AMERICAN INTERNATIONAL UNIVERSITY-BANGLADESH

FACULTY OF SCIENCE AND TECHNOLOGY

Crime Pattern Detection In San Francisco Using Apriori and Clustering Algorithm

A Thesis Presented to the
DEPARTMENT OF COMPUTER SCIENCE
In Partial Fulfillment of the Requirements for the Degree BSc. in CSE

Supervised By

Md. Tohedul Islam
Assistant Professor
Department of Computer Science
Faculty of Science and Technology

Submitted By

17-33324-1	Abdullah-Al-Mahadi
17-33100-1	Islam, Md. Asiful
17-33302-1	Iftekhar Rahman
17-33458-1	Hossain Imran

Declaration

We announce that this thesis is our unique work and has not been submitted in any shape for another degree or diploma at any college or other organized of tertiary instruction. Data determined from the distributed and unpublished work of others has been recognized within the content and a list of references is given.

Abdullah-Al-Mahadi
17-33324-1
BSc. CSE

Islam, Md. Asiful
17-33100-1
BSc. CSE

Iftekhar Rahman
17-33302-1
BSc. CSE

Hossain Imran
17-33458-1
BSc. CSE

Approval

The thesis titled "Crime Detection using data mining" has been submitted to the taking after regarded individuals of the board of inspectors of the faculty of Computer Science in fractional fulfillment of the necessities for the degree of Bachelor of Science in Computer Science & Engineering on December 2020 by Abdullah-Al-Mahadi (17-33324-1), Islam Md.Asiful (17-33100-1), Iftekhar Rahman (17-33302-1), Hossain Imran (17-33458-1) has been accepted as satisfactory.

Md. Tohedul Islam
Assistant Professor & Supervisor
Department of Computer Science
American International University-Bangladesh

Juena Ahmed Noshin
Lecturer & Co-Supervisor
Department of Computer Science
American International University-Bangladesh

Dr. Md Mahbub Chowdhury Mishu
Assistant Professor and Department Head
Department of Computer Science
American International University-Bangladesh

Professor Dr. Tafazzal Hossain
Dean
Faculty of Science & Information Technology
American International University-Bangladesh

DR. Carmen Lamagna
Vice Chancellor
American International University-Bangladesh

Acknowledgement

To begin with and preeminent, we would like to much obliged the all-powerful Allah for the great wellbeing and favouring required to wrap up this book**.** We would like to precise our profound and earnest appreciation to our respectable research supervisor, Md. Tohedul Islam, Assistant Professor, Department of CSE, American International University-Bangladesh (AIUB) for his extraordinary back and direction all through the complete work. His dynamism, vision, earnestness and inspiration offer assistance us to go through the correct track It was a incredible benefit and respect to work and consider under his direction. We want to thank our external Juena Ahmed Noshin, Lecturer, Faculty of science and technology, American International University-Bangladesh (AIUB) for giving us her valuable time. With many thanks to Dr. Carmen Z. Lamagna, honourable Vice Chancellor, American International University-Bangladesh (AIUB) specially for her encouragement. We would like to thank all of our companions and relatives for their love, prayers, caring and penances for teaching and planning us for the longer term.

Abstract

The aim of this thesis is to find out whether the information in existing Police recording frameworks can be utilized by existing data mining techniques in a proficient way to realize the outcome about that are more accurate than those accomplished by Police masters when analyzing crime. Data mining technique is the most important technique for analyzing data in different fields. The main aim of this study to design some pattern of crime such as which days in the week the crime occurs. The dataset for this thesis used from DataSF which is an open data portal of San Fransisco and the data of criminal records was submitted by the police department. The apriori algorithm which is a part of association role mining have been used in this study for predicting crime pattern. To enhance this study the k-means Clustering have been also used. This study discloses the relationship between attributes of the criminal records and analyzing this pattern of crime predict the result. We believe that by this study the police department will be benefited and they can analyze the crime pattern easily and also, they will be alert before the crime occurs.

Keywords: Crime pattern, Apriori algorithm, K-means clustering, Weka.

