Journal of Advanced Research in Social Sciences and Humanities
Volume 2, Issue 3 (176-187)
DOI: https://dx.doi.org/10.26500/JARSSH-02-2017-0303

Mining crime instance records of Philippine National Police District Vi Province of Cavite, Philippines: An Exploratory Study to Enhance Crime Prevention Programs

MENGVI P. GATPANDAN*, SHANETH C. AMBAT
Jose Rizal University, Mandaluyong, Philippines
School of Graduate Studies, AMA University, Quezon City, Philippines

Abstract

Aim: This analysis aimed to assess the effectiveness, productivity, and management of one of the country's national police forces. In this study, especially in the data envelopment analysis, crime management was taken into account as the primary function of the police, which necessitated the use of resources that were viewed as decision-making units.
Methodology: To identify productive DMUs, the study combined a Rate-to-Scale (RTS) metric based on scale efficiency with an input-oriented radial measure of efficiency. In the analytical tasks, clustering was used as part of a structured approach to planning data mining activities known as CRISP-DM.
Findings: Municipality C is the most efficient DMU based on the 3-year scale efficiency result.
Implications/Novelty: Given the current push by the government and the PNP to eradicate criminality and illegal activity in the Philippines, this research couldn't come at a better time. This research aided police departments and crime scene investigators in identifying patterns of criminal activity and performing geospatial analyses.

Keywords: Data Envelopment Analysis, Data Mining, Clustering Techniques, Crime Management

Received: 7 February 2017 / **Accepted:** 13 April 2017 / **Published:** 30 June 2017

INTRODUCTION

The Philippine National Police (PNP) is the country's armed and civilian police force, with headquarters at Camp Crame in Quezon City. The PNP Director oversees a force of 160,000 armed officers who maintain order across the vast territory. The Philippine National Police is a significant branch of the Philippine Armed Forces. There are a total of twelve main sections, including one devoted to the safety of aviation systems. SAF, or the PNP Mobile Unit, is the common name for the Special Action Forces. The PNP also includes the police administration office known as the Program Management Office. To protect the marine ecosystem within Philippine territorial waters, the Philippine National Police Maritime Group was given the authority to do so. The Intelligence Group of the Philippine National Police provides nationwide assistance to PNP units in pursuit of their intelligence aims. The Public Safety Battalion is a Regional Force as well. For every group that works to undermine police-community relations, others do everything they can to support it. The Anti-Cybercrime Group leads anti-cybercrime campaigns. The Anti-kidnapping Group is the functional group for kidnapping and other menaces in the country; the Highway Patrol Group puts traffic management and traffic laws into practice; and the Civil Security Group protects businesses and employees.

As of 2012, the Philippines had a police-population ratio of 1:651, with 79,878 crimes solved and a crime-solving efficiency of 36.67 percent, according to a Senate document (Francisco 2016). April 1, 2016, Rappler report found that out of all the cities in the Philippines, Quezon City had the most index crimes with 65,514, followed by Manila City with 54,689, Cebu City with 38,797, Davao City with 37,684, and Cagayan de Oro City with 31,345. In 2015, 733,853 people were living in the PNP District VI Province of Cavite, where the study was conducted. The municipalities of Amadeo, General Trias, and Tanza, along with the component city of Trece

*Corresponding author: Mengvi P. Gatpandan
†Email: mengvi.gatpandan@jru.edu

Martires, are all within the purview of PNP District VI. According to the Police Regional Office 4A January-May 2016 report (pro4a.pnp.gov.ph) under the Crime Statistics 2016 menu, the Province of Cavite has the highest crime volume, at an average monthly rate of 34.60 percent, totaling 5,987 crimes. Four (4) DMUs will be evaluated in terms of organizational resources and index crime rate to determine efficient DMU through Data Envelopment Analysis because performance efficiency is a crucial factor in determining an organization's capacity to reach targets or produce the desired results. DEA is a robust analytic tool for evaluating the efficiency of a business or government department, as pointed out by Norman and Stoker (1991). Decision analysis (DEA) was used to assess the effectiveness of decision-making units by Charnes, Cooper, and Rhodes (1978). The ineffective DMUs can reduce their index crime rate by using their organization's resources better. To reduce crime, clustering-based data mining helps identify crime patterns. According to Devi and Rajagopalan (2012), clustering analysis is a method of data mining that groups together unidentified objects that share many characteristics. Numerous clustering methods and algorithms exist, but the study employed the partition clustering technique and the k-means algorithm. These data mining techniques were used to determine the crime patterns based on the historical data of the specifically chosen cities and municipalities in the Province of Cavite. Foreseeing criminal activity, recognizing recurring crime trends, and addressing security concerns in the neighborhood are all goals of the forthcoming software application.

Since the government and the Philippine National Police are currently making an effort to root out criminality and illegal activity, this study couldn't be more pertinent. The results of this research were useful to law enforcement agencies, crime scene investigators, and others interested in analyzing crime patterns geographically. Using the study results, law enforcement officials were better able to prioritize the clearance and resolution of the index crimes, which increased their effectiveness in crime prevention.

LITERATURE REVIEW

DEA is a methodology being used in measuring the productivity and efficiency of several organizations (Boussofiane, Dyson, and Thanassoulis 1991; Vyver and Gillies 2017; Izhar et al. 2016). Several scholarly works were using data envelopment analysis to measure the organization's efficiency of its different Decision-Making Units (DMUs).

