International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 6 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195617

Longitude and Latitude Prediction Using ARIMA

Vina Ayumi

Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia
vina.ayumi@mercubuana.ac.id

ABSTRACT

As an initial research in human mobility, human mobility prediction can be done by using a time-series predictor algorithm, one of which is ARIMA. ARIMA is short of the integrated moving-average autoregressive. The order of the ARIMA model is represented by the ARIMA symbol (p, d, q), where p is the order of the autoregressive part, d is the order of differencing and q is the order of moving-average process. The research regarding the application of human mobility conducted through five phases, including data collection, data pre-processing, data model building, data prediction and data evaluation. We conducted three times of experiments with different parameters. We defined different value for D, Seasonality, MALags, SMALags and Variance. Based on experiment as conclusion of this research obtained that the best parameter values to get the best MAPE Longitude and MAPE Latitude are Constant = 0, D= 1, Seasonality= 12, MALags = 1, SMALags=12 with MAPE Lon: 0.037433% and MAPE Lat: 0.11632%
Keywords: ARIMA, neural network, genetic algorithm, human mobility

I. INTRODUCTION

The role of information of technology in pattern recognition has been improved in many aspects, such as batik pattern[1], [2], human mobility pattern [3] and many more. Human mobility is a pattern of how a person moves in an urban area, for example, such as walking, driving to a place using a private vehicle or public transportation, etc. [4]. This research is important to understand patterns of human mobility, one of which is to control epidemics [5], [6]

As an initial research in human mobility, human mobility prediction can be done by using a time-series predictor algorithm, one of which is ARIMA.

ARIMA is short of the integrated moving-average autoregressive. The order of the ARIMA model is represented by the ARIMA symbol (p, d, q), where p is the order of the autoregressive part, d is the order of differencing and q is the order of moving-average process [7].

Research that applies the ARIMA model has been carried out by several researchers as a solution for various cases [7]–[15].

However, the use of the ARIMA model for solving cases in the field of human mobility is still little. A recent study in 2012, a time-series based prediction was proposed by Li et. al [16]. Li et. al investigating patterns of human mobility in urban taxi transportation systems.

Based on the background above, this study will use the ARIMA model to predict human mobility in Indonesia and determine the performance of the ARIMA model.

II. LITERATURE REVIEW

Studies that applied ARIMA model has been done by some researchers to solve problem in many study fields, for example [7]–[15]. However, the severe lack of research regarding application of ARIMA model for solving the case of human mobility can be opportunities for this research. The recent research about human mobility is conducted by [16] which investigated pattern of human mobility for transportation system.

III. METHODS AND MATERIAL

The research regarding the application of human mobility conducted through five phases which is shown in Figure 1.

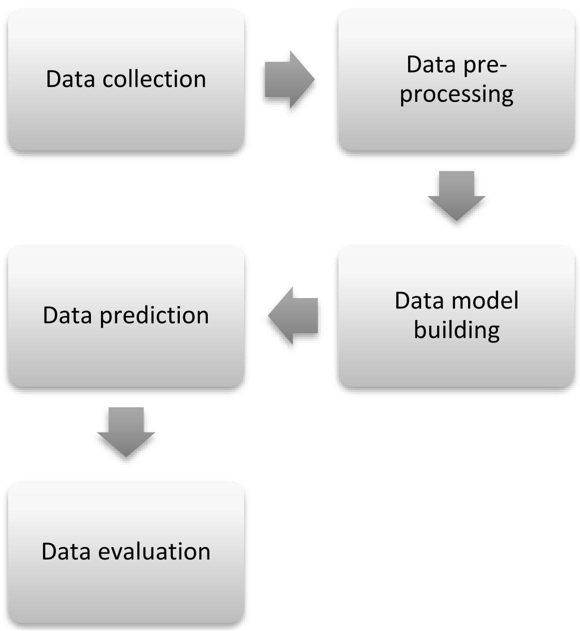

Fig. 1 Research phases

Based on Figure above, here are the details of each stage of the research:

1. The dataset used is the GeoLife GPS Trajectories dataset which can be obtained on the following page of Microsoft.com
2. After the data is obtained, then the preprocessing of the data includes cleaning to eliminate inconsistencies in the data, removing the sample data that crosses the boundary, and eliminating overlapping data. After further data cleaning is preprocessed by mapping human mobility data and identifying geospatial features such as building blocks, roads, or the environment. This mapping will be done using the R-tree [17].
3. The next stage is the construction of the model using training data using the ARIMA algorithm.
4. After the model is made, the model is then tested by predicting the test data. Predictions are made to estimate the next location that someone will visit in a city.
5. Furthermore, the performance evaluation of the predicted results will be carried out in the previous stage. The evaluation metric error used is symmetric Mean Absolute Percentage Error (sMAPE) [18]. sMape is an error metric that is commonly used to evaluate errors in the prediction of a prediction of ground truth data.

