

## Solution Development to Prevent Crime Rate

Dr. A.S Poornima  
Associate Professor  
Department of Computer Science and Engineering  
Siddaganga Institute of Technology, Tumkur

Harshith.R.N  
Mtech Student  
Computer Network Engineering  
Siddaganga Institute of Technology, Tumkur  
harshi.neelakant@gmail.com

**Abstract:-** Crime is one of the major problems encountered in a society. Thus, there is an urgent need for security agents and agencies to battle and eradicate crime. Preventive are taken to reduce the increasing number of cases of crime. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analysing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Study of crime data can help us analyse crime pattern, inter-related clues& important hidden relations between the crimes. That is why data mining can be great aid to analyse, visualize and predict crime using crime data set. We analyse data objects using machine learning techniques. Dataset is classified on the basis of tree based algorithm. In this prediction is done using random forest algorithm according to various types of crimes taking place in different states and cities. Crime mapping will help the administration to plan strategies for prevention of crime, further using Random forest algorithm technique data can be predicted and visualized in various form in order using leaflet and shiny package to provide better understanding of crime patterns.

**Keywords:** *Random forest, Machine learning, leaflet and shiny package, prediction*

\*\*\*\*\*

### 1. INTRODUCTION

#### 1.1 INTRODUCTION TO CRIME DATA ANALYTICS

According to Chicago data portal Records Bureau, crime a has significantly increased in recent years. It has become the most prior to the administration to enforce law & order to reduce this increasing rate of the crime. This is where criminology comes into picture . Criminology is scientific study of crime and criminal behaviour in order to detect crime characteristics. Use of machine learning techniques can produce important results from crime dataset.The very step in study of crime is crime analysis .Crime analysis is exploring, inter relating and detecting relationships between various crimes and characteristics of crimes .Police department maintains crime data at the record. This data contains huge amount of data set with complex relationships which needs use of data mining techniques in order to be transformed into useful information.

#### 1.2 OBJECTIVES

Crime reduction and prevention challenges in today's world are becoming increasingly complex and are in need of a new technique that can handle the vast amount of information that is being generated.

Traditional police capabilities mostly fall short in depicting the original division of criminal activities, thus contribute less in the suitable allocation of police services.

By studying the geographical areas which are at greater risk and outside the traditional policing limits. The developed method makes the use of a geographical crime mapping algorithm to identify areas that have relatively high cases of crime. The term used for such places is hot spots.

The identified hotspot clusters give valuable data that can be used to train the artificial neural network which further can model the trends of crime.

#### 1.3 BENIFITS

Police will greatly benefit by a software that will be able to intelligently analyse a constantly updating database of crime incidences and its description, providing accurate predictions of where the crime is most likely occur and at what time

This will help in optimum police resource allocation. In doing so, one of the drawback is the fact that crime occurrences are generally sparse with respect to the type of crime, time and space at which the incidence occurs, and the randomness subjected to it. Apart from this the ability to process unstructured data was limited till now but with the advent of big data we can explore a new approach for making predictions.

A crime analysis tool must be able to identify crime patterns accurately and efficiently for future forecasting and accurate crime pattern.

#### 1.4 CHALLENGES

- 1) Increasing size of the information that has to be stored and analyzed.
- 2) Different techniques that can analyze with accuracy and efficiency of this increasing volume of data on crime.
- 3) Varied methods and infrastructure that are used for recording data on crime.
- 4) The available data are inconsistent and incomplete and are making the task increasingly difficult formal analysis
- 5) Due to complex nature, it takes more time

#### 1.5 FIND OUT HIDDEN FACTS THROUGH DATA

We could find answers for a lot of questions by bringing out lot of different insights based on each and every minute aspects derived from data analysis. Many researchers have tried explaining why crimes occur in certain areas or is there any pattern that can be concluded from the past events. One such theory that answers these questions is the crime prevention theory . According to it crime does not happen in random fashion, it is either opportunistic or planned.

Find crime pattern

Predict when, where and what kind of Crime may happen.

Build a complete data driven solution to prevent the number of crimes happening

#### 1.6 PREDICT AND ACT

- Predict where and what type of crime may happen
- What is the probability for a person to become a victim of a particular crime
- How to prevent it
- End to end data driven solutions

## 2: LITERATURE SURVEY

There are number of papers which we have analyzed in order to determine the technology used and data mining techniques. The review of literature will involve efficient and usable techniques such as Tree based system and R studio tool.

In the studied paper tool for forecasting the crime was created. The discrete choice model will be used in this case. The discrete choice model takes into account the choice of the criminal and location in order to forecast the crime. In this project a specific location Malaysia is considered. The crime location and mindset of the criminal

will be considered in this case. Crime Forecasting is rarely used globally by police including Malaysia. In practice usually the police would target persons with their criminality and studying their strategy of implementing crime.

