ISSN 2278-3091
Volume 9, No.1.3, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse2291.32020.pdf
https://doi.org/10.30534/ijatcse/2020/2291.32020

Implementation of Predictive Crime Analytics in Municipal Crime Management System in Calauan, Laguna, Philippines

Jonardo R. Asor[1], Sherwin B. Sapin[2]
[1]Laguna State Polytechnic University, Philippines, asor.jonardo@lspu.edu.ph
[2]Laguna State Polytechnic University, Philippines, sbsapin@lspu.edu.ph

ABSTRACT

Predictive crime analytics is a process of evaluating a dataset to discover veiled patterns that can be useful in forecasting crime occurrence. The application of a machine learning algorithm is widely used in developing artificial intelligence to integrate into computer systems. This project aims to develop a management system for crime records implementing predictive crime analytics in forecasting crime occurrences. Through predictive crime analytics and software development life cycle, a management information system that is capable of forecasting crime was successfully developed. A predictive model was designed using multinomial logistic regression, which acquires a totality accuracy of 86.60% and prediction confidence. It is also shown in this paper that regression is better than other classification algorithms in terms of predicting crime occurrence. Moreover, discussed in this research, among all the barangay in Calauan, Laguna. Dayap is the most vulnerable in different index crimes like rape, murder, robbery, theft, homicide, and illegal drugs.

Key words: Machine learning, multinomial logistic regression, web-based police record system, visual map

1. INTRODUCTION

Crime is an illegal activity that is punishable by the law, and it signifies the quality of government progress in the country. It serves to be objective to improve moral awareness over time. Crimes are not uniformly distributed in an area; it may happen unpredictably, and in fact, some geographic areas have less crime than others [1]. The widely spread of criminality to society may cause various problems to economic growth. The crime defined as the casualties included in an event; it classify into two groups, which is the Index and Non-index crime.

Index crime known as the major case in an area, the USA's FBI crime is compose of seven index crimes: murder and non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny, and motor vehicle theft [2], while non-index crime known as a petty or minor crime. Non-Index Crime includes careless unlawful death, non-aggravated assault, forgery & counterfeiting, fraud, embezzlement, stolen property, vandalism, weapons, prostitution and ordinary vice, sex offense, narcotics, gambling, offense against family and children, driving under the influence of liquor, unruly behavior, and other crimes not indicated in the index crimes. Offenses are the unlawful acts reported to a law enforcement agency. Arrests are those individuals seized, held, summoned, or cited by law enforcement agencies for taking part in an illegal act [3].

In the Philippines, Index Crime is the most leading delinquency records as per the Philippines National Police (PNP); this is because of unplanned urbanization, migration, high rate of unemployment, and poverty. As for the strategies of the PNP, they are taking immediate action if ever there is a concerned citizen that is reporting and accusing someone that doing an illegal activity or suspicious for something and also patrol some places where there is a potential crime in an area. According to the survey, 58% or almost 500 thousand in 1 million respondents believe that local law enforcement agencies are not capable of combating economic crimes [4]. The lack of preparedness and awareness of public safety is one of the factors in the Government issue. Despite these increasing rates, crime prevention is one of the focus on this project. It may be eliminating the criminality in its percentage for a particular area, to the risk before it happens by the help of the data mining technology.

Data mining is identified to determine veiled patterns of a precise dataset that may be helpful for improving a number of events such as product sales, strategic planning, crime, etc. [13] [14] [15]. These methods paved the thought of the authors to come up with the factors that can affect data mining to the criminality in the Philippines [10][11]. Data mining techniques are used to extract important information from the data. These techniques are classified into two methods, and the first one is supervised learning; this method the dataset divided into two, namely parts training datasets and Test datasets. The second one will be the unsupervised learning. In this method dataset are not focus on pre-determined attributes, nor does it predict a target value. Instead, unverified data mining finds hidden structure and relative between data. [12]

Project intended to develop a system that can contribute to lower the increasing crime rate in Calauan, Laguna, through Predictive analytics. Primarily the project must satisfy the following goals: a.) Develop a web-based police record

system with GIS and forecasting systems through time series. b.) Provide a time series model using machine learning algorithm. c.) Provide a probabilistic report of possible crime in the premises of Calauan, Laguna.

2. RELATED WORKS

Crimes are ordinary social problem touching the quality of life and the economic development of a society. It is considered as a necessary factor to determine whether people move to a new city and what places should be avoided when they travel [5]. With these crimes, law enforcement agencies are enduring to insist to acquire advanced geographic information systems and innovative data mining approaches to enhance crime analytics and improved protecting their communities.