A. Table of Contents

Chapter 1

1.Introduction

In our today's life crime is a common word, we hear about different crime all the time around us. for the most part acknowledged definition of crime comprises of actions which are exterior of society's ethical values, condemned by the society and requiring legitimate punishment by a government [3]. People arrested and/or convicted of a crime may not have engaged in a very harmful behaviour or even in the behaviour of which they are suspected, and people with no criminal record have in fact engaged in harmful and even criminal behaviour. Crime analysis is a function of law enforcement that provides systemic analysis for the detection and analysis of crime and disorder patterns and trends. Pattern information will assist law enforcement agencies to distribute resources in a more productive way and help detectives locate and arrest offenders. Now-a-days various criminal activities are on the rise in our society or in the country or abroad. Concept of crime is considered to be interior the study fields of distinctive sciences such as humanism, brain research and criminology. Moreover, preferences of innovation and analyzing the components of wrongdoing with the back of computer sciences have gotten to be imperative in fathoming the crime and identifying the wrongdoers. Different challenges are faced police department, one of the challenges faced by the police divisions is to play down dangers to society by exploring huge volumes of information. Different activity has been taken by researchers to analyze crime information utilizing information mining procedures. Criminal cases have become more dangerous day by day. Criminals were using new technique. By utilizing modern and up-to-date approaches within the examination of cases and nitty gritty discovery of similitudes between wrongdoing records constitute an imperative issue for criminology. A few ponders have been conducted within the past to plan an efficient framework that can speed up the wrongdoing investigations. Suitable data mining tactic must also be created to perform data capture, and as cluster analysis is a data gathering method that groups a collection of data analysis techniques.

Objects are much more closely related to objects in the very same collective than to objects in many other groups and involve different algorithms that differ significantly in their understanding of what constitutes a cluster and how to find them effectively. However due the changing wrongdoing design and growing crime information it has ended up more vital to upgrade the crime design discovery and avoidance framework utilizing latest data mining advances. It is possible to use information mining calculations to extract covered information from colossal data volumes [5]. Indeed, in spite of the fact that we cannot anticipate who all may be the victims of wrongdoing but can anticipate the put that has likelihood for its occurrence. The anticipated comes about cannot be guaranteed of 100% precision but the comes about appears that our application helps in diminishing wrongdoing rate to a certain degree by providing security in wrongdoing delicate zones. The classification algorithms utilised have made utilize of the spatiotemporal observations helping to more productively anticipate future wrongdoing occasions. This thesis paper is used unsupervised data mining technique to investigate criminal's data. We implemented this paper combination of Association Rules Analysis (apriori algorithm) and K-Means Clustering algorithms to conduct a comparative study of various crime patterns from the dataset [4]. Whether it is called data mining, predictive analytics, sense making, information discovery, or data science, our environment has been changed in many respects by the rapid growth and increased availability of advanced computational techniques. There are very few, if any, non-monitored, recorded, aggregated, evaluated, and modelled electronic transactions. Data on everything from our financial practices to our shopping habits, is collected. There is also casino gambling.

In this study, by utilizing wrongdoing dataset which incorporates several different sorts of wrongdoing that are based on genuine information, the relations between qualities of criminal cases are uncovered by association rules. For establishing association laws, the Apriori calculation is used. The reason of this consider is to anticipate the unknown characteristics of a particular case such as offender profile, wrongdoing weapon, casualty profile and geological zone by taking into thought the known characteristics of past criminal cases. By the use of the crime dataset in this report which includes several.

The relationships between characteristics in criminal cases are revealed by association rules and are focused on various types of crime based on real data. For establishing association rules, the Apriori algorithm is used. The purpose of this research is to predict the Unknown features of a particular case, such as the perpetrator, Profile, weapon of crime, victim profile and geographical area by taking into account the known features of the past, Crime occurrences. When the literature is reviewed, most criminal offences are Profiling experiments are only performed for one crime and Penal profiles. With this study, a large number of different kinds of studies Together, crime cases were investigated and it was shown that the clear relations between the features can be revealed with the help of technology. A survey on crime analysis found that 10% of the criminals commit 50% of the crimes [4][10]. When the writing is checked on, most crime profiling thinks about are carried out for as it were one wrongdoing and criminal profiles. With this consider, a expansive number of different crime occasions were inspected together and it was appeared that the solid joins between the highlights can be uncovered with the help of innovation. We collected a data set and apply data mining algorithm and produced a pattern this pattern helps to prevent future crime.

1.1. Problem Statement

Mining of data could be a strategy of managing with sweeping information files to see diagrams and set up an affiliation to handle issues through data examination. The gadgets utilized, allow endeavours to accept future illustrations. Although the public is worried about crime, at least some of this concern could go beyond what would justify the facts about crime. For instance, while much of the population assumes that the crime rate has been increasing, as we have just mentioned, this rate has actually been decreasing since the early 1990s.It is very difficult to identify criminals belabour that’s why crime rate could not reduce. Data mining could be a strategy to analyze information from an informational collection to alter it into a sensible structure for extra utilization. Criminals changing their behaviour day by day. Increasing their activities police department don’t know which day crime occur or which type crime may be occurred by the criminals. Sometimes Police department could not complicate their investigation because of lacking some information. To prevent this issue, we generate a pattern. This pattern show that which day crime is occurred to much and which type. Crime is indeed a real problem, but public concern about crime may be higher than the facts warrant. If police department follow this study, we hope that crime will be reduced and they can take necessary actions early.