The police will also monitor the current crime situation and will take necessary action when the crime index increases. Both of these scenarios require action taken after crime incurred. Therefore if crime forecasting can be adopted perhaps early crime prevention can be enforced. The aim of this study is to identify crime patterns in Kedah using univariate forecasting technique. Seventy six recent monthly data (January 2006 – April 2012) were obtained from IPK Alor Star with the permission from PDRM Bukit Aman. Exploratory Data Analysis (EDA) and adjusted decomposition technique were conducted in order to fulfill the objective of the study. The findings revealed that total crimes in Kedah were mainly contributed by type of property crime (80-85%) while violent crime has a small proportion only. Fortunately due to the productiveness of the police the property crime trend indicated curve declining pattern. Special section of the crime forecasting is considered in this case. The crime forecasting is used so that crime can be controlled. No specific location is considered in this case. The GIS(Geographical Information System) is used in order to detect the location where crime is happened. The mentality of criminal is considered in this case. The help of crime forecasting is also considered in this case.

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. Crime control tends to use knowledge from the analyzed data to control and prevent the occurrence of crime, while the criminal suppression tries to catch a criminal by using his/her history recorded in data mining. Brown (1998) constructed a software framework called ReCAP (Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques. Data fusion was used to manage, fuse and interprets information from multiple sources.

The main purpose was to overcome confusion from conflicting reports and cluttered or noisy backgrounds. Data mining was used to automatically discover patterns and relationships in large databases. Crime detection and prevention techniques are applied to different applications ranging from cross-border security, Internet security to household crimes. Abraham et al. (2006) proposed a method to employ computer log files as history data to search some relationships by using the frequency occurrence of incidents. Then, they analyzed the result to produce profiles, which can be used to perceive the behavior of criminal. De Bruin

et. al. (2006) introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. This method also provided a visual clustering of criminal careers and identification of classes of criminals. From the literature study, it could be concluded that crime data is increasing to very large quantities running into zeta bytes (1024 bytes). This in turn is increasing the need for advanced and efficient techniques for analysis. Data mining as an analysis and knowledge discovery tool has immense potential for crime data analysis. As is the case with any other new technology, the requirement of such tool changes, which is further augmented by the new and advanced technologies used by criminals. All these facts confirm that the field is not yet mature and needs further investigations.

### 3. SYSTEM DESIGN

#### 3.1 Data Collection and description:

Enormous amount of crime data is collected at the end of year at police records. This data is made available by Chicago data portal. This data is in the form of number of cases recorded all over the nation throughout the year.

The data is in raw form and also contains some wrong as well as missing values. Hence preprocessing of data becomes very necessary in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and PreProcessing.

**3.2 Random forest algorithm:** In the random forest approach, a large number of decision trees are created. Every observation is fed into every decision tree. The most common outcome for each observation is used as the final output. A new observation is fed into all the trees and taking a majority vote for each classification model. An error estimate is made for the cases which were not used while building the tree. That is called an OOB (Out-of-bag) error estimate which is mentioned as a percentage.

**The R package "randomForest" is used to create random forests.**

Random forest is like bootstrapping algorithm with Decision tree (CART) model. Say, we have 1000 observation in the complete population with 10 variables. Random forest tries to build multiple CART model with different sample and different initial variables. For instance, it will take a random sample of 100 observation and 5 randomly chosen initial variables to build a CART model. It will repeat the process (say) 10 times and then make a final prediction on each observation. Final prediction is a function of each prediction. This final prediction can simply be the mean of each prediction.

Random forest gives much more accurate predictions when compared to simple CART/CHAID or regression models in many scenarios. These cases generally have high number of predictive variables and huge sample size. This is because it captures the variance of several input variables at the same time and enables high number of observations to participate in the prediction.

Random Forest is a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is planted & grown as follows:

1. If the number of cases in the training set is N, then sample of N cases is taken at random but with replacement. This sample will be the training set for growing the tree.
2. If there are M input variables, a number  $m \ll M$  is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

#### Python

7

#### R Code

```
library(randomForest)
x <- cbind(x_train, y_train)
# Fitting model
fit <- randomForest(Species ~ ., x, ntree=500)
summary(fit)
#Predict Output
predicted= predict(fit, x_test)
```

#### 3.3 Visualizing and Analysing Result:

The leaflet package and Shiny app used for visualizing and analysing result.

The Shiny package comes with ten built-in examples that demonstrate how Shiny works. This reviews the first three examples, which demonstrate the basic structure of a Shiny app.

Using visualization tools we visualized the result of Prediction as shown below:



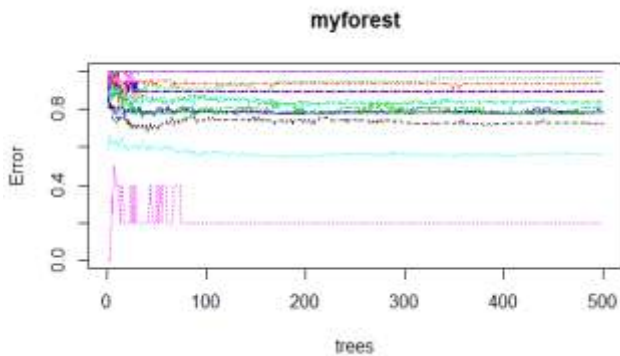
#### 4. CONCLUSION AND FUTURE ENHANCEMENT

Result of this project analyse and predict the crimes from huge data set available. Results will be in the form of correlation between various crime and location of crime i.e. state/city. Prediction of the crime will be displayed using R graph plot and webbrowser.

