

## SPECIAL ARTICLE

MATHEMATICAL APPROACH TOWARDS RECENT INNOVATION IN COMPUTATION AND ENGINEERING SYSTEM (MATRICS)

# PREDICTION OF SOCIAL MEDIA BASED ON ARIMA AND ANN MODEL

Priyadarshini E.<sup>1\*</sup>, Manjula T.<sup>2</sup>, Krishnareddy Deepika<sup>3</sup>, Anuhya Bandi<sup>4</sup>

<sup>1,2</sup>Department of Mathematics, Sathyabama Institute of Science & Technology, Chennai, INDIA

<sup>3,4</sup>Department of ECE, Sathyabama Institute of Science & Technology, Chennai, INDIA

### ABSTRACT

One of the popular approaches for prediction of time series is Auto Regressive Integrated Moving Average abbreviated as ARIMA. The main objective of this paper is to predict the percentage of Facebook users with the time series approach. Methods: To obtain optimum accuracy, statistical support like autocorrelation and partial autocorrelation has been used. Standard statistical techniques are used to verify the validity of the model. The prediction power of ARIMA model was used to predict the percentage of Facebook users for succeeding month and to estimate the mean absolute percentile error (MAPE). Results: The performance of Artificial Neural Network using Multi Layer Perceptron (MLP) in the prediction of percentage of Facebook users was evaluated. In this study, a total of 122 data points (monthwise) of past 11 years from April 2009 to April 2019 has been taken to explore and predict Facebook users based on statistical and computational techniques. Conclusions: The Mean Absolute Percentage Error (MAPE) was used to evaluate the accuracy of the models.

### INTRODUCTION

In today's advanced world, social media play a humungous part. Social media are the interactive technologies that are mediated by computer. They facilitate sharing and creation of information via virtual communities and networks. One such huge platform of social media is Facebook. It was founded in 2004 by Mark Zuckerberg and his fellow Harvard students. It is considered as one among "Big Four" technology companies which are Amazon, Apple and Google. Though founders initially confined websites membership to students of Harvard, later it was extended to IVY LEAGUE schools, MIT and higher education institutions in Boston area. In 2006 began the outgrowth when anyone who claimed to be at least 13 years old were allowed to register as a member. Facebook can be accessed from devices which have internet connectivity. By 2007, Facebook had 100,000 pages on which various companies promoted themselves. Pages of organizations began rolling out in the year 2009. The company had announced 500 million users in the month of July in 2010. Half of the people on Facebook used it daily for an average of 34 minutes. In 2015 Facebook announces that it had reached 2 million active advertisers with most gain coming from small businesses. In 2016, the number had reached 3 million with more percentage coming from outside United States. As they say that anything overused harms, Facebook's over usage did have multiple drawbacks. Spreading of fake messages was easy using such platform. In 2015, Facebook's algorithm was revised trying to filter out fake news stories and hoaxes. In addition to spreading fake news, the perpetrator of March 2019 New Zealand attacks used Facebook for live streaming of footage. Though Facebook is advantageous, it also has drawbacks of vulnerability. The outcome depends on proper and purpose of usage.

Peng et al. [1] applied the ARIMA model to forecast the crime which is helpful for the local police stations and municipal government. Hong et al. [2] predicted the traffic flow using ARIMA models. Using ARIMA models Weng Dongdong [3] had predicted the consumer price index (CPI). In the paper [4] Chinese mobile user forecast was discussed using ARIMA models. Babu and Reddy [5] in their paper analysed the average the global temperature using ARIMA models. Hussain et.al [6] made three types of forecasts such as historical, ex-post and ex-ante. Amin et al. [7] made analysis for potato prices and forecasts the nine future points. Priyadarshini et al. made forecasts of foreign exchange rates, silver rates and crude oil rates using ARIMA models [8, 9,]. A comparative analysis has been done using Artificial Neural Network and Auto Regressive Integrated Moving by Priyadarshini et al. [10].

