Forecasting Model for Criminality in Barangay Commonwealth, Quezon City, Philippines using Data Mining Techniques

Bertiz Armenico R, Vinluan Albert A, Laureta Marc P, Paltad Anthony B and Marques Westley
College of Computer Studies
New Era University
Quezon City Philippines
nicolocco05@gmail.com and aavinluan@neu.edu.ph

Abstract— **Crime is an act or a behaviour that is forbidden and punishable by law and its occurrences are problems that the society wants to avoid. Crimes are actually expected to happen to people who live in a city or state because of a popularity or the high rate within the area like Barangay Commonwealth, Quezon City. In order to prevent crime occurrences, predictions are can be performed. Prediction of crime occurrences may lead to police enforcement to make tactical preparation that will help to solve crimes and avoiding it to happen. With the arising world of technology and large amount of data recorded in police deparments, data mining techniques and algorithm are the processes used to forecast crime occurrences. Data mining is a computer assisted way of finding the meaning of patterns within large amount of data. The researchers extracted raw data from blotter crime records from Quezon City Police District so the researchers created a datasets where data mining techniques and algorithm are applied. Different Data Mining techniques such as regression analysis, Neural Network (1 Hidden Layer) and Decision Tree (ID3) are used to apply in the datasets and remarkably resulted accurate prediction of crime occurrences.**

Keywords-Crime forecasting, Data Mining, Time series analysis, Regression analysis, Neural Network, Decision tree

I. INTRODUCTION

Crime is a forbidden act punishable by law. It is an expected event that no one can say when it will happen. There are many reasons for crimes to happen: poverty, drugs, alcoholism and many others that may affect human behaviour. Crime can occur in any community.

Police forces are responsible for law enforcements to prevent crime, protect the community and property by targeting of resources [2]. They use tactical and strategic rules of preventive measures against crime based on criminological theories and crime analysis.

Another method for preventing crime is by forecasting future events like crime trends to improve crime prevention and reduction measures [3]. The major targets of these tactics are persons and their crminality. For example analysis of modus operandi or serial criminals lead to their apprehensions

[4]. Crime forecasting helps the authorities to take strategic actions such as targeting hotspots, and deploying special units. This can be also used for police trainings.

Forecasting can be performed easily with the advancements of technology. For data that are previously collected, data mining is the primary key for accumulating crime records, computer assisted digging and analyzing enormous sets of data. Temporal data mining is a data mining method for temporal data primarily to find data patterns with respect to time. Maps are also used to develop practical crime prevention solutions corresponding to specific time and place [5].

There is no single technique for data mining. Data mining should be thought as a set of tools to extract information from large amounts of data. It is fairly common to experiment with a number of algorithms and parameters to select the best modelling approach [6]. Some available techniques for data mining are: Classification, Clustering, Segmentation, Decision Tree and Neural Networks.

II. STATEMENT OF THE PROBLEM

The researchers aimed to demonstrate the benefits of this research to the authorities for forecasting crimes in Barangay Commonwealth. Specifically this research sought to answer these questions:

- What are the predictors to consider in forecasting crimes?

- What are the relationships of these predictors to the occurrence of crimes

- What Data Mining Techniques and algorithm to be used in developing the forecasting model for occurrence of crimes

What is the level of acceptance for developing forecasting model?

III. CONCEPTUAL FRAMEWORK

The Figure 1.0 depicted a sequential and iterative data

mining techniques and algorithm stage of developing the study. The process starts with the gathering of data, the data came from Quezon City Police District Station 6 which the researchers gathered and also managed to get all of the needed flat files or the daily crime report that only occurred in the area of Barangay Commonwealth. Blotter record was in a format of free text, means that it was in the structure of sentence.
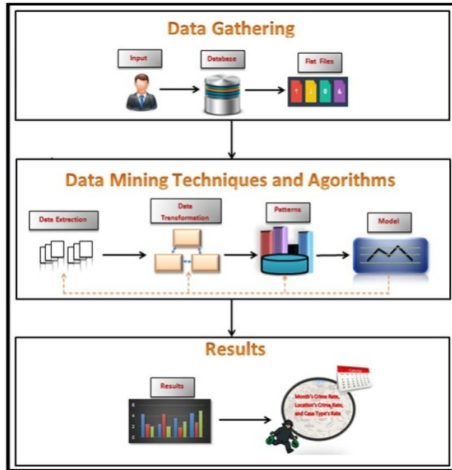

Figure 1.0: Conceptual framework

Data extraction comes after the data gathering. The researchers remove the unnecessary fields not needed in forecasting. This process is time consuming since the fields have to be properly formatted, and edited for missing, null, unidentified data entry and typographical errors.