Kinaci, Najjari, and Alp (2016) used the Stochastic Frontier Analysis and Data Envelopment Analysis to measure the efficiency of hydroelectricity centers of 32 Iranian electricity industries. The study used an input-output oriented model using CCR-model and BCC-model in DEA. Chakraborty (2016) used the output-oriented nonparametric DEA to determine the inconsistencies in the operational efficiencies of the 18 life insurance firms in India and for the improvement of the operations. Tran and Bhaiyat (2016) used the parametric and nonparametric tools in the assessment of the Vietnamese commercial banks' technical efficiency and scale efficiency. The finding of the study showed that larger commercial banks in Vietnam were more efficient than the smaller ones. Baran et al. (2016) used the comparison analysis in their study related to the manufacturing sector of basic metals and metal products in Poland. The comparison analysis was used to measure the technical efficiency of 12 sectors being studied. The study revealed that the manufacturing sectors were inefficient. Therefore, plan improvements and directions were needed to produce efficient results. Caballero, Cadavid, and Toloza (2015) employed the DEA model during their empirical investigation on the technical, pure technical, and scale efficiency of the 44 public institutions in the district of Santa Marta. The study used CCR basic approach in DEA. Further, these contributors used the BCC model during the relative efficiency evaluation of the educational institutions in two groups.

Tsai and Chou (2015) employed DEA in order to facilitate the determination of operational efficiency after Enterprise Resource Planning system has been installed and operationalized. This contribution advocated in using the multiple outputs, patents, and net sales which were inputted to the DEA. In the banking industry, the input-oriented CCR model and the super efficiency model were used to identify banking efficiency (Ab Rahim 2015). The study pegged to figure out the technical efficiency of the banks and the super efficiency index of the Malaysian commercial banks from the year 2000 to 2010. Mitrovic, Vujosevic, and Savic (2017) used Data Envelopment Analysis in evaluating the efficiency of a healthcare system in Serbia and other countries in Europe. The research helped the public health practitioners to measure the efficiency of the health care services. Narimatsu et al. (2015) used DEA to measure the feasibility for predicting obesity and establishing effective risk models for

the individuals susceptible to obesity. The study calculates the efficiency scores and assesses the effectiveness of the risk models. Fixler, Paradi, and Xiaopeng (2014) used DEA to measure the performance efficiency of the Canadian hospital. It helps to attain the high standard and increase the quality standards of health care services in Canada.

Yang (2014) postulated that among the most regulated industries in the developed economies in the world is the banking industry. In Yang's (2014) contribution, a slacks-based DEA model was utilized to measure the bank's efficiency and risk regulation settings. DEA was applied to assess and evaluate the performance of the profit-making efficiency and risk-controlling efficiency. Kazan and Kitapci (2013) used DEA as a tool for measuring the performance of the Banking industry in Turkey when it comes to productivity and efficiency. Kazan and Kitapci (2013) employed the input-oriented Data Envelopment Analysis in measuring and identifying the bank's efficiency. There were 13 out of 21 DMUs labeled to be inefficient banks and 8 were identified below the efficiency level. The results of the study showed that inefficient banks need to consider the number of personnel, number of branches, their interest, and non-interest expenses to be an efficient bank. Chhikara and Rani (2012), on the other hand, used DEA in measuring the performance of the public banks in terms of their financial capabilities. The study analyzed the efficiency of the 26 PSBs operational processes of those with "variable-return-to-scale" on 2010-2011 in India. Hu et al. (2011) used an output-oriented meta-frontier DEA method in their study which has a focus on the Taiwan's regional police forces' relative efficiency. Their study has pegged and advocated in developing and built an appropriate DEA model to measure the efficiency. Liu and Liu (2010) examined how to improve the performance of the higher private educational universities and institutions. It utilized benchmarking processes, and the study determined the overall efficiency performance of the institutions and universities. Estrada et al. (2009) demonstrated a research focusing on a dynamic method in benchmarking to identify and measure the Canadian Bank branches' inefficient decision-making units, thereby improve the performance and efficiency. The literature cited significant improvement of those identified inefficient DMUs.

