IMPLEMENTATION OF CRIME DETECTION AND ANALYSIS USING STATISTICAL ANALYSIS ALGORITHMS

S. S. Aravinth, AP/CSE, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu

B. Kalaiselvan, IV-CSE, V. Rajagopal, I. M.E (CSE)

S. Beeshmakumar, IV-CSE, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu

V. Karthikeyan, IV-CSE, Dhirajlal Gandhi College of Technology, Salem, Tamilnadu

Contact E-mail: aravinthkrithick@gmail.com

ABSTRACT:

Data science is the art of extracting meaningful insights from hidden data. More statistical algorithms are proposed and used to do the crime detection and analysis in recent days in the era of data science. These predominantly used algorithms are giving accurate results in the prediction of the solutions. This paper clearly explains about Crime Rate data of the particular city with respective years. This proposed idea focuses of crime occurring frequency in particular place as well as it shows the necessary study reasons, factors and relations between occurrence of different crimes and presenting the most appropriate solutions to control and avoid the unwanted crimes. This paper is divided into multiple sections with the results of crime detection. Those sections are explained in a detailed a view. Initially the dataset is extracted and prepared with the usual attributes then data cleaning technique is applied to get the appropriate format. Followed by the age group prediction, gender identification and mapping the crimes with cities are carried out. Finally as a result, the heat map is generated to identify the crime attack areas and the major places are highlighted.

I. INTRODUCTION:

To track crimes, computer data analysts have started helping the law enforcement officers and detectives to speed up the process of solving crimes. Here we take an approach between computer science and criminal justice to develop a data mining paradigm that can help solve crimes faster. This project will use clustering based models to help in identification of crime patterns. This idea discusses some terminology that is used in criminal justice and police departments and compare and contrast them relative to data mining systems. Suspect refers to the person that is believed to have committed the crime. The suspect may be identified or

unidentified the densely populated group of crime is used to visually locate the 'Hot-spots' of crime.

II. DATA PREPARATION & DATA CLEANING :

The data used in this project contains 1.8 million observation and 26 variables. The dataset date range from 2010 up until recent respective years. Before the analysis, a simple data analysis such as convert data into a corrected data type, recode the variable into a readable format and select relevant variables.

III. APPLYING STATISTICAL TECHNIQUES:

The following techniques are used

1. Data Preparation
2. Data Cleaning

Fig 1. Life Cycle of Data Cleaning

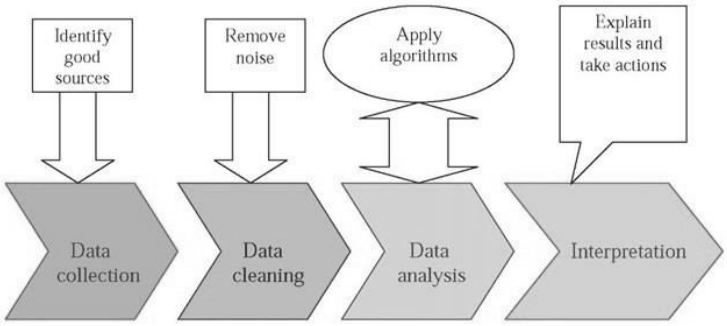


Fig 2. Schematic of Data Cleaning

TECHNIQUE I

Data preparation is the act of preparing (or pre-processing) raw data or disparate data sources into refined information assets that can be used effectively for various purpose. As figure 1 depicts, data preparation is necessary to manipulate and transform raw data so that the information content enfolded in the data set can be exposed, or made more easily accessible. The

components of data preparation include pre-processing, profiling, cleansing, validation and transformation; it often also involves pulling together data from different internal systems and external sources. Fig 2 narrates the big data applications, data preparation is largely an automated task, since it could take years of work by data analysts to manually correct every field in every file that's due to be used in an analysis. In this proposed work, the dataset are collected from social media. This dataset contains variables and observations along with the range of year 2018.

TECHNIQUE II

Any data which tend to be incomplete, noisy and inconsistent can affect the results. Data cleaning in data mining is the process of detecting and removing corrupted or inaccurate records from a record set, table or database.

The ability to understand and correct the quality of data is imperative in getting to accurate final analysis. Data mining is considered exploratory, data cleaning in data mining gives the user the ability to discover inaccurate or incomplete data prior to the business analysis and insights. In most cases, data cleaning in data mining can be a laborious process and typically requires IT resources to help in the initial step of evaluating your data. Without proper data quality, the final analysis will suffer in accuracy or you could potentially arrive at the wrong conclusion. In this paper, a simple data analysis is performed on dataset such as the conversion of raw data variable into readable codes.

IV. IMPLEMENTATION DETAILS

AGE GROUP

The age group is most likely to become victim of crime. Factoring the age into different group and examine the crimes targeted to different age group is carried out. Here the age group is divided into teenager, young adult, middle age and elderly. The following chart (Fig 3) gives the exact figure of age group and their influences.

