

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 7, July 2018, pg.145 – 148

Crib Analysis in Bangladesh

Sarna Majumder¹, Tamanna Niger Tuly²

¹Computer and Communication Engineering Department, Patuakhali Science and Technology University, Bangladesh

²Computer Science and Engineering Faculty, Patuakhali Science and Technology University, Bangladesh

¹sarna.majumder90@gmail.com; ²tniger53@gmail.com

Abstract— *Crib is a criminal act which harms people by taking away someone's property. This crime can't be removed over a night. A smooth flow of exact process can predict these crime occurrences and thus may help the crime investigators to remove this crime. Here, an approach of machine learning such as decision tree has been applied. Data's are collected from Dumki police station, web sources and Facebook.*

Keywords— *Data mining, Decision tree, Machine learning, Crime, Data science.*

I. INTRODUCTION

Crime is a common fact in our society nowadays. This harmful act is handled by various Crime Investigation force. In Bangladesh crimes are filed by the victim or complainant in local police station or court. At the first stage a police employee in charge writes a General Diary (GD) about the crime. The next day he investigates whether the crime had been happened or not, if so this becomes a case then. A case against the criminal is filed. Officer-in-Charge (OC) the head of the police station is responsible for all the decisions that are made in the police station. OC convey the case to an investigating officer according to the priority of the case. According to the priority, the most perfect man of handling the case is given the responsibility to solve this. Officer starts his investigation from the next day and investigates until the case has been solved. When he completes his investigation he presents a report on the case. That is presented in the court then. Judge justify the case and declare the result of crime and ordered the police to necessary steps thus take the case is solved. If the complainant is not happy with the case he can apply for rethinking the case. Court then gives the case in various crime investigating sector to re-investigate the case. In general, these cases are given about 80% to the police station, 5% to the CID (crime investigating department), 10% to the DB, 4% to the PBI, 1% to others. Not every case is getting solved. Some cases remain unsolved for good. We are approaching to get an analysis in which situation crib occurs. The investigation officer differentiates some categories when he files a case. From simple criminal record, we categorize the crib in some features and analyse them in Decision tree to make a probability of occurrence for crib. Features are converted into numerical value and those values are taken for further analysis. Here we are taking a week occurrence record of crib of an area for analysis. In this preliminary stage we are training data to analysis the crib with decision tree.

II. RELATED WORK

In the progress of technology the crime analysis and prediction procedures are becoming more popular by using the existing methods. These methods are implemented by Data Mining, Machine Learning, Big Data analysis, Artificial Intelligence and so on. As crime data's are private so sometimes the result of analysis becomes less probable.

In [1], linear regression is used with gradient descent for analysing crime and predicting a future crime number which was compared with Bangladesh police websites result. In [2], an overview of existing Data

Mining process has been covered for introducing the crime analysis with Data Mining and the another field and Decision tree working principle using Hunt’s algorithm is shown which has discovered in many areas of crime. For [3], a case for burglary is introduced by some factors which give a result of 80% accuracy.

For [1], the result is shown for the crime occurrence prediction which calculates the total number of crimes in general but our attempt is to train the data set for which situation the crib will occur and which situation it won’t, thus it will also identify the outlier situation. For a series Crib the investigator will be able to find some query from those previous dataset and predict when and how the crib may occur.

III.METHODOLOGY

Machine learning is such a process which trains, learns, builds model and gives an excellent feedback. Decision tree gives the most appropriate result for analysis. We are using Decision tree for our crib analysis. Methodologies in this paper are divided into two sections, Data Building, Implementation of the data into decision tree. Purpose of this methodology is:

- We are attempting to categorize crib according to some characteristic in which situation it occurs.
- Analyse those characteristic using Decision tree for detecting future situation in which crib may occurs.

A. Data Building

The main step of the crime analysis is to collect data. A large number of data is becoming garbage because of proper use of those data. Cases are filed for the crime. These cases may be the source of data. In this paper we generate some characteristic from the cases and categorized them into different types of classes. Then we convert them into numerical values and thus analyse them. The case history of crib has been analysed here. Some characteristics are taken from various sources and categorize them into various classes. This data’s are will be pre-processed and analysis for further steps.

B. Implementation of Data into Decision Tree

In this section we process the data and implement those data’s characteristic into Decision tree. Characteristics for crib analysis are taken according to the factors as shown in table I. In which situation crib was occurred that has taken and in which situation crime was not occurred are calculated. From there we calculate Entropy, Gain and information Gain. Characteristics are categorized according to their priority in which factors are popular for this crime.