The process continues with data transformation for time series analysis followed by finding patterns in the data using data mining softwares such as regression analysis, neural network and decision tree. These are then used to develop the forecasting model. Predictions are then used to compute monthly rates and compared with actual crime occurrence using the records for the year 2012.

IV. METHODOLOGY

This study used Constructive Research which encompasses the area of theory not requiring research based on solidity. This kind of research is lined upon testing theories, hypotheses or case studies.

Data mining is a computational process of discovering patterns in large data sets involving the three methods that intersect with each other. The researchers' chosen few data mining techniques which are related to the previous study and those data mining techniques are Regression Analysis, Neural Network (1 Hidden Layer), and Decision Tree [10].

For measuring the correlation of forecasting crime occurrence's predictors towards each other, the researchers used Bivariate Correlation also known as Bivariate Analysis. And for statistical tool, the researcher will used Exponential Smoothing, Multiple Linear Regression, Mean Absolute Percentage of Error (MAPE),

The researcher also used software packages such as IBM SPSS Amos and Rapid Miner.

V. RELATED WORKS

This study explored a methodology to reliably predict location, time, and/or likelihood of future residential burglary. From original crime records, architectural datasets are generated. These datasets contain aggregated counts of different types of crimes and related events as categorized by the city's police department. Spatial and temporal information of different types of crime are embedded in these architected datasets. Each original data entry is a record for an individual crime or related event, the location in longitude and latitude, and the time and date of the incident [7] .

The details about each reported crime incident are entered into numerous crime databases controlled by various government and law enforcement agencies. These databases contain rich data about all crime activities across the country. Data include location, time, description of the perpetrators and victims. With these information, citizens could infer the degree of habitability of a particular area. This study aimed to apply myriad of tools to mine for information that is otherwise hidden from human perception [8] .

Much work on crime analysis have been reported. These are mostly confined to predicting crime occurrence based on location frequently chosen by criminals [9] . In addition, analyses are done on frequencies of occurrences by type of crime, time committed, and social status of people involved, crime links, etc. These are information help law enforcers to predict crime and take necessary measures.

VI. RESULTS AND DISCUSSION

The researchers presented, analyzed and interpreted every data produced by the models for forecasting crime occurrences and also to elaborate on how the researchers came up with having the right algorithm and data mining to be used in the study.

PREDICTORS FOR FORECASTING CRIME OCCURRENCESS

Since that the study comprises with time series analysis, or methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Every phenomenon recorded according to the occurrence of crime was significant in order to forecast month's crime rate of Year 2012.

Therefore, the researchers did not do any detection of outliers for it may affect or became bias but instead, the researchers just do normalizing by detecting and correcting the misformatted values from data such as null, missing, unidentified, or misspelled [10].

The researchers had sum numbers of crime in Months for analyzing the sequence of observations [11].

The researchers considered as predictor the Years 2008, 2009, 2010, 2011 by being significant for time series analysis that was used in order to forecast crime occurrences because time series forecasting was the use of a model to predict future values based on previously observed values.

EXPONENTIAL SMOOTHING

Before measuring the correlations of the predictors with each other, the researchers transformed the datasets first by the use of Exponential Smoothing were the noises and intervals in the datasets are reduced. The time series data themselves were a sequence of observations. The observed phenomenon may be an essentially random process, or it may be an orderly, but noisy, process. Whereas in the simple moving average the past observations were weighted equally, exponential smoothing assigns it exponentially decreasing weights over time. Alpha was set or the smoothing factor to 0.5 because if it was set it higher than 0.5, the datasets would be over-weighted equally, exponential smoothing assigns exponentially decreasing weights over time [12].

