

Intelligent Investigation on Crime Incident Reports in the Province of Laguna through Predictive Model Development

Francis F. Balahadia¹, Jonardo R. Asor², Gene Marck B. Catedrilla³, Mia Villarica⁴, Joseph M. Cabiente⁵

¹Laguna State Polytechnic University, Philippines, balahadiafrancis@gmail.com

²Laguna State Polytechnic University, Philippines, asor.jonardo@lspu.edu.ph

³Laguna State Polytechnic University, Philippines, catedrilla.genemarck@lspu.edu.ph

⁴Laguna State Polytechnic University, Philippines, mia.villarica@lspu.edu.ph

⁵Laguna State Polytechnic University, Philippines, josephcabiente@yahoo.com



ABSTRACT

In the past years, crime becomes one of the main concerns in the Philippines for it affects drastically in the economic growth of the country. Awareness was one of the key factors that a police officer must possess to effectively reduce crime in particular location. Many criminologists study on the number or occurrence of a crime to resolve the problem, however, number vagueness and possible source are often encountered that compromises the possible real effects or pattern. Machine learning is well-known to produce new knowledge and discover hidden patterns intelligently in particular databases which can be used to produce data-driven reasoning or policy recommendation. The key objective of this research is to develop a predictive model in investigating crime records in the province of Laguna. Following the famous concept of knowledge discovery in databases, the researchers found out that decision tree algorithm is the best machine learning algorithm in classifying crime occurrence. Furthermore, date, time and place have a significant correlation in crime occurrence. Also shown in this paper, that the bigger district in the province of Laguna is more vulnerable in different crimes.

Key words: Crime incident, decision tree, crime reports, hotspot, mapping system, prediction

1. INTRODUCTION

Crime has a high impact in economic growth of particular community. It showed an undisputed negative effect on investment, outflows of human capital and/or increased in production cost [1]. Community-based approaches on criminal or crime prevention is more effective rather than enforcement alone which shows a greater impact on criminals' behavior. This program works to increase the sense of safety and security of the residents and soon will promote increase in economic and social growth [2].

Criminal justice officials, journalists and even academic criminologists count crimes occur in particular places and unintentionally compromised the key source of vagueness of that number. It is a common criminological knowledge that not all the crimes are being reported, however, that knowledge ended-up only no contribution or role in the discourse about crime patterns in local communities [3]. Thus, analyzing crime patterns will be a big help in

desolating crime in a community. It plays a vital role in developing crime resolution and other programs that may help to improve the crime rates in certain communities.

Police consciousness is mainly a significant factor that may lead to lessening the crime rate. Police departments cannot simply imagine where and what time certain crimes frequently happen [4], if not systems are implemented to aid and support them. Data mining is the practice of discovering patterns in the data. Data mining is well known as an expert-based semi-supervised learning method and develops the scheme for weighing the substantial attributes. It was already proven itself on discovering significant models and/or patterns on different aspects such as student and employee performance, marketing, expertise retrieval, crime patterns and many more [5]. With the power of data mining, it is possible to identify hidden patterns through analysis of chronological crime occurrences in particular places which can help law enforcement agencies and the government itself.

In the Philippines, one of the priorities of the law enforcement is the decrease and avoidance in crime rates as stated in the Philippine National Development Plan 2017 – 2022, national security and public order are vital elements in building the foundation for inclusive growth, a high trust and resilient society, and a globally competitive knowledge economy. PNP has cleared the crime tagging as an index and non-index crimes. Index crimes involve crimes against persons such as murder, homicide, physical injury and rape, and crimes against property such as robbery, theft, car-napping/carjacking and, cattle rustling while non-index crimes are violations of unusual and private laws such as local ordinances.