1.2. Objectives

The main objective is that to get a pattern which day crime will occur by using some data attribute so department of local government can get information and take necessary steps earlier for saving people life and resources. aside from supported the collected information some observations can make to help individuals for taking an astute choice.

1.3. Research Questions

After analyzing various issues in problem statement portion, the main purposes of this paper found out and this research speaks to the reply of the taking after questions:

1. How can we predict crime pattern before it happens?
2. What attributes should consider and to detect crime pattern?
3. Does this research paper fulfil the objective?

Chapter 2

2. Literature Review

Crime has been rising day by day and everyone in the world. The world is trying to find out how the crime rate can be handled and how Most individuals try to store in order to work on such cases, Data for comparison in the future. A great deal of scientific study and research on crime data mining has been done over the past decade.

Chao Yangt et al, [1] discussed the Rough-Fuzzy C-means algorithm for violent analysis Crime, rough setup, and entropy of data. It was combined to boost the ability so that ambiguity, vagueness, and incompleteness could be handled. For the resolution of overlapping data, this algorithm was used.

Chao Yang et al., [2] The swarm rough algorithm was suggested to investigate the mixing components of Brutal crime and three kinds of mixed causes are broken down, i.e. Genetic, natural and psychological variables and evaluated the technique of execution and fuzzy swarm optimization by obtaining multiple decreases for the datasets of the mix factor.

Jorge E et al., [3] External discussion on outdoor physical activity and violent crime youth of the City. Outdoor physical actions were used to perform multiple regression analysis. This study was conducted to show links between outdoor physical activity among adolescents and to measure violent crime densities along with other main natural variables.

Adesola Falade, [4] … The key motivation of this study is to explore the different techniques and challenges of crime prediction and data mining published in literature and to know which state of the arts is important. The aims of the research are to systematically examine the methods, problems and challenges of crime prediction and data mining encountered in existing studies. This analysis will assist researchers in acquiring state of the arts method of crime prediction and data mining techniques and help to highlight research gaps.

Naeimeh Laleh et al., [5] In order to detect the type of fraud and compare the various techniques, this paper addressed supervised methods, semi-supervised methods, unsupervised methods, and real time approaches.

G.C. Oatley, [6] For matching and predicting crime events, the authors provide a general overview of the application of intelligent crime analysis approaches, including genetic algorithms, Bayesian networks and neural networks.

R. William Adderley, [7] In this paper neural networks were used for the clustering of crime data and the classification of crime data by using both unsupervised and supervised learning methods.

National project of COPLINK [8-11], originally built by the COPLINK project [8-11] With support from the National Institute of Justice, the University of Arizona Artificial Intelligence

Lab represents a prominent platform for text mining, classification and clustering of crime data aimed at achieving reasonably high levels of crime. Complex study of crimes. The project consists of two basic components: 1) COPLINK CONNECT and 2) COPLINK DETECT. The former performs the responsibility of data pre-processing and data collection and the latter deals with data pre-processing and data collection. Extracting trends by using data mining and artificial intelligence from vast amounts of crime data.

MohammadReza Keyvanpoura [12] This paper is divided into 5 main parts. The second section was devoted to a brief survey on relevant research and current software utilities for intelligent crime analysis, both of which are popular in the application of intelligent crime analysis. In the field of crime detection and research, data mining techniques. The basic elements of crime analysis were explored in section 3, including the principles of crime variables and crime matching. Introduces segment 4 A systemic approach through standard clustering techniques, SOM and MLP neural networks for crime matching. Finally, section 5 is dedicated to the writers' observations and future works.

Neetu Singh, [13] In order to analyze crime data, unsupervised, supervised data mining techniques are applied in this paper. To conduct a relative study of different crime patterns from different dataset, we introduced Association Rules, multiple linear regression and K-Means Clustering algorithms here. The above approaches can assist us define the underlying trends of offence and produce useful insights from that dataset. The purpose of this paper is to assist the police of Gujarat in the reduction of offence and to further help understanding the benefits of data mining in this field. Different data mining algorithms have been incorporated in this paper and their subsequent findings have been addressed, that can be utilized in various enforcement agencies and as a reference point for future crime analysis research initiatives.

