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/ijatcse52832019.pdf
https://doi.org/10.30534/ijatcse/2019/52832019

Crime Trend Analysis Using Data Mining Technique

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

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. These paper clusters municipalities in Surigao del Norte using K-Means algorithm. This is instrumental in finding identical traits, patterns, and values in categorizing municipalities with much, more, and most number of recorded index and non-index crimes from 2013-2017. Simulation results showed that the Surigao City topped as the municipality in Surigao del Norte, Philippines with the most number of reported index and non-index crimes which belongs to Cluster 1 as depicted by the algorithm. It is suggested that the output of this study may be used as input for new studies about crime identification. Future researchers may also utilize other data mining techniques supported by a better accuracy result.

Keywords: Crime forecasting, Clustering, Data mining, K-Means algorithm

1. INTRODUCTION

Police agencies around the world are anxious and eager to reduce crime [1]. Crime is a result of actions which do not conform with the society's norm and moral values [2], 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 [3].

Data Mining, as the method of analyzing different type of data to extract interesting patterns and knowledge [4]-[6], is also used to discover critical information which can help local authorities detect crime [7] as well as predicting areas which have high probability for occurrence of crime and indicate crime-prone areas [8] including the type of crimes.

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

Data mining, as coined with the term Knowledge Discovery in databases (KDD), used diverse approaches of DM analysis such as decision tree, Bayes classifiers, association rules, clustering, neural networks, genetic algorithms, support vector machines, predicting or forecasting and many more. These are instrumental in extracting important information from the database [13]. Along with DM analytics, big data and predictive analysis are relatively new concepts in criminology, while they have become standard practice in disciplines such as business intelligence, biomedical sciences, finance, and marketing. The use of such techniques will also inadvertently impact social sciences and humanities in general [14], and criminology in particular [15] and is proved to be very useful as it is employed to perform in crime forecasting [16].

The paper of [7] 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.

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 [3]. 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.

Furthermore, [17] 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.

As stated by [18], based on the data source, predicting crimes can be categorized into two strategies. One approach is to use usual statistic techniques and models such as STL [19], 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.

Meanwhile, precise crime prediction for small areas like police precincts is proposed in [20]. In [21], 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.

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 [22].

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 Clustering

Clustering algorithms divide the group of objects into clusters, where objects in each cluster are similar to each other [23]. Index and non-index crime data from the province of Surigao del Norte are shown in Table 3.

Table 3: Index and Non-Index crime dataset per city and municipality in the province of Surigao del Norte

MUNICIPALITY	INDEX CRIME	NON-INDEX CRIME
ALEGRIA	151	313
BACUAG	148	297
BURGOS	27	61
CLAVER	366	617
DAPA	222	437
DEL CARMEN	151	216
GENERAL LUNA	161	294
GIGAQUIT	183	264
MAINIT	194	303
MALIMONO	76	149
PILAR	98	138
PLACER	316	801
SAN BENITO	65	121
SAN FRANCISCO	103	298
SAN ISIDRO	88	111
SISON	120	308
STA. MONICA	85	185
SOCORRO	88	124
SURIGAO CITY	3562	3705
TAGANAAN	165	367
TUBOD	139	329

In K-means algorithm, the user specifies the k centroids. This K centroid refers to the number of the wanted clusters. Each cluster must have a centroid that is a mean of a cluster. Then the nearest centroid is assigned to each data record. When all input data records have been assigned, the centroid changes in each cluster and is updated by calculating the mean cluster. These processes will be repeated until the latest centroids do not change [24].

The experimental result for clustering was implemented using KNIME (Konstanz Information Miner) [25] analytics platform. Figure 1 shows the node structure of the K-means clustering executed in KNIME. The node for the K-Means is connected and then positioned after the node of the imported CSV file of the dataset. The node color manager comes after as it put distinctions to the results to be generated later. The node scatter plot shows the scatter plot of the clusters while the interactive table is used to view the result in a table manner

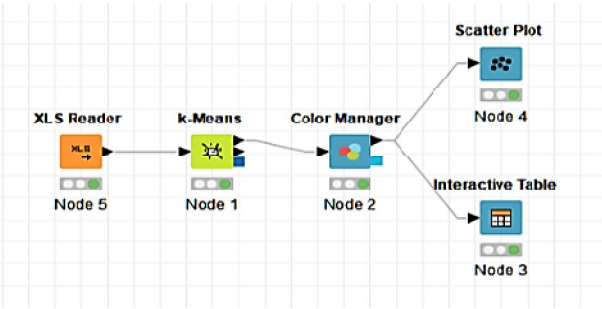

Figure 1: Node structure of K-means clustering in KNIME

Surigao City, as the only place that belongs to Cluster 1, denotes that it does not possess any similar traits with the other places. It is evident in Table 4 that having the highest value makes the algorithm categorized it differently from the others. Municipalities under Cluster 2 is then interpreted as next to Surigao City to have the highest value of recorded crimes. Meanwhile, municipalities that are under Cluster 3 has the least among the groups.

Table 4: Clustering result in KNIME

MUNICIPALITY	INDEX CRIME	NON-INDEX CRIME	CLUSTER
ALEGRIA	151	313	2
BACUAG	148	297	2
BURGOS	27	61	3
CLAVER	366	617	2
DAPA	222	437	2
DEL CARMEN	151	216	3
GENERAL LUNA	161	294	2
GIGAQUIT	183	264	2
MAINIT	194	303	2
MALIMONO	76	149	3

PILAR	98	138	3
PLACER	316	801	2
SAN BENITO	65	121	3
SAN FRANCISCO	103	298	2
SAN ISIDRO	88	111	3
SISON	120	308	2
STA. MONICA	85	185	3
SOCORRO	88	124	3
SURIGAO CITY	3562	3705	1
TAGANAAN	165	367	2
TUBOD	139	329	2

Figure 2 shows that among the city and all municipalities in Surigao del Norte, Surigao City has the highest crime rate with a total of 7,267 recorded index and non-index crimes from the year 2013-2017.

