ISSN 2278-3091
Volume 8, No.3, May - June 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse59832019.pdf
https://doi.org/10.30534/ijatcse/2019/59832019

Crime Prediction using Autoregressive Integrated Moving Average (ARIMA) Algorithm

Elvis P. Patulin, Ronita E. Talingting
College of Arts and Sciences, Surigao State College of Technology, Surigao City, Philippines
elvispatulin@yahoo.com
Vice President for Academic Affairs, Surigao State College of Technology, Surigao City, Philippines
talingting2013@gmail.com

ABSTRACT

Data mining (DM), as coined with the term Knowledge Discovery in databases, used diverse approaches of DM analysis such as clustering, prediction or forecasting and many more. These are instrumental in extracting important information from the database. Along with DM analytics, big data and predictive analysis are relatively new concepts in criminology. This paper analyzes the index and non-index crimes in the province of Surigao del Norte, Philippines using the crime dataset from years 2013-2017. The prediction of its occurrence for the year 2018-2022 was also provided using ARIMA(1,0,7) model. Results showed that physical injury, homicide, violation of special laws, car napping, reckless imprudence resulting to physical injury, and other non-index crimes has 26%, 25%, 25%, 24%, 24%, and 23% forecasted increase for the year 2018-2022 with the highest occurrence in years 2018, 2018, 2020, 2018, 2020, and 2020 respectively.

Key words: ARIMA, Crime prediction, Data mining, Philippines

1. INTRODUCTION

Data Mining, as the method of analyzing different type of data to extract interesting patterns and knowledge [1]-[6], is also used to discover critical information which can help local authorities detect crime [4] as well as predicting areas which have high probability for occurrence of crime and indicate crime-prone areas [5] including the type of crimes.

Crime is a result of actions which do not conform with the society's norm and moral values [6], condemned by the society for the violation of any law. There is a continuous increase in the analysis of crime in the past decades hence, allowing many possibilities of studying and extracting crime data from different disciplines through different perspectives [7]. Police agencies around the world are anxious and eager to reduce crime [8].

The goal of this paper is to create clusters to identify as to what areas and municipalities within the province of Surigao del Norte established the highest recorded index and non-index crimes from 2013-2017. This is done by using the K-Means clustering algorithm. Data clustering is an unsupervised classification method used in creating groups of objects, or clusters, in such a way that objects in the same cluster are very similar, and objects in different clusters are quite distinct [9]-[12]. The police authority can identify and come up with strategic actions and preventive maintenance using the output of this study.

2. LITERATURE REVIEW

An increasing incidence of crime has led to the development and use of computer-aided diagnosis system, tools, and methods in analyzing, classifying, and predicting crimes. To name some, an approach based on auto-regressive models to reliably forecast crime trends in areas in Chicago was also performed. In particular, ARIMA as a model to forecast the number of crimes that are likely to occur in rolling time horizons was used to predict the number of crimes with an accuracy of 84% on the one-year-ahead forecast and of 80% on two-year-ahead forecasts [13].

Moreover, a method to execute a problem-oriented fitness function to predicted spatiotemporal patterns of criminal activity, using the memetic differential fuzzy approach which is a unique time series approach from fuzzy clustering, in the city of San Francisco, USA was presented in the paper of [7]. Results show that the series approach of fuzzy clustering for criminal patterns is a feasible method and is effective in producing a forecast of criminal patterns.

As stated by [14], predicting crimes can be categorized into two strategies. One approach is to use usual statistic techniques and models such as STL [15], ARIMA, kernel density distribution, and more in identifying hotspots of crime. Crime hotspots refer to geospatial locations where the probability for a crime to occur is high.

The paper of [4] analyzed crimes such as theft, homicide, and various drug offenses along with suspicious activities, noise complaints, and burglar alarm by using qualitative and quantitative approach. Using K-means clustering data mining approach on a crime dataset from the New South Wales region of Australia, crime rates of each type of crimes and cities with high crime rates have been found.

Meanwhile, precise crime prediction for small areas like police precincts is proposed in [16]. In [17], autoregressive integrated moving average (ARIMA) is employed for near future prediction of property crime. Based on 50 weeks of property crime data, an ARIMA model was built to predict crime number of 1 week ahead. It is found that the ARIMA model has higher fitting and prediction precision than exponential smoothing.

Furthermore, [18] implemented data mining techniques to understand certain trends and pattern of terrorist attacks in

India. K-means clustering was used to determine the year wherein the terrorist groups were most active and also which terrorist group has affected the most. The experimental result is implemented in Rapid miner tool to determine the active group and the affected year.

