Legal File Crime Analysis using Data Mining Techniques

Riya I[1], Dr. R. Kavitha Jaba Malar[2]

"[1]Post Graduate Scholar in Computer Science, St. john's College of Arts and Science, Ammandivilai

"[2]Associate Professor, Department of Computer Science, St. John's College of Arts and Science, Ammandivilai

Abstract: Crime analysis is a methodical approach for identifying and analyzing patterns and trends in crime. With the increasing origin of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining, we can analysis previously unknown, useful information from an unstructured data. Predictive policing means, using analytical and predictive techniques, to identify criminal and it has been found to be pretty much effective in doing the same. Because of the increased crime rate over the years, we will have to handle a huge amount of crime data stored in warehouses which would be very difficult to be analyzed manually, and also now a day's, criminals are becoming technologically advance, so there is need to use advance technologies in order to keep police ahead Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. of them. In this paper, the main focus is on the review of algorithms and techniques used for identify the criminals.

Keywords: Crime analysis, Crime prediction, Classification, Regression.

I. INTRODUCTION

Crime is an important social problem, which not only uninterruptedly occupies the mind of all the students of criminology, but which continuously alarms our courts and horrifies our peaceful and unsuspecting people. Detection of crime requires an exact knowledge of many practical and theoretical scientific subjects which cannot be mastered by one single individual. Specialization in various branches of criminological science is of primary importance and must be vigorously insisted upon in order to prevent any amateur work. The analysis of the determinants of crime is always at the forefront of public debate. Effective crime fighting strategies are widely debated with some commentators in policy circles and the popular press highlighting the role of law enforcement and punishment, and others emphasizing the role of socio-economic factors. Often one of the two views have been emphasized while neglecting the other. Those who would look for the so called root cause of crime (the dominant academic view among criminologists at one time) argue that social circumstances lead to criminal behaviour and tend to dismiss policing as an effective crime fighting tool while others have pushed for harsher penalties for crime and neglected the socioeconomic environment while the recent budget cuts for the police have led to worries of lowered detection rates leading to an increase in crime. The police are the officers of the state who have the task of the investigation of crime. Indeed, they see it as central to their job, even though, in reality, non-investigative work takes up most of their time. In carrying out this work, the police have a great deal of discretion. The basic powers of a police officer arise from the status of the office constable, and this means that the police officer does not simply act as directed like a normal employee. In addition, the task in hand also lends itself to the exercise of discretion. Though the police are expected to investigate crime, not every crime which is detected is expected to result in formal action. In addition, a basic function of the police is to keep the peace, which again requires sensitivity and common sense rather than legalistic intervention at all times. When investigating crime, the main choice of strategies has been presented as between reactive and proactive policing. The reactive approach involves the police in responding to public calls for help. It has the advantages that the police operate openly and in response to real public demand and with the consent of the public. When not answering calls, the police are expected to be patrolling openly to deter wrongdoing. The police have traditionally approached policing in this way, and it is important to realise that most crime is reported by, and detected on the basis of information from, members of the public. The police are heavily dependent on public cooperation - it is far more important than any legal powers to detect crime. But it has been pointed out that the strategy, especially patrolling, is very inefficient - the police rarely bump into criminals who are on their way home from a burglary. The proactive approach involves building up pictures of threats to the peace and potential criminality through the targeting of potential criminals and the surveillance of them. Intelligence is vital so that threats can be identified and appropriate counter-measures taken. But this information may or may not come from the general public.