According to a news article by William Balasa in hagupit.ph, PNP or the Philippine National Police said that Laguna is one of the places in the country that has a very high crime rate. Sr. Supt. Gilbert Cruz, Provincial Director of PNP-Laguna mentioned the different towns in Laguna that have the highest crime rates. These cities are San Pedro, Binang, San Pedro, Binang, Cabuyao, Calamba City, San Pablo, Los Baños and Sta. Cruz. There are several big crime incidents that happened in Laguna such as Bank Robbery resulted to killing 10 people last 2008 and hijacking last 2011 [6]. Based on the news report by [7] The Philippine Star (2015), Roxas and the PNP announced a plan to

implement Oplan Lambat-Sibat in Central Luzon and in Region IV-A due to the increasing crime rate up by 46% last 2015. Under Oplan Lambat-Sibat, police officers set up bolt from the blue checkpoints, raid or visit the homes of delinquent gun owners, intensify intelligence-gathering and operations against organized crime group and assertively pursue fugitives.

Novel manner of knowledge illustration and information mining are needed in the multifaceted and dynamic crime investigation domain. In crime investigation, the data that pattern the criminal event and its investigation, known as the crime analysis information is very important in revealing the criminal offense. Though, the present system of a recording nature only keeps the criminal records, card files and databases. It does not construct the information meaningful and therefore it is insufficient in providing the information for crime investigation. As a result, without a proper knowledge representation model through data mining to represent the crime investigation knowledge in an organized and meaningful ways, it would be hard for the investigation officer to analyze the related information and make a proper decision during the investigation process.

Pattern analysis can be used in developing improved crime resolution, better crime prediction and prevention of offenses [8]. This will be used to produce an artificial intelligent for the forecasting process. Intelligent system shows its capabilities in predicting crime occurrence. International Association of Crime Analysts has recognized four types of crime analysis which is crime intelligence analysis, tactical crime analysis, strategic crime analysis and administrative crime analysis [9,10]. Business analytics supported that types of crime analysis where there are several factors need to be identified such as the purpose, scope, data sources and analyses. Also, the tactical crime analysis is an analysis of data for long range problem and it is a must for a police department to have this kind of analysis for their future preparation. These days, the crime intelligence analysis is a fresh development of crime analysis [11].

This paper aims to create a crime pattern model that can analyze non-index crime incidents in terms of the following attributes: date, month, day, year, time of the crime incident, district which pertains to the municipality of the crime incident, longitude, latitude, crime committed, victim's age, sex, nationality suspect's age, sex, nationality, status and actions taken. Patterns and information extracted from the analysis of crime data can be applied in the crime investigation domain to preempt and stop possible crime occurrences.

2. METHODS

Research has shown that crime pattern analysis systems can provide useful insights and knowledge about crime incidence, which are helpful in designing crime prevention activities especially for known crime hotspots. The current research project has as an additional objective the goal of enabling the police authorities in Laguna to increase dissemination of crime reports and knowledge throughout the province. For the above purposes the common process of

knowledge discovery in databases (KDD) was adopted in this study.

2.1 Data Gathering

Data from police records were obtained with permission from the provincial headquarters of the PNP in Laguna province. The records include details for every crime responded from 2014 to 2017 (n=7,260) such as the date and time of the crime incident, address/location of crime, type of crime, alert level, and status of the report.

2.2 Data Preprocessing

Two important transformations were performed on the crime data set to make it more suitable for analysis. The first involved the use of Google Maps API to identify the corresponding geographical coordinates of each address/location in the dataset. The second transformation involved splitting the dates and times of each incident into year, month, and day of the week, organizing them into separate columns, and converting time into 24-hour format. Visualization of the results using graphs and charts to facilitate further analysis and interpretation was done using Microsoft Excel. As the result of the pre-processing the following attributes were generated:

Table 1: Dataset attributes

Attributes	Description
Date	Exact date when the crime occurred.
Month	Month of the year when the crime occurred.
Day	Day of the week when the crime occurred.
Year	Year when the crime occurred.
Time	Time when the crime occurred.
Address	Barangay where the crime occurred.
City	City where the crime occurred.
Violation	Type of crime occurred.
Sus_Sex	Suspect's sex.
Vic_Sex	Victim's sex.
Remarks	Action taken by the officer-in-charge in the case.