One of the significant causes of insecurity and fear is crime and violence, particularly in urban areas [6]. Most of the place committed high percent of crime are happening in Cities, Townhall, and in a popular locality that causes different people offense in our law. A main effect of the rising poverty and inequality found in urban areas has been the rising level of crime, especially violent crime [7].

In the Philippines, many factors that contribute to higher the crime rate in the Philippines urban centers, these are Rapid Urbanization, Industrialization, and migration to the cities. These factors are the effect of unplanned urbanization, inequalities in resources. So that there are numbers of crimes directly associated to urbanization that pose serious concern for the government and civil society, leading of these are street crimes, illegal drug trafficking, robbery and theft, aggressive crimes against women and children, and terrorism [8].

With analyses of crime data, the project can find meaningful patterns in and extract valuable knowledge from such records to give a technical basis for developing efficient crime reduction and prevention strategy [16]. Spatial decision making of offenders and their social standing on how they interact with others, as of interest. Predictive policing pertains to any policing strategy or approach that develops and uses information and advance analysis to inform forward-thinking crime prevention-predicting offenders. The rationale is predicting future offenders using the history of individuals, such as features of their living environment and behavioral patterns. Predictive crime analytics is the analysis that detects crimes; however, crimes could occur everywhere; commonly, criminals work on crime opportunities they face in most common areas for them [9]. Predictive Crime Analytics for Policy Planning Predictive policing draws how data can be utilized to influence a favored result. The challenge with the police using this information stems from the blend of a high rate of crime in areas that are normally poverty-stricken, uneducated, and have above ethnic populations [10].

Logistic regressions fall under the classification of supervised learning; it identifies the correlation between the specific dependent variable and one or more independent variable by estimating probabilities using a sigmoid function. Despite the logistic regression, this is not utilized for the regression problem where the task is to predict the real-valued output. It is a classification problem that is used to predict a binary outcome (1/0, -1/1, True/False) given a set of independent variables. Logistic regression is like the linear regression, or serves it as a generalized linear model. In linear regression, predict a real-valued output 'y' based on a weighted sum of input variables. [11]

Logistic regression is utilized to get the chances relative amount in the occurrence of more than one explanatory variable. The process is quite like multiple linear regressions, with the exception that the response variable is binomial. The outcome is the impact of each variable on the odds ratio of the observed event of interest. The major benefit is to keep away from confusing effects by analyzing the association of all variables together. In this article, we explain the logistic regression procedure using examples to make it as easy as possible. After the description of the technique, the necessary analysis of the results is presented, and then some unusual issues are mentioned [12].

3. METHODOLOGY

In this project, two different methods were applied; first is the model development–the method was used to develop an artificial intelligence model that can be integrated into the management system. Typical, software development life cycle was also followed in this project to observed the reliability of developing the system.

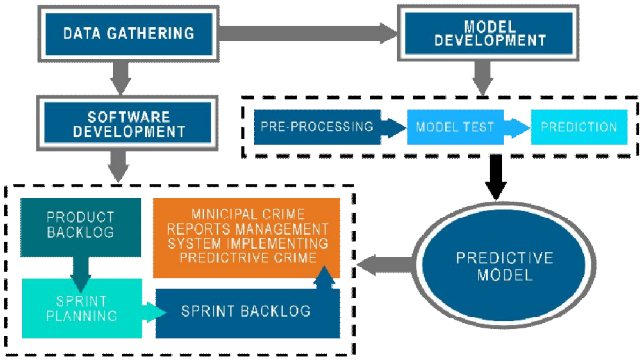


Figure 1: Project Methodology

Collaboration with the Philippines National Police (PNP) in Calauan, Laguna was observed in this research project to have an actual process and also crime visualization. The interview was done to have a clear pattern on what are the modules that can be included in the system.

2.3 Model Development

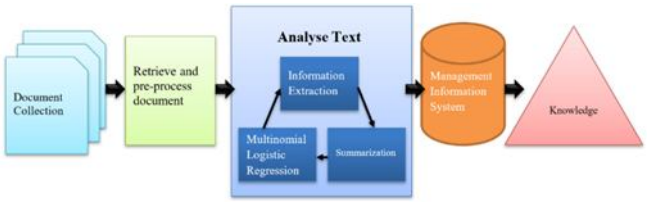


Figure 2: Text Mining Process

Text mining was integrated into the system for continuous training of the model. Likewise, the process is also done in this project to develop a model that is implemented in the project.