### ARIMA METHODS

ARIMA models are most commonly used to predict time series that can be stated by transformation lags of differenced series that appear in forecasting equation are called auto regressive terms . The lags of forecast errors are called moving average terms and a time series that needs to be differenced to be made stationary is called an integrated version of a stationary series. Random trend models, exponential smoothing model, auto regressive models and random walk are special cases of ARIMA models. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model where:

- p is the number of auto regressive terms
- d is the number of non-seasonal difference

**KEY WORDS**  
Forecast, Auto  
Correlation, Partial Auto  
Correlation, Hidden  
layers, Multi Layer  
Perceptron

Received: 25 Oct 2019  
Accepted: 24 Feb 2020  
Published: 3 Mar 2020

\*Corresponding Author  
Email:  
priyaeb@gmail.com

- $q$  is the number of moving average terms

The following is how simple regression model can be expressed

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + e_t \quad [1]$$

Where  $Y_t$  is the forecast variable  $y_{t-1}, \dots, Y_{t-p}$  are explanatory variables and  $e_t$  is the error term. The name auto regression is used to denote the above equation due to the time-lagged values of the explanatory variable. The moving average model uses the past errors as the explanatory variables. A simple moving average model is represented as follows:

$$Y_t = b_0 + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_p e_{t-p} + e_t \quad [2]$$

Similarly, a seasonal model can be represented as ARIMA (p,d,q) (P,D,Q). Basically, this method has three phases: model identification, parameters estimation and diagnostic checking. To identify the appropriate ARIMA model for a time series, the order(s) of differencing the series and remove the gross features of seasonality.

### Artificial neural network (ANN)

Artificial Neural Network (ANN) uses a variety of optimization tools to learn from past experiences and uses this to predict and find new patterns. Using training algorithms, ANN systems are made to learn the percentage of Facebook users. Learning involves the extraction of rules or pattern from the historic data. In this neural network, using Multilayer Perceptron (MLP), models have been used for the prediction of Facebook users.

### Predictors for the ANN

The five predictors for ANN are the technical indicators namely three monthly moving averages, six monthly moving averages, one yearly moving averages and two yearly moving averages respectively [Table 4].

### Experimental results of multi-layer perceptron (MLP) neural network

The MLP computes the new value of the weights and biases using the gradient descent. It quickly adjusts the network weights giving good performance. The space denoting the error for every combination of weights and biases is called as error space. The feed forward neural network architecture used in this experiment consists of two hidden layer along with one input and output layer respectively. The transfer function which is used in the hidden layer and output layer are hyperbolic tangent and identity [Table 5].

### Backward pass:

- Compute the error
- Compute
- Compute
- Keep updating the weights connecting the input layer to the hidden layer using the rule. After determining all the  $\mu$  factors, the weights are adjusted for all layers simultaneously.

$x$  :  $p \times 1$  input vector

$h$  : Weighted sum of input stimuli

$r$  :  $m \times 1$  output vector of hidden layer

$g$  : Weighted sum of  $v_j$

$Y$  :  $n \times 1$  output vector of output layer

$w_{ij}$  : weight connecting  $i$ th unit of output layer and  $j$ th unit of hidden layer

$w_{jk}$  : Weight connecting  $j$ th unit of hidden layer to  $k$ th unit of input layer

$y$  : Actual output

$y_d$  : Desired output.

where  $i, j$  and  $k$  indices refers to the neurons in the output, hidden and input layers respectively,  $p, m$  and  $k$  are the number of neurons in input, hidden and output layer respectively.

### Data analysis using arima

The analysis is taken for the data from the monthly percentage of Facebook users for a period of 10 years from 2009 to 2019 [Fig 1]. X-axis denotes the months from April 2009 to April 2019. The percentage of Facebook users was forecasted for the month of May 2019 using ARIMA modelling [Fig. 3].