Naturally, the least complex approach to smooth a time series is to figure a basic, or unweighted, moving average. This is known as utilizing a rectangular or "boxcar" window capacity. The smoothed measurement st is then simply the last's mean k observations:

$$s_t = \frac{1}{k}\sum_{n=0}^{k-1} x_{t-n} = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-k+1}}{k} = s_{t-1} + \frac{x_t - x_{t-k}}{k},$$

(1)

SIGNIFICANT RELATIONSHIP OF THE PREDICTORS

The researchers used IBM SPSS Amos' Bivariate Correlation function in order to determine if two or more variables were linearly related to each other. Bivariate Correlation is a function where it could produce an output called Correlation Box. Pearson Correlation, Significance (2-Tailed) and an

(N) values were the results that can be found inside the Correlation Box.

When Pearson's correlation is close to 1, this means that changes in one variable were strongly correlated with changes in the second variable. But if close to 0, the changes in one variable are weakly correlate with changes in the second variable.

If the Significance (2-Tailed) value is greater than 0.05, this means, increases or decreases in one variable do not significantly relate to increases or decreases in your second variable. But if close to less than or equal to 0.05, this means, increases or decreases in one variable do not significantly related to increases or decreases in your second variable.

Table 1.0: Sample Relationship of the Predictors (Month's Crime Rate Forecasting)

		@2008	@2009	@2010	@2011
@2008	Pearson Correlation	1	.518	.579*	.597*
	Sig. (2-tailed)		.084	.048	.041
	N	12	12	12	12
@2009	Pearson Correlation	.518	1	.327	.282
	Sig. (2-tailed)	.084		.300	.374
	N	12	12	12	12
@2010	Pearson Correlation	.579*	.327	1	.107
	Sig. (2-tailed)	.048	.300		.741
	N	12	12	12	12
@2011	Pearson Correlation	.597*	.282	.107	1
	Sig. (2-tailed)	.041	.374	.741	
	N	12	12	12	12
*. Correlation is significant at the 0.05 level (2-tailed).					

Table 1.0 presents the correlation output of month's crime rate's predictors using IBM SPSS Amos' Bivariate Correlation function.

Results shows for forecasting month's crime rate, the year 2009, 2010, and 2011 were strong correlated with changes towards each other except Year 2008, the Year 2008 towards 2009 and Year 2010 towards 2011 had no statistically significance with each other but the remaining years towards each other were statistically significant.

DATA MINING TECHNIQUES AND ALGORITHM

REGRESSION ANALYSIS

The researchers performed a function of IBM SPSS Amos called Regression Analysis that was applied on the datasets for forecasting crime occurrences such as month crime rate, location crime rate, and case type rate. By performing regression analysis, the researchers were able to focus on estimating the changes in relationship when independent variables Years 2008, 2009, 2010, and 2011 are varied, while dependent variable Year 2012 is held fixed [13].

Table2.0: Sample Coefficients of the Predictors (Month's Crime Rate Forecasting)

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	33.654	12.708		2.648	.033
	@2008	-1.922	.771	-1.120	-2.491	.042
	@2009	.217	.344	.191	.631	.548
	@2010	.948	.522	.617	1.815	.112
	@2011	.234	.367	.220	.637	.545
a. Dependent Variable: @2012						

Displayed in Table2.0, the coefficient output of month's crime rate's predictors using IBM SPSS Amos' Regression Analysis function. The 2008 predictor variable is negative (-), this means for every unit increase in 2008 coefficient, a unit decrease in the dependent variable 2012 was predicted,

holding all other variables constant. The coefficient for constant and predictor variable 2009, 2010, and 2011 is positive (+), this means for every unit increase in the said coefficients, a unit increased in the dependent variable 2012 was predicted, holding all other variables constant.

The B or Beta was the standardized coefficients. These were the coefficients that you would obtain if you standardized all of the variables in the regression, including the dependent and all of the independent variables, and ran the regression.