2.3 Building and Training the Model

Three (3) different well-known machine learning algorithm were used in developing the model.

Naïve Bayes which was based from Bayes theorem: $P(A/B) = P(B/A) P(A)/P(B)$ applies probabilistic approach in the dataset. In this paper, the Bayes theorem can be written as:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (1)$$

wherein X is set as the element to find in the line of subset which are the attributes of the dataset and represented as x_n :

$$X = (x_1, x_2, x_3, \dots, x_n) \quad (2)$$

From here $x_1, x_2, x_3, \dots, x_n$ represents the attributes such as Date, Time, Month, Year, Address, etc. After substituting the X, it will be expanded using chain rule and following will be produced:

$$P(y|x_1, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)} \quad (3)$$

All the values will be acquired from the dataset. Upon substituting all the entries from the dataset, the denominator will not transform and remains stationary, therefore, it can be removed and proportionality will be produced:

$$P(y|x_1, \dots, x_n) \propto P(y) \prod_{i=1}^n P(x_i|y) \quad (4)$$

Since the class variable (y) has more than two outcomes, the classification will be multivariate and finding y must be done will maximum probability:

$$y = \operatorname{argmax}_y P(y) \prod_{i=1}^n P(x_i|y) \quad (5)$$

The Naïve Bayes classifier is basically compared with KNN classifier for crime analysis and is used for crime analysis[11][12]. KNN is a machine learning algorithm which assumes that similar objects appear in close proximity. Therefore, those attributes which appears in close proximity are considered true and useful.

Decision tree is also used in this paper. Decision tree is famous for its result is understandable and readable in human language. Tree algorithm is represented by a flowchart-like structure in which external nodes are the outcome while the internal nodes are the attributes which are considered correlated to the class attributes.

Decision tree figure is being read as conditional statements for acquiring pattern in dataset. The combination of all related internal nodes will be written in an if-else statement for decision analysis.

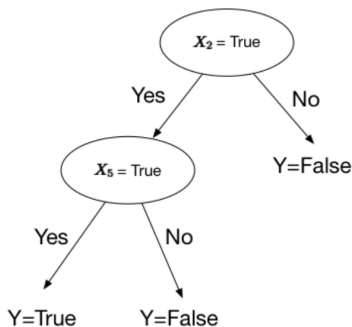


Figure 1: Decision Tree

The data set was divided into two parts, the training set and the testing set, for use in validation and evaluation. The training set constituted 80% of the data set while the testing set consisted of the remaining 20%. This process was necessary to ensure the accuracy and precision of the classifier model.

Table 2: Train and Test Set

Labels	Train Set	Test Set
	Number of Instance	Number of Instances
Child Abuse	1008	252
Comprehensive Firearms and Ammunition	83	20
Dangerous Drug	2417	604
Illegal Gambling	701	175
Land Transportation	287	71
Violence Against Women	1314	328
Total	5810	1450

2.4 Prediction Evaluation

The best classifier was chosen based on the performance of each model in terms of accuracy, precision, sensitivity, kappa statistics, and specificity which will be based on the confusion matrix. Confusion matrix is well-known in evaluating the performance of machine learning in terms of classification, it is represented as:

	Predicted: 0	Predicted: 1
Actual: 0	TN	FP
Actual: 1	FN	TP

Figure 2: Confusion Matrix

It contains four (4) different attributes namely TN which is true negative, FP = false positive, FN = false negative and TP which is true positive. In confusion matrix, TN and TP are values with correct prediction wherein FP and FN are incorrect prediction.

Accuracy is measured in terms of the percentage of features that are predicted correctly; a rating of 70% is considered acceptable, meaning the algorithm shows a promising correctness in prediction. Accuracy can be computed as:

$$Accuracy = (TN + TP) / (TP + TN + FP + FN) \quad (6)$$

Precision is measured based on the number of positive features classified by the model as positive.

$$Precision = TP / (TP + FP) \quad (7)$$

Recall gives the number of positive features classified correctly by the model.

$$Sensitivity(recall) = TP / (FN + TP) \quad (8)$$

Kappa is the correctness of the prediction of the algorithm, measured from 0 to 1.0, where a score of at least 0.50 is considered acceptable.

