

# *Crime Prediction, and Detection Using Data Mining Technique*

Md. Sumon Rony<sup>1</sup>, Sagor Chandro Bakchy<sup>2</sup>, Hadisur Rahman<sup>3</sup>  
Department of Computer Science and Engineering, Varendra University<sup>1, 2, 3</sup>  
sumoncse2395@gmail.com, sagorchandro.10@gmail.com, hudacse6@gmail.com

**Abstract—** As crime rates keep spiral on a daily basis, new challenges area unit sweet-faced by enforcement agencies. They need to stay there on the lookout for any signs of criminal activity. The enforcement agencies ought to thus be ready to predict such increase or decrees or trends in crime. Like: larceny, Killing, fraud, murder, drug proof, kidnapping. The crime that will occur in a very explicit space in a very explicit month, year, any timespan. The method is outlined as a process of discovering hidden valuable information by analyzing giant amounts of information, that is kept in information bases or data warehouse, exploitation varied data processing techniques like machine learning, computing, applied mathematics. Several algorithms for data processing approach to assist find the crimes patterns. Information assortment, information Preprocessing section, information Filtering, Linear Regression. Wekasoft area unit used for an assortment of information analyzing. Visualization finally gets results. The advantage of exploitation this tool is that bunch are going to be performed mechanically. For example, in data processing, you want to apprehend your information examine it thoroughly in each attainable manner. Weka visualizes panel enables you to inspect a dataset and choose completely different attributes ideally numeric ones for the x and y-axes.

**Keywords—** *Crime patterns; Data mining; Decision tree; Wekasoft; Classification; Prediction.*

## I. INTRODUCTION

Data mining could be a method of helpful info and patterns from immense information. It's additionally known as information discovery method, information mining from information, information extraction or information /pattern analysis. This term originally noted the algorithmic step within the data mining process, that at first was called the information Discovery in Databases (KDD) method. Data mining method could be a process that takes information as input and outputs erudition. Crimes area unit a standard social drawback poignant the standard of life and therefore the economic process of a society [1]. Though crimes might occur all over, it's common that criminals work on crime opportunities they face in most acquainted areas for them. By providing an information mining approach to work out the foremost criminal hotspots, and notice the kind, location and time of committed crimes, we tend to hope to lift people's awareness relating to the harmful locations inbound time periods [2]. Crime analysis uses past crime information to predict future crime locations and times. Sociology is a part that focuses on the scientific study of crime and criminal behavior. It's a method that aims to spot crime characteristics. Sociology could be a method that aims to spot crime patterns. The high volume of crime information sets, and additionally the quality of relationships between these types of data have created sociology an acceptable field for applying data processing techniques [7]. Then exiting data processing techniques area unit applied to induce patterns of crime information, and a brand-new formula is projected to boost the accuracy of the crime pattern detection system. The varied data processing

techniques like bunch and classification area unit accustomed get the patterns of crime information.

## II. METHODOLOGY AND ALGORITHM

### A. Data Preprocessing Phase

Real world data usually have the following drawbacks: Incompleteness, Noisy, and Inconsistence. So, these data need to be preprocessed to get the data suitable for analysis Purpose. The preprocessing includes the following tasks: Data cleaning: fill in missing values, smooth noisy Data, identify or remove outliers, and resolve Inconsistencies, Data integration: using multiple database, Data transformation, Data reduction, Data discretization.

### B. Data Filtering

An abstract instance filter that assumes instances form time-series data and performs some merging of attribute values in the current instance with attribute values of some previous (or future) instance. An instance filter that adds a new attribute to the dataset [4]. A filter that adds a new nominal attribute representing the cluster assigned to each instance by the specified clustering algorithm.

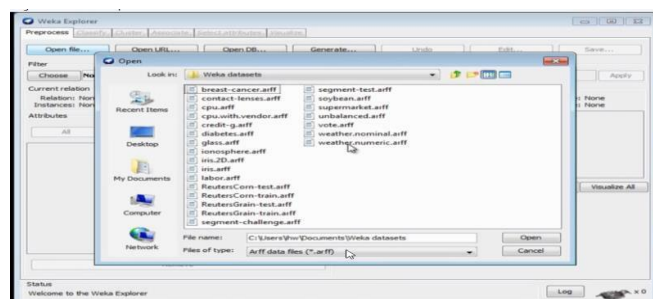


Fig.1. Data Filtering

Either the clustering algorithm gets built with the first batch of data or one specifies a serialized clusterer model file to use instead.