3. METHODOLOGY

3.1 Datasets

The data that were used are the following but are not limited as to wit: Index and non-index crime datasets of the city and municipalities from the province of Surigao del Norte, Philippines.

Table 1: Actual data of index crimes from 2013-2017 in the province of Surigao del Norte

Year	MURDER	HOMICIDE	PHYSICAL INJURY	RAPE	ROBBERY	THEFT	CAR NAPPING	CATTLE RUSTLING
2013	69	57	911	49	338	821	74	2
2014	53	15	329	62	355	871	107	4
2015	49	13	272	72	278	490	140	5
2016	62	10	198	56	89	170	29	0
2017	57	10	145	52	62	109	23	0

Table 2: Actual data of non-index crimes from 2014-2017 in the province of Surigao del Norte

Year	RECKLESS IMPRUDENCE RESULTING TO			VIOLATION OF SPECIAL LAWS	OTHER NON INDEX CRIMES
	HOMICIDE	PHYSICAL INJURY	DAMAGE TO PROPERTY		
2014	37	365	297	1024	739
2015	42	283	145	772	570
2016	48	262	165	672	385
2017	39	213	153	506	256

3.2 Forecasting Index Crimes

The ARIMA(1,0,7) model was used in this paper in determining the occurrence of index crimes for the next five years. Figure 1 to Figure 8 showed the graph of the predicted index crimes from 2018 to 2020 having 80% and 95% interval.

An autoregressive integrated moving average (ARIMA) model makes prediction of time series values based upon prior values (AR terms) as well as the errors made by previous predictions (MA terms). This allows the model to adjust itself to sudden changes in the time series. Therefore, the ARIMA forecasting equation for a stationary time series is a linear regression equation in which the predictors are the lags of the dependent variable and/or lags of the prediction errors. This model is explained in more detail in [18] [24]. In this paper, this method was implemented in R Studio using R language.

It is shown in Figure 1 that there is a decrease of recorded murder cases in the province of Surigao del Norte since 2016 to 2017. The highest predicted rate of murder is 57 in the years 2019 and 2022. The total predicted murder case for the year 2018-2022 is 279 which is 4% lesser from the 290 actual murder cases from the past five years.

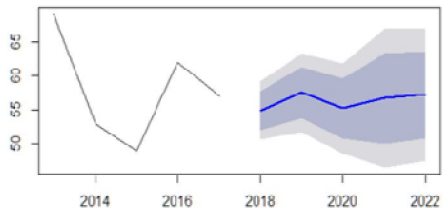

Figure 1: Forecasted murder from 2018-2022

The total predicted homicide from 2018-2022 in the province of Surigao del Norte is 140 which is 25% higher from the actual homicide data from the past five years. It can be seen in Figure 2 that there is a predicted increase of homicide from year 2017 to 2018 but a decreasing trend from 2020 to 2022.

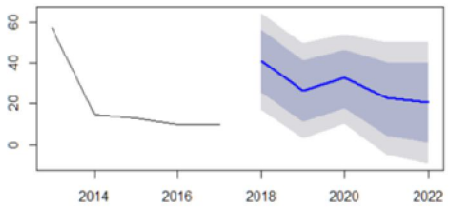

Figure 2: Forecasted homicide from 2018-2022

It is evident in Figure 3 that there is a rapid increase rate of physical injury from 2017 to 2018 with crime data of 608 which is also considered as the highest. A decrease is forecasted in year 2018 to 2019 and an increase a year after. The total forecasted crime data of physical injury from 2018 to 2022 in the province of Surigao del Norte is 2,508 which is 26% higher from the past five years.

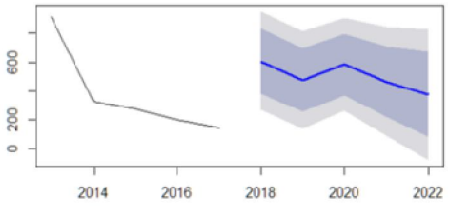

Figure 3: Forecasted physical injury from 2018-2022

The total forecasted rape from 2018-2022 in the province of Surigao del Norte is 306 which is 5% higher from the actual rape data of year 2013-2017. It is forecasted that there is a considerable increase of rape cases from 2017 to 2018 and a decrease years after. A slightly increase pattern from year 2020 to 2022 is also evident in Figure 4.