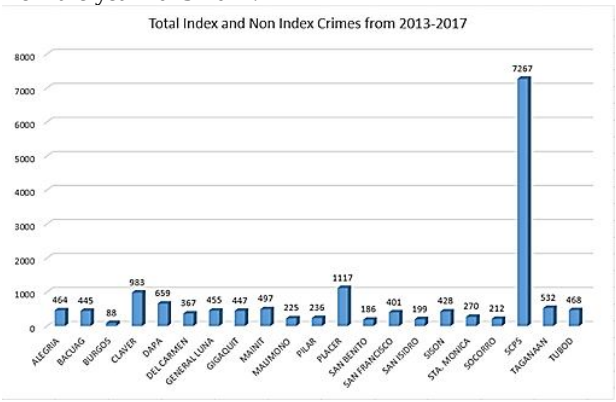

Figure 2: Crime rates in Surigao del Norte

In Figure 3, violation of special laws is the top recorded crime in Surigao City. It is followed by other non-index crimes, theft, physical injury, and robbery. The least recorded crimes in the city are the cattle rustling, homicide, and reckless imprudence resulting in homicide.

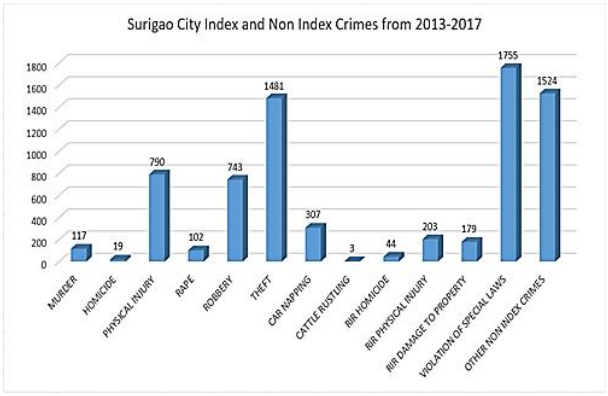

Figure 3: Index and non-index crime rates in Surigao City

Municipalities that belong to Cluster 2 is shown in Figure 4. The municipality of Placer has the highest recorded incident of the index and non-index crimes next to Surigao City of Cluster 1. It is followed by the municipality of Claver and Dapa making them as the second, third, and fourth most number of recorded crime incidents from the whole province of Surigao del Norte from the year 2013-2017.

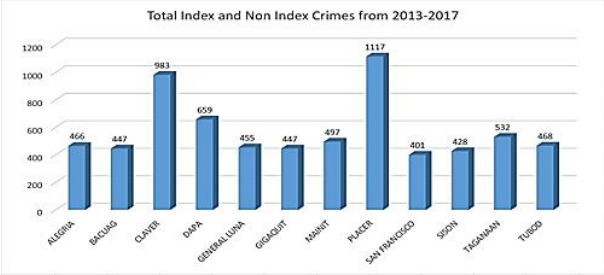

Figure 4: Crime rates of each municipality from Cluster 2

In Cluster 3, the municipality of Del Carmen has the most number of recorded index and non-index crimes from the year 2013-2017 followed by Sta. Monica and Pilar. The municipality of Burgos has the lowest number of recorded index and non-index crimes for the past five years, as shown in Figure 5.

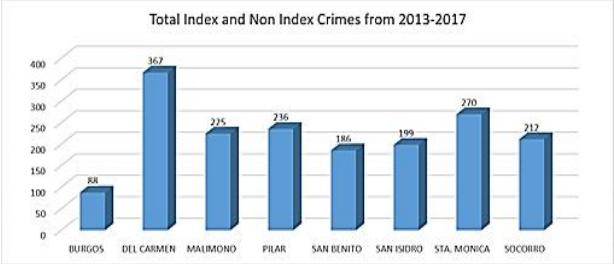

Figure 5: Crime rates of each municipality from Cluster 3

4. CONCLUSION

With the use of the K-Means clustering algorithm, determining the groupings of the municipality with identical traits and values became possible. In cluster 1, the Surigao City topped as the municipality in Surigao del Norte with the most number of reported index and non-index crimes. In cluster 2, the municipality of Placer, Claver, and Dapa has the highest crime rate. Meanwhile, in cluster 3, Del Carmen, Sta. Monica and Pilar were identified.

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 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.

It is recommended that since groupings were already identified using the K-Means algorithm, another knowledge extraction may be conducted using the same dataset as future research.

REFERENCES

[1] 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

[2] 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

[3] 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

[4] 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.

[5] 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

[6] 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.

[7] 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

[8] 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.

[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. Susnea, "Using data mining techniques in higher education," in *The 4th International COnference on Virtual Learning ICVL 2009*, 2009, vol. 1, no. 1, pp. 371–375.

[14] R. Kitchin, "Big Data, new epistemologies and paradigm shifts," *Big Data Soc.*, vol. 1, no. 1, pp. 1–12, 2014. https://doi.org/10.1177/2053951714528481

[15] J. Chan and L. Bennett Moses, "Is Big Data challenging criminology?," *Theor. Criminol.*, vol. 20, no. 1, pp. 21–39, 2016. https://doi.org/10.1177/1362480615586614

[16] C. Yu, M. W. Ward, M. Morabito, and W. Ding, "Crime Forecasting Using Data Mining Techniques," 2011. https://doi.org/10.1109/ICDMW.2011.56

[17] 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

[18] 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.

[19] 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

[20] 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

[21] 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

[22] 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

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

[24] 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

[25] 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