Shyam Varan Nath [14] In this paper the authors try to find out the answer of "Could crimes be modelled as problems with data mining?" In order to assist in the process of detecting crime trends, they looked at k-means clustering with some changes. They applied these approaches to actual crime data from the office of a sheriff and checked their outcomes. Here they have used semi-supervised learning methods to uncover information from crime data and to help improve predictive accuracy. They still have A weighting scheme for attributes was built here to deal with the limitations of different tools and techniques for clustering out of the box.

Peng Chen, [15] This paper seeks to discuss the issue of recognizing suspected serial offenses patterns are using previously underused police attributes that have been reported Data on violence. A crime data analysis technique that brings out three variables from police reported offence event data is suggested to accomplish this: (1) setting; (2) modus operandi; and (3) time Widely used for repeated item set pattern mining is apriori algorithm, is practiced to model each offence-event attribute. Learning from association rule learning and complex datasets. Results from the model indicate that Apriori can detect important correlations and can thus illustrate trends in crime patterns nested within larger cops-filed crime databases, which may contribute to more efficient cop responses than conventional empirical approaches currently offer.

Mehmet Sevri, Hacer Karacan, [16] In this analysis, the associations between characteristics of criminal cases are revealed by association rules by the help of a crime dataset that contains many different forms of crime based on real data. For establishing association rules, the Apriori algorithm is used. The purpose of this analysis is to predict the unknown characteristics of a particular case, like the profile of the suspect, the weapon of crime, the profile of the sufferer and the geographical region, by taking into account the known characteristics of previous criminal cases.

Arvind Venkatesh, [17] This paper proposed crime modelling to find effective crime detection algorithms, precise detection, preparation and transformation of data, and processing time. The paper describes crime activity, crime prediction, effective detection, and vast quantities of data collected from different sources are controlled. This system also is an automation for registering complaints, forecasting crime trends based on the previous information of crime obtained from different sources.

With **De Bruin et al**. [18] Using a new distance metric, developed a method for crime patterns to compare all individuals based on their profiles and then cluster them accordingly.

Gupta's Manish et al. [19]. It highlights the current frameworks used as e-governance initiatives by the Indian police and also suggests an open query-based interface as a platform for crime detection to support police in their operations. He suggested an app to extract valuable data from the National Crime Record Bureau (NCRB)'s crime database and identify offence hot spots using crime data mining techniques such as clustering, etc. Indian crime reports have shown the efficacy of the proposed interface.

A. Malathi et al. [20] look at the use of the missing meaning and clustering algorithm for a data mining technique to help predict the trends of crime and speed up the crime solving process.

Malathi. A et. al. [21] used a model focused on clustering/classify to predict patterns in crime. Data mining tools are used to analyze the Police Department's city crime data. The outcome of this data mining could theoretically be used in the next few years to mitigate and even deter crime.

Malathi and Dr. S. Santhosh Baboo. A [22] Research work cantered on the development of an Indian scenario crime analysis method using various data mining techniques that can help law enforcement agencies manage crime investigations effectively. The proposed instrument helps agencies to clean, characterize and evaluate crime data quickly and economically to identify actionable patterns and trends.

The.Kadhim B. Swadi Al-Janabi [23] Introduces a proposed method for the analysis and identification of offence and offensive data using Data Classification Decision Tree Algorithms and the Basic KMeans algorithm for data clustering. The paper focuses to help experts identify trends and patterns, make predictions, find relationships and possible theories, map criminal networks and identify potential suspects.

Chapter 3

3. Methodology

The first thing we've done is check several websites to find out about the criminal history of forms of crime over the past few years. We have searched different types of dataset related website.