Data/Document Collection & Preprocessing

Data on crime reports were collected with the span of three (3) years (2016-2018), from the database of PNP–Calauan, which has the classes of robbery, rape, car accident, thief, which is the main focus of the forecasting. Also, it includes the location (barangay) to find the area and provide a tag label.

Table 1: Attributes of Crime in Calauan, Laguna dataset

Attributes	Description
Date Committed	Exact Date of Crime Events
Time Committed	Exact Time
Location	The Barangay where the Crime happens
Crime types	Type of Crime

The dataset was cleaned by removing all the noises, especially information that may provide error and other problems such as the unique character, symbols, and redundant words. This process is integrated into the system for the future prediction that is continuously learning.

Analyze Text

Information extraction is the first step of training then undergone summarization for clustering the data into two category parts, Crime Prediction and Prediction for the position. Logistic regression is well-known to illustrate data and to make clear the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. It is the machine learning algorithm used in this project.

Management Information System

For the last part of the Model Development, the system shows the aggregate function, including the visual map and the predicting method for the crime.

The agile method was followed in the development for improving the system functions and provides proper documentation. This is to assure the delivery of the project in minimal expansion on time without compromising its reliability.

Three modules are provided in the system, which consists of the Classifying module, Visual Map, and Information Management module.

In the classifying module, the data has been undergone model development wherein the inputted data will undergo preprocessing, clustering, removal, and transformation of the raw data. As for the model testing, the researchers used the classification algorithm, namely Multinomial Logistic Regression. For showing a graphical representation of the predicted location, Visual Map will provide coordinates for each site (barangay) and directly undergone in Information Management wherein the results of predicted information have been shown.

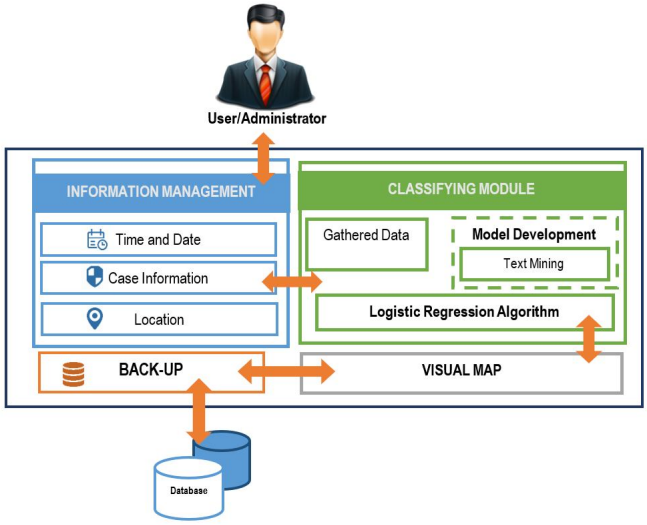


Figure 3: System Architecture of the Crime Report System

3. RESULTS AND DISCUSSION

At this point, all the results of different experiments that the researchers were conducted, including the development of the software will be shown and discuss. The results will be divided into two-part, the results of the experiment for each objective of the project and the evaluation of the system.

One of the main reasons for this research is to prove that what is the best classifier to provide predictive information, and because of that, table 2 depicts that the researcher observed that the highest-ranked classifier is Logistic Regression. Therefore, the Logistic Regression delivers a more precise and more robust output of giving accurate predictive results for crime events. In this project, two (2) models are developed and integrated (crime and crime location), to observe an excellent learning performance.

With the results shown in table 2, all three (3) machine learning algorithms have acquired an acceptable outcome, however, multinomial logistic regression shows a promising effect in terms of accuracy. To further evaluate the performance of multinomial logistic regression, a classification report was created.

Table 2: Algorithm Accuracy in Predicting Crime

Classifier	Accuracy
Multinomial Logistic Regression	86.60%
SVM	71.02%
Neural Network	72.10%

Table 3 or Classification Report for Crime shows that the highest precision among the crime event happens in the municipality of Calauan is the Comprehensive Dangerous Drug Act. This means that these crime events can provide correct predictions and provide a strong statement for forecasting each of barangay. Also, among all of the crime, the Comprehensive Dangerous Drug Act is leading again for delivering the highest percentage that can provide for the prediction of the project. These show that among the crime, the algorithm performs better in predicting the Comprehensive Dangerous Drug Act. Furthermore, it is also indicated in table 3 that multinomial logistic regression has passed all three (3) essential metrics in predicting crime, which are precision, recall, and f1-score.

Precision shows the capability of Logistic Regression in classifying the correctness of the existence of each crime type to be predicted. It also indicates that the highest percentage of precisely predicted crime is the Comprehensive Drug Act.