II. LITERATURE SURVEY

Shabbier et al., [1] described Generic algorithm for preventing credit card frauds.I t was used for improving the computing cost with time by creating complex systems. It could analyse a fraudulent transaction in few second. The probability of misrepresentation exchanges could anticipate not long after credit card exchanges and arrangement of hostile to extortion systems could be received to keep banks from incredible misfortunes and minimize dangers. Naeimeh Laleh et al., [2] discussed supervised methods, semi-supervised methods, unsupervised methods, and real time approaches to detect the type of fraud and compare the different techniques. Abhinav Srivastava et al., [3] described hidden Markov model. It showed the execution and adequacy of the device. It also demonstrated the needfulness of taking the spending profile. The accuracy of the system was 80 %.Sammaes et al., [4] proposed Bayesian and Neural networks that provide computation all earner which consist of training set having feature and data for detecting fraud so that it can correctly classify the new data as fraud or not. It is concluded that both the technique can beused for detecting fraud.4.2. Survey on violent crime Chao Yangt et al., [5] discussed about rough- fuzzy c-means algorithm for analysis of violent crime, rough set and information entropy. It was combined to upgrade the capacity so that it could deal with the uncertainty, vagueness, and incompleteness. This algorithm was used for resolving overlapping data. Chao Yang et al., [6] proposed swarm rough algorithm to investigate the mix components of brutal crime and break down three sorts of mix factors, i.e. Genetic, natural and psychological factors and assessed the execution and the fuzzy swarm optimization technique by getting numerous diminishments for the mix factor datasets. It works better in a mixed dataset group. Jorge E et al., [15] discussed about outdoor physical actions and violent crime among internal city youth. Multiple regression analysis was performed using outdoor physical actions. This was performed for demonstrating connections between connections between c physical action and for measuring violent crime densities along other natural key variables. Survey on Traffic Violence Sachin Kumar et al.,[13] discussed k- mode clustering and association rule mining algorithm which were used to examine various design or pattern of accidents occurred in the road. After applying the algorithm EDS was made basis of month and hour to monitor the accidents occurred. Aaron Christian et al., [7] proposed genetic algorithm. The system provided detection for both violation but detected swerving violations faster than blocks the pedestrian lane violation and process one data at a time but runtime of the system is slow but can be improved. Jieling jin et al.,[8] described about cumulative logistics model, neural network model and Bayesian network model and used for analyzing the traffic violation and compared different model. Accuracy of Bayesian networks was about 70%, the cumulative logistic model was about 47%, and the neural network model was about 51%. Bayesian networks model better predicted the level of traffic violations. Sachin Kumar et al.,[12] proposed k-means clustering and association rule mining algorithm. It was used for indicating areas i.e. high, low and moderate. Association rule mining was used for finding the association between various attributes that frequently happened together when an accident takes place. Both the algorithm could be used for recognizing factors related with road accidents. Clougherty et al., [9] discussed kernel density estimation, logistic regression and random forest modelling was used to conduct spatial and temporal analysis of sexual assault. Kernel density estimation was used to compare the probability density functions of sexual assaults over daily, weekly, and monthly time periods. They constructed time series using logistic regression, and random forest models to assess correlation between point-locations of sex crimes, weather conditions.

III. PROPOSED METHODOLOGY

Crime analysis has become one of the most essential activities in the world, because the technology development and the high growth of community have resulted a high magnitude of crimes, most of the time with bizarre patterns. Due to the advancement of technology, lot of criminals has become very intelligent and, therefore they conduct crimes in an untraceable manner. Majority of those crimes evolve in a long period of time making them even more difficult to predict. Therefore, manual techniques of analyzing such data with a vast variation have resulted in lower productivity and ineffective utilization of manpower. This is one of the most dominant problems in many law enforcement institutes. Existing in this world, a lot of unemployment, hunger, starvation, together with the mind goes the wrong way. Terrorism, murder, robbery, theft and rape are growing every day. The amount of data being produced in modern society is growing at an accelerating pace. One of the areas where information plays an important role is that of law enforcement. Obviously, the amount of crime data gives rise to many problems in areas like data storage, data warehousing, data analysis and privacy.
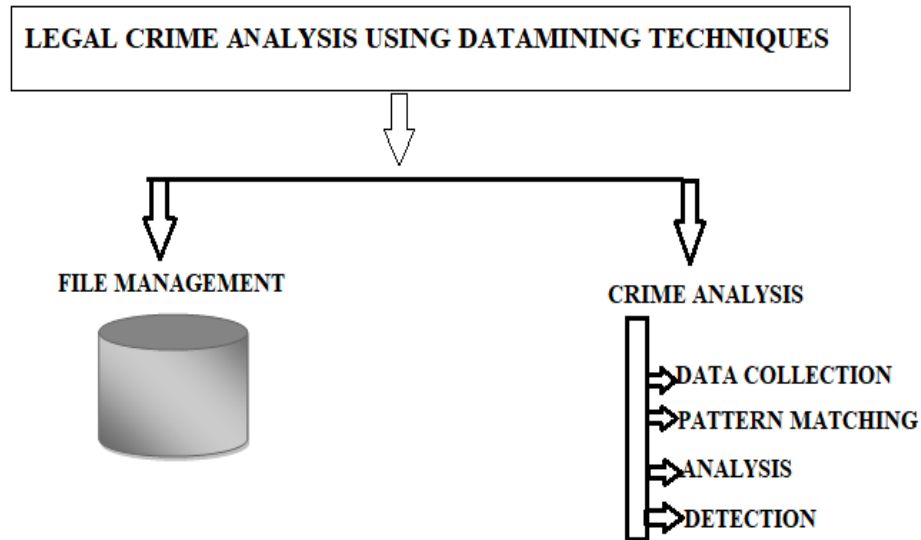
Figure 1: Proposed Architecture