Data Mining has been considered that it can also be used to identify crime patterns. Reich and Porter (2015) had shown in the study the model-based clustering analysis that helped the investigators in the identification of crime series and eventually classified the type of crimes. In order to cluster the unlawful acts of the criminal being utilized by the crime investigator or analyst, there were lists of potential suspects of those unresolved crimes. The study also provided and determined the group of crimes or suspects of the category of crimes or the same flock of individuals having had committed the crime. The study provided a demographic and criminal profile for forecasting crime incidents or crime events. Tayebi et al. (2014) focused on spatial crime analysis that focused on crime hotspot areas with disproportionality in which location has higher crime density. The crime tracer can be used in all other method locations in crime prediction. Renushe, Joshi, and Rasal (2011) highlighted the importance of data mining techniques or methods applied in crime and criminal investigation to lessen or decrease the number of incidences. The data mining technique used was time series analysis to forecast crime. Groff and La Vigne (2002) determined the various current methods used in predicting crime mapping from the basic approach to sophisticated models. The study assessed the forecasting approach used to describe or identify the measure in predicting future crime hotspots. Bini and Mathew (2016) provide an analysis system to help the people in the stock market to predict the future value and identify the most financial key player companies by employing data mining. The regression analysis and clustering techniques were used in prediction. Baser and Saini (2013) research focus was on the exploration of various clustering techniques in data mining and its applications in various domains. The study discussed the various pros and cons of different clustering techniques. The study focused on comparative analysis of various clustering techniques. K-means algorithm was used and resulted in a nearby cluster as compared to others. Ferreira and Zhao (2015) aimed to determine groups that have similar time series in a database. The study proposed the use of time series clustering by using the community detection in a complex type of networks. Okamoto (2015) focused on clustering methods that identified the effective Life Cycle Assessment (LCA) system boundary and investigated the instability of the clustering methods. Wu, Duan, and Du (2015) proposed a study for medical experts diagnosing diseases using the multiple fuzzy c-means which are amongst the algorithms that can be used in medical diagnosis. Fuzzy c-means clustering method is the best common mechanism for medical diagnosis that helps the medical practitioners to formulate a treatment plan by obtaining useful and accurate information of the patient's diagnostic data. Banu and Gomathy (2014) used classification, clustering, and association rule mining to

predict the outcomes in heart diseases of the health care industry. Ni et al. (2014) used a classification and clustering data mining techniques for energy customers. A dataset of 350 Swedish houses applied a customer classification and clustering tasks on actual electricity consumption. The finding of the study presented the comparison results and examined the added value of the high-frequency measurements, e.g. 10-minute measurements, in terms of its influence on customer classification and clustering.

Kaur and Krishan (2013) analyzed the student behavior in the E-learning environment using the clustering techniques. The clustering techniques analyzed the relationship between the usage of the course and the student's performance. The data mining techniques were used to analyze the behavior of students using e-learning data from Greek University. Peckham and McCalla (2012) study utilized k-means clustering method to determine the students' behavior patterns being extracted.

K-means clustering was integrated with the Blooms Taxonomy in order that the positive or negative cognitive skills set will be determined with the reading comprehension task wherein positive cognitive skills mean good grade and negative cognitive skills mean poor grade. Maimon and Rokach (2005) presented a tutorial overview of the main clustering methods used in data mining. The clustering methods are presented in performing large sets and how to determine the number of clusters.

All the above-cited DEA literature is similar to the proponents study wherein the study aim is to assess and evaluate the different decision-making units using DEA model but it focused on efficiency analysis of the DMUs and did not use other tools to enhance the performance of the organizations. The proposed study will also be compared to other DEA models such as a non-radial measure of efficiency and hybrid model to identify the model that best determines the efficiency of the Philippines National Police-District VI in the province of Cavite.

The proponent's study is similar to the mentioned scholarly works on Data Mining, wherein the study also aimed to identify crime patterns and trends. The data mining literature did not measure the efficiency of the decision-making units before mining the data of the organizations.

In the existing literature of data envelopment analysis, no research study was undergone with data mining techniques at the same time. In the existing literature of data mining, no research study applied data envelopment analysis.

The proponent combined the concepts of measuring the efficiency of the organizations using data envelopment analysis and data mining using clustering analysis to enhance the performance of the Philippines National Police District VI in the province of Cavite.

In data envelopment analysis, the Philippines National Police District VI in the Province of Cavite will utilize their organizational resources to be more efficient in terms of increasing the number of crime cleared and crime solved in relation to the total crime incidence. Data mining will help the Philippines National Police District VI in the Province of Cavite to identify crime patterns to develop comprehensive crime prevention programs to countervail the crime for the security of an individual, community, and the State.

METHODOLOGY

Research Design

The researcher embarked on using the DEA in identifying the efficiency of the four (4) police stations being referred to as DMUs. Computation is as follows:

$$\text{Performance} = \frac{\text{virtual output}}{\text{virtual input}} = \frac{u_1 y_{1o} + \ldots + u_s y_{so}}{v_1 x_{10} + \ldots + v_m x_{mo}} \tag{1}$$

Where
x and y are resp. the input and output vectors and
u_s output$_s$'s weight, v_m output$_m$'s weight.

Data Clustering is a fundamental task in data mining. There are many clustering techniques and algorithms but the research study will use the partition as the clustering technique and the k-means method as the clustering algorithm.

Software Engineering and Processes

A Rational Unified Process was used as a refined software development and software Engineering Process. The Rational Unified Process follows a disciplined approach in task assignments and the roles and responsibilities in the software development organization. The end-goal of RUP Model is to deliver a value, high-quality software which meets and has the fitness for use of the end-users. The RUP consists of 4 phases, the Inception Phase, Elaboration, Construction, and the Transition Phase.