IV. RESULTS AND DISCUSSION

We conducted three times of experiments with different parameters. In the first experiment, we defined the value or degree of non-seasonal integration (in the linear time series model) named D with value 0. For the value or the difference degree of the seasonal between a *polynomial* in the *lag operator* for the linear time series named seasonality, we defined it with value 12. We also define the value of integer lags related to the coefficient of MA named

MALags with value 1. Moreover, we defined SMALags with value 1 for integer lags vector related to the coefficient of SMA. Then, we defined the model of conditional variance with garch(0,1). As the result, we found the mean absolute percentage error (MAPE) for latitude data is 0.040058% and or longitude data is 0.23519%. The result also depicted on Figure 1.

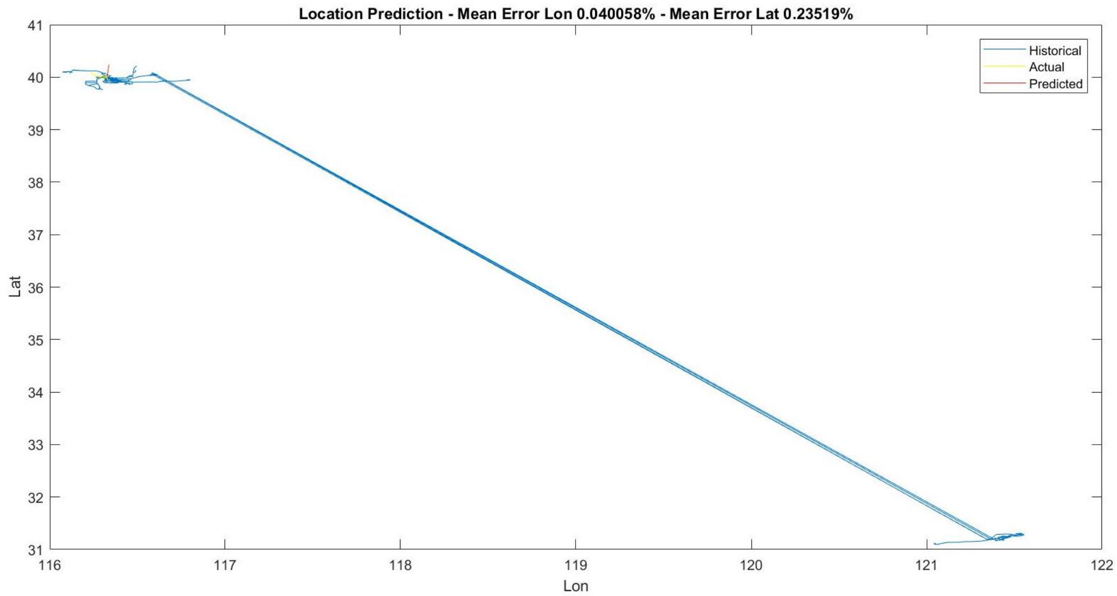

Fig. 2 The first experiment result for location prediction based historical, actual and predicted

In the second experiment, we set the value or degree of non-seasonal integration (in the linear time series model) named D with value 1. For the value or the difference degree of the seasonal between a *polynomial* in the *lag operator* for the linear time series named seasonality, we defined it with value 12. We also set the value of integer lags related to the coefficient of MA named MALags with value 1. Moreover, we defined SMALags with value 12 for integer lags vector related to the coefficient of SMA. Then, we defined the model of conditional variance with garch(0,1). As the result, we found the *mean absolute percentage error* (*MAPE*) for latitude data is 0.037433% and or longitude data is 0.11632%. The experiment result is shown in Figure 3.

