International Journal of Technology (2015) 5: 872-880
ISSN 2086-9614

FORECASTING ANALYSIS OF CONSUMER GOODS DEMAND USING NEURAL NETWORKS AND ARIMA

Arian Dhini[1*], Isti Surjandari[1], Muhammad Riefqi[1], Maya Arlini Puspasari[1]

[1]*Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Kampus Baru UI Depok, Depok 16424, Indonesia*

(Received: June 2015 / Revised: September 2015 / Accepted: October 2015)

ABSTRACT

Accurate forecasting of consumer demand for goods is extremely important as it allows companies to provide the right amount of goods at the right time. Autoregressive integrated moving average (ARIMA) is a popular method for forecasting time series data, and previous studies have shown that ARIMA can produce fairly accurate forecasting results. On the other hand, the neural network method has advantages in detecting non-linear patterns in data. In addition to these methods, the hybrid method, which combines the ARIMA and neural network methods, was applied in this study. A comparison analysis was conducted to determine the best performing model. In this study, the neural network model was found to be the most accurate.

Keywords: ARIMA; Consumer goods; Forecasting; Neural network

1. INTRODUCTION

Demand management is one of the critical parts of the supply chain system because it determines what, how many, where, and when goods should be provided (Rexhausen et al., 2012). In order to conduct an effective demand management, an accurate demand forecasting should be performed. Adebanjo and Mann (2000) stated that the ability to forecast consumer demand accurately is of great importance to companies in the consumer market, as this enables companies to determine the number of items that need to be produced. The absence of an accurate forecasting method can lead to higher safety stock to prevent goods scarcity. Higher levels of safety stock in turn lead to higher inventory levels. Inventory is considered waste in operations management, so its levels should be kept as low as possible (Liker, 2004). Conversely, lower safety stock reduces inventory costs, which has a positive impact on a company's performance. Therefore, accurate forecasting is critical to achieving effective and efficient production.

The accuracy of time series forecasting is challenging for scientists (Taskaya-Temizel & Ahmad, 2005). Time series data often comprise linear as well as non-linear components (Faruk, 2010). In some cases, linear-based approaches might be more suitable than non-linear ones due to the data characteristics. Neural network methods provide an excellent alternative as they can be applied in non-linear time series forecasting. In addition, some mixture models can have benefits over the single models (Taskaya-Temizel & Ahmad, 2005). Nevertheless, none of these approaches has been found to be superior for all types of data.

In this study, three approaches for time series data forecasting were applied, namely

* Corresponding author's email: arian@ie.ui.ac.id, Tel. +62-21-78888805, Fax. +62-21-78885656
Permalink/DOI: http://dx.doi.org/10.14716/ijtech.v6i5.1882

autoregressive integrated moving average (ARIMA), including seasonal ARIMA (SARIMA), neural network, and a hybrid of the neural network and ARIMA. This study aimed to obtain an accurate forecasting model for consumer goods demand to minimize inventory. The performance of these three approaches was compared by their error values.

2. METHODOLOGY

The study comprised three time series forecasting methods. The first method, ARIMA, is a univariate forecasting method and is quite popular for its time series data analysis (Ong et al., 2005). ARIMA is an empirical method that systematically identifies, estimates, diagnoses, and forecasts time series data. In addition, no basic assumptions are applied, therefore ARIMA is very flexible (Ho & Xie, 1998); however, ARIMA models are based on the assumption of linearity and errors in the original data (Babu & Reddy, 2014).

The neural network method is a mathematical model that resembles the way the brain works. It is widely used in business applications that require pattern recognition, prediction, classification, forecasting, and optimization (Bennell et al., 2006). Kotsialos et al. (2005) suggested that the predictions generated by neural networks have greater accuracy compared to classical forecasting methods, and Zhang (2001) indicated that neural networks are useful for modeling and predicting the properties of time series data. Cybenko (1989) described neural networks as having a universal non-linear function and a relatively good degree of forecasting accuracy. In addition, according to Hill et al. (1996), neural network forecasting provides better results than traditional forecasting methods over monthly as well as quarterly periods.