#### a) Model identification:

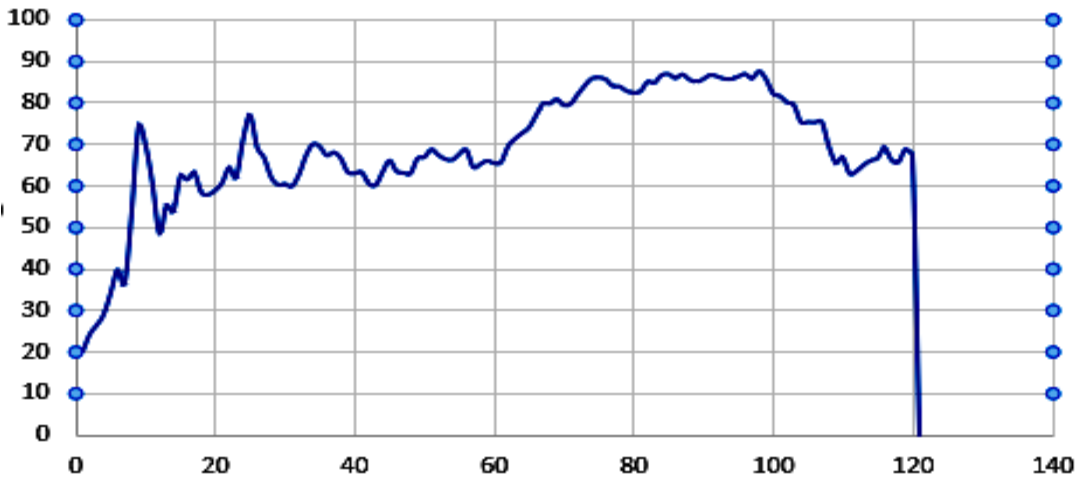
The variable is transformed into a time series under forecasting. The value which varies over time around a constant variance and mean is the stationary series. The approach is to test stationary through checking the time plot. Using appropriate differencing non stationary is corrected. To obtain the stationarity

difference of order one was enough. Then the values of p and q are identified. The autocorrelation and partial autocorrelation are calculated for different orders [Fig 2]. The order p and q can be one. Four ARIMA models were calculated and the one which has minimum Bayesian Information Criterion (BIC) was chosen.

**b) Model estimation and verification:**

SPSS package is used to estimate the model parameters. Results are given in table. The model examination is concerned with testing the residuals then observe whether they have any systematic pattern which can be removed to improve. This can be done by verifying the autocorrelation and partial autocorrelation of various orders. Many correlation of upto 16 lags are calculated and this significance is verified by Box-Ljung test. This shows us that chosen ARIMA model is appropriate model.

**Percentage of Facebook Users**



**Fig. 1:** Chart showing the percentage of Facebook users from April 2009 to March 2019.

**Table 1:** Model Description

FACE BOOK USERS	Model ID	Model Type
	Model 1	ARIMA(0,1,1)

**Table 2:** Model Fit

Fit Statistic	Mean	Min	Max	Percentile						
				5	10	25	50	75	90	95
Stationary R-squared	.043	.043	.043	.043	.043	.043	.043	.043	.043	.043
R-squared	.916	.916	.916	.916	.916	.916	.916	.916	.916	.916
RMSE	3.796	3.796	3.796	3.796	3.796	3.796	3.796	3.796	3.796	3.79
MAPE	3.931	3.931	3.931	3.931	3.931	3.931	3.931	3.931	3.931	3.93
MaxAPE	33.436	33.436	33.436	33.436	33.436	33.436	33.436	33.436	33.436	33.4
MAE	2.406	2.406	2.406	2.406	2.406	2.406	2.406	2.406	2.406	2.40
MaxAE	17.851	17.851	17.851	17.851	17.851	17.851	17.851	17.851	17.851	17.8
Normalized BIC	2.708	2.708	2.708	2.708	2.708	2.708	2.708	2.708	2.708	2.70