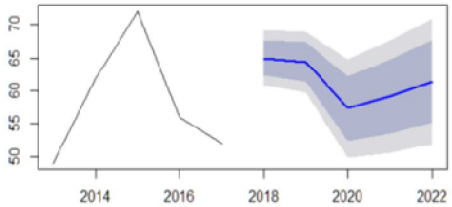

Figure 4: Forecasted rape from 2018-2022

It is presented in Figure 5 that there is a decreasing trend of robbery from year 2013 to year 2017. Meanwhile, the forecasted crime data of robbery shows a rapid increase from 62 actual data for the year 2017 to 146 forecasted data in year

2018. An increasing and decreasing pattern is shown from 2017 to 2020 and 2020 to 2022, respectively. The total forecasted crime data of robbery from 2018-2022 in the province of Surigao del Norte is 1,203 which is 7% higher from the past.

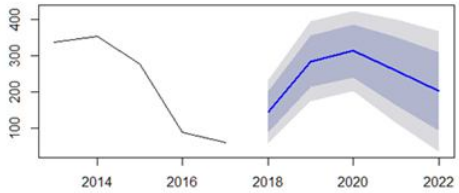

Figure 5: Forecasted robbery from 2018-2022

The total forecasted theft from 2018-2022 in the province of Surigao del Norte is 2,565 which is 4% higher from the actual theft data of 2,461 from 2013-2017. An increasing trend from 2017 to 2020 is evident in the Figure 6.

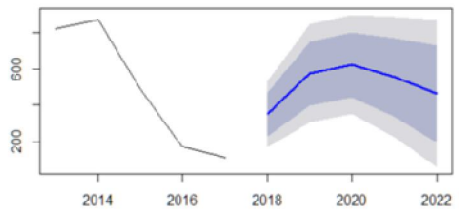

Figure 6: Forecasted theft from 2018-2022

It is shown in Figure 7 that there is a rapid increase from 23 recorded data of car napping in 2017 to forecasted data of 157 in the year 2018. A slightly decrease pattern is shown from year 2018 all the way to year 2020. Meanwhile, an increasing pattern from 2020 to 2022 is also visible. The total forecasted crime data of car napping from 2018-2022 in the province of Surigao del Norte is 492 which is 24% higher from the past five years.

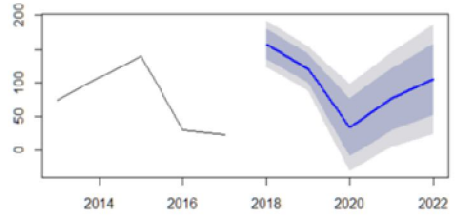

Figure 7: Forecasted car napping from 2018-2022

Among all the index crimes in the province of Surigao del Norte, the cattle rustling has the lowest number of crime reported from the year 2013-2017. The forecasted crime data of cattle rustling from 2018-2022 is 14 which has an estimated increase of 3 from the actual data of 11 from year 2013-2017. An increase from the year 2017 to 2018 and a decrease in the next two years is shown in Figure 8. Further, a slightly increase trend is evident in 2020-2022.

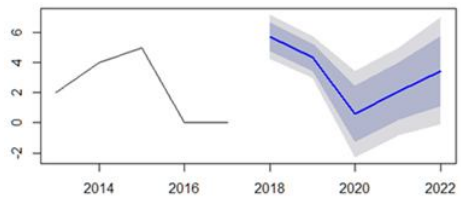

Figure 8: Forecasted cattle rustling from 2018-2022

3.3 Forecasting Non-index Crimes

ARIMA(1,0,6) model was used to forecast the occurrences of non-index crimes for the next five years. Figure 9 to Figure 13 showed the graph of the predicted non-index crimes from 2018 to 2020 having 80% and 95% interval.

It is shown in Figure 9 that there is an increase of recorded reckless imprudence resulting to homicide cases in the province of Surigao del Norte since 2014 to 2016 and a decreasing pattern thereafter. The highest predicted rate of reckless imprudence resulting to homicide is 45 in the year 2019. Further, a decreasing pattern is shown from year 2019 to 2020 and a steady trend right after. The total predicted case for the year 2018-2022 is 209 which is 20% higher from the 166 actual cases from the past five years.

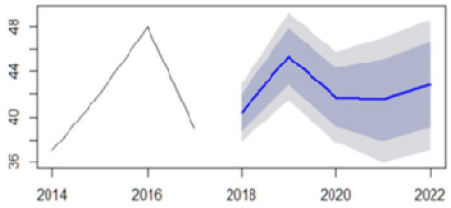

Figure 9: Forecasted reckless imprudence resulting to homicide from 2018-2022

The total predicted data for reckless imprudence resulting to physical injury from 2018-2022 in the province of Surigao del Norte is 1,482 which is 24% higher from the 1,123 actual data from the past five years. It can be seen in Figure 10 that there is a predicted increase from year 2017 to 2020 but a decreasing trend from 2020 to 2022.

