Information Sciences, vol. 174, no. 3-4, pp. 219–235, 2005.

[35] F. Giordano, M. La Rocca, and C. Perna, "Forecasting nonlinear time series with neural network sieve bootstrap," Computational Statistics and Data Analysis, vol. 51, no. 8, pp. 3871–3884, 2007.

[36] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," Applied Soft Computing Journal, vol. 7, no. 2, pp. 585–592, 2007.

[37] I. N. Tansel, S. Y. Yang, G. Venkataraman, A. Sasirathsiri, W. Y. Bao, and N. Mahendrakar, "Modeling time series data by using neural networks and genetic algorithms," in Smart Engineering System Design: Neural Networks, Fuzzy Logic, Evolutionary Programming, Data Mining, and Complex Systems: Proceedings of the Intelligent Engineering Systems Through Artificial Neural Networks, C. H. Dagli, A. L. Buczak, J. Ghosh, M. J. Embrechts, and O. Erosy, Eds., vol. 9, pp. 1055–1060, ASME Press, New York, NY, USA, 1999.

[38] C. K. Lee, Y. Sehwan, and J. Jongdae, "Neural network model versus SARIMA model in forecasting Korean stock price index (KOSPI)," Issues in Information System, vol. 8, no. 2, pp. 372–378, 2007.

[39] N. Merh, V. P. Saxena, and K. R. Pardasani, "A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting," Journal of Business Intelligence, vol. 3, no. 2, pp. 23–43, 2010.

[40] J. Sterba and K. Hilovska, "The implementation of hybrid ARIMA neural network prediction model for aggregate water consumption prediction," Aplimat—Journal of Applied Mathematics, vol. 3, no. 3, pp. 123–131, 2010.

[41] A. G. Lahane, "Financial forecasting: comparison of ARIMA, FFNN and SVR," 2008, http://www.it.iitb.ac.in/~ashishl/files/MTechProjectPresentation.pdf.

[42] J. T. Yao, C. L. Tan, and H. L. Poh, "Neural networks for technical analysis: a study on KLCI," International Journal of Theoretical and Applied Finance, vol. 2, no. 2, pp. 221–241, 1999.

[43] D. K. Bebarta, R. Bisoi and P. K. Dash, "Locally recurrent Functional Link Fuzzy neural network and unscented H-infinity filter for shortterm prediction of load time series in energy markets," 2015 IEEE Power, Communication and Information Technology Conference (PCITC), 2015, pp. 663-670, doi: 10.1109/PCITC.2015.7438080.

[44] V. R. Prybutok, J. Yi, and D. Mitchell, "Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations," European Journal of

Operational Research, vol. 122, no. 1, pp. 31–40, 2000.

[45] J. F. Lorenzato de Oliveira and T. B. Ludermir, "A distributed PSO-ARIMA-SVR hybrid system for time series forecasting," 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2014, pp. 3867-3872, doi: 10.1109/SMC.2014.6974534.

[46] Santika, G. D. (2018). Enabling external factors for consumption electricity forecasting using hybrid genetic algorithm and fuzzy neural system. Paper presented at the Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017, , 2018-January 1-6. doi:10.1109/CAIPT.2017.8320708 Retrieved from www.scopus.com

[47] T. H. Roh, "Forecasting the volatility of stock price index," Expert Systems with Applications, vol. 33, no. 4, pp. 916–922, 2007.

[48] H. Al-Qaheri, A. E. Hassanien, and A. Abraham, "Discovering stock price prediction rules using rough sets," Neural Network World, vol. 18, no. 3, pp. 181–198, 2008.

[49] S. Bano, N. Nida and S. A. Irtaza, "Financial Temporal Patterns and Deep FinanicalNet For Crude-Oil Prices Forecasting," 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-6, doi: 10.1109/INMIC50486.2020.9318086.

[50] S. K. Mitra, "Optimal combination of trading rules using neural networks," International Business Research, vol. 2, no. 1, pp. 86–99, 2009.

[51] G. Bathla, "Stock Price prediction using LSTM and SVR," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 211-214, doi: 10.1109/PDGC50313.2020.9315800.

[52] M. M. Mostafa, "Forecasting stock exchange movements using neural networks: empirical evidence from Kuwait," Expert Systems with Applications, vol. 37, no. 9, pp. 6302–6309, 2010.

[53] C. R. Madhuri, M. Chinta and V. V. N. V. P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA," 2020 7th International Conference on Smart Structures and Systems (ICSSS), 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9202042.