### C. Data Classification

Data mining techniques classification is the most commonly used data mining technique which contains a set of pre classified samples to create a model which can classify the large set of data. In this process the data is used to measure the precision of the classification rules [5]. There are ongoing works done to improve document clustering techniques such as extractions and clustering approaches to overcome the difficulty in designing a general purpose document clustering for crime investigation and the ill posed problem of extraction and clustering [7].

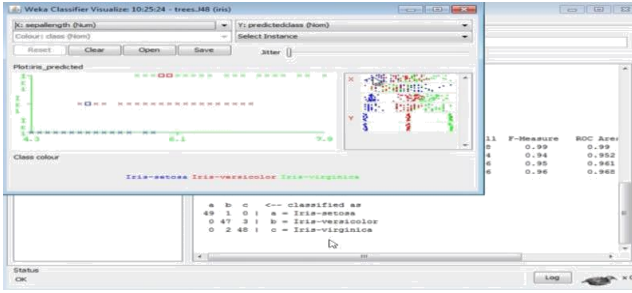


Fig.2. Data Classification

#### D. Cluster Algorithm

We will show a simple clustering example here. Let us take an oversimplified case of crime record. A crime data analyst or detective will use a report based on this data sorted in different orders, usually the first sort will be on the most important characteristic based on the detective's experience.

Crime Type	Suspect Race	Suspect Sex	Suspect Age	Victim Age	Weapon
Robbery	B	M	Middle	Elderly	Knife
Robbery	W	M	Young	Middle	Bat
Robbery	B	M	?	Elderly	Knife
Robbery	B	F	Middle	Young	Piston
Murder	W	M	Young	Middle	Knife

Fig.3. Cluster System

We look at table 1 with a simple example of crime list. The type of crime is robbery and it will be the most important attribute [4]. The rows 1 and 3 show a simple crime pattern where the suspect description matches and victim profile is also similar. The aim here is that we can use data mining to detect much more complex patterns since in real life there are many attributes or factors for crime and often there is partial information available about the crime

#### E. Linear Regression

Numeric prediction is called regression. Regression is a data mining function that predicts a number. Age, weight, distance, temperature, income, or sales could all be predicted using regression techniques. . For example, a regression model could be used to predict children's height, given their age, weight, and other factors [5]. A regression task begins with a data set in which the target values are known. Regression is a data mining technique used to predict a range of numeric values (also called *continuous values*), given a particular dataset. For example, regression might be used to predict the cost of a product or service, given other variables.

The simplest and oldest form of regression is linear regression used to estimate a relationship between two variables.

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k$$

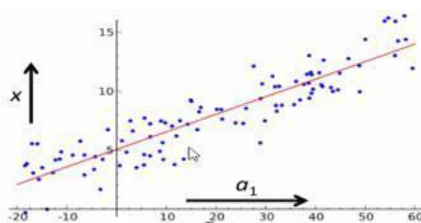


Fig.4. Regression Process

#### F. Decision Tree

In this technique, each branch of the tree is viewed as a classification and the leaves of the trees are considered as partitions of the dataset related to that particular classification. This technique can be used for exploration analysis, data pre-processing and prediction work. Decision tree can be considered as a segmentation of the original dataset where segmentation is done for a particular reason. Decision tree technique can be used for Prediction and Data pre-processing. [10]

#### G. Association Algorithm

Association algorithms find to correlations between different attributes in a dataset. The most common application of this kind of algorithm is for creating association rules, which can be used in a market basket analysis. Utilizing big data approach and data mining techniques, big data stacks that include previously recorded crimes can be analyzed in order to reveal the associations and relations between the attributes of records [6]. Which A and B occur together in the dataset to the amount of records in which A occurs? An association rule is stated as  $A \Rightarrow B$  and the association rules which have minimum support and minimum confidence values are accepted as strong rules –

$$\text{Support}(A) = \frac{|T \cap A|}{|T|}$$

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{Support}(A)}$$

#### H. Fuzzy System

This is one of the most commonly used tools for the data collection and filtering. The filtering mechanism will be accomplished by the use of rules of fuzzy. [7] These rules are based upon the IF-THEN rules. The conditions if satisfied then result will be obtained. If the condition is false then the result will not be obtained. Fuzzy sets are created in this case. A triangular fuzzy number can be defined by a triplet  $(a_1, a_2, a_3)$  with  $a_1 < a_2 < a_3$ . The membership function is defined as-

$$\mu_A(x) = \begin{cases} 0 & \text{if } x > a_3, \\ \frac{x-a_1}{a_2-a_1} & \text{if } a_1 \leq x \leq a_2, \\ \frac{a_3-x}{a_3-a_2} & \text{if } a_2 \leq x \leq a_3, \\ 0 & \text{if } x < a_1. \end{cases}$$