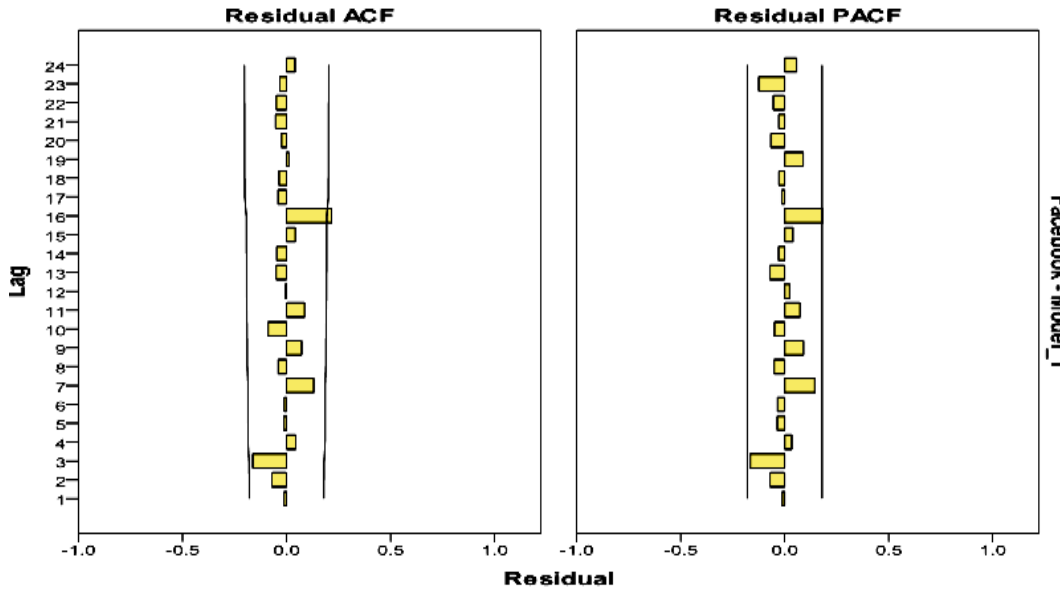


Fig. 2: Chart showing residual ACF and PACF for the model ARIMA (0,1,1).

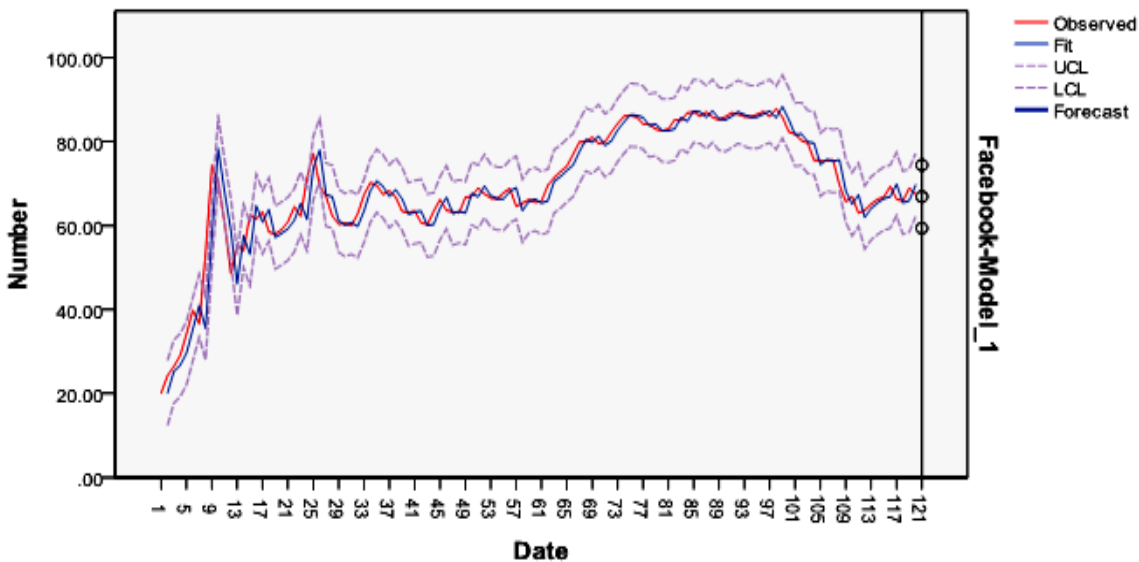


Fig. 3: Chart showing the actual and predicted values of the percentage of Facebook users.