A major challenge facing all law-enforcement and intelligence- gathering organizations is accurately and efficiently analyzing the growing volumes of crime data. As information science and technology progress, sophisticated data mining and artificial intelligence tools are increasingly accessible to the law enforcement community. These techniques combined with state- of-the-art Computers can process thousands of instructions in seconds, saving precious time. In addition, installing and running software often costs less than hiring and training personnel. Computers are also less prone to errors than human investigators, especially those who work long hours.

This Project work focus on developing a crime analysis tool using different data mining techniques that can help law enforcement department to efficiently handle crime investigation. The proposed tool enables agencies to easily and economically clean, characterize and analyze crime data to identify actionable patterns and trends. The proposed tool, applied to crime data, can be used as a knowledge discovery tool that can be used to review extremely large datasets and incorporate a vast array of methods for accurate handling of security issues.

The development of the crime analysis tool has four steps, namely, data cleaning, clustering, classification and outlier detection. The data cleaning stage removed unwanted records and predicted missing values. The clustering technique is used to group data according to the different type of crime. From the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future. The classification of data is mainly used predict future crime trend. The last step is



#### SYSTEM DESIGN

The block diagram of system design is as follows:

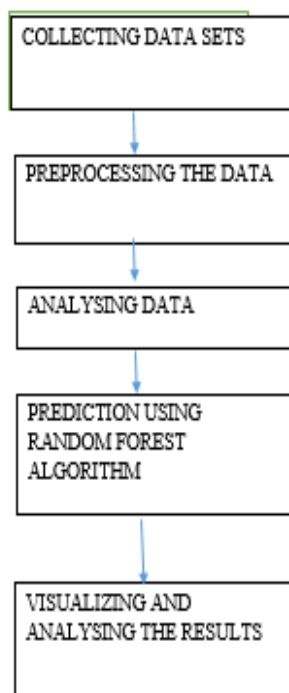


Fig1: Methodology for crime pattern analysis

mainly used to identify future crimes that are emerging newly by using outlier detection on crime data. Experimental results prove that the tool is effective in terms of analysis speed, identifying common crime patterns and future prediction. The developed tool has promising value in the current changing crime scenario and can be used as an effective tool by police and enforcement of law organizations for crime detection and prevention.

#### Future Enhancement:

The artificial neural network specification and estimation approach is going to use in order to enhance by processing capability of Hadoop platform

#### References

- [1] Amarnathan, L.C. (2003) Technological Advancement: Implications for Crime, The Indian Police Journal, April June.
- [2] Abraham, T. and de Vel, O. (2006) Investigative profiling with computer forensic log data and association rules," in Proceedings of the IEEE International Conference on Data Mining (ICDM'02), Pp. 11 – 18.
- [3] Brown, D.E. (1998) The regional crime analysis program (RECAP): A frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, Pp. 2848-2853.
- [4] Corcoran J.J., Wilson I.D. AND Ware J.A. (2003) Predicting the geo-temporal variations of crime and disorder, International Journal of Forecasting, Vol. 19, Pp.623–634.
- [5] David, G. (2006) Globalization and International Security: Have the Rules of the Game Changed?, Annual meeting of the International Studies Association, California, USA, [http://www.allacademic.com/meta/p98627\\_index.html](http://www.allacademic.com/meta/p98627_index.html).
- [6] de Vel, O., Anderson, A., Corney, M. and Mohay, G. (2001) Mining E-Mail Content for Author Identification Forensics, ACM SIGMOD Record, Vol. 30, No. 4, Pp. 55-64.
- [7] de Bruin, J.S. , Cocx, T.K. , Kusters, W.A. , Laros, J. and Kok, J.N. (2006) Data mining approaches to criminal career analysis," in Proceedings of the Sixth International Conference on Data Mining (ICDM'06), Pp. 171-177. International Journal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011
- [8] Hauck, R.V. Atabakhsh, H., Ongvasith, P., Gupta, H. anChen, H. (2002) Using Coplink to Analyze Criminal- Justice Data, Computer, Volume 35 Issue 3, Pp. 30-37.
- [9] Krishnamorthy, S. (2003) Preparing the Indian Police for 21st Century, Puliani and Puliani, Bangalore, India.
- [10] Keyvanpour, M.R., Javideh, M. and Ebrahimi, M.R. (2010) Detecting and investigating crime by means of data mining: a general crime matching framework, Procedia Computer Science, World Conference on Information Technology, Elsevier B.V., Vol. 3, Pp. 872-830.
- [11] <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>