NEURAL NETWORK (1 HIDDEN LAYER)

To get the definition of a function or a distribution over dependent and independent variables, the researchers used a simple mathematical model in Rapid Miner called Neural Network (1 Hidden Layer). The implementation of neural network is almost identical with the application of regression analysis for it has the neural network was also applied in to the datasets for forecasting crime occurrences. Sigmoid function, biases, linear regression output, and threshold are the functions that were defined by neural network through receiving the input independent variables such as Year 2008, 2009, 2010, and 2011 that were summed to produce an output of dependent variable Year 2012 by varying parameters, connection weights, or specifics of the architecture such as the number of neurons or their connectivity [14].
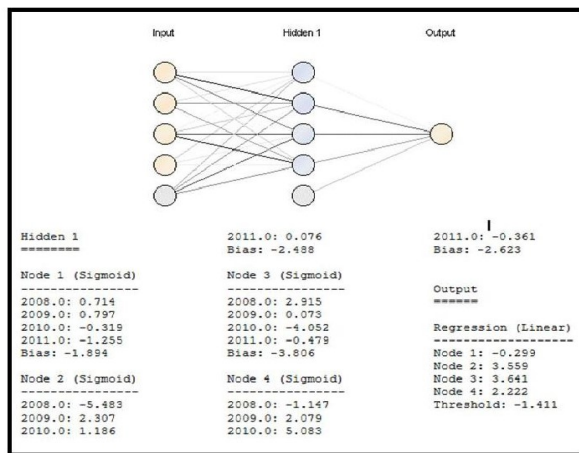

Figure 2.0: Sample Neural Network's Output (Month's Crime Rate Forecasting)

Presented in the Figure 2.0is the figural and computational output of Rapid Miner's Neural

Network applied in the dataset for forecasting month's crime rate of the Year 2012. Figural output was where the illustration of the nodes and hidden patter are can be seen. While computational output was composed of the result of sigmoid function and bias for each hidden node and regression output and its threshold for output node. The regression output showed the results of the nodes for neural network were Node 1 or Year 2008 is -0.299, Node 2 or Year 2009 is 3.559, Node 3 or Year 2010 is 3.641, Node 4 or Year 2011 is 2.222, and Threshold value of -1.411.

DECISION TREE (ID3)

Before proceeding into the application of decision tree, the researchers clustered the number of crime in the datasets for forecasting crime occurrences into hi and low rate by using k-means clustering. The prediction through decision tree had a different approach. The prediction pertains to what class (hi or low) on where the outcome was.

Since that the datasets were clustered in to two classes, the researchers used Rapid Miner's Decision Tree function into it. The Decision tree learning used a decision tree as a predictive model which maps observations about an item to conclusions about the item's target value. There were many specific decision-tree algorithms notable ones includes ID3. The researchers selects the independent variable Years 2008, 2009, 2010, and 2011 as attributes in to the ID3 which iterates and analyze every unused attributes and selected the smallest entropy and partitioned the subsets until it found the best attributes for concluding the class outcome for the target item or the dependent variable Year 2012 [15].
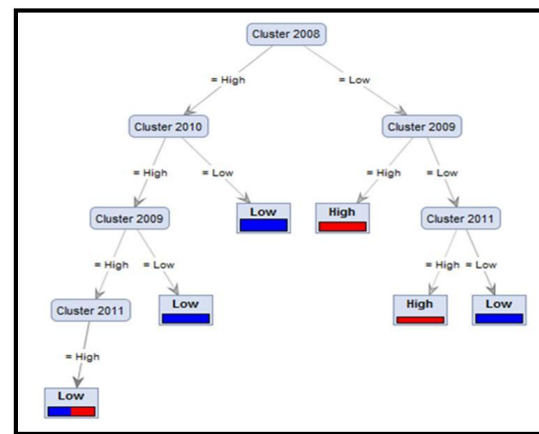

Figure 2.1: Sample Decision Tree Output (Month's Crime Rate Forecasting)

Figure 2.1 shows the root node for forecasting Year 2012's month's crime rate was Year 2008, which burst nodes downward by having a pointed line to leaf nodes such as Years 2011, 2010, and 2009. And for every end of each leaf nodes, the outcome of classification can be found.

LEVEL OF ACCEPTABILITY OF THE FORECASTING MODEL

The level of acceptability was set to 95.00% percent of accuracy of the forecasting model, if the accuracy of the forecasting model was greater than or equal to 95.00% meaning the model was used was acceptable and if lower than 95.00% vice versa the forecasting model was not acceptable.