Fig.5. Fuzzy System

#### I. Crime Matching

The process of assigning crimes or criminals to the previous solved or unsolved crime incidents is known as crime matching. Assuming that one or more offenders responsible for a specific crime have been arrested. Crime matching process is used for assigning the previous unsolved crimes to the arrested offenders [3]. Considering a situation that police has been warned about a new unsolved crime incident so corresponding criminals are not discovered yet. In such a situation, crime matching is applied to suggest some prolific offenders as probable suspects based on the offender profiles and the method of committing criminal act.

### J. Terrorist-Behavior

Predicting terrorist attacks by group networks is an important but difficult issue in intelligence and security informatics. [14] According to our analysis, almost half (44 percent) of all terrorists examined lived within 30 miles of their targets. (See spatial analysis “Distance from Terrorist Residence to Target (All Groups).”) When the types of terrorist groups are examined separately, however the findings are so much different. Terrorists may stay close to home because of new immigration status, lack of transportation. 1 percent of the preparatory among single-issue terrorists in particular, 71 percent of the preparatory acts occurred within 12 miles and 92 percent within 28 miles of the target. [4]

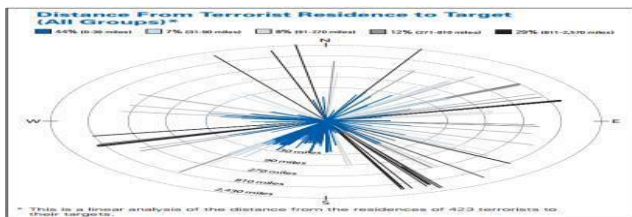


Fig.6. Restricted *Zone*

### III. IMPLEMENTATION

### A. Data Collection

It includes criminal offenses and crime incidents in the city and county. The exploratory factor analysis was also used to reduce the large number of related variables to a more efficient number to avoid redundancy. Since crime is increasing at an alarming rate globally it is important to control it. [7] In order to reduce crime rates we need to study the crime rates of various places of a country. In this research all the states and union territories crime rate (Fig) is studied in detail for different types of crimes.

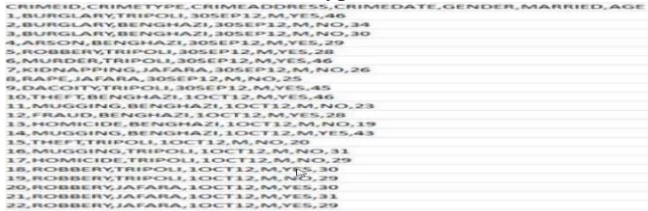


Fig.7. Data Collection

### B. Work Flow Diagram

Comparing crime statistics between different jurisdictions can be misleading without sufficient knowledge about local data and methods. . In this article, direct comparisons between jurisdictions have been avoided and the focus is instead on developments over time. Such comparisons rely on the assumption that the characteristics of the national recording systems, etc. are fairly constant over time.

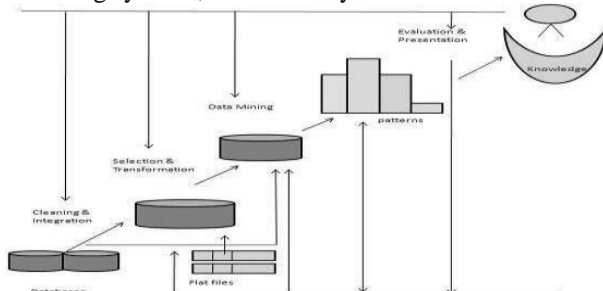


Fig.8. Work Flow Diagram

### C. Compare a Crime Period Chart

Comparing crime statistics between different jurisdictions can be misleading without sufficient knowledge about local data and methods. In this article, direct comparisons between jurisdictions have been avoided and the focus is instead on developments over time.[8] Such comparisons rely on the assumption that the characteristics of the national recording systems, etc. are fairly constant over time.