The last method, the hybrid method, is a combination of ARIMA and the neural network method. According to Faruk (2010), hybrid methods have a higher degree of accuracy than neural networks. ARIMA can recognize time series patterns well but not non-linear data patterns. On the other hand, neural networks only handle non-linear data. Hybrid models therefore combine the advantages of ARIMA with respect to linear modeling and neural networks in terms of non-linear edge modeling (Cybenko, 1989). Notwithstanding, in some circumstances, the single model approach can outperform hybrid models (Taskaya-Temizel & Ahmad, 2005).

2.1. Data Collection

The data for this study was obtained from one of the largest consumer goods companies in Indonesia for 81 periods over a span of two years. The weekly consumer goods demand data (in cartons with the same dimensions) for three kinds of products, namely cooking oil, margarine, and butter, were analyzed. Figures 1, 2, and 3 show the time series demand plot of the three types of consumer goods data.

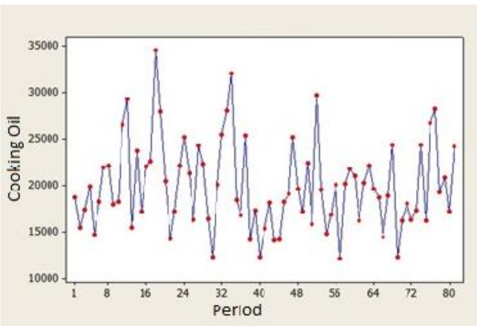

Figure 1 Demand plot of cooking oil

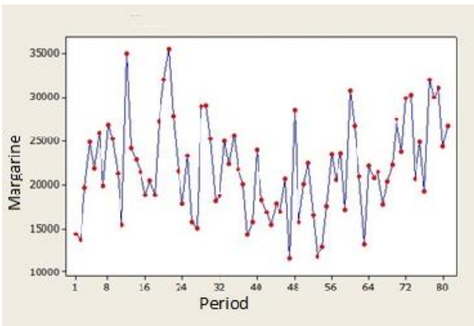

Figure 2 Demand plot of margarine

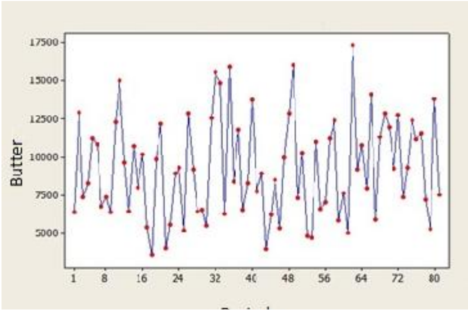

Figure 3 Demand plot of butter

The plots of the three sources of data reveal no trend in the data, thus it can be said that the data are stationary. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test further confirmed that all the data were stationary. As a result, no differentiation process was required.

The data processing was conducted by applying the three identified approaches (i.e., ARIMA, the neural network method, and a hybrid of ARIMA and the neural network method). The following subsections explore each of these approaches.

2.2. ARIMA Model

ARIMA is a model of time series forecasting that combines autoregression, an integrated process, and the moving average. Autoregression is the development of simple linear regression. A model of ARIMA (1, 0, 0), often called the AR (1) model, is represented in Equation 1:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + e_t \tag{1}$$

where θ and ϕ are the selected coefficients to minimize the total of the squared errors. The equation shows that the autoregression variable, *Yt*, is the linear functions of the past (Y_{t-1}). The ARIMA moving average processes have similarities with exponential smoothing. The moving average in the ARIMA equation can be written as follows:

$$Yt = \mu - \theta_1 e_{t-1} + e_t \tag{2}$$

where θ_1 is the estimated coefficient and the value of *Yt* only correlates with the value of the previous period forecasting error (e_{t-1}). The integrated process is not stationary and involves random walks and trends. As we know, if a sequence is a random walk, then the actual value of the previous period will be the best predictor of the future value. This can then be expressed in a model of ARIMA (0, 1, 0) in the form of a random walk model:

$$Y_t = Y_{t-1} + e_t \tag{3}$$

Modeling of ARIMA uses a simple and versatile notation, which is compiled based on the current level of autoregression, integration, and the moving average. This standard notation identifies the order of autoregression with p, integration or differencing with d, and the moving average with q. There are several stages in ARIMA research, namely model identification, parameter estimation, and diagnostic checking.