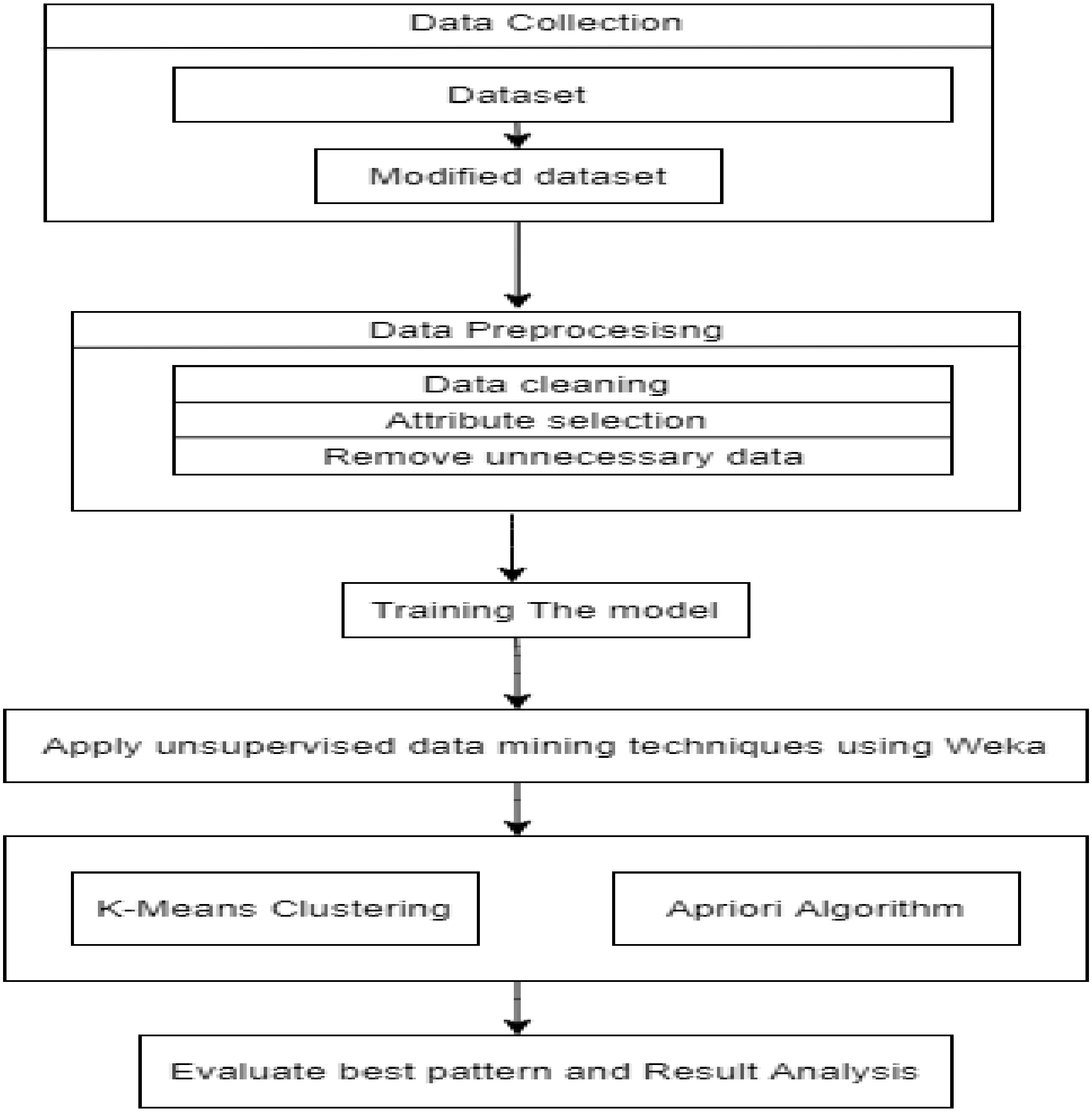
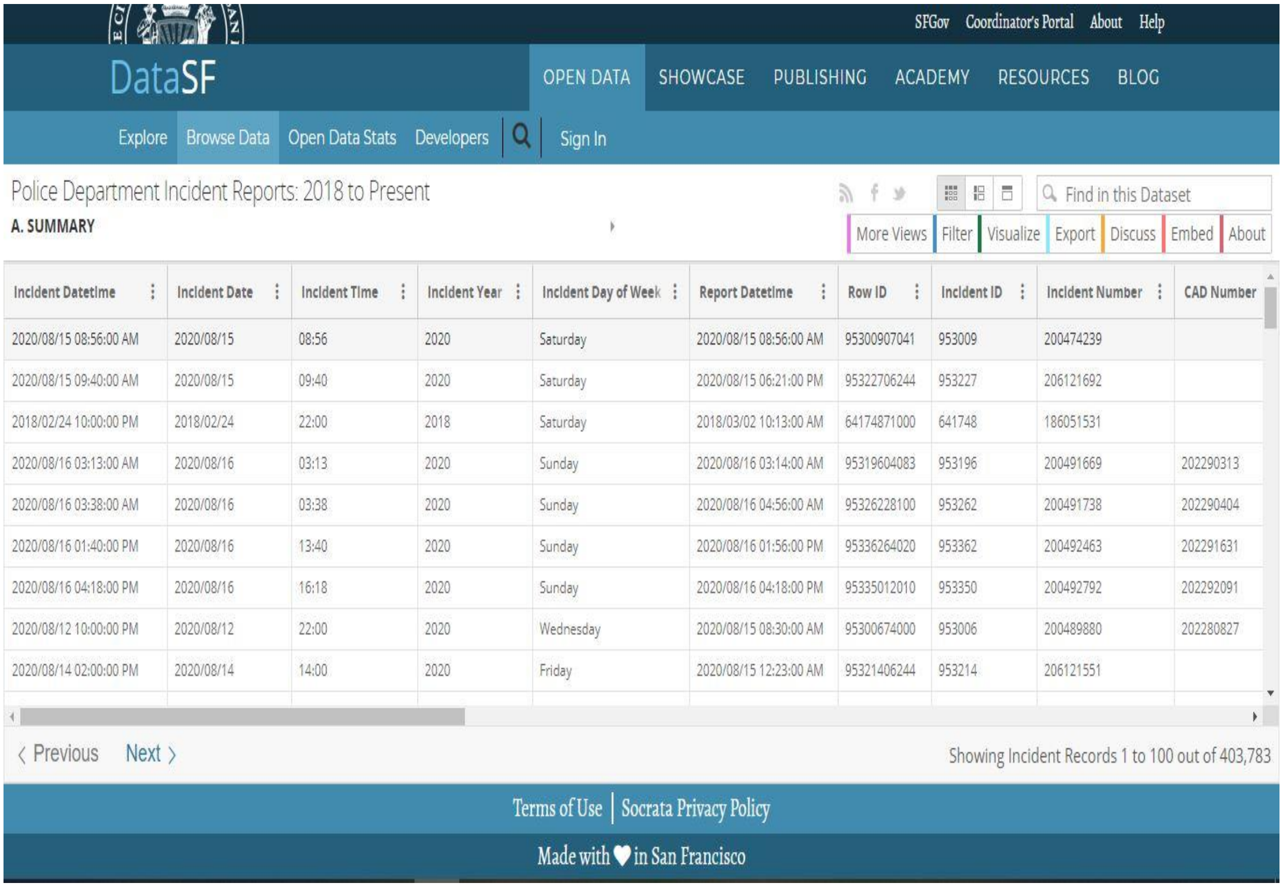