Table 3: Case Processing Summary

		N	Percent
Sample	Training	67	69.1%
	Testing	30	30.9%
Valid		97	100.0%
Excluded		25	
Total		122	

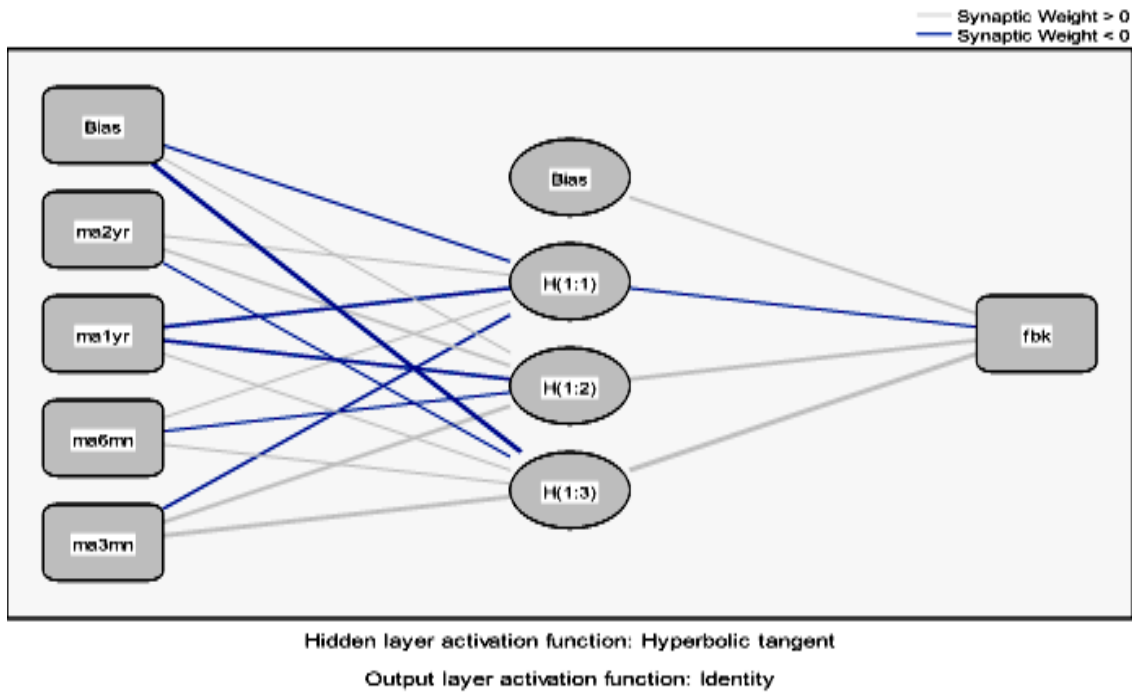
## RESULTS

In 2018, recent time’s researchers conducted a survey on social media users, in which Facebook and YouTube dominate the social media landscape. Facebook remains the primary pattern for most of the Americans. It has been reported that roughly two thirds of U.S adults use Facebook and among them three quarters of those use Facebook on a daily basis. A majority of Americans with exception of those 65 and older now use Facebook. Prediction is made about the outcome of a future event based on a pattern of evidence. Using ARIMA methodology, we have estimated the percentage of users using Facebook for the month of June. In this study, 122 data points were taken from April 2009 to April 2019 and have been analyzed with help of ARIMA (0,1,1)[Table1]. From the analysis carried out the Mean

Absolute Error (MAE) was found to be 2.406, Root Mean Square Error (RMSE) was found to be 3.796 and Mean Absolute Percentage Error (MAPE) as 3.931[Table 2].

**Table 4:** Network Information

<b>Input Layer</b>	<b>Covariates</b>	1	2yr MA
		2	1yr MA
		3	6mth MA
		4	3mth MA
<b>Hidden Layer(s)</b>	Number of Units <sup>a</sup>		4
	Rescaling Method for Covariates		Standardized
	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 <sup>a</sup>		3
<b>Output Layer</b>	Activation Function		Hyperbolic tangent
	Dependent Variables	1	Facebook
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
Error Function		Sum of Squares	