ARIMA model identification is usually conducted by analyzing the value of the autocorrelation and the partial autocorrelation in the data. At this stage, we usually also initiate the differentiation process to make the data stationary. Stationary data are data that have a constant mean and variance over time.

After the identification of the model has been completed, the next step is to conduct parameter estimation. The coefficient of the parameter is estimated by finding the value of the coefficient parameters that gives the smallest error value. Parameter estimation can be undertaken using a non-linear optimization procedure. Lastly, diagnostic checking is performed to validate the forecasting model. Ljung Box is selected to detect any correlation among residuals.

2.3. Neural Network Model

Neural networks are formed from the simple process units that are connected to each other to form a structure, which can then be studied to determine the relationships among a set of variables. The neural network model is thus able to predict the non-linear properties of data.

One advantage of neural networks compared to other non-linear models is its universal model, which is capable of predicting fairly extensive functions with a high degree of accuracy. No assumptions are required for neural networks, thus neural networks conform to the characteristics of the data.

While conducting time series forecasting using a neural network, usually a single hidden layer feedforward network is used. The model consists of three interconnected layers (i.e., the input layer, the hidden layer, and the output layer). Relationships that occur in the output (yt) and input (y_{t-1} y_{t-2}, ...,p_{y-t}) layers follow mathematical equations like Equation 4:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right) + \varepsilon_t \tag{4}$$

where αj (j = 1,2,3, ..., q) and βij (i = 1,2,3, ..., p; j = 1,2,3, ..., q) are the parameters of the model (often called the weights), p is the number of input points (input nodes), and q is the number of hidden nodes. The activation function used in the hidden layer is the logistic sigmoid function and the linear function is the output layer.

Based on Equation 4, the neural network model is a non-linear function that maps the values of observation (y_{t-1}, y_{t-2},…,y_{t-p}) on the future value of the yt, which can be expressed as follows:

$$Yt = f\,(y_{t-1}, y_{t-2},\ldots,y_{t-p}, w) + \varepsilon_t \tag{5}$$

where w is a vector of the parameters and f is a function that is determined based on the network structure and weight. The equation form in Equation 5 is non-linear autoregressive (NAR).

The last method, the hybrid method, is a combination of ARIMA and the neural network method. According to Kaastra and Boyd (1996), a neural network with a single hidden layer comprising a sufficient number of hidden neurons has the ability to perform almost all functions. In fact, neural networks with one or two hidden layers are widely used and offer good performance.

In this study, three layers were used: one input layer, one hidden layer, and one output layer. In addition, three hidden neurons were used (i.e., 1, 2, 3 hidden neurons). There is still no single function that can be used to determine the optimal number of hidden neurons, but some experts have provided their opinions, including Kaastra and Boyd (1996) who stated that the number of hidden neurons is the square root of the number of input layers times the number of output layers. In this study, the optimal number of hidden neurons was set at one and a half to three times the amount of the input layer. Since one input layer and one output layers were used in the study, the number of hidden neurons was one, two, and three hidden neurons. Nevertheless, as yet there is no reliable formula to determine the optimal number of hidden neurons. By following the rule of thumb, an analysis of four and five hidden neurons was conducted in an attempt to determine the model offering the best results.

Too many hidden neurons can lead to several problems. First, too many hidden neurons in the hidden layer may result in overfitting. Overfitting occurs when the neural network has too much capacity for information processing so that the amount of information in the training data is not sufficient to train all the neurons in the neural network. Second, increasing the number of hidden neurons lengthens the time taken to process the data. In both cases, this will produce substandard results in the training and lead to poor neural network performance.

The activation functions used were the sigmoid function and the linear function. The sigmoid function was used because it has a gradient proportional with its output reflection and is the most widely used function in research using neural networks. The proportion of data used was 70% for the training data, 15% for data validation, and the remaining 15% for the data testing. The training method used was Levenberg–Marquardt backpropagation. This method was selected due to the fast computation of backpropagation neural networks and its use of supervised training.