CATEGORY	2014	2015	VARIANCE	
<b>A. CRIME AGNST PERSON</b>	<b>4,292</b>	<b>3,377</b>	<b>-915</b>	<b>OR -21.32%</b>
1. MURDER	499	446	-53	OR -10.62%
2. HOMICIDE	118	24	-94	OR -79.66%
3. PHYSICAL INJURY	3,281	2,407	-874	OR -26.64%
4. RAPE	396	384	-12	OR -3.03%
<b>B. CRIME AGNST PROPERTY</b>	<b>6,663</b>	<b>4,918</b>	<b>-1,745</b>	<b>OR -26.19%</b>
1. ROBBERY	1,423	1,036	-387	OR -27.20%
2. THEFT	4,670	3,435	-1,235	OR -26.45%
3. CARJACKING	465	405	-60	OR -12.88%
4. CATTLE RUSTLING	76	41	-34	OR -45.33%
INDEX CRIMES	10,955	8,295	-2,660	OR -24.28%
NON-INDEX CRIMES	18,884	15,588	-3,296	OR -17.45%
CRIME VOLUME	29,839	23,883	-5,956	OR -19.96%
CRIME CLEARED	9,817	11,350	2,059	OR 21.33%
CRIME SOLVED	7,597	10,243	2,646	OR 34.83%
CCER	27.99%	47.36%	19.37%	69.20%
CSER	26.46%	42.89%	16.43%	68.46%
AMER	5.02%	59.89%	54.87%	-16.03%

Fig.9. Comparative to Crime Period

Out of 23,883 total crime incidents recorded from January to November 2015, City Police Office recorded the highest number of crime incident at 14,328 or 59.99%, followed del Norte Police Provincial Office at 5,053 or 21.15%, del Sur Police Provincial Office at 3,783 or 15.83% while Police Provincial Office recorded the lowest crime incidents at 1,719 or 7.19%.

## IV. RESULTS

Illustrate in the following, some experimental results for testing the qualitative combinations rules defined previously. These tests are based on the comparison between the results obtained by combining data using the Classic combination rules and the results obtained by combining the same data using our qualitative combination rules.

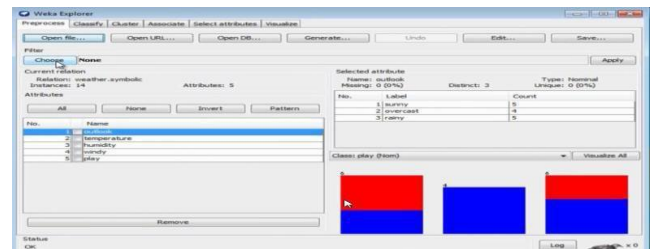


Fig.10. Wekasoft Work Flow

### A. WekaSoft

This tool is very common in order to gather the information about the particular topic. The advantage of using this tool is that clustering will be performed automatically. It is a collection of machine learning algorithm which will produce efficient result based upon the gathered information. I also collected huge data for experiment used by wekasoft software. This tool is very common in order to gather the information about the particular topic. The advantage of using this tool is that clustering will be performed automatically.

### B. Visualization

For successful data mining you must “know your data”; examine it in detail in every possible way. Weka’s Visualize



panel lets you look at a dataset and select different attributes – preferably numeric ones – for the x- and y-axes. Instances are shown as points, with different colors for different classes. You can sweep out a rectangle and focus the dataset on the points inside [10]. You can also apply a classifier and visualize the errors it makes by plotting the “class” against the “predicted class”.

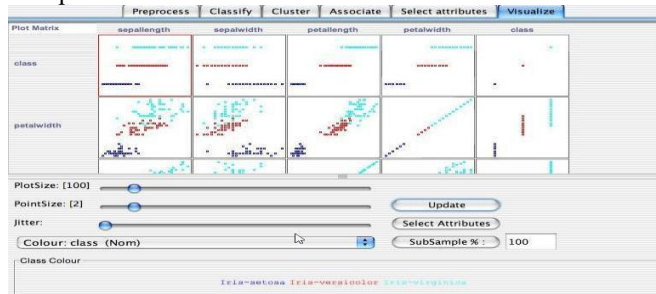


Fig.11. Visualization

## V. FUTURE WORK

As a future extension of our work, we have a tendency to arrange to apply a lot of classification models to extend crime prediction accuracy and to boost the performance. It's conjointly a useful extension for our study to think about the financial gain data for neighborhoods so as to ascertain if there are relationships between neighborhoods financial gain level and their rate. Moreover, we have a tendency to shall analyze demographics data with its crime findings.