Commonly, Recall shows the number of all the results that providing the true statement for the prediction. As the figure is shown, the highest percentage with the numerical value of 0.98 is the crime such as Traffic Rules, Hit and Run, and the Act of Lasciviousness.

F1-score is the sub-contrary mean of exactness and recall. It is therefore the quantifier to class the capability of the algorithm to confidently seek for the actual crime events, with the numerical value of 0.83.

Table 3: Classification Report of Multinomial Logistic Regression for Crime Prediction

Crime Type	Precision	Recall	F1-score
Acts of Lasciviousness	0.85	0.98	0.87
Alarm and Scandal	0.76	0.87	0.89
Carnapping	0.87	0.86	0.76
Child Abuse	0.91	0.94	0.90
Comprehensive Dangerous Drug Act	0.99	0.95	0.89
Grave Threat	0.84	0.89	0.87
Hit and Run	0.85	0.98	0.76
Homicide	0.00	0.81	0.78
Illegal Disposition of Firearms	0.89	0.94	0.76
Kidnapping	0.85	0.87	0.95
Murder	0.78	0.89	0.87
Parricide	0.71	0.91	0.78
Rape	0.83	0.87	0.85
Robbery	0.98	0.90	0.83
Theft	0.78	0.94	0.79
Traffic Rules	0.90	0.98	0.81
VAWC	0.97	0.81	0.79
Total	0.80	0.90	0.83

Table 4 represents the results of the three (3) chosen machine learning algorithms' performance in predicting crime location. It is determined that multinomial logistic regression stills got the highest result in terms of accuracy. It is shown in table 4 that the neural network did not pass the required accuracy rate in predicting crime, which is 70%. However, accuracy alone still not acceptable as metrics in choosing a machine learning algorithm for model development, and so, it is required to evaluate further the algorithm that has the highest accuracy rate.

Table 4: Algorithm Accuracy in Predicting Crime Location

Classifier	Accuracy
Multinomial Logistic Regression	82.03%
SVM	78.09%
Neural Network	52.00%

For providing Predictive Location on a specific crime, Table 5 indicates that Brgy. Dayap in Calauan Laguna, is leading to the highest percentage of giving Precision or providing a true statement and precise tagged location on this research. While for the Recall or presenting the actual results for giving true statement, the Brgy. Prinza takes leading barangay for that, followed by Brgy. Mabacan, Brgy. Bangyas, and Hanggan.

Table 5: Classification Report of Multinomial Logistic Regression for Crime Location Prediction

Location	Precision	Recall	F1-Score
Dayap	0.87	0.89	0.45
Balayhangin	0.85	0.90	0.93
Imok	0.86	0.90	0.37
Kanluran	0.79	0.79	0.89
Limao	0.93	0.88	0.79
Lamot 1	0.84	0.90	0.97
Lamot 2	0.63	0.81	0.90
Silangan	0.54	0.97	0.88
Santo Tomas	0.93	0.89	0.94
Paliparan	0.85	0.66	0.89
Perez	0.11	0.75	0.91
Hanggan	0.69	0.93	0.84
Bangyas	0.85	0.93	0.87
Masiit	0.85	0.00	0.87
Mabacan	0.89	0.94	0.86
San Isidro	0.90	0.88	0.84
Prinza	0.83	0.96	0.79
Total	0.78	0.82	0.82

With further observation of the predicted crime events on the municipal crime report management system, the top four (4) crimes were provided based on the systems data analytics result.

Legend:

Figure 4 shows the rating of illegal drugs in different barangay of Calauan, Laguna. It also indicates that barangay Dayap has the highest rate of illegal drugs, followed by barangay Balayhagin which has the second-highest rate. Moreover, Wednesday and Saturday have the highest number of occurrences.

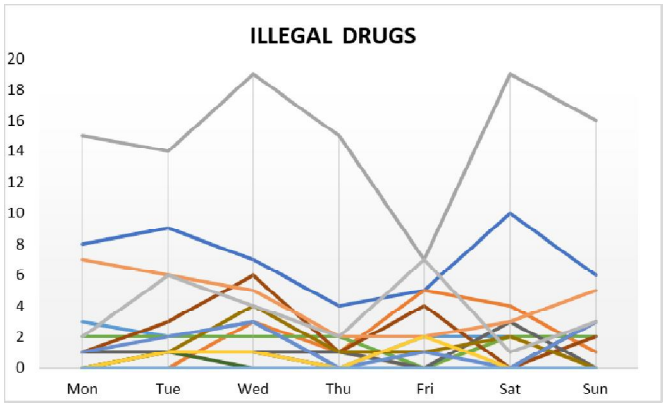


Figure 4: System's Data Analytics for Illegal Drugs Violation in Calauan, Laguna

Figure 5 shows that the leading rates were also barangay Dayap, which is the grey line, Balayhangin has the same rate in the day of Monday, and for Tuesday Barangay Perez, Bangyas and Dayap has the same rate. Balayhangin has the highest rate on the day of Wednesday, and for the following day, Dayap is the highest rate. For the day of Sunday, the highest rate is Bangyas and Lamot 2.