Figure: Conceptual model

We have found several criminal record datasets. There are so many research work have done with the datasets, but we are looking for some real data for better analysis. Then we found a criminal database in DataSF which is submitted by the police department. SF OpenData is the City and County of San Francisco's official open data repository and is a component of DataSF, the official open data program. The platform, launched in 2009, provides developers, researchers, residents, and more with hundreds of city datasets to use. We try to find out some pattern of crime such as crime day, crime type.

3.1. Data Collection

Sample: Finally, we select our datasets from DataSF. It has so many data rows, which are near about 0.45 million. We evaluated the entire dataset in our dataset to extract relevant insights and inferences.



Police Department Incident Reports: 2018 to Present

A. SUMMARY

Incident Datetime	Incident Date	Incident Time	Incident Year	Incident Day of Week	Report Datetime	Row ID	Incident ID	Incident Number	CAD Number
2020/08/15 08:56:00 AM	2020/08/15	08:56	2020	Saturday	2020/08/15 08:56:00 AM	95300907041	953009	200474239	
2020/08/15 09:40:00 AM	2020/08/15	09:40	2020	Saturday	2020/08/15 06:21:00 PM	95322706244	953227	206121692	
2018/02/24 10:00:00 PM	2018/02/24	22:00	2018	Saturday	2018/03/02 10:13:00 AM	64174871000	641748	186051531	
2020/08/16 03:13:00 AM	2020/08/16	03:13	2020	Sunday	2020/08/16 03:14:00 AM	95319604083	953196	200491669	202290313
2020/08/16 03:38:00 AM	2020/08/16	03:38	2020	Sunday	2020/08/16 04:56:00 AM	95326228100	953262	200491738	202290404
2020/08/16 01:40:00 PM	2020/08/16	13:40	2020	Sunday	2020/08/16 01:56:00 PM	95336264020	953362	200492463	202291631
2020/08/16 04:18:00 PM	2020/08/16	16:18	2020	Sunday	2020/08/16 04:18:00 PM	95335012010	953350	200492792	202292091
2020/08/12 10:00:00 PM	2020/08/12	22:00	2020	Wednesday	2020/08/15 08:30:00 AM	95300674000	953006	200489880	202280827
2020/08/14 02:00:00 PM	2020/08/14	14:00	2020	Friday	2020/08/15 12:23:00 AM	95321406244	953214	206121551	