MEANABSOLUTE PERCENTAGE ERRORAND SYMMETRIC MEAN ABSOLUTE PERCENTAGE ERROR

The researchers used MAPE for measuring the accuracy of Regression Analysis and Neural Network (1 Hidden Layer)

for forecasting month's crime rate of Year 2012 for it has no zero values in its actual and forecasted data [16].

$$M = \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{At - Ft}{At} \right\| \qquad (2)$$

ACCURACY OF THE DATA MINING MODELS

Table 3.0: Accuracy and Error of Regression Analysis and Neural Network (1 Hidden Layer)

Data Mining Model	Forecasting Month's Crime Rate of Year 2012	
	Accuracy	Error
Regression Analysis	97.75%	2.24%
Neural Network (1 Hidden Layer)	99.92%	0.07%

Table 3.0 appears that the accuracy and error of Regression Analysis and Neural Network (1 Hidden Layer) are acceptable for forecasting crime occurrences for having accuracies that were definitely higher than 95.00% which is the accepted level of accuracy.

Since the Decision Tree had its own algorithm in solving the accuracy of the model in accordance with precision and recall, the researchers present the table of model's performance below:

	Accuracy	Precision	Recall
Forecasting Month's Crime Rate of Year 2012	91.67%	100.00%	75.00%

Figure 4.0: Accuracy, Precision, and Recall of Decision Tree

Table shows 4.0 shows the accuracy, precision, and recall of Decision Tree was not acceptable for forecasting month's crime rate that only got 91.67% of accuracy which was definitely lower the 95.00% level of acceptability.

BEST FIT FOR CRIME OCCURRENCES FORECASTING

The researchers compared all the accuracy and saw the best fit data mining techniques that were used in the datasets to forecast the rate of crime occurrences using Regression Analysis and Neural Network. Since that Decision Tree had different approaches in forecasting crime occurrences, there was no reason to compare it with Regression Analysis and Neural Network. The best fit was the one that the forecasting result's accuracy was highest compared to others.
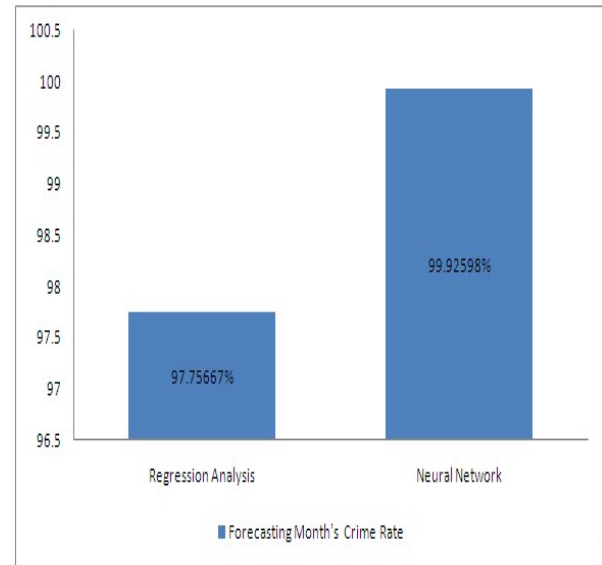

Figure 3.0 Best Fit Forecasting Model

Figure 3.0 reveal that the best fit for forecasting month's crime rate is Neural Network that got 99.92598% of accuracy which is higher than Regression Analysis that got 97.75667% of accuracy.

VII. CONCLUSION

The Exponential Smoothing really increased the accuracy and reduced the error of the forecasting model by eliminating the noise in the data where the intervals between the values of data were reduced that also gave a strong relationship between the predictors and the model.

In the research, three data mining algorithms were applied on the assessment data to predict the occurrence of crimes. The best data mining technique for forecasting the month's crime rate is Neural Network that got 99.92598%.

Computers Assisted procedures completed through the use of software packages such as IBM SPSS Amos and Rapid Miner absolutely aided the researchers with ease in most of the activities needed to be perform to create a forecasting model especially in working with the data mining techniques and algorithm that had been performed in the study

VIII. RECOMENDATION

The researchers came up to this mathematical model to forecast the month's crime rate of year 2012, which the researchers use the results of predictors produced by neural network that got an acceptability rate of 99.92598%.