2.5 Visualization and Knowledge Extraction

The chosen classification model will be utilized in extracting and producing a visualization of knowledge, and generating specific recommendations based on this knowledge. The viability of the developed model can be taken advantage of to present new information through data visualization, which can be used for intelligent reasoning, policy recommendation and for other relevant purposes.

3. RESULTS AND DISCUSSIONS

Results of evaluation showed the Decision Tree algorithm to be the best classifier, with both Naïve Bayes and k-NN getting less than the desired accuracy and kappa scores. The Decision Tree classifier model was then further evaluated in terms of precision and recall through the use of a Confusion Matrix table.

Table 3: Classifier Performance Parameter

Classifier	Accuracy	Kappa
Naïve Bayes	70.34%	0.44
k-NN	64.45%	0.16
Decision Tree	84.75%	0.64

The confusion matrix table4 for the Decision Tree classifier shows a high confidence score in classifying *Child Abuse*, *Dangerous Drugs*, and *Violence against Women* violations, which are all common crimes in the Philippines. *Comprehensive Firearms and Ammunition* had a low confidence score in classification while *Illegal Gambling* had a low recall value but high precision rate, which means that this result can be improved with a bigger sample of data shown in Table 3. Overall, the results shown in the table indicate that further preprocessing may be needed before creating an artificial intelligence (AI) model using the data set.

Table 4: Confusion Matrix

	true Child Abuse	true Dangerous Drugs	true Violence against Women	true Illegal Gambling	true Comprehensive Firearms	true Land Transportation
pred. Child Abuse	1176	31	335	5	18	13
pred. Dangerous Drugs	50	2799	218	570	83	131
pred. Violence against Women	29	33	1047	19	0	13
pred. Illegal Gambling	1	92	25	226	1	22
pred. Comprehensive Firearms	1	1	0	0	0	0
pred. Land Transportation	3	65	17	56	1	179

Based on table 4, the model is having problem in classifying *Comprehensive Firearms*. Nonetheless, it acquires acceptable classifying confidence to the rest of the class-label. To further investigate the model’s performance, table V represents other evaluation results. It is observed that the model is having problem in classifying *Comprehensive Firearms* and *Illegal Gambling*. Furthermore, in table IV it can be observed that most of *Comprehensive Firearms* and *Illegal Gambling* were classified as *Dangerous Drugs*.

Table 5: Summary of Decision Tree’s Performance

Class-label	Precision	Recall	Specificity	F-score
Child Abuse	74.52	93.33	0.99	0.83
Dangerous Drugs	72.68	92.65	0.93	0.81
Violence against Women	91.76	63.76	0.90	0.75
Illegal Gambling	61.58	25.80	0.58	0.36
Comprehensive Firearms	0.00	0.00	0.58	0.00
Land Transportation	55.76	50.00	0.97	0.53
Average	59.38	54.26	0.83	0.55

In order to examine the dataset for knowledge discovery, the researchers clustered the induction model of the Decision Tree classifier according to the type of crimes included in the data set. The following six tables show the patterns discovered using the algorithm. Six (6) induction models were created from the data set to show patterns of crime incidence in Laguna based on police records (see following figures). Majority of the most conspicuous patterns shown were produced by the attributes of location, date, and time of crime incidents.

Table 6: Child Abuse

If	Then
Sex = Female and Date > Sep 29	Child Abuse
District = Cabuyao and Date > Oct 30 and Action Taken = At Large	Child Abuse
District = Cabuyao and Date ≤ Oct 30 and Sex = Female and Action Taken = At Large	Child Abuse

Table 7: Comprehensive Fire Arms and Ammunition

If	Then
District = Los Banos and Sex = Male and Date > Feb 22 and Action Taken = Deceased	Comprehensive Fire Arms and Ammunition