Furthermore, we wish to check alternative crimes datasets from new cities at the side of their demographic's datasets. Lastly, however, not least, we have a tendency to hope by business enterprise this paper beginning a trend of crimes prediction, which may facilitate law enforcement and keep our community safer for everybody.

## VI. DISCUSSIONS AND CONCLUSION

Data mining may be a new discipline lying at the interface of statistics, information technology, pattern recognition, machine learning, and alternative areas. We tend to check out the employment of information mining for distinctive crime patterns crime pattern exploitation the agglomeration techniques. We tend to know the numerous attributes; exploitation skilled primarily based semi-supervised learning methodology and developed the theme for weight the numerous attributes. Our modeling technique was able to determine the crime patterns from an oversized range of crimes creating the task for crime detectives easier. Detected some feature for criminal determine system. like a non-educated person, record all crimes, determine somebody with criminal background in any event or cluster, Voice analysis, non-secure space [9]. Like as just in case of against the law against youngsters, get a fast list of all suspects & criminals for similar activities within the past year to the prime location. The opposite approach is to use a tiny low set of recent crime knowledge and score it against the present cluster's exploitation tracers or better-known crime incidents injected into the new knowledge set and so compare the new clusters relative to the tracers. We tend to cluster the crimes supported our advisement technique, to return up with crime teams (clusters in data processing

terminology), that contain the doable crime patterns of crime spreeds. Produce Multi-Biometric knowledge base of all criminals or suspects, create geographic location primarily based searches in the information.

## VII. REFERENCES

- [1] Shakir Khan, Dr. Arun Sharma, Abu Sarwar Zamani, Ali Akhtar, "Data Mining For Security Purpose & Its Solitude Suggestions", University of Warsaw, Polna 40 00-635 Warsaw Poland International Journal of Engineering Research & Technology, Vol. 1, ISSUE 7, August 2012 ISSN 2277-8616.
- [2] Mrs. Bharati M. Ramageri, "Data Mimning Techniques and Applications", "International Conference – 2008", Modern Institute of Information Technology and Research, Yamunanagar, Nigdi Pune, Maharashtra, India-411044. Vol. 1 No. 4 301-305.
- [3] N.Sudha Bhuvaneswari, S.Yamuna, "Datamining Techniques to Analyze and Predict Crimes", Dr.G.R.Damodaran College of science Coimbatore, Avinashi Road, Civil Aerodrome Post, Peelamedu, Coimbatore, Tamil Nadu 641014, India, The International Journal of Engineering And Science (IJES), ISSN: 2319 – 1813 ISBN: 2319 – 1805, Vol. 2-2016.
- [4] Viraj Samant , Sindhu Nair , Vineet Pande , "Crime Detection using Data Mining ",D. J. Sanghvi College of Engg , Mumbai, Maharashtra, India, Published by : International Journal of Engineering Research & Technology (IJERT) <http://www.ijert.org>, ISSN: 2278-0181, Vol. 5 Issue 01, January-2016 .
- [5] Levine, N., 2002. CrimeStat: A Spatial Statistic Program for the Analysis of Crime Incident Locations (v 2.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May 2002.
- [6] D. E. Brown, "The regional crime analysis program (RECAP): A framework for mining data to catch criminals," International Proceedings of IEEE International Conference on Systems and Cybernetics, pp. 2848- 2853, 1998.
- [7] M. Strano, "A neural network applied to criminal psychological profiling: An Italian initiative,"International Journal of Offender Therapy and Comparative Criminology, vol. 48, no. 4, pp. 495-503,2004.
- [8] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in Proc. of the Twenty-eighth Australasian Computer Society Conference on Computer Science, pp. 333-342, vol. 38, 2005.
- [9] M. Orakcı and B. Cıylan, "Prediction of unknown elements in rape cases: Machine learning approach in crime analysis," International Proceedings of The International Conference on Research in Education & Science, Muğla/Bodrum, May 19-22, 2016.
- [10] Y. Xiang, M. Chau, H. Atabakhsh, H. Chen, "Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list," Decision support systems, Elsevier Science Publishers, vol. 41 no.1, pp: 69-83, Nov. 2005.