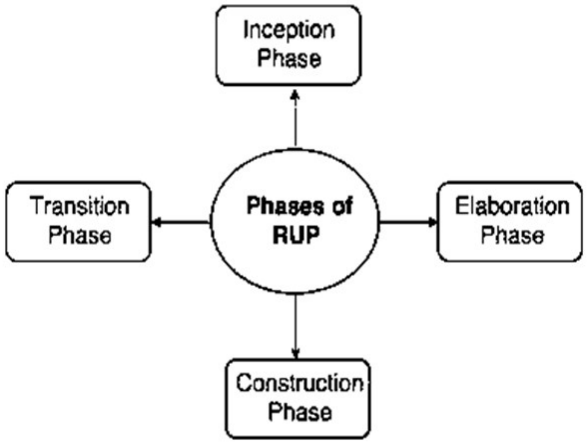

Figure 1. The Rational Unified Process (RUP) model (Kruchten 2004)

Phases of Rational Unified Process (RUP)

Inception phase

In order to consider the end-users' requirement, a business case has been established; hence the scope was defined. Actors including all other entities have been identified. The study defines the nature and level of interaction of these entities and actors with those of the system or project at the high-level. In relation to the inception phase, the proponent gathered the data, and discussed the requirements and the scope of the user of the system.

Elaboration phase

In this phase, a problem domain is being analyzed in order to establish and devise a sound architectural rudimentary within the corollaries and context of the project. During this phase, a project plan is being developed and project risks are identified in order that it could be addressed along with other elements in the project ecosystem. The researcher will design and develop a working prototype and has had a prototype demonstration with the end-users of the system. A project plan was designed with project tasks, dependencies, milestones, deliverables, and the resources.

Construction phase

Based on the prototype, all components with the application features will be developed. The development will be patterned based on the project plan, the timeline and scope, and the deliverable that has to be done during the development. It includes the database design, application front end, unit testing, module testing, integration testing, and system testing.

Visual C# will be used as the programming language for the application development.

Transition phase

Purposely, the transition phase is installation and making the application available at the user's end. It includes user training, fixes for bugs, corrections, and some preventative measures. Because the application will be made available to the user community, the development team may develop new versions, schedule version release, and upgrade the application. The purpose of the transition phase is to transition the software product to the user community. Once the product has been given to the end user, issues usually arise that require you to develop new

releases, correct some problems or finish the features that were postponed. In relation to the transition phase, the proponent conducted system training, turned over a system manual, and assisted in the application deployment and installation.

Other Tools Used in the Research Study

The researcher will use several technologies during the development of the software application for Crime Management System (CriMS). Microsoft.net 4.5 will be used as the environment framework and ASP.NET MV5 as the development framework. Microsoft Visual C# will be used as the programming language. SMS technology will also be used to integrate the mobile application in software application.

RESULTS AND DISCUSSION

Based on Table 1, Municipality C got a score of 1.0 for the years 2014-2016. Therefore, Municipality C is the efficient DMU.

Table 1: 3-year scale efficiency results of the Philippines national police-district VI in the province of cavite

DMU	2014		2015		2016		Average	Ranking
	Technical Efficiency Score	Scale Efficiency Score	Technical Efficiency Score	Scale Efficiency Score	Technical Efficiency Score	Scale Efficiency Score		
Municipality A	0.27	0.27	0.59	0.59	1.00	1.00	0.62	3
Municipality B	0.82	0.82	1.00	1.00	0.72	0.72	0.85	2
Municipality C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1
Municipality D	1.00	1.00	0.56	0.67	0.94	0.94	0.85	2

In the scholarly work of Hu et al. (2011), an output-oriented meta-frontier Data Envelopment Analysis method was used to measure the relative efficiency of regional police agencies in Taiwan to assess the technical efficiency of 23 regional police agencies.

Table 2: District VI statistics on criminal cases (index crime) from January to December 2014

City/Municipality	Crime Vs Persons						Crime Vs Propoerty				Index Crimes		
	Mur	Hom	Phy Inj	Rape	Tot	Rob	Theft	Ra 6539	Pd 533	Tot	Total	Cleared	Solved
Municipality A	6	1	7	2	16	14	13	2	0	29	45	5	3
Municipality B	34	5	81	29	149	54	124	5	0	183	332	100	49
Municipality C	23	1	60	31	115	26	55	3	0	84	199	68	40
Municipality D	15	2	38	10	65	20	38	2	0	60	125	58	35

Table 2 presents the statistics on criminal cases of District VI from January to December 2014. Municipality A has 45 total index crimes with 5 cleared cases and 3 solved cases. Municipality B has 332 total index crimes with 100 cleared cases and 49 solved cases. Municipality C has 199 total index crimes with 68 cleared cases and 40 solved cases. Municipality D has 125 total index crimes with 58 cleared cases and 35 solved cases.

Table 3: District VI statistics on criminal cases (index crime) from January to December 2015

City/Municipality	Crime Vs Persons						Crime Vs Propoerty				Index Crimes		
	Mur	Hom	Phy Inj	Rape	Tot	Rob	Theft	Ra 6539	Pd 533	Tot	Total	Cleared	Solved
Municipality A	7	0	5	4	16	9	6	1	0	16	32	11	4
Municipality B	35	5	83	36	159	67	94	34	0	195	354	206	112
Municipality C	34	3	37	18	92	18	33	16	0	67	159	75	53
Municipality D	20	3	26	10	59	31	33	14	0	78	137	41	25

Table 3 presents the statistics on criminal cases of District VI from January to December 2015. Municipality A has 32 total index crimes with 11 cleared cases and 4 solved cases. Municipality B has 354 total index crimes with 206 cleared cases and 112 solved cases. Municipality C has 159 total index crimes with 75 cleared cases and 53 solved cases. Municipality D has 137 total index crimes with 41 cleared cases and 25 solved cases.