Table 8: Land Transportation

If	Then
District = Cabuyao and Date > Oct 30 and Action Taken = Arrested and Time > 11:52 PM	Land Transportation
District = Cabuyao and Date ≤ Oct 30 and Sex = Male and Date > Apr 17 and Date > May 24	Land Transportation
District = Cabuyao and Date ≤ Oct 30 and Sex = Male and Date ≤ Apr 17 and Date ≤ Feb 14	Land Transportation
District = Liliw and Date ≤ May 21 and Time > 6:30 AM	Land Transportation
District = Magdalena and Date ≤ Sep 16 and Date > Feb 23 and Date > Mar 2	Land Transportation
District = Nagcarlan and Date ≤ Oct 15 and Time > 4:15 AM and Date > Feb 14 and Sex = Female and Date ≤ Mar 28	Land Transportation
District = Nagcarlan and Date ≤ Oct 15 and Time > 4:15 AM and Date > Feb 14 and Sex = Male and Action Taken = At Large	Land Transportation
District = Nagcarlan and Date ≤ Oct 15 and Time > 4:15 AM and Date ≤ Feb 14	Land Transportation

Table 9: Illegal Gambling

District = Bay and Action Taken = Arrested and Date > Feb 17 and Date ≤ March 10 and Date > 25	Illegal Gambling
District = Bay and Action Taken = At Large and Sex = Male and Date > April 17	Illegal Gambling
District = Binan and Action Taken = At Large and Date ≤ Apr 6	Illegal Gambling
District = Cabuyao and Date ≤ Oct 30 and Sex = Male and Date > Apr 17 and Date ≤ May 24	Illegal Gambling
District = Calamba and Sex = Female and Date ≤ Mar 23	Illegal Gambling
District = Calauan and Date ≤ Mar 23 and Time ≤ 3:50 PM	Illegal Gambling
District = Cavinti and Date ≤ May 21	Illegal Gambling
District = Kalayaan and Sex = Female	Illegal Gambling
District = Liliw and Date > May 21 and Sex = Female	Illegal Gambling
District = Los Banos and Sex = Female and Date ≤ Mar 27	Illegal Gambling
District = Los Banos and Sex = Male and Date > Feb 22 and Action Taken = Arrested and Date ≤ Jul 16 and Time > 9:00 AM	Illegal Gambling
District = Los Banos and S_Sex = Male and Date ≤ Feb 22 and Date ≤ Feb 11	Illegal Gambling
District = Mabitan and Date > Mar 19 and Date ≤ Apr 24	Illegal Gambling
District = Magdalena and Date > Sep 16 and Date > Nov 15 and Date > Jul 19	Illegal Gambling
District = Magdalena and Date > Sep 16 and Date ≤ Nov 15	Illegal Gambling
District = Nagcarlan and Date ≤ Oct 15 and Time > 4:15 AM and Date > Feb 14 and S_Sex = Female and Date > Mar 28 and Time ≤ 10:45 PM	Illegal Gambling

Table 10: Dangerous Drugs

If	Then
District = Bay and Action Taken = Arrested and Date > Feb 17 and Date > March 10	Dangerous Drugs
District = Bay and Action Taken = Arrested and Date > Feb 17 and Date ≤ March 10 and Date ≤ 25	Dangerous Drugs
District = Cabuyao and Date > Oct 30 and Action Taken = Arrested and Time ≤ 11:52 PM	Dangerous Drugs
District = Cabuyao and Date > Oct 30 and Action Taken = Deceased	Dangerous Drugs
District = Cabuyao and Date ≤ Oct 30 and S_Sex = Female and Action Taken = Arrested and Date > March 5	Dangerous Drugs
District = Cabuyao and Date ≤ Oct 30 and S_Sex = Male and Date ≤ Apr 17 and Date > Feb 14	Dangerous Drugs
District = Calamba and S_Sex = Female and Date > Mar 23	Dangerous Drugs
District = Calamba and S_Sex = Male	Dangerous Drugs
District = Calauan and Date > Mar 23	Dangerous Drugs
District = Calauan and Date ≤ Mar 23 and Time > 3:50 PM	Dangerous Drugs
District = Cavinti and Date > May 21	Dangerous Drugs
District = Kalayaan and S_Sex = Male and Date > Sep 16	Dangerous Drugs
District = Liliw and Date > May 21 and S_Sex = Male	Dangerous Drugs
District = Los Banos and S_Sex = Female and Date > Mar 27	Dangerous Drugs
District = Los Banos and S_Sex = Male and Date > Feb 22 and Action Taken = Arrested and Date > Jul 16	Dangerous Drugs
District = Los Banos and S_Sex = Male and Date > Feb 22 and Action Taken = Arrested and Date ≤ Jul 16 and Time ≤ 9:00 AM	Dangerous Drugs
District = Los Banos and S_Sex = Male and Date > Feb 22 and Action Taken = At Large	Dangerous Drugs
District = Luisiana and Date > Feb 26 and Date > May 27 and Date Mar y 31	Dangerous Drugs
District = Mabitan and Date > Mar 19 and Date > Apr 24	Dangerous Drugs