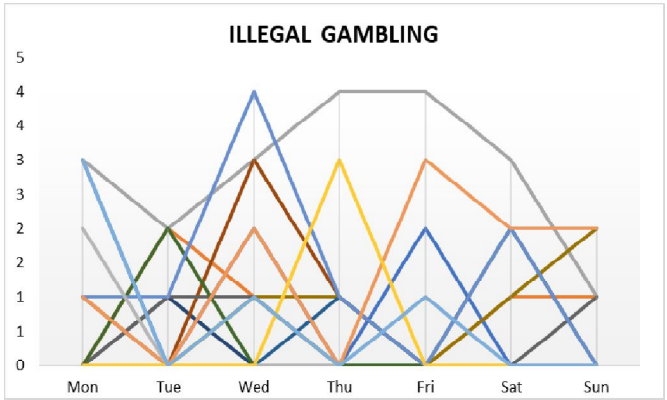


Figure 5: System's Data Analytics for Illegal Gambling Violation in Calauan, Laguna

Figure 6 shows the VAWC was the 3rd crime rate that has an unstable statistic of results that indicate that percentage of the women that got assaulted has a lack of chance to happen. Likewise, among all the records in the system, violence against women occurs typically in different barangay of Calauan, Laguna, during weekends.

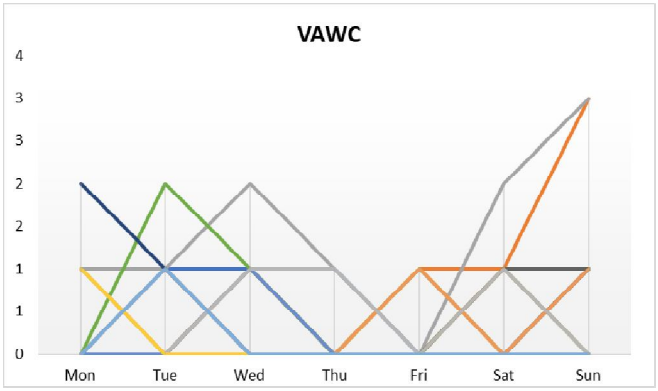


Figure 6: System's Data Analytics for VAWC in Calauan, Laguna

Homicide was the 4th highest crime rate on Calauan. Based on three years of records, it shows that the Dayap has the most significant number of homicides, while the Silangan has a Zero rate. Furthermore, shown in figure 7 that most of the homicide from different barangays happened on Tuesday; however, it was not showing any pattern in terms of daily occurrences.

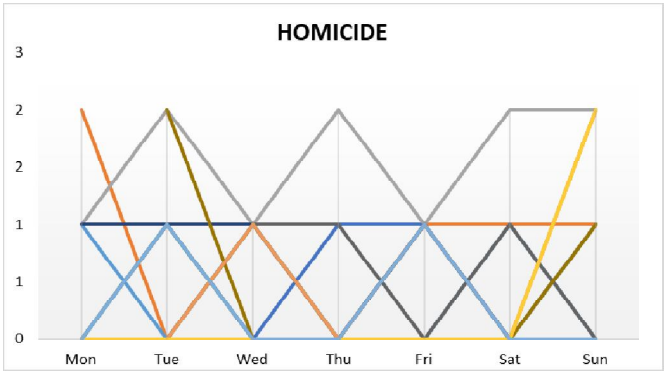
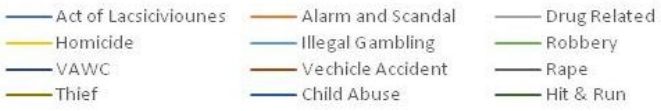


Figure 7: System's Data Analytics for Homicide Cases in Calauan, Laguna

In this part, the top five (5) barangay in terms of crime occurrences will be presented based on what crime is always occurred in every one of them.

Legend:



On figure 8, Drug-related crime is the most noticeable crime problem on barangay Dayap among others. This crime was prevalent in this barangay. It can quickly notice that most of the records are on Wednesday. However, it does not correlate drug violation since the file is recorded after drug operation, unlike any other crimes that are recorded during its occurrence. Moreover, the figure can be used as a reference that most of the illegal drug violators can quickly be arrested on Wednesday since most of the recorded was from this day.