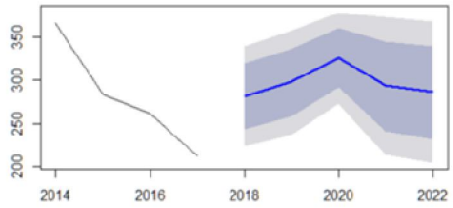

Figure 10: Forecasted reckless imprudence resulting to physical injury from 2018-2022.

It is shown in Figure 11 that there is almost a steady trend from 153 recorded data of reckless imprudence resulting to damage to property in 2017 to forecasted data of 210 in the year 2018. A decrease and increase pattern is shown from year 2018 to 2019 and 2019-2020 respectively and then repeat all the way to year 2022. The total forecasted crime data of reckless imprudence resulting to damage to property from 2018-2022 in the province of Surigao del Norte is 958 which is 20% higher from the 760 actual data from 2014-2017.

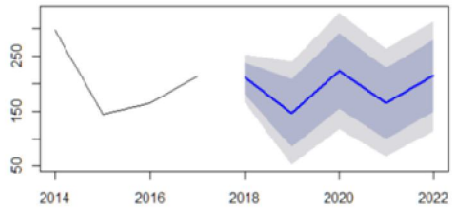

Figure 11: Forecasted reckless imprudence resulting to damage to property from 2018-2022.

The predicted crime data for the violation of special laws from 2018-2022 in the province of Surigao del Norte is considerably high from the year 2017 to 2020 with forecasted 893 crime records. Meanwhile, there is a decreasing pattern from 2020 to 2022 as shown in Figure 12. The total forecasted crime data of violation of special laws from 2018-2022 in the province of Surigao del Norte is 3,959 which is 25% higher from the 2,974 actual data from 2014-2017.

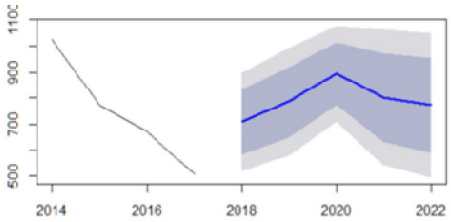

Figure 12: Forecasted violation of special laws from 2018-2022

It is presented in Figure 13 that there is a decreasing trend of non-index crimes from year 2014 to year 2017. Meanwhile, the forecasted crime data shows a rapid increase from 256 actual data for the year 2017 to 344 forecasted data in year 2018. An increasing and decreasing pattern is shown from 2017 to 2020 and 2020 to 2022, respectively. The total forecasted crime data of other non-index crimes from 2018-2022 in the province of Surigao del Norte is 2,530 which is 23% higher from the 1,950 actual data from 2014-2017.

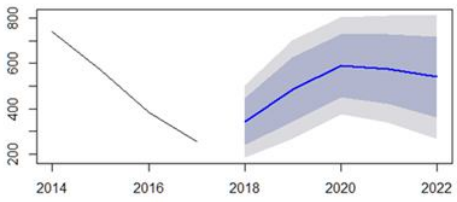

Figure 13: Forecasted other non-index crimes from 2018-2022

4. CONCLUSION

Among the index crimes in the province of Surigao del Norte, theft was identified as the highest number of recorded crime with a total of 2,565 from 2013-2017 with the highest occurrence in 2014. Furthermore, the highest predicted crime for the year 2018-2022 is the physical injury having the predicted value of 2,508 or 26% increase from 2014-2017. Moreover, the least reported crime in the province is cattle rustling.

For the non-index crimes, violation of special laws was identified as the highest reported incident in the province with the highest occurrence in 2014. Moreover, violation of special laws has the highest predicted value of 3,959 or 25% increase from the data of year 2014-2017 with the highest occurrence in 2020.

REFERENCES

[1] K. Rajalakshmi, S. S. Dhenakaran, and N. Roobini, "Comparative Analysis of K-Means Algorithm in Disease Prediction," *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 7, pp. 2697–2699, 2015.

[2] A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable Reduction-based Prediction through Modified Genetic Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 356–363, 2019. https://doi.org/10.14569/IJACSA.2019.0100544

[3] A. J. P. Delima, "An Experimental Comparison of Hybrid Modified Genetic Algorithm-based Prediction Models," *Int. J. Recent Technol. Eng.*, vol. 8, no. 1, pp. 1756–1760, 2019.