Showing Incident Records 1 to 100 out of 403,783

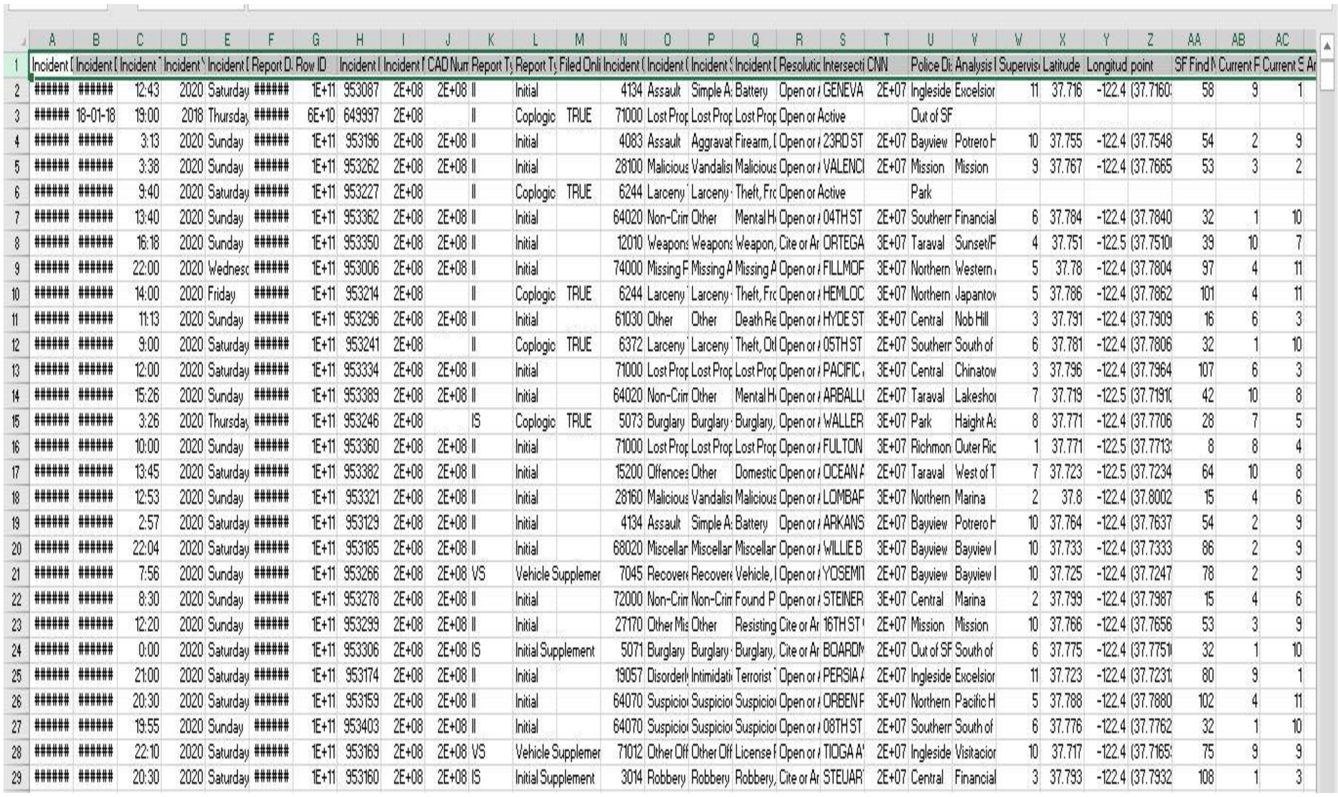
Figure: Sample Dataset

Explore: By visualizing the dataset in Weka, we conducted exploratory analysis to research descriptive statistics and define relationships in the dataset between the attributes that have enabled us to understand the dataset a concise way.

Model: Using unsupervised learning techniques such as the K-Means Clustering algorithm and the Apriori algorithm, we developed predictive models to analyze underlying data patterns and associations.