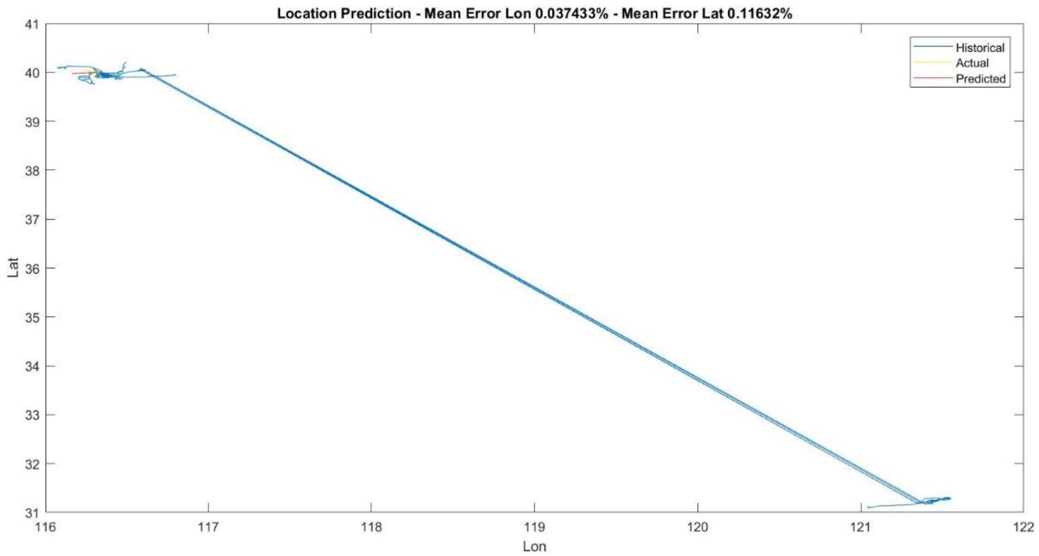

Fig. 3 The second experiment result for location prediction based historical, actual and predicted.

In the third experiment, we defined the value or degree of non-seasonal integration (in the linear time series model) named D with value 1. For the value or the difference degree of the seasonal between a *polynomial* in the *lag operator* for the linear time series named seasonality, we defined it with value 12. We also set the value of integer lags related to the coefficient of MA named MALags with value 1. Moreover, we defined SMALags with value 12 for integer lags vector related to the coefficient of SMA. Then, we defined the model of conditional variance with garch (0,1). As the result, we found the *mean absolute percentage error* (*MAPE*) for latitude data is 0.23371% and or longitude data is 0.039338%. The experiment result is shown in Figure 4.

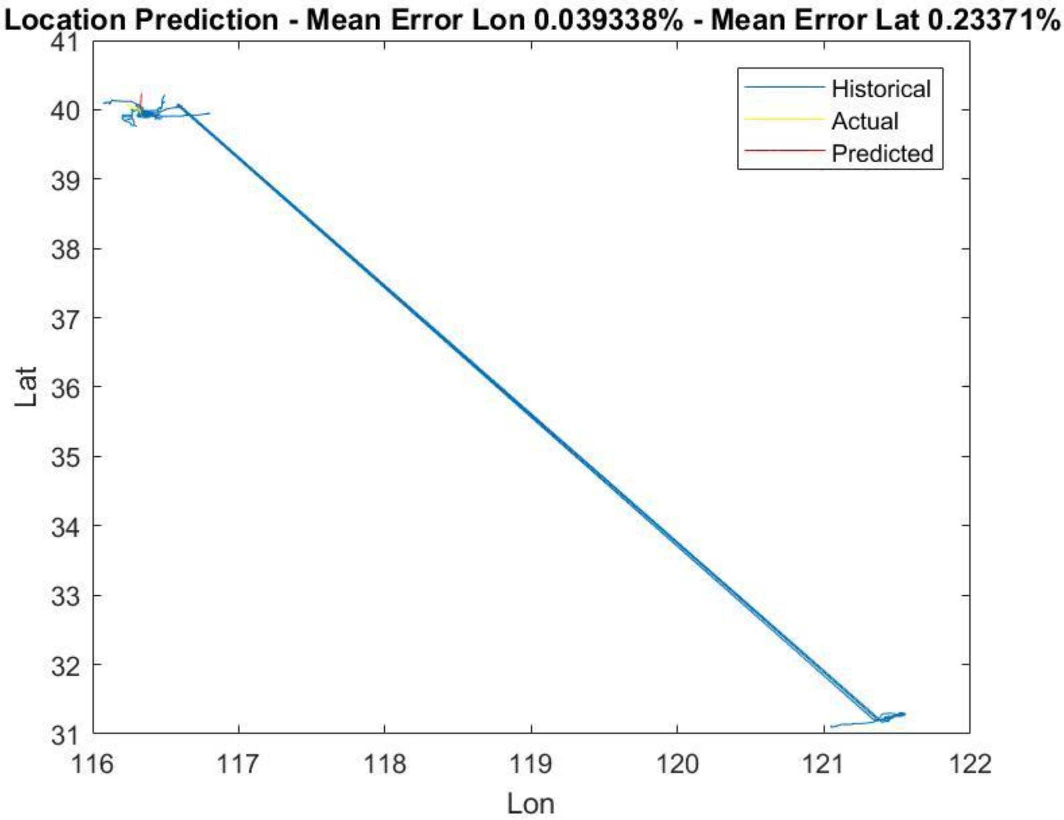

Fig. 4 The third experiment result for location prediction based historical, actual and predicted.