Cross-correlation (CC) is the statistical approaches used for pattern matching. A simple sum of pair wise multiplications of corresponding pixel values of the template image and regions of the same size in the captured image is computed, yielding a similarity value between the images. However, because this value is subject to changes in reflectivity or illumination in the captured image, the approach has been replaced by normalized gray-scale correlation (NGC), in which the correlation value is invariant to global brightness changes. One of the sub problems that occur in the specification above is calculating the similarity measure of the aligned template image and the overlapped segment of the input image, which is equivalent to calculating a similarity measure of two images of equal dimensions. This is a classical task, and a numeric measure of image similarity is usually called image correlation.

$$\text{Cross-Correlation}(\text{Image1}, \text{Image2}) = \sum_{x,y} \text{Image1}(x,y) \times \text{Image2}(x,y)$$

A Geographic Information System (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. GIS is the merging of cartography, statistical analysis,and computer science technology. It digitally makes and "manipulates" spatial areas that may be jurisdictional, purpose, or application-oriented. The clustering algorithm used is

Keep K points into the space denoted by the objects that are being clustered.
These points specify the initial group centroid.
Assign each object to the group that has the closest centroid.
When all objects have been assigned, recalculate the positions of the K centroid.
Repeat Steps 2 and 3 until the centroid no longer move.
This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

To classify the crime incidents based on the similarity between the crime objects stored in the class, structure crime classification is used. Classification is the hierarchy of these attributes. These attributes are represented by classification in three ways,

1) Classification of crime place
2) Classification crime types
3) Classification of crime time

IV. RESULT AND DISCUSSION

The system can be implemented only after through testing is done and if it found to working according to the specification. This method also offers the greatest security since the old system can take over if the errors are found or inability to handle certain type of transactions while using the new system.

Figure 2: Implemented Result

To identify the criminal activities and criminals there are methodologies available to reduce it. Where more number of criminal activities have happened that location is identified as hotspot. By identifying the hotspot of the criminal activities, this will help the police department to avoid such kind of activity in future in all the location. Based on the result in which location more number of crimes occurred, for the particular location more protection will be given by the police department. Clustering is data mining technique for grouping the similar type of crimes will be grouped together.

Table 1: Groupings of crime data

Factor Scores	Low Crime Activity	High Crime Activity	Intermediate Crime Activity
1	–0.129	4.959	–0.685
2	–0.148	0.768	4.128
3	–0.022	1.622	–0.886
4	–0.038	– 0.344	1.626
Total	33	01	01