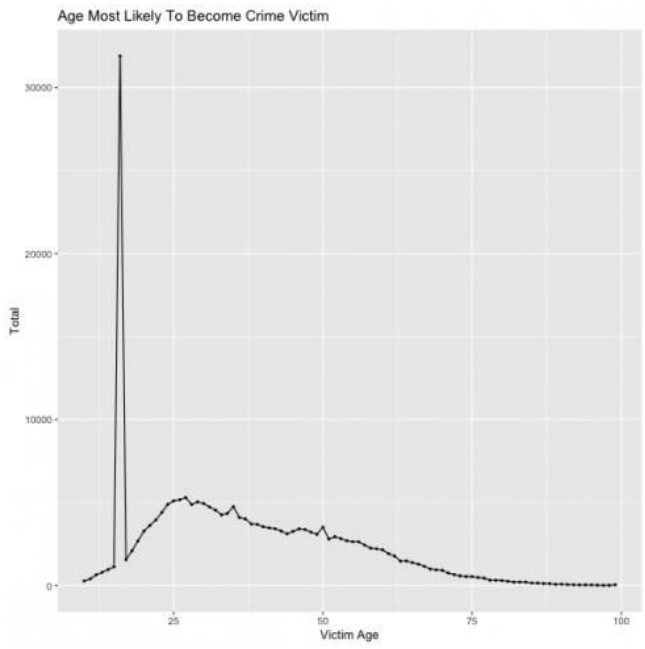


Fig 3. Age Group Prediction

GENDER

This section examines the type of crime targeted to different gender. And the various factors had influenced by different genders. The following table calculates the crime list separated by Gender wise.

Table 1. Gender Influenced Rate

S.No	Types of Crime	Male	Female
1.	Simple Assault	9242	9835
2.	Burglary	1225	1552
3.	Bike stolen	268	259
4.	Robbery	2987	2986
5.	Theft	102	108

MAPPING THE CRIMES

In this section the mapping of crimes with the respective cities are implemented. As per the illustration, the mapping results yield the highest crime committed in respective year are assault chain snatching, Kidnap Simple assault, rape and sexual assault and so on.

The fig 4. Clearly explains and positions the various cities that are affected by crimes. This map is generated with the help of info graphics design. Various actors are labeled with symbols and shapes. The following shapes are being used in this diagram.

Such as rectangle boxes are indicating the theft area with most influencing parameters.

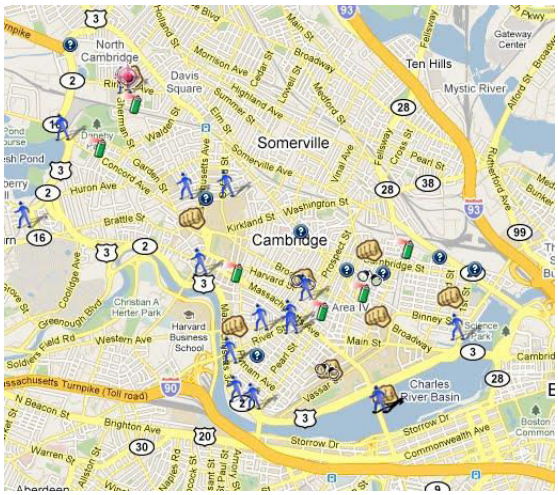
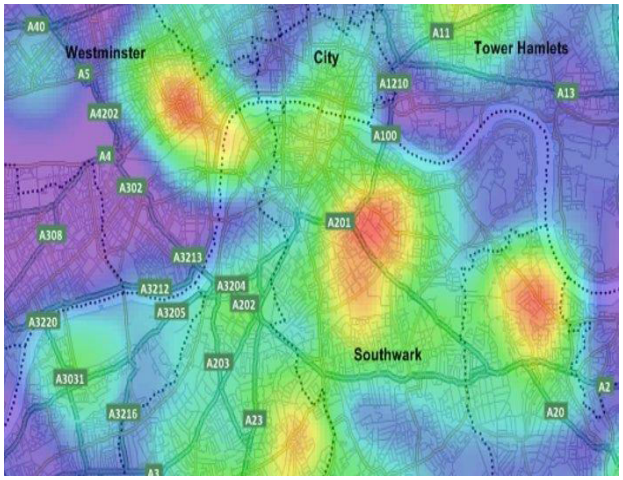
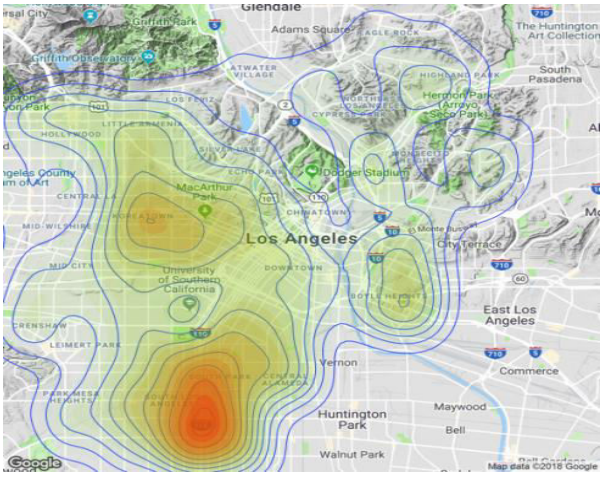


Fig 4. Info graphic of city

HEAT MAP





Next we are going to map the crime. Based on crime occurs in respective place. If crime rate is less then it shows green color in map. If crime rate is high then it shows red orange color in map.

CONCLUSION:

This is just a simple demonstration of how to gain insight of the data and mapping the crime in respective year. Project focuses on crime analysis by implementing clustering algorithm on crime dataset using rapid miner tool.

REFERENCES:

[1] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: An Overview and Case Studies", AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: http://ai.bpa.arizona.edu/

[2] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: A General Frameworkand Some Examples", IEEE Computer Society April 2004.

[3] C McCue, "Using Data Mining to Predict and Prevent Violent Crimes", available at: age&zone=rtsidebar