Table 4: District VI statistics on criminal cases (index crime) from January to December 2016

City/Municipality	Crime Vs Persons					Crime Vs Propoerty					Index Crimes		
	Mur	Hom	Phy Inj	Rape	Tot	Rob	Theft	Ra 6539	Pd 533	Tot	Total	Cleared	Solved
Municipality A	4	0	4	7	15	2	4	2	0	8	23	4	10
Municipality B	42	8	75	29	154	80	120	50	0	250	404	79	47
Municipality C	37	3	20	27	87	9	25	10	0	44	131	53	33
Municipality D	13	3	40	10	66	37	73	16	0	126	192	57	36

Table 4 presents the statistics on criminal cases of District VI from January to December 2016. Municipality A has 23 total index crimes with 4 cleared cases and 10 solved cases. Municipality B has 404 total index crimes with 79 cleared cases and 47 solved cases. Municipality C has 131 total index crimes with 53 cleared cases and 33 solved cases. Municipality D has 192 total index crimes with 57 cleared cases and 36 solved cases.

Table 5: 3-year crime efficiency results of the Philippines national police- district VI in the province of Cavite

Province/City	Crime Efficiency For 2014					Crime Efficiency For 2015					Crime Efficiency For 2016				
	Total Crime Vol	Total Crime Cleared	Total Crime Solved	Crime Clearance Sol Eff	Crime Sol Eff	Total Crime Vol	Total Crime Cleared	Total Crime Solved	Crime Clearance Sol Eff	Crime Sol Eff	Total Crime Vol	Total Crime Cleared	Total Crime Solved	Crime Clearance Sol Eff	Crime Sol Eff
Municipality A	45	5	3	11.11	6.67	32	11	4	34.38	12.50	23	4	10	17.39	43.48
Municipality B	332	100	49	30.12	14.76	354	206	111	58.19	31.36	404	79	47	19.55	11.63
Municipality C	199	68	40	34.17	20.10	159	75	53	47.17	33.33	131	53	33	40.46	25.19
Municipality D	125	58	35	46.40	28.00	137	41	25	29.93	18.25	142	57	36	40.14	25.35
Total	701	231	127	32.95	18.12	682	333	193	48.83	28.30	700	193	126	27.57	18.00

Table 5 presents the 3-year crime efficiency results of the Philippines National Police District VI in the Province of Cavite. Crime clearance solution efficiency is determined by dividing total crime cleared over total crime volume. On the other hand, crime solution efficiency is determined by dividing total crime solved over total crime volume. For the year 2014, the crime clearance solution efficiency is 32.95 and crime solution efficiency is 18.12. For the year 2015, the crime clearance solution efficiency is 48.83 and crime solution efficiency is 28.30. For the year 2016, the crime clearance solution efficiency is 27.57 and crime solution efficiency is 18.00.

Table 6: Clustering characteristics of municipality C by crime category years 2014-2016

Variables	Population (All)					
	2014		2015		2016	
	Values	Probability	Values	Probability	Values	Probability
Crime Category	Physical Injuries	31%	Physical Injuries	26%	Murder	29%
Crime Category	Theft	24%	Murder	21%	Rape	23%
Crime Category	Rape	16%	Theft	18%	Theft	18%
Crime Category	Murder	14%	Robbery	11%	Physical Injuries	14%
Crime Category	Robbery	12%	Rape	10%	Robbery	9%
Crime Category	Carnapping	2%	Motornapping	7%	Motornapping	4%
Crime Category	Homicide	1%	Carnapping	5%	Carnapping	1%
Crime Category			Homicide	2%	Homicide	1%

Table 6 shows the clustering characteristics of Municipality C by crime category for years 2014-2016. The summary shows that Physical Injuries ranked the highest in the year 2014 and 2015 with 31% and 26% probability while Murder ranked the highest in the year 2016 with 29% probability.

Table 7: Clustering characteristics of municipality C by Barangay year 2014-2016

Population (All)						
Variables	2014		2015		2016	
	Values	Probability	Values	Probability	Values	Probability
Barangay	Barangay I	17%	Barangay I	13%	Barangay I	25%
Barangay	Barangay II	14%	Barangay II	13%	Barangay VII	11%
Barangay	Barangay III	8%	Barangay VI	9%	Barangay III	11%
Barangay	Barangay IV	8%	Barangay III	8%	Barangay II	8%
Barangay	Barangay V	7%	Barangay X	7%	Barangay VI	8%
Barangay	Barangay VI	6%	Barangay VIII	7%	Barangay V	7%
Barangay	Barangay VII	6%	Barangay VII	6%	Barangay XIII	5%
Barangay	Barangay VIII	5%	Barangay IV	6%	Barangay IV	5%
Barangay	Barangay IX	4%	Barangay XVI	4%	Barangay XVII	3%
Barangay	Barangay X	4%	Barangay XVII	4%	Barangay XV	3%
Barangay	Barangay XI	4%	Barangay V	4%	Barangay IX	3%
Barangay	Barangay XII	3%	Barangay XII	4%	Barangay X	3%
Barangay	Barangay XIII	3%	Barangay XI	3%	Barangay VIII	2%
Barangay	Barangay XIV	3%	Barangay XIII	3%	Barangay XI	2%
Barangay	Barangay XV	2%	Barangay IX	3%	Barangay XII	1%
Barangay	Barangay XVI	2%	Barangay XIV	3%	Barangay XIV	1%
Barangay	Barangay XVII	2%	Barangay XVIII	2%	Barangay XVI	1%
Barangay	Barangay XVIII	2%	Barangay XV	2%		

Table 7 shows the clustering characteristics of Municipality C by barangay year 2014-2016. The summary shows that Barangay I ranked the highest in the years 2014-2016 with 17%, 13%, and 25% probability.