[4] J. Agarwal, "Crime Analysis using K-Means Clustering," *Int. J. Comput. Appl.* , vol. 83, no. 4, pp. 975–8887, 2013. https://doi.org/10.5120/14433-2579

[5] O. Vaidya, S. Mitra, R. Kumbhar, S. Chavan, and R. Patil, "Comprehensive Comparative Analysis of Methods for Crime," *Int. Res. J. Eng. Technol.*, pp. 715–718, 2018.

[6] M. Sevri, H. Karacan, and M. A. Akcayol, "Crime Analysis Based on Association Rules Using Apriori Algorithm," vol. 7, no. 3, 2017. https://doi.org/10.18178/IJIEE.2017.7.3.669

[7] C. D. R. Rodriguez, D. M. Gomez, and M. A. M. Rey, "Forecasting time series from clustering by a memetic differential fuzzy approach: An application to crime prediction," *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc.*, vol. 2018-Janua, pp. 1–8, 2018. https://doi.org/10.1109/SSCI.2017.8285373

[8] G. Dudfield, C. Angel, L. W. Sherman, and S. Torrence, "The 'Power Curve' of Victim Harm: Targeting the Distribution of Crime Harm Index Values Across All Victims and Repeat Victims over 1 Year," *Cambridge J. Evidence-Based Polic.*, vol. 1, no. 1, pp. 38–58, 2017. https://doi.org/10.1007/s41887-017-0001-3

[9] D. Kaur and K. Jyoti, "Enhancement in the Performance of K-means Algorithm," vol. 2, no. 1, pp. 29–32, 2013.

[10] A. Bansal, M. Sharma, and S. Goel, "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining," *Int. J. Comput. Appl.*, vol. 157, no. 6, pp. 975–8887, 2017. https://doi.org/10.5120/ijca2017912719

[11] J. Goyal and B. Kishan, "Progress on Machine Learning Techniques for Software Fault Prediction," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 2, pp. 305–311, 2019. https://doi.org/10.30534/ijatcse/2019/33822019

[12] S. B. B, K. K. V, and A. N. Ahmed, "Semantically enriched Tag clustering and image feature based image retrieval system," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1, pp. 138–141, 2019.

[13] E. Cesario, C. Catlett, and D. Talia, "Forecasting Crimes Using Autoregressive Models," *Proc. - 2016 IEEE 14th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2016, 2016 IEEE 14th Int. Conf. Pervasive Intell. Comput. PICom 2016, 2016 IEEE 2nd Int. Conf. Big Data*, pp. 795–802, 2016. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.138

[14] J. Azeez and D. J. Aravindhar, "Hybrid approach to crime prediction using deep learning," *2015 Int. Conf. Adv. Comput. Commun. Informatics*, pp. 1701–1710, 2015. https://doi.org/10.1109/ICACCI.2015.7275858

[15] A. Malik, R. Maciejewski, S. Towers, S. Mccullough, and D. S. Ebert, "3B - Proactive Spatiotemporal Resource Allocation and Predictive Visual Analytics for Community Policing and Law Enforcement," vol. 20, no. 1, pp. 1863–1872, 2014. https://doi.org/10.1109/TVCG.2014.2346926

[16] W. Gorr, A. Olligschlaeger, and Y. Thompson, "Short-term forecasting of crime," *Int. J. Forecast.*, vol. 19, no. 4, pp. 579–594, 2003. https://doi.org/10.1016/S0169-2070(03)00092-X

[17] P. Chen, H. Yuan, and X. Shu, "Forecasting crime using the ARIMA model," *Proc. - 5th Int. Conf. Fuzzy Syst. Knowl. Discov. FSKD 2008*, vol. 5, pp. 627–630, 2008. https://doi.org/10.1109/FSKD.2008.222

[18] P. Gupta, A. S. Sabitha, and T. Choudhury, "Terrorist Attacks Analysis Using Clustering Algorithm," © *Springer Nat. Singapore Pte Ltd.*, pp. 317–328, 2018. https://doi.org/10.1007/978-981-10-5547-8_33

[19] D. Hand, D. Hand, H. Mannila, H. Mannila, P. Smyth, and P. Smyth, *Principles of data mining*, vol. 30. 2001.

[20] A. Thammano and A. K. Algorithm, "Enhancing K-means Algorithm for Solving Classification Problems," pp. 1652–1656, 2013. https://doi.org/10.1109/ICMA.2013.6618163

[21] L. Feltrin, "KNIME an Open Source Solution for Predictive Analytics in the Geosciences [Software and Data Sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 4, 2015. https://doi.org/10.1109/MGRS.2015.2496160