Figure 3 shows the three most common non-index crimes in Laguna to be *Dangerous Drugs* (n=3851), which constitute around 53% of all cases, *Child Abuse* (1578, or 21.74%) and *Violence against Women* (1141, or 15.72%). *Illegal Gambling* (n=367) and *Land Transportation*-related (n=321) violations combined represent around 10% of all cases, while only 2 cases of comprehensive firearms and ammunition cases were recorded. Analysis revealed a number of important patterns in the non-index crimes data set. For example, crime incidence in the urban areas of Calamba, Sta. Rosa, San Pablo, Cabuyao, Biñan, San Pedro, and Sta. Cruz made up more than 63% of the total figure for Laguna province (Please see Table 7).

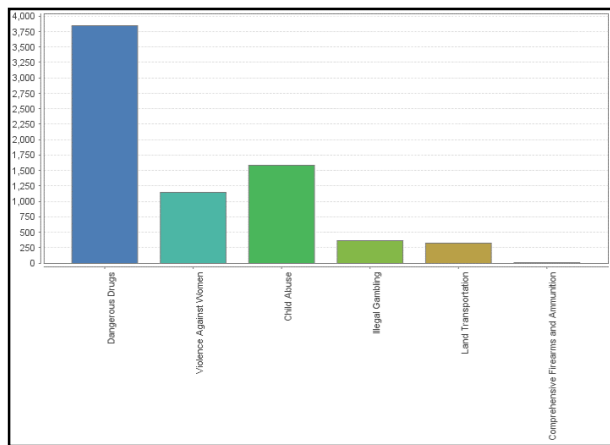


Figure 3: Crime commonality in the Laguna Province.

The greatest number of crime incidents took place in August (857 cases), while the least number happened in January (184). Another interesting pattern revealed that the period from June to August had the greatest number of crimes recorded, while the period from December to February had the least number of cases. Generally speaking, the monthly figures were relatively close to each other except for the colder months (December to February) shown in Figure 4.

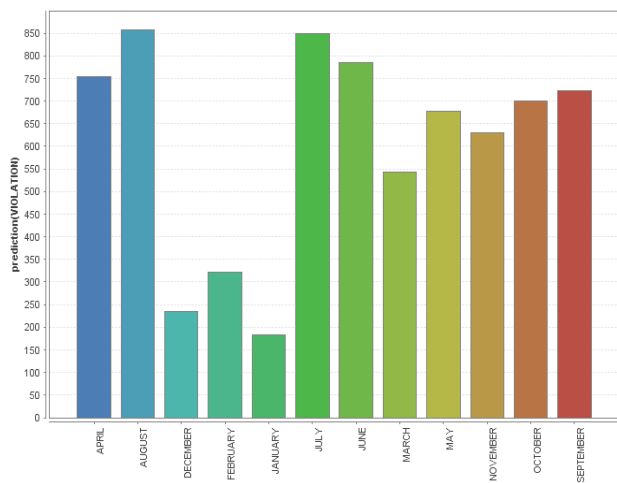


Figure 4: Crime Occurrence per district (predicted)