2.4. Hybrid Method

The hybrid method is a combination of the two methods described in the previous section (i.e., ARIMA and the neural network method). Both ARIMA and neural networks have their respective advantages for certain data domains. ARIMA is suitable for modeling linear data while the neural network method can accommodate non-linear properties in the data.

The application of ARIMA for analyzing non-linear problems in a complex analysis would yield unexpected results. Conversely, the use of neural networks for linear problems would give poor results in the analysis. Previous studies have shown that if there is an outlier or multicolinearity in the data, this can be analyzed much better with a neural network than linear regression.

Mathematically, time series data can be expressed as a combination of linear and non-linear components:

$$Yt = Lt + Nt \tag{6}$$

where *Yt* shows the time series data, *Lt* indicates the linear components, and the non-linear components are represented by *Nt*. In this case, the linear components were modeled using ARIMA; the residual of the estimated ARIMA model would then only contain non-linear data. Assuming e_t is the residual of the linear model at period *t*, then

$$e_t = y_t - \hat{L}_t \tag{7}$$

where *Lt* is the predicted value in the period *t* and y_t is the actual value in that period. Because ARIMA is unable to detect non-linear relationships in the data, the residual from the linear model must be analyzed using a method, such as the neural network method, which is capable of detecting non-linear properties. The neural network method is thus used to predict the error of the ARIMA model.

Mathematically, the neural network model for residual of n input nodes can be expressed as the following:

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t \quad (8)$$

where *f* is a non-linear function that is specified by the neural network. With regard to the results of the prediction error of *Nt*, the combination forecast using the hybrid method can be expressed as:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \quad (9)$$

2.5. Comparison of the Models

Two kinds of analysis can be used to determine the level of accuracy of forecasting: the mean squared error (MSE) and the mean absolute percentage error (MAPE). The MSE is the error rate calculation method derived by summing the squares of the error and then dividing this total by the amount of data or the period used. Mathematically, the MSE can be expressed as:

$$MSE = \frac{\sum_{k=0}^{n}(At - Ft)^2}{n} \quad (10)$$

The MAPE, which is an error calculation method in forecasting, is calculated by finding the percentage error of each forecast period, then dividing this by the number of data periods used. Mathematically, it is represented as follows:

$$MAPE = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\frac{Ft - At}{At}\right| \quad (11)$$

where *n* is the number of data, *Ft* is the forecasted value of demand, and *At* is the actual demand.

3. RESULTS AND DISCUSSION

3.1. Results

In this study, since the coefficient parameters of ARIMA for all three products were not significant, an additional analysis was performed. An analysis using the SARIMA method was presented since the ARIMA result was suboptimal due to the seasonal characteristics of the data. Therefore, a special model of ARIMA that took into account seasonality was applied to process the data. The experimental results using all four methods are shown on Tables 1 and 2 below.

Table 1 MSE analysis results comparison

Methods	Cooking oil	Margarine	Butter
ARIMA	21,780,889	30,503,529	10,758,400
SARIMA	23,181,510	24,041,001	11,493,141
Neural network	12,296,906	16,207,306	6,678,047
Hybrid	19,521,349	22,541,955	9,358,159

Table 2 MAPE analysis results comparison

Methods	Cooking oil	Margarine	Butter
ARIMA	19.41%	20.14%	34.94%
SARIMA	19.60%	18.98%	35.41%
Neural network	17.54%	17.41%	27.89%
Hybrid	17.68%	17.48%	30.01%

From Tables 1 and 2, it can be seen that the neural network method was superior compared to the other two methods in terms of forecasting accuracy. Indeed, both the MSE and MAPE values show that the neural network method managed to outperform the other methods.

3.2. Discussion

Based on the results of the data processing in this study, it is clear that the neural network method is more accurate than the other models. This may be due to the non-linearity of the data. The data spreads in Figures 4–6 show a non-linear relationship between the period and the demand for each type of consumer goods. Traditional methods such as the Box–Jenkins model and ARIMA require the assumption that the time series data used in forecasting are linear, therefore they are not suitable for predicting data that are non-linear (Khashei & Bijari, 2010).