Table 8: Clustering characteristics of municipality C by season (month) for years 2014-2016

Population (All)						
Variables	2014		2015		2016	
	Values	Probability	Values	Probability	Values	Probability
Month Comttd	December	17%	September	12%	September	23%
Month Comttd	April	13%	April	12%	August	12%
Month Comttd	June	9%	November	11%	May	11%
Month Comttd	January	9%	August	10%	March	9%
Month Comttd	February	9%	July	9%	July	8%
Month Comttd	March	8%	February	8%	April	8%
Month Comttd	August	6%	January	8%	February	7%
Month Comttd	November	6%	March	8%	October	7%
Month Comttd	July	6%	December	7%	December	5%
Month Comttd	October	6%	October	6%	January	4%
Month Comttd	September	6%	May	5%	June	4%
Month Comttd	May	5%	June	4%	November	3%

Table 8 shows the clustering characteristics of Municipality C by season (month) for years 2014-2016. The summary shows that December ranked the highest in the year 2014 with 17% probability and September in the year 2015-2016 with 12% and 23% probability.

Table 9: Clustering characteristics of municipality C by season (Day) for years 2014-2016

Population (All)						
Variables	2014		2015		2016	
	Values	Probability	Values	Probability	Values	Probability
Day Comttd	Sunday	20%	Wednesday	22%	Sunday	30%
Day Comttd	Tuesday	19%	Saturday	17%	Thursday	18%
Day Comttd	Saturday	16%	Sunday	17%	Wednesday	15%
Day Comttd	Wednesday	14%	Monday	17%	Friday	14%
Day Comttd	Thursday	11%	Friday	11%	Saturday	11%
Day Comttd	Monday	11%	Tuesday	10%	Tuesday	5%
Day Comttd	Friday	9%	Thursday	6%	Monday	5%

Table 9 shows the clustering characteristics of Municipality C by season (day) for years 2014-2016. The summary shows that Sunday ranked the highest in the year 2014 and 2016 with 20% and 30% probability and Wednesday in the year 2015 with 22% probability.

Scholarly works of Yamuna and Bhuvaneswari (2012), Yu et al. (2011), Herrera, Sosa, and Delgado (2015), and Ceccato and Uittenbogaard (2014) used data mining for criminal and intelligence analysis.

Summary of the Results

DEA as a powerful, analytical technique for evaluating and measuring the performance of organizational units was used to assess and measure the efficiency of the four (4) police stations being referred to as DMUs in District VI in the province of Cavite. The efficiency of the four (4) police stations was measured using MAXDEA software in terms of organizational resources. Based on DEA, a DMU should obtain a score of 1.00 in order to be considered efficient. The result of Table 1.0 shows that only Municipality C got a score of 1.00 for three (3) years. Therefore, Municipality C is the efficient DMU.

In the academic criminology, the commission of a crime is accounted as sociological, psychological, biological or economic to present an explanation about a criminal's behavior (Becker 1968). The fundamental dimension of crime is both a behavioral and a social problem (Reckless 1967). Every violation of the law does not mean crime and is not counted as a crime. In essence, breaches of private law (torts and bridge of contract) are not automatically penalized by the state but can be imposed through civil procedure. The most prevalent crimes in the Philippines are 1) organized crime, 2) petty crime, 3) violent crime, 4) prostitution, 5) human trafficking, 5) corruption, 6) police misconduct, and 7) the illegal drug trade. Philippines National Police (PNP) has defined crime classification as index and non-index crimes. Index crimes involve 1) crimes against persons such as a) murder, b) homicide, c) physical injury, and d) rape, and 2) crimes against property such as a) robbery, b) theft, c) car-napping/carjacking commonly known as RA 6539, and d) cattle rustling commonly known as PD 533 while non-index crimes are violations of special and private laws such as local ordinances. Table 2.0, 3.0, and 4.0 presented the index crimes of the four (4) municipalities. The tables also presented the crimes cleared and crimes solved. Crimes cleared according to law enforcement lingo means a criminal suspect was identified and charge was made but the criminal suspect is at-large. Crimes solved, on the other hand, means that a criminal suspect was identified, charged, and arrested.

Crime solution efficiency is the percentage of solved cases out of the total number of crime incidents handled by law enforcement agencies. Table 5.0 presented the 3-year crime efficiency results of the Philippines National Police District VI in the province of Cavite. Among the four (4) municipalities, Municipality A obtains the highest crime solution efficiency of 43.48.

Data mining is a grouping of similar or dissimilar characteristics of objects. Data mining clustering technique using K-means algorithm was used to extract crime pattern recognition. Table 6, 7, 8, and 9 presented

the crime patterns by crime category, barangay, month, and day. Based on the crime patterns extracted, crime detectives and law enforcement agencies can develop comprehensive crime prevention programs, to improve the crime prevention and crime efficiency, for a safe and secure environment and strong commitment and leadership of the local government for peace and order.