Figure 5 is a visual representation of the crime occurrence prediction for each city/municipality in Laguna generated by the Decision Tree classifier model. Incidence of *dangerous drugs*-related crimes were predicted for each city/town in

the province with the exception of the cities of Cabuyao and Biñan, and the municipalities of Magdalena, Calauan, Los Baños, Lumban, Majayjay and Kalayaan. Only the towns of Magdalena, Lumban and Majayjay turned out to have prediction confidence values within the acceptable range for *violence against women*-related crimes. On the other hand, the figures suggest that *child abuse* has a high probability of occurrence in Biñan, Los Baños, Sta. Cruz, Pangil, and Kalayaan.

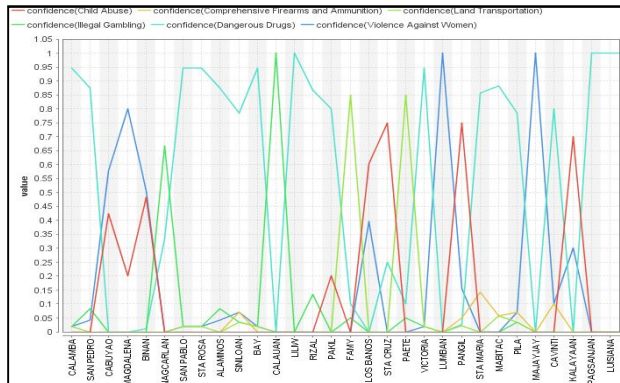


Figure 5: Crime Occurrence per district (predicted)

The three most common non-index crimes (*dangerous drugs*, *violence against women*, and *child abuse*) have almost similar density rates, as the DT algorithm showed low confidence levels for these. As for the other criminal offenses, the algorithm showed high confidence for the density levels of *comprehensive firearms and ammunition* cases in Alaminos and Los Baños, as well as for *land transportation*-related crimes in Cabuyao and in Liliw. For *illegal gambling* in Nagcarlan and San Pablo City. The algorithm showed almost zero confidence for crime incidence in the town of Kalayaan, however.

In terms of prediction correctness, incidence of *comprehensive firearms and ammunition* violations were predicted only for the towns of Siniloan, Pangil, Sta Maria, Mabitac, Pila, and Cavinti, though there is little probability of occurrence. The towns of Famy and Paete were predicted to highly likely have incidence of land transportation-related crimes while Nagcarlan and Calauan were predicted to have a high turnout of illegal gambling violations.

As for the predicted crime occurrence per month, *dangerous drugs*-related crimes have very high likelihood of taking place from March to June, August, October, and November. Incidence of *child abuse* and *violence against women* were predicted to have high probability of occurring during the months of December to February. *Land transportation*-related criminal offenses were predicted to occur during the months of July and September, while comprehensive firearms and ammunition violations, though with a very low probability, are likely to happen during November.

Lastly, results show that non-index criminal events most commonly took place between 10 in the morning and 7 in the evening. Specifically, dangerous drugs crime incidents mostly were reported between 12 PM and 4:20 PM, violence against women incidents between 11 AM and 3:30 PM, child abuse between 12 PM and 6 PM, illegal gambling

between 2 PM and 3 PM, land transportation-related criminal violations between 1 PM and 4 PM, and comprehensive firearms and ammunition criminal offenses between 10 AM and 3 PM.

4. CONCLUSIONS AND RECOMMENDATIONS

This study was designed to show that the Decision Tree algorithm is an efficient and effective machine learning algorithm highly capable of knowledge discovery and pattern extraction from data sets. This suggests that Decision Tree is a promising algorithm for use in model development in terms of its viability in intelligently analyzing data sets.

The results generated by the predictive analytical process revealed a number of important insights about the incidence of non-index crimes in the province of Laguna. For example, urban areas tend to be more prone to crime incidents than rural areas; criminal offenses have the highest probability of occurring around August and the least in January; and that criminal offenders are more likely to commit non-index crimes between 10 AM and 7 PM, which are business hours for most people and organizations.