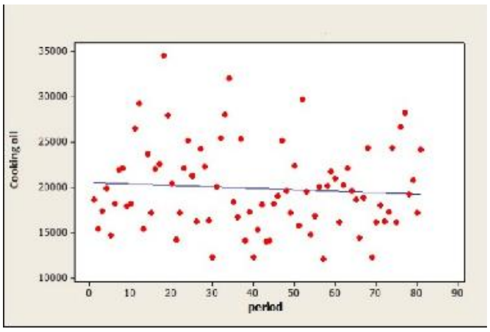

Figure 4 Scatterplot of cooking oil demand

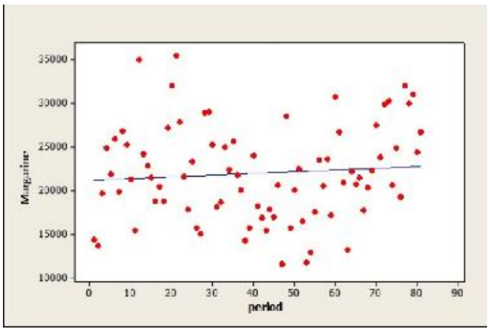

Figure 5 Scatterplot of margarine demand

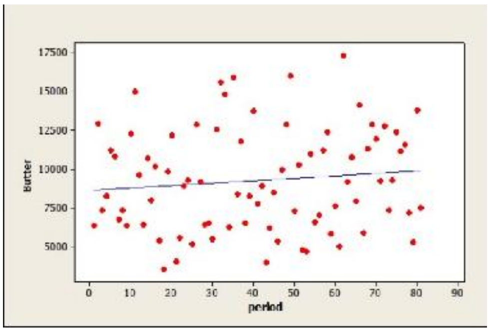

Figure 6 Scatterplot of butter demand

The underlying mechanism for ARIMA is linear process (Zhang, 2003, Khashei & Bijari, 2010; Jian-Chang et al., 2004). On the other hand, the neural network method has the ability to detect the properties of non-linear data and thus produces better results. It is also apparent from the results of this study that, when data are non-linear, the neural network method has advantages over both the hybrid method and ARIMA. Contrary to our findings, Zhang (2001) found that the hybrid method can outperform ARIMA and the neural network method. Nevertheless, Taskaya-Temizel and Ahmad (2005) also found that the hybrid method does not always give better results compared to single methods.

According to Taskaya-Temizel and Ahmad (2005), there are two factors that prevent the hybrid ARIMA–neural network method from delivering better results. First, the assumption of the existence of a relationship between the components of the linear and non-linear components in the data can cause performance degradation as other model relationships (e.g., multiplicative) may exist within the data instead of linear/non-linear relationships. Second, no one can guarantee that the residual of the linear components will have valid non-linear patterns.

Of the three kinds of products assessed in this study, butter had quite an interesting characteristic of error as the forecasting model for the butter data produced the smallest value of the MSE compared to the other products. In contrast, the butter model also produced the greatest value of the MAPE compared to the other products. This could have been caused by the basic equation of the calculation related to the error evaluation method. The MSE is the average result of the squared error, which is the difference between the actual values and the values predicted by the model. The smaller value will give the smaller results of the squared data (MSE). In our study data, the demand for butter was much smaller with an average demand of 9,276 cartons compared to the average demand for cooking oils and margarine (19,864 and 21,969 cartons, respectively). In the other hand, MAPE is the average percentage value of the difference between the actual values and the predicted value compared to the actual value itself. The smaller value of divisor, as in butter, will likely produce the greater value of MAPE.

4. CONCLUSION

In this study, the hybrid method, which combines elements of the neural network method and ARIMA, was applied. The advantages of this hybrid method were that it merged the ARIMA method for the linear data and the neural network method for the non-linear properties with the aim of producing the highest possible accuracy rate compared to the other two methods individually; however, based on the results of the data analysis, the neural network method had better accuracy than the hybrid method and ARIMA. The basic assumption of the hybrid method was that the time series data consisted of linear and non-linear components, but this was identified as a possible weakness in the hybrid method. All three of these methods can improve the performance of forecasting but could also lower the level of accuracy of forecasting.