CONCLUSION, RECOMMENDATIONS AND IMPLICATIONS

This study aimed to assess and evaluate the efficiency of the Philippines National Police- District VI in the province of Cavite. The proponent applied an input-oriented radial measure of efficiency and RTS using scale efficiency to determine the efficient DMUs. Based on the 3-year scale efficiency result, Municipality C is the efficient DMU. A clustering data mining technique of MS SQL Server 2012 using K-means algorithms were applied to extract data from the DMUs that have high crime instance but a smaller percentage of crime cleared and crime solved, hence patterns were extracted and identified. Application software that can be used by the Philippines National Police-District VI in the province of Cavite to develop a comprehensive crime prevention program based on the pattern identified is developed and tested.

REFERENCES

Ab Rahim, R. 2015. "Ranking of Malaysian Commercial Banks: Super-Efficiency Data Envelopment Analysis (DEA) Approach." ***Asian Academy of Management Journal of Accounting & Finance*** 11(1): 123-143.

Banu, M. N., and Gomathy, B. 2014. "***Disease Forecasting System Using Data Mining Methods.***" Presented at the International Conference on Intelligent Computing Applications (ICICA), (pp. 130-133), March 06-07 Coimbatore, India.

Baran, J., Wysokinski, M., Stas, D., Samolejova, A., and Lenort, R. 2016. "Efficiency of Polish Metallurgical Industry based on Data Envelopment Analysis." ***Metalurgija*** 55(2): 245-248.

Baser, P., and Saini, J. R. 2013. "A Comparative Analysis of Various Clustering Techniques used for Very Large Datasets." ***International Journal of Computer Science & Communication Networks*** 3(5): 271-275.

Becker, G. S. 1968. "Crime and Punishment: An Economic Approach." ***Journal of Political Economy*** 76(2): 169-217.

Bini, B. S., and Mathew, T. 2016. "Clustering and Regression Techniques for Stock Prediction." ***Procedia Technology*** 24: 1248-1255.

Boussofiane, A., Dyson, R. G., and Thanassoulis, E. 1991. "Applied Data Envelopment Analysis." ***European Journal of Operational Research*** 52(1): 1-15.

Caballero, E. R., Cadavid, V. D., and Toloza, A. J. M 2015. "Efficiency in the Public Educational Institutions of the City of Santa Marta (Colombia) Through Enveloping Data Analysis." ***Ingeniare, Chilean Engineering Journal*** 23(4): 579-593.

Ceccato, V., and Uittenbogaard, A. C. 2014. "Space-time Dynamics of Crime in Transport Nodes." ***Annals of the Association of American Geographers*** 104(1): 131-150.

Chakraborty, J. 2016. "Efficiency Analysis of Indian Life Insurance Firms: A DEA Investigation." ***SCMS Journal of Indian Management*** 13(3): 5-24.

Charnes, A., Cooper, W. W., and Rhodes, E. 1978. "Measuring the Efficiency of Decision Making Units." ***European Journal of Operational Research*** 2(6): 429-444.

Chhikara, K. S., and Rani, S. 2012. "Data Envelopment Analysis: A Study Measuring Efficiency of Public Sector Banks in India by using Data Envelopment Analysis: A Study." ***Journal of Institute of Public Enterprise*** 35(3/4): 2-15.

Devi, P. A., and Rajagopalan S. P. 2012. "Analysis of Customer Behavior using Clustering and Association Rule." ***International Journal of Computer Applications (IJCA)*** 43(23): 19-26.

Estrada, S. A., Song, H. S., Kim, Y. A., Namn, S. H., and Kang, S. C. 2009. "A Method of Stepwise Benchmarking for Inefficient DMUs based on the Proximity-based Target Selection." ***Expert Systems with Applications*** 36(9): 11595-11604.

Ferreira, L. N., and Zhao, L. 2015. "A Time Series Clustering Technique based on Community Detection in Networks." ***Procedia Computer Science*** 53: 183-190.

Fixler, T., Paradi, J. C., and Yang, X. 2014. "A Data Envelopment Analysis Approach for Measuring the Efficiency of Canadian Acute Care Hospitals." ***Health Services Management Research*** 27(3-4): 57-69.

Francisco, K. 2016. "PNP: Quezon City has the Highest Number of Index Crimes." Rappler. Retrieved 19 June 2016. (https://goo.gl/lCSSNr).

Groff, E. R., and La Vigne, N. G. 2002. "Forecasting the Future of Predictive Crime Mapping." ***Crime Prevention Studies*** 13: 29-58.

Herrera, F., Sosa, R., and Delgado, T. 2015. "***GeoBI and Big VGI for Crime Analysis and Report.***" Presented at the 3rd International Conference on Future Internet of Things and Cloud (FiCloud), (pp. 481-488), August 24-26 Rome, Italy.

Hu, J. L., Tung, Y. Y., Shieh, H. S., and Lo, K. 2011. "A Meta-Frontier Efficiency Analysis of Regional Police Agencies in Taiwan." ***Journal of Management Research*** 11(3): 168-182.