**Fig. 4:** Architecture of the ANN Model.

**Table 5:** Model Summary

<b>Training</b>	Sum of Squares Error	2.807
	Relative Error	.085
	Stopping Rule Used	1 consecutive step(s) with no decrease in error <sup>a</sup>
<b>Testing</b>	Training Time	00:00:00.029
	Sum of Squares Error	1.832
	Relative Error	.146

**Table 6:** Parameter Estimates

Since the MAPE, RMSE and MAE are very less, this model can be used for future prediction of Facebook users. The Artificial Neural Network model for training the data the sum of the squares of the error (SSE) was found to be 2.807[Table 3] and the relative error was 0.085 and for testing the data, the sum of the squares of the error was found to be 1.832 and the relative error was found to be 0.146[Table 6].The architecture of ANN consists of 5 input layers, hidden layers and output layer,the hidden layer activation function was hyperbolic function and output layer function was identity function.

Predictor		Predicted			
		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	Facebook
Input Layer	(Bias)	-.274	.139	-.638	
	ma2yr	.266	.280	-.145	
	ma1yr	-.427	-.440	.031	
	ma6mn	.050	-.042	.020	
	ma3mn	-.276	.609	.870	
Hidden Layer 1	(Bias)				.296
	H(1:1)				-.191
	H(1:2)				.752
	H(1:3)				1.070

**Table 7:** Prediction of Facebook users for the month of May, 2019

Model	ARIMA(0,1,1)	ANN	ACTUAL VALUE
Percentage of FACE BOOK Users	66.93	67.8	69.52

### CONCLUSIONS

Using Arima model, the percentage of Facebook users for the month of May 2019 was found to be 66.9 and using ANN model ,it was found to be 67.8[Table 7].This type of analysis enables us to understand how a time series model can predict and provide us with appropriate result to help us modify the technology suitably and approach the possible upcoming challenges and scenarios. It also helps us to give the future forecast of such time series.

#### CONFLICT OF INTEREST

There is no conflict of interest.

#### ACKNOWLEDGEMENTS

None.

#### FINANCIAL DISCLOSURE

None.

### REFERENCES

- [1] Peng C, Hongyong Y, Xueming S. [2008] Forecasting Crime using the ARIMA model, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 5: 627- 630.
- [2] H. Dong, L. Jia, X. Sun, C. Li and Y. Qin. [2009] Road Traffic Flow Prediction with a Time-Oriented ARIMA Model, Fifth International Joint Conference on INC, IMS and IDC, 2009:1649-1652.
- [3] Dongdong W. [2010] The Consumer Price Index Forecast Based On ARIMA Model, WASE International Conference on Information Engineering , Beidaihe, Hebei, 1: 307-310.
- [4] Xu Y. [2010] The Application of ARIMA Model in Chinese Mobile User Prediction, IEEE International Conference on Granular Computing, San Jose, CA, 2010:586-591.
- [5] Babu NC, Reddy EB. [2012] Predictive Data Mining On Average Global Temperature Using Variants of ARIMA models, IEEE- International Conference On Advances In Engineering, Science And Management, 259-263
- [6] Hossain Z, Abdus QS, Ali Z. [2006] ARIMA Model and Forecasting with Three Types Of Pulse Prices In Bangladesh: A Case Study, International Journal of Social Economics, 33(4):344-353.
- [7] Amin F, Razzaque MA, [2000] Autoregressive Integrated Moving Average (ARIMA) Modelling for Monthly Potato Prices in Bangladesh, Journal of Financial Management and Analysis, 13(1):74-80.
- [8] Priyadarshini E, Chandrababu A. [2011] Modeling And Forecasting Of Foreign Exchange Rates Using ARIMA Models, Proceedings of the National Conference on Recent Developments in Mathematics and its Applications at, SRM University, Chennai, Excel India Publishers:379 -383.
- [9] Priyadarshini E, Chandrababu A. [2011] Forecasting of Crude Oil Rates using ARIMA Models, International Journal of Statistics and Systems(IJSS), Research India Publications, 6(3):287-29.
- [10] Priyadarshini E. [2015] A Comparative Analysis Of Prediction Using Artificial Neural Network and Auto Regressive Integrated Moving, ARPJN Journal of Engineering and Applied Sciences, 10(7):3078-3081.