A crime data analyst or detective will use a report based on the data sorted in different orders, usually the first sort will be on the most important characteristic based on the detective's experience.
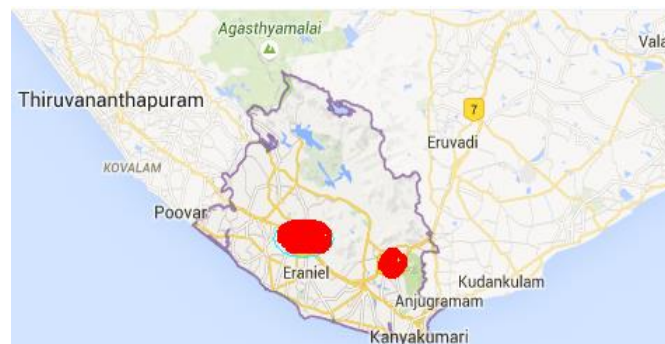

Figure3: Crime hotspot

V. CONCULSION

A clustering algorithm is used to detect the crime and crime hotspot.10 real samples are taken from a detective of police department. Experiments are done on the real dataset. Based upon the criminal data, crime pattern detection considers the data mining technique and image processing technique that is used to support police detectives in solving crimes. Results from geographic data analysis conducted on various tweets provided a clear picture of the criminal trends in several different cities. The crime intensity day-wise positively correlated with crime statistics from cops, which ultimately prove the hypothesis. This type of study would help with informing others of the crime pattern both within and around their location, ultimately assisting them with staying in a safe zone. Monitoring various social media outlets would improve accuracy.

REFERENCES

[1]. J. Agarwal, R. Nagpal, and R. Sehgal, ―Crime analysis using k-means clustering, International Journal of Computer Applications, Vol. 83 – No4, December 2013.
[2]. J. Han, and M. Kamber, ―Data mining: concepts and techniques, Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000.
[3]. P. Berkhin, ―Survey of clustering data mining techniques,‖ In: Accrue Software, 2003.
[4]. W. Li, ―Modified k-means clustering algorithm, IEEE Congress on Image and Signal Processing, pp. 616-621, 2006.
[5]. D.T Pham, S. Otri, A. Afifty, M. Mahmuddin, and H. Al-Jabbouli, ―Data clustering using the Bees algorithm, proceedings of 40th CRIP International Manufacturing Systems Seminar, 2006.
[6]. J. Han, and M. Kamber, ―Data mining: concepts and techniques,‖ 2nd Edition, Morgan Kaufmann Publisher, 2001.
[7]. S. Joshi, and B. Nigam, ―Categorizing the document using multi class classification in data mining, International Conference on Computational Intelligence and Communication Systems, 2011.
[8]. T. Phyu, ―Survey of classification techniques in data mining, Proceedings of the International Multi Conference of Engineers and Computer Scientists Vol. IIMECS 2009, March 18 -20, 2009, Hong Kong.
[9]. S.B. Kim, H.C. Rim, D.S. Yook, and H.S. Lim, ―Effective Methods for Improving Naïve Bayes Text Classifiers, In Proceeding of the 7th Pacific Rim International Conference on Artificial Intelligence, Vol.2417, 2002.
[10]. S. Sindhiya, and S. Gunasundari, ―A survey on Genetic algorithm based feature selection for disease diagnosis system, IEEE International Conference on Computer Communication and Systems(ICCCS), Feb 20-21, 2014, Chermai, INDIA.
[11]. P. Gera, and R. Vohra, ―Predicting Future Trends in City Crime Using Linear Regression, IJCSMS (International Journal of Computer Science & Management Studies) Vol. 14, Issue 07Publishing Month: July 2014.
[12]. L. Ding et al., ―PerpSearch: an integrated crime detection system,‖ 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.
[13]. K. Bogahawatte, and S. Adikari, ―Intelligent criminal identification system,‖ IEEE 2013 The 8th International Conference on Computer Science & Education (ICCSE 2013) April 26-28, 2013. Colombo, Sri Lanka.
[14]. A. Babakura, N. Sulaiman, and M. Yusuf, ―Improved method of calssification algorithms for crime prediction, International Symposium on Biometrics and Security Technologies (ISBAST) IEEE 2014.
[15]. S. Sathyadevan, and S. Gangadharan, ―Crimeanalysis and prediction using data mining, IEEE 2014.