This mathematical model is based on Time Series Analysis that has natural temporal ordering. This makes time series analysis distinct from other data, in which there is no natural ordering of the observations. The researchers

considered as predictor the years 2008,2009,2010,2011 that will be used to forecast the month's crime rate of year 2012.

$$y = \frac{B_0 + B_1 + B_2 + B_3}{B_n} \qquad (3)$$

Where:

y   -   Forecasted value

B_0, B_1, B_2, B_3 -   Predictors

B_n -   Total number of Predictors

1. Tally the data based on the scales used.

2. Add the sum in every scale and divide it by the number of data to obtain the forecasted value.

3. Interpret the results

The future researchers are recommended to use the study as a basis on how to have knowledge and to develop a forecasting model for forecasting crime occurrences.

The beneficiaries of the study such as barangay, police forces, individual, etc. are suggested to use the forecasting model to predict future crime occurrences in order to prevent different kinds of crimes.

Adding more features into the forecasting model would greatly evolve the study by getting more unique information like suspects and victims' description in order to make new forecast made out of these information

REFERENCES

[1] Polat, E. (2007, September). Spatio Temporal CRIME Prediction Model Based on Analysis of Crime Clusters. Retrieved from http://etd.lib.metu.edu.tr/upload/1260881/index.pdf

[2] Gorr, W.,Olligchlaeger, R., Thompson, Y. 2003. Short Term Forecasting Crime. International Journal of Forecasting. Retrieved from http://www-afs.secureendpoints.com/afs/ANDREW.CMU/supa/wpapers/2000-08.pdf

[3] Perry, W., McInnis. B., Price, C., Smith, S., Hollywood, J.(2013).Predictive Policing, The Role of Crime Forecasting in Law Enforcement Operations.Retrieved From http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf

[4] Yu, C., Ward, M.W., Morabito, M., & Ding, W.(2011). Crime Forecasting Using Data Mining Techniques. Retrieved from http://www.cs.umb.edu/~ding/papers/CrimeForecasting_10_12_2011.pdf

[5] VijayaKumar, M., Karthick, S., &Prakash, N. 2013. The Day-To-Day Crime Forecasting Analysis of Using Spatial and Temporal Clustering Simulation. International Journal of Scientific & Engineering Research, Volume 4, Issue 1.

[6] Mahendiran, A., Shuffet, M.,Muthiah, S., Malla, R.,Zhang, G.,(2012). Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks. Retrieved from http://filebox.vt.edu/users/shuffet/analytics/assets/ProjectCheckpoint1_Mahendiran_Shuffet_Muthiah_Zhang_merged.pdf

[7] Mande, U., Srinivas, Y.,Murthy, J.V.R.(2012). Feature Specific Criminal Mapping Using Data Mining Techniques and Generalized Gaussin Mixture Model. International Journal of Computer Science & Communication Networks, Vol 2(3), 375-379. Retrieved from http://www.ijcscn.com/Documents/Volumes/vol2issue3/ijcscn2012020315.pdf

[8] João, P., Lobo, V., Bação, F. 2008. Predictive Model for Criminality in Lisbon

[9] VijayaKumar, M., Karthick, S., &Prakash, N. 2013. The Day-To-Day Crime Forecasting Analysis of Using Spatial and Temporal Clustering Simulation. International Journal of Scientific & Engineering Research, Volume 4, Issue 1.

[10] Exponential Smoothing. (n.d.). In Wikipedia. Retrieved September 1, 2015, from http://en.wikipedia.org/wiki/Exponential_smoothing

[11] Regression Analysis. (n.d.). In Wikipedia. Retrieved July 24, 2015, from http://en.wikipedia.org/wiki/Regression_analysis

[12] Artificial Neural Network. (n.d.). In Wikipedia. Retrieved September 13, 2015, from http://en.wikipedia.org/wiki/Artificial_neural_netw ork

[13] Decision tree learning. (n.d.). In Wikipedia. Retrieved July 14, 2015, from http://en.wikipedia.org/wiki/Decision_tree_learning

[14] Mean absolute percentage error. (n.d.). In Wikipedia. Retrieved September 10, 2015, from http://en.wikipedia.org/wiki/Mean_absolute_percentage_error