This suggests, therefore, that date, time, and location factors are the best predictors of crime occurrence based on the patterns generated from the criminal records data set. The above information may be useful for the police authorities in Laguna in their bid to reduce criminal activities in terms of resource utilization and preemptive crime counter-operations. Developing online and mobile systems integrated with artificial intelligence technology for analyzing crime incident reports with a view to generating up-to-date crime statistics, crime hot spots mapping, and maintaining dynamic databases is highly recommended.

REFERENCES

1. V. Rios. **The impact of crime and violence on economic sector diversity**. Research Fellow, The Wilson Center. Retrieved from http://scholar.harvard.edu/files/vrios/files/riosv_crimeseconomicdiversitydec212015.pdf, 2016
2. Crime Prevention Ottawa. **Why Community-Based Crime Prevention Works: Case Study of three Ottawa Communities**. Retrieved from crimepreventionottawa.ca
3. R. Brame, M.G. Turner & R. Paternoster. **Surveying Residential Burglaries: A Case Study of Local Crime Measurement**. Retrieved from <https://arxiv.org/pdf/1204.6735.pdf>, 2013
4. J.D.E. Sandig, R.M. Somoba, M.B. Concepcion, & B.D. Gerardo, B. D. **Mining Online GIS for Crime Rate and Models based on Frequent Pattern Analysis**. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 2, pp. 23-27), 2013
5. S.V. Nath. **Crime pattern detection using data mining**. In 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops (pp. 41-44). IEEE, 2006
<https://doi.org/10.1109/WI-IATW.2006.55>
6. B. Sison, & Felipe. **Philippine crime rate up by 46%**. Retrieved from:
7. <https://www.philstar.com/metro/2015/08/02/1483869/philippine-crime-rate-46#RD8PXOLZoKQ0HdH.99>, 2015
8. J. Hays. **Crime in the Philippines**. Retrieved from: http://factsanddetails.com/southeast-asia/Philippines/sub5_6f/entry-3908.html, 2015
9. V. Kollepara & S. Ananyan. **Crime Pattern Analysis: Megaputer Case Study in Text Mining**.
10. S. Elder, C.W. Bruce, R. B. Santos. E. Rodriguez, & L. Wyckoff. **Definition and Types of Crime Analysis**. International Association Crime Analysis. (IACA) Denver, Colorado, 2015
11. S. Tripathi, K. Amendola & L. Wyckoff. **Crime Analysis with Crime Mapping. Crime Mapping and Analysis News: A Police Foundation Publication**. Summer Issue (3), 2015
12. H. Jantan & A. Jamil. **Association Rule Mining Based Crime Analysis using Apriori Algorithm**. International Journal of Advanced Trends in Computer Science and Engineering, Vol. 8, No. 1.5, 2019, pp 18-24, May 2019.
<https://doi.org/10.30534/ijatcse/2019/0581.52019>
13. J. L. D. Mercara. **Prediction of Employees' Lateness Determinants using Machine Learning Algorithms**. International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 1, pp. 779-783, 2020.
<https://doi.org/10.30534/ijatcse/2020/111912020>
14. J. S. Gil. **Predicting Students Dropout Indicators in Public School using Data Mining Approaches**. International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 1, pp. 774-778, 2020.
<https://doi.org/10.30534/ijatcse/2020/110912020>
15. S. Sivaranjani, S. Sivakumari, M. Aasha. **Crime prediction and forecasting in Tamilnadu using clustering approaches**. Emerging Technological Trends (ICETT) International Conference on, pp. 1-6, 2016.
<https://doi.org/10.1109/ICETT.2016.7873764>
16. C. Wu, F. Yang, Y. Wu & R. Han. **Prediction of crime tendency of high-risk personnel using C5.0 decision tree empowered by particle swarm optimization**. Math. Biosci. Eng. 2019, 16, 4135-4150.
<https://doi.org/10.3934/mbe.2019206>