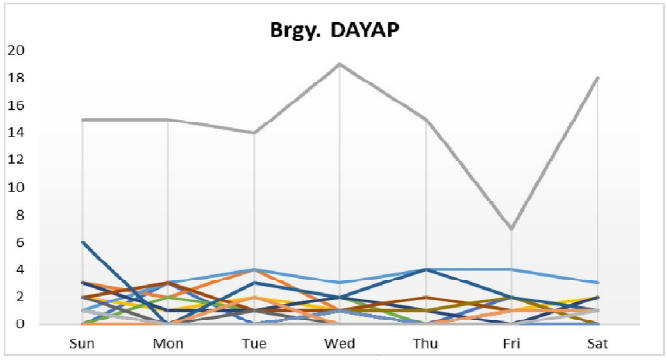


Figure 8: System's Data Analytics for Barangay Dayap Crime Occurrences

Illegal Drug is still the most crime recorded in barangay Silangan. It is clearly shown in figure 9, that illegal drug has the highest number of crimes recorded in the second barangay. However, unlike barangay Dayap, barangay Silangan has a smaller number of illegal drugs, and most of the recorded operations were from Friday, Tuesday, and Wednesday.

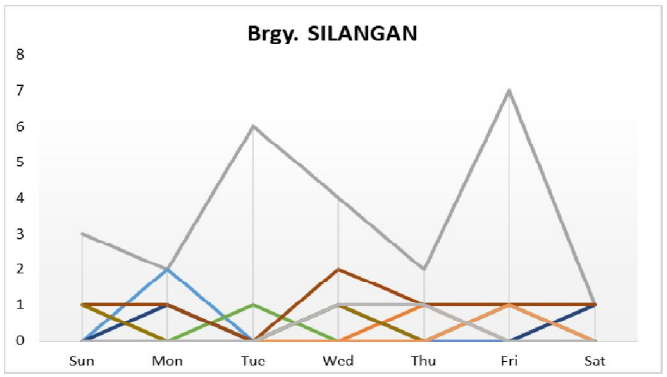


Figure 9: System's Data Analytics for Barangay Silangan Crime Occurrences

Figure 10 shows that in barangay Sto. Tomas, illegal drug still the most common crime recorded by the PNP-Calauan. Moreover, unlike the first two barangay, in this barangay, illegal gambling has more numbers than illegal drugs during Friday. Likewise, illegal drugs are more successful on Monday, followed by Tuesday, Sunday, and Wednesday.

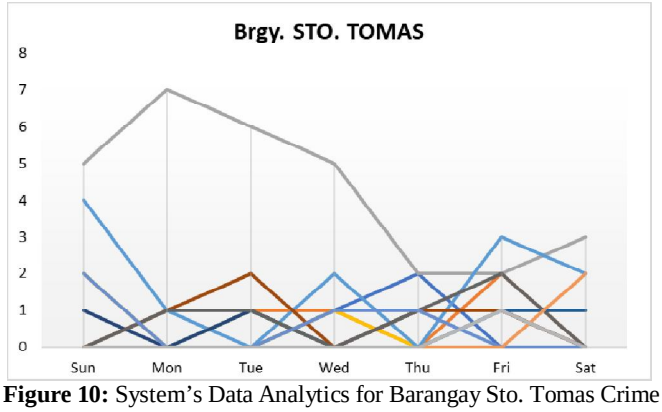


Figure 10: System's Data Analytics for Barangay Sto. Tomas Crime Occurrences

In figure 11, it is noticed that all crimes have almost the same numbers, unlike the first three (3) barangay. However, even though the numbers were nearly the same, it can still be observed that illegal drugs have the most significant number especially during Wednesday and Saturday.

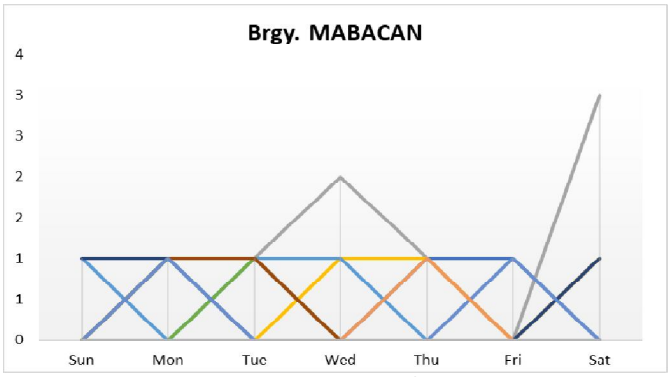


Figure 11: System's Data Analytics for Barangay Mabacan Crime Occurrences

In barangay Balayhangin, illegal drugs were once again having the most significant number of crimes recorded, as shown in figure 12. Just like the other barangay, barangay Balayhangin has a daily record of illegal drugs. It is presented in figure 12 that most illegal drugs crime are recorded on Monday, Tuesday, Sunday, and Wednesday. Moreover, some other days were not different in terms of numbers.