TABLE I. SUMMARY OF EXPERIMENT

Exp.	D	Seasonality	MALags	SMALags	Variance	MAPE Lon	MAPE Lat
1	1	12	1	1	garch(0,1)	0.040058%	0.23519%
2	1	12	1	12	garch(0,1)	0.037433%	0.11632%
3	1	12	1	1	garch(0,1)	0.039338%	0.23371%

V. CONCLUSION

Based on experiment as conclusion of this research obtained that the best parameter values to get the best MAPE Longitude and MAPE Latitude are Constant = 0, D= 1, Seasonality= 12, MALags = 1, SMALags=12 with MAPE Lon: 0.037433% and MAPE Lat: 0.11632%

VI. REFERENCES

[1]. I. Nurhaida, R. Manurung, and A. M. Arymurthy, “Extraction Methods for Batik Recognition,” in 2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2012, pp. 978-979.

[2]. I. Nurhaida, H. Wei, R. A. M. Zen, R. Manurung, and A. M. Arymurthy, “Texture Fusion for Batik Motif Retrieval System,” Int. J. Electr. Comput. Eng., vol. 6, no. 6, pp. 3174-3187, 2016.

[3]. B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, “Crowd sensing of traffic anomalies based on human mobility and social media,” Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - SIGSPATIAL’13, pp. 334-343, 2013.

[4]. K. Zhao, S. Tarkoma, S. Liu, and H. Vo, “Urban human mobility data mining: An overview,” 2016 IEEE Int. Conf. Big Data (Big Data), pp. 1911-1920, 2016.

[5]. W. Ni, Shunjiang and Weng, “Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks,” Phys. Rev. E, vol. 79, no. 1, p. 016111, 2009.

[6]. V. Belik, T. Geisel, and D. Brockmann, “Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases,” Phys. Rev. X, vol. 1, no. 1, pp. 1-5, 2011.

[7]. Y. Liang, “Combining seasonal time series ARIMA method and neural networks with genetic algorithms for predicting the production value of the mechanical industry in Taiwan,” Neural Comput Applic, vol. 18, no. 1, pp. 833-841, 2009.

[8]. C. Yuan, S. Liu, and Z. Fang, “Comparison of China ’ s primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model,” Energy, vol. 100, pp. 384-390, 2016.

[9]. K. Soni, S. Kapoor, K. Singh, and D. G. Kaskaoutis, “Statistical analysis of aerosols over the Gangetic - Himalayan region using ARIMA model based on long-term MODIS observations,” Atmos. Res., vol. 149, pp. 174-192, 2014.

[10]. S. Purnomo, S. Koshio, and V. Oktaferdian, “Implementation of ARIMA Model to Asses Seasonal Variability Macrobenthic Assemblages,” Aquat. Procedia, vol. 7, pp. 277-284, 2016.

[11]. E. Abounoori and S. Price, “Forecasting Stock Price Using Macroeconomic Variables : A Hybrid ARDL , ARIMA and Artificial Neural Network,” 2009.

[12]. M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, “Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks,” 2014.

[13]. M. Qin, Z. Li, and Z. Du, “Knowle dge-Base d Systems Red tide time series forecasting by combining ARIMA and deep belief network,” Knowledge-Based Syst., vol. 125, pp. 39-52, 2017.

[14]. P. Ramos, N. Santos, and R. Rebelo, “Robotics and Computer-Integrated Manufacturing Performance of state space and ARIMA models for consumer retail sales forecasting,” Robot. Comput. Integr. Manuf., vol. 34, pp. 151-163, 2015.

[15]. P. Sen, M. Roy, and P. Pal, “Application of ARIMA for forecasting energy consumption and GHG emission : A case study of an Indian pig iron manufacturing organization,” Energy, vol. 116, pp. 1031-1038, 2016.

[16]. X. Li et al., “Prediction of urban human mobility using large-scale taxi traces and its applications,” Front. Comput. Sci. China, vol. 6, no. 1, pp. 111-121, 2012.

[17]. A. Guttman, “R-trees,” Proc. 1984 ACM SIGMOD Int. Conf. Manag. data - SIGMOD ’84, p. 47, 1984.

[18]. S. Makridakis, “The M3-Competition : results , conclusions and implications,” vol. 16, pp. 451-476, 2000.

Cite this article as :

Vina Ayumi, "Longitude and Latitude Prediction Using ARIMA", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 6, pp. 57-61, November-December 2019. Available at doi : https://doi.org/10.32628/CSEIT195617
Journal URL : http://ijsrcseit.com/CSEIT195617