The results of this study indicated that the hybrid method does not always give better results than the single methods as the neural network method outperformed the hybrid method. Some of the possible causes for this are the basic assumptions used in the method as well as the possibility that the residual from the linear components may not be non-linear.

Although the ARIMA method delivered the worst performance among all the methods due to the non-linear properties of the data, it has advantages when applied to linear data. The forecasting performance of ARIMA in this study however was unexpectedly poor due to the seasonality of the data. Experiments using SARIMA for the same data provided more satisfactory results and also generated significant parameters. Moreover, the forecasting accuracy of SARIMA was not significantly different from that of ARIMA.

The neural network method has the potential to be the best model. Optimization of the neural network parameters, such as the number of layers, the number of hidden neurons, and the training methods, could lead to the development of the best model and should therefore be undertaken in future research. Hybrid methods are also useful as they combine the characteristics of single methods, including the advantages of each method. Future research could potentially combine a wider range of methods, each of which has its own characteristics and advantages, to produce optimal forecasting accuracy.

5. REFERENCES

Adebanjo, D., Mann, R., 2000. Identifying Problems in Forecasting Consumer Demand in the Fast Moving Consumer Goods Sector. *Benchmarking: An International Journal*, Volume 7(3), pp. 223–230

Babu, C.N., Reddy, B.E., 2014. A Moving-average Filter Based Hybrid ARIMA–ANN Model for Forecasting Time Series Data. *Applied Soft Computing*, Volume 23, pp. 27–38

Bennell, J.A., Crabbe, D., Thomas, S.,ApGwilym, O., 2006. Modelling Sovereign Credit Ratings: Neural Networks Versus Ordered Probit. *Expert Systems with Applications*, Volume 30(3), pp. 415–425

Cybenko, G., 1989. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals and Systems*, Volume 2(4), pp. 303–314

Faruk, D.Ö., 2010. A Hybrid Neural Network and ARIMA Model for Water Quality Time Series Prediction. *Engineering Applications of Artificial Intelligence*, Volume 23(4), pp. 586–594

Hill, T., O'Connor, M., Remus, W., 1996. Neural Network Models for Time Series Forecasts. *Management Science*, Volume 42(7), pp. 1082–1092

Ho, S.,Xie, M., 1998. The Use of ARIMA Models for Reliability Forecasting and Analysis. *Computers and Industrial Engineering*, Volume 35(1–2), pp. 213–216

Jian-Chang, L., Dong-xiao, N., Zheng-Yuan, J., 2004. A Study of Short-term Load Forecasting based on ARIMA-ANN. *Proceedings of International Conference on Machine Learning and Cybernetics*, Volume 5, pp. 3183–3187

Kaastra, I., Boyd, M., 1996. Designing a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing*, Volume 10(3), pp. 215–236

Khashei, M.,Bijari, M., 2010. An Artificial Neural Network (p, d, q) Model for Timeseries Forecasting. *Expert Systems with applications*, Volume 37(1), pp. 479–489

Kotsialos, A., Papageorgiou, M.,Poulimenos, A., 2005.Holt-winters and Neural-Network Methods for Medium-term Sales Forecasting. *The 16th IFAC World Congress*, Prague, Czech Republic

Liker, J., 2004. *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*, McGraw-Hill

Ong, C.-S., Huang, J.-J.,Tzeng, G.-H., 2005.Model Identification of ARIMA Family using Genetic Algorithms. *Applied Mathematics and Computation*, Volume 164(3), pp. 885–912

Rexhausen, D., Pibernik, R., Kaiser, G., 2012. Customer-facing Supply Chain Practices—The Impact of Demand and Distribution Management on Supply Chain Success. *Journal of Operations Management*, Volume 30(4), pp. 269–281

Taskaya-Temizel, T., Ahmad, K., 2005. Are ARIMA Neural Network Hybrids Better than Single Models?, In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, Volume 5, pp. 3192–3197

Zhang, G.P., 2001. An Investigation of Neural Networks for Linear Time-series Forecasting. *Computers and Operations Research*, Volume 28(12), pp. 1183–1202

Zhang, G.P., 2003. Time Series Forecasting using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, Volume 50, pp. 159–175