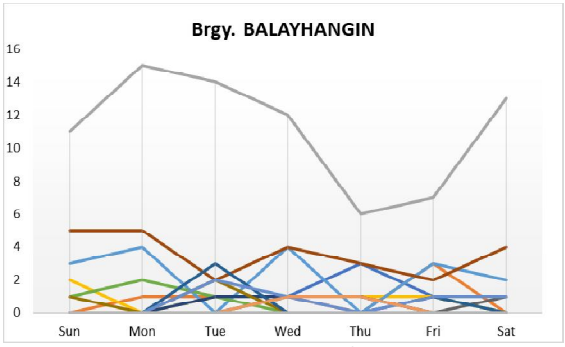


Figure 12: System's Data Analytics for Barangay Balayhangin Crime Occurrences

Web-based Police Crime Report System. For updating the data of the system, the admin must input the code correctly and frequently. All the input data will be added to the model for continuous improvement of the models' learning. This will also provide new knowledge or patterns that are significant in the future operation of PNP-Calauan. As for the adding page of the data, it consists of the month, day, hour, location of the crime, and the crime events.

Figure 12: Web-based Police Crime Reports System

Forecasting System with GIS. The User must select a specific date, time, and crime before system begin, on the prediction, it shows the map of Calauan sorted by highest to lowest threat percentage to know their ranking. As the figure shows, it displays that the admin wants to see the rate of crime of the "Comprehensive Dangerous Drug Act." The system generated predictive information of what are the possible barangay that has a possible crime event.

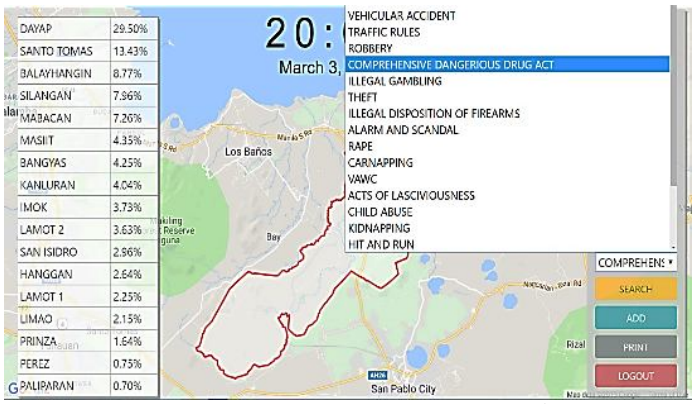


Figure 13: Forecasting System using geographical information system

For providing Probabilistic Report of possible crime event, the admin must select a specific date, time, and especially location before the system begins, the prediction would show the top-five possible crime with percentage. The System also provides a Pie graph for a better understanding of representing the crime events.

As for providing the predicted crime and location, the data gathered must be organized and categorized so that it can easily be managed. The raw data from the CSV file already connected to the system.

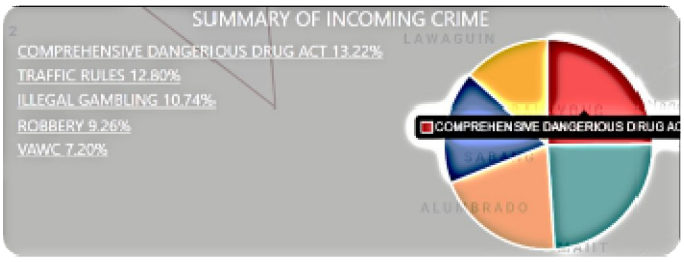


Figure 14: Probabilistic Report of Crime Occurrence

5. CONCLUSION AND RECOMMENDATION

The project generates a primary method to produce an idea that can help the community. With that, the development of a Municipal Crime Report Management System with Predictive Crime Analytics for the possibility of policy recommendation was implemented. It was found out that using predictive crime analytics can really help for the PNP in producing strategic planning based on the history data. As for the forecasting in crime, it is observed that multinomial logistic regression would help for predicting patterns of location and crime events. And generally, for showing the predicted crime, the data set must be equal from each category so that there will be no more bias from analytics. As for GIS or Visual Map that represents the tagged location, it is concluded that using the plain map can help and understand better the possible crime. On illustrating the probabilistic report of a possible crime, some graphical representation records such as a table, ranking by the highest to the lowest rate, and providing pie graph as showing the comparable data of analytics crime makes the system more understandable to any kind of user.