Izhar, T. A. A., Baharuddin, M. F., Mohamad, A. N., ... Wan Hasnol, W. M. H. (2016). "Using ontology for goal-based query to evaluate social media data." ***Journal of Advances in Humanities and Social Sciences*** 2(2): 108-118.

Kaur, E., and Krishan, K. 2013. "Cluster Analysis of Behavior of E-learners." ***International Journal of Soft Computing and Engineering (IJSCE)*** 3(2): 344-346.

Kazan, H., and Kitapci, H. 2013. "Performance Measurement with Data Envelopment Analysis in Service Industry: Banking Application." ***Business Management Dynamics*** 3(5): 37-50.

Kinaci, H., Najjari, V., and Alp, I. 2016. "Using Data Envelopment Analysis and Stochastic Frontier Analysis Methods to Evaluate Efficiency of Hydroelectricity Centers." ***Gazi University Journal of Science*** 29(1): 167-176.

Kruchten, P. (2004). ***The Rational Unified Process: An Introduction.*** Boston, MA: Addison-Wesley Professional.

Liu, C. Y. A., and Liu, W. H. 2010. "***Performance Evaluation on Private Higher Education using Data Envelopment analysis.***" Presented at the Industrial Engineering Research Conference, December 07-10 Macao, China.

Maimon, O., and Rokach, L. 2005. ***Data Mining and Knowledge Discovery Handbook.*** New York, NY: Springer.

Mitrovic, Z., Vujosevic, M., and Savic, G. 2017. "Data Envelopment Analysis for Evaluating Serbia's Health Care System." ***Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*** 20(75): 39-46.

Narimatsu, H., Nakata, Y., Nakamura, S., Sato, H., Sho, R., Otani, K., ... and Yamashita, H. 2015. "Applying Data Envelopment Analysis to Preventive Medicine: A Novel Method for Constructing a Personalized Risk Model of Obesity." ***PloS One*** 10(5): 1-11.

Ni, Y., Engstrom, C., Malyarenko, A., and Wallin, F. 2014. "***Investigating the Added Values of High Frequency Energy Consumption Data using Data Mining Techniques.***" Presented at the AIP Conference Proceedings (pp. 734-743), July 15-18, Narvik, Norway.

Norman, M., and Stoker, B. 1991. ***Data Envelopment Analysis: The Assessment of Performance.*** New York, NY: John Wiley & Sons, Inc.

Okamoto, S. 2015. "Analyzing Instability of Industrial Clustering Techniques." ***Environmental Economics and Studies*** 17(3): 389-406.

Peckham, T., and McCalla, G. 2012. "***Mining Student Behavior Patterns in Reading Comprehension Tasks.***" Presented at the proceedings of the 5th International Conference on Educational Data Mining. June 19, 2012 - June 21, 2012 Panorama Hotel, Chania, Greece.

Reckless, W. C. 1967. ***The Crime Problem.*** New York, NY: Appleton-Century-Crofts.

Reich, B. J., and Porter, M. D. 2015. "Partially Supervised Spatiotemporal Clustering for Burglary Crime Series Identification." ***Journal of the Royal Statistical Society: Series A (Statistics in Society)*** 178(2): 465-480.

Renushe, H. N., Joshi, M. J., and Rasal, P. R. (2010). "Importance of Data Mining Time Series Technique in Crime and Criminal Investigation: A Case Study of Pune Rural Police Stations." ***International Journal of Computer Applications*** 30(9): 38-42.

Tayebi, M. A., Ester, M., Glasser, U., and Brantingham, P. L. 2014. "***Crimetracer: Activity Space Based Crime Location Prediction.***" Presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), (pp. 472-480), August 17-20 Beijing, China.

Tran, T. T. H., and Bhaiyat, F. 2016. "***Evaluating the Efficiency of Vietnamese Commercial Banks using Data Envelopment Analysis.***" Presented at the 6th Annual International Conference on Accounting and Finance, May 30-31 Singapore.

Tsai, B. H., and Chou, S. B. 2015. "Application of Multiple Output Data Envelopment Analysis in Interpreting Efficiency Improvement of Enterprise Resource Planning in Integrated Circuit Firms." ***The Journal of Developing Areas*** 49(1): 285-304.

Vyver, A. G. D. V., and Gillies, D. 2017. "Analysing public perceptions of international events by using geo-located twitter data." ***International Journal of Humanities, Arts and Social Sciences*** 3(2): 64-70.

Wu, Y., Duan, H., and Du, S. 2015. "Multiple Fuzzy C-means Clustering Algorithm in Medical Diagnosis." ***Technology and Health Care*** 23(s2): S519-S527.

Yamuna, S., and Bhuvaneswari, N. S. 2012. "Data Mining Techniques to Analyze and Predict Crimes." ***The International Journal of Engineering and Science*** 1(2): 243-247.

Yang, C. C. 2014. "Evaluating the Performance of Banking Under Risk Regulations: A Slacksbased Data Envelopment Analysis Assessment Framework." ***Expert Systems*** 31(2): 176-184.

Yu, C. H., Ward, M. W., Morabito, M., and Ding, W. 2011. "***Crime Forecasting using Data Mining Techniques.***" Presented at the IEEE 11th International Conference on Data Mining Workshops (ICDMW), (pp. 779-786), December 11 Vancouver, Canada.