[54] T. H. Yu and K. H. Huarng, "A neural network-based fuzzy time series model to improve forecasting," Expert Systems with Applications, vol. 37, no. 4, pp. 3366–3372, 2010.

[55] A. O. Adewumi and A. Moodley, "Comparative results of heuristics for portfolio selection problem," in Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr '12), pp. 1–6, New York, NY, USA, March 2012.

[56] P. Areekul, T. Senjyu, H. Toyama and A. Yona, "Combination of artificial neural network and ARIMA time series models for short term price forecasting in deregulated market," 2009 Transmission & Distribution Conference & Exposition: Asia and Pacific, 2009, pp. 1-4, doi: 10.1109/TD-ASIA.2009.5356936.

[57] Z. Junlan and Y. Chengke, "Data-analysis-based discussion on COVID-19 Pandemic Shocks to the Economy and Policy Responses : Cases in the United Kingdom," 2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID), 2020, pp. 538-541, doi: 10.1109/MSIEID52046.2020.00109.

[58] G. Mengov and I. Nenov, "Towards an Intelligent Neural System for Economic and Business Forecasting," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 340-344, doi: 10.1109/IS48319.2020.9200124.

[59] R. Sallehuddin, S. M. Shamsuddin and S. Z. M. Hashim, "Hybridization Model of Linear and Nonlinear Time Series Data for Forecasting," 2008 Second Asia International Conference on Modelling & Simulation (AMS), 2008, pp. 597-602, doi: 10.1109/AMS.2008.142.

[60] Wibowo, W., Dwijantari, S., & Hartati, A. (2017). Time series machine learning: Implementing ARIMA and hybrid ARIMA-ANN for electricity forecasting modeling doi:10.1007/978-981-10-7242-0_11.

[61] Y. Peng, K. He and Q. Yu, "Stock Index Prediction Method based on ARIMA-ELM Combination Model," 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), 2021, pp. 637-642, doi: 10.1109/AEMCSE51986.2021.00132.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[62] V. Hryhorkiv, L. Buiak, A. Verstiak, M. Hryhorkiv, O. Verstiak and K. Tokarieva, "Forecasting Financial Time Sesries Using Combined ARIMA-ANN Algorithm," *2020 10th* International Conference on Advanced Computer Information Technologies (ACIT), 2020, pp. 455-458, doi: 10.1109/ACIT49673.2020.9208859.

[63] Panigrahi, S., & Behera, H. S. (2019). An adaptive fuzzy filter-based hybrid ARIMA-HONN model for time series forecasting doi:10.1007/978-981-10-8055-5_74.

[64] Belmahdi, B., Louzazni, M., & Bouardi, A. E. (2020). A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in morocco. European Physical Journal Plus, 135(11) doi:10.1140/epjp/s13360-020-00920-9

[65] Büyükşahin, Ü. Ç., & Ertekin, Ş. (2020). A feature-based hybrid ARIMA-ANN model for univariate time series forecasting. [Tek değişkenli zaman serileri tahmini için öznitelik tabanlı hibrit ARIMA-YSA modeli] Journal of the Faculty of Engineering and Architecture of Gazi University, 35(1), 467-478. doi:10.17341/gazimmfd.508394.

[66] R. M. Kapila Tharanga Rathnayaka, D. M. K. N. Seneviratna, W. Jianguo and H. I. Arumawadu, "A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models," 2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC), 2015, pp. 54-60, doi: 10.1109/BESC.2015.7365958.

[67] S. Chen et al., "Cash Flow Forecasting Model for Electricity Sale Based on Deep Recurrent Neural Network," 2019 IEEE International Conference on Power Data Science (ICPDS), 2019, pp. 67-70, doi: 10.1109/ICPDS47662.2019.9017192.

[68] G. D. Santika, "Enabling external factors for consumption electricity forecasting using hybrid genetic algorithm and fuzzy neural system," 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), 2017, pp. 1-6, doi: 10.1109/CAIPT.2017.8320708.

[69] A. Garlapati, D. R. Krishna, K. Garlapati, N. m. Srikara Yaswanth, U. Rahul and G. Narayanan, "Stock Price Prediction Using Facebook Prophet and Arima Models," 2021 6th International Conference for Convergence in Technology (I2CT), 2021, pp. 1-7, doi: 10.1109/I2CT51068.2021.9418057.

[70] L. Yang and H. Yang, "A Combined ARIMA-PPR Model for Short-Term Load Forecasting," 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Chengdu, China, 2019, pp. 3363-3367, doi: 10.1109/ISGT-Asia.2019.8881315.