Through the project, the following recommendations are generated:

1. It is recommended to the PNP-Calauan to have a further and more extensive investigation of the Brgy. Dayap Calauan, Laguna, since the system shows the Barangay having a high probability rate of a possible crime.
2. Since illegal drugs are present in every barangay of Calauan, Laguna, the project would like to give stress to the local municipality of Calauan to conduct programs regarding illegal drugs.
3. This project must be implemented in other police departments in the Philippines to have a superb model for forecasting crime occurrences.
4. The development and implementation of a Centralized Management System, including other variables that are necessary for criminal investigations for the future, are strongly recommended.
5. Lastly, the application of Multinomial Regression in other intelligent systems that intends to implement forecasting must be considered since it shown a very favorable result in the project.

ACKNOWLEDGEMENT

The authors would like to extend their gratitude to the Laguna State Polytechnic University – Los Baños and Philippines National Police – Calauan, for their help to successfully finish this research study.

REFERENCES

1. J. E. Eck and D. Weisburd. **Crime Places In Crime Theory**, *Crime Control Institute*, pp. 1-33, 2014.
2. Encyclopedia.com. **Index Crime**, 2016. [Online]. Available:https://www.encyclopedia.com/social-sciences/dictionaries-thesauruses-pictures-and-press-releases/index-crime.
3. M. U. C. Report. **Non-Index Crime**, 2015. [Online]. Available:https://www.michigan.gov/documents/Non-IndexStateTotals_106263_7.pdf.
4. PwC's 2016 Global Economic Crime Survey - The Philippines. **Economic Crime at Work?,** *PWC*, pp. 1-16, 2016.
5. T. Almanie, R. Mirza and E. Lor. **Crime Predictive Based on Types and Using Spatial and Temporal Criminal Hotspot**, *International Journal of Data*

Mining& Knowledge Managemnt Process (IJDKP) Vol. 5, pp. 1-19, 2015.
https://doi.org/10.5121/ijdkp.2015.5401

6. United Nations Human Settlements Programme. **Enhancing Urban Safety and Security:Global Report on Human Settlements**, *London: Earthscan,* 2010.
7. M. Shaw and K. Travers. **Strategies and Best Practices in Crime Prevention in particular in relation to Urban Areas and Youth at Risk**, *International Center for the Prevention of Crime*, pp. 1-189, 2012.
8. C. V. Sanidad-Leones. **The Current Situation of Crime Associated with Urbanization: Problems Experienced and Countermeasures Initiated In The Philippines**, *129th International Senior Seminar Visiting Experts' Papers*, pp. 133-150, 2010.
9. M. Tayebi, F. Richard and G. Uwe. **Understanding the Link Between Social and Spatial Distance in the Crime World**, Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12), Redondo Beach, California, pp. 550-553, 2012.
https://doi.org/10.1145/2424321.2424412
10. D. Pager and H. Shepherd. **The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets**, *Ann Rev of Sociol*, p. 181–209, 2008.
11. S. Tahsildar. **Why is logistic regression so important for machine learning?**, 20 Febuary 2018. [Online]. Available:https://www.quora.com/Why-is-logistic-regression-so-important-for-machine-learning.
12. S. Sperandei. **Understanding logistic regression analysis**, 24-1 Febuary 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3936971/.
13. D. K. S. M.Tawarish. **A Review on Pricing Prediction on Stock Market by Different Techniques in the Field of Data Mining and Genetic Algorithm**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1, pp. 23-26, 2019.
https://doi.org/10.30534/ijatcse/2019/05812019
14. A. Z. Maghuyop. **A Response Assessment on the Implementation of Senior High School TVL Track through Data Mining Technique**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 2721 - 2725, 2019.
https://doi.org/10.30534/ijatcse/2019/06862019
15. D. B. K. Bhardwaj. **A Critically Review of Data Mining Segment: A New Perspective**, *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 2984-2987, 2019.
https://doi.org/10.30534/ijatcse/2019/50862019
16. J. Asor, G.M. Catedrilla, M. Villarica, J. Cabiente and F. Balahadia. **Intelligent Investigation on Crime Incident Reports in the Province of Laguna through Predictive Model Development**, Manuscript sumbitted for publication.