TABLE II
DATA PROCESSING

Priority	Characteristics(for crib crime)		
	Timing	Entry	Exit
1	Night	Window	Door
2	Day	Door	Window
3	Unpredictable	Rooftop, Unpredictable	Unpredictable

From calculation we analyse the highest Gain and we implement that section for further analysis and this continues until the calculation is reached in an appropriate stage. Decision Tree gives most accurate result behind all of those existing method.

The terms are,

$$Information\ Gain\ I(p, n) = ((-p)/p + n)log_2((p)/(p + n))-(n/p + n)log_2(n/p + n); \dots\dots\dots (1)$$

$$Entropy (E) = sum\ of((p_i + n_i)/(p + n)(I(p_i, n_i)); \dots\dots\dots (2)$$

$$Gain = I(p, n)-E; \dots\dots\dots (3)$$

Here,

Crib has occurred= p

Crib has not occurred= n;

This p and n are converted into numerical value which are represents by 1 and 0 respectfully. The final results for crib occurrences are designed by the decision tree’s final stage. From Equation (1), (2), and (3) we calculate the Entropy, Information Gain and Gain for p and n and those characteristic like if the crib occurs at night, places of entry is window, places of exit is door, weather is overcast, and temperature is hot then we calculate each characteristic by taking the value of p and n.

IV. RESULTS AND DISCUSSION

The Decision tree algorithm is used here. It is the most appropriate procedures in machine Learning. Characteristics are based on preliminary training data set for this crib crime. Considering one month primary analysis on each and every day in which situation crib has been occurred or not we are doing our procedures. Here some calculations are given:

TABLE II
CHARACTERISTIC FOR ONE WEEK CRIB

SL. No.	Timing	Entry	Exit	Weather	Temperature	Theft/not theft
1.	Night	Window	Door	Rainy	Cool	yes
2.	Night	Roof	Door	Overcast	Mild	yes
3.	Day	Door	Door	Overcast	Hot	no
4.	Day	Window	Window	Sunny	Hot	no
5.	Night	Roof	Door	Rainy	Mild	yes
6.	Day	Door	Door	Overcast	Mild	yes
7.	Night	Window	Window	Overcast	Mild	yes

In Table II one week analysis for crib factors in which situation that has been took place in various aspect.

Here $I(p, n) = 0.87$.

				Entry	p	n	I(p, n)
				Window	2	1	0.92
				Door	1	1	1
				Roof	2	0	0
timing	P	n	I(p, n)				
Day	4	0	0				
Night	1	2	0.92				

Fig. 1 Mathematical calculation for some of the factors

The highest information gain is from weather so further analysis will be weather factors and they will be broadened to the next label of decision tree.

TABLE III
WEATHER FACTORS

Weather	Timing	Entry	Exit	Crib/Not
Overcast	night	roof	door	yes
Overcast	day	door	door	no
Overcast	day	door	door	yes
Overcast	night	window	window	yes

From this calculation we repeat the procedures until an exact analysis is predicted. Using the Decision Tree label wise detection we can get an appropriate flow structure in which situation the crib will occur next. For serial cribbing case an exact prediction can be made by using this procedure. As criminal Data is private and can't be accessed at ease, so the result is a probable approach for making such decision when this types serial cribbing is occurring.

V. CONCLUSIONS

Now a day's machine learning is a challenging task to do and the crime prediction too. But the exact approach for making decision can give a probable result. Some characteristics are considered here found from local police station which are analyzed by decision tree and thus gives a prediction for serial cribbing for Bangladesh perspective. The tree structure thus helps the investigating officer for further analysis of cribbing.

ACKNOWLEDGEMENT

First of all, we thank the almighty. Thanks to our university for giving such opportunity and unconditional support for completing this research.

REFERENCES

- [1] M. D. Abdul Awal, Jakaria Rabbi, Imran rana, "Using Data mining technique to analyse crime of Bangladesh," *IJCSN - International Journal of Computer Science and Network*, vol. 6, issue 4, Aug. 2017.
- [2] Hossein Hassani, Xu Huang, Emmanuel S. Silva, Mansi Ghodsi, "A Review of Data Mining Application in Crime," in *Statistical Analysis and Data Mining*, April 2016.
- [3] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri, "Detecting patterns of crime with series finder," in *Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD 2013)*, 2013.