3.2. Data Preprocessing

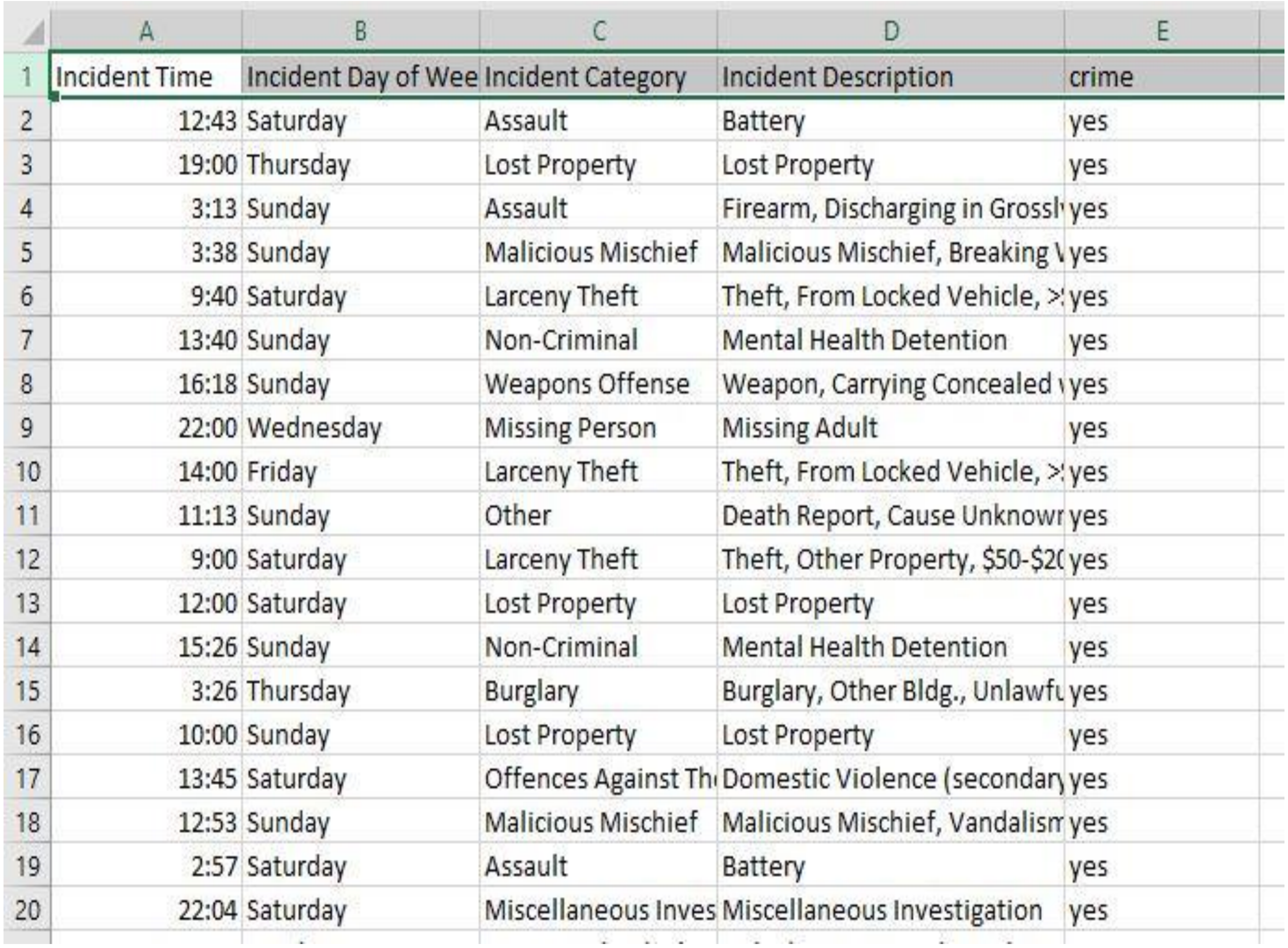
There are so many attributes in our sample datasets. The attributes are Incident Date, Incident Time, Incident Year, Incident Day of Week, Report Datetime, Incident ID, Incident Number, CAD Number, Report Type Description, Incident Code, Incident Category, Incident Subcategory, Incident Description, Resolution, Intersection, Police District, Analysis Neighborhood and 24 others attributes also.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	
1	Incident [	Incident [	Incident '	Incident '	Incident [	Report D.	Row ID	Incident I	Incident [	CAD Num	Report Ty	Report Ty	Filed Onli	Incident (	Incident (	Incident S	Incident [	Resolutic	Intersecti	CNN	Police Di:	Analysis [	Supervis	Latitude	Longitud	point	SF Find N	Current F	Current S	Ar
2	######	######	12:43	2020	Saturday	######	1E+11	953087	2E+08	2E+08	II	Initial		4134	Assault	Simple A:	Battery	Open or /	GENEVA	2E+07	Ingleside	Excelsior	11	37.716	-122.4	(37.7160:	58	9	1	
3	######	18-01-18	19:00	2018	Thursday	######	6E+10	649997	2E+08		II	Coplogic	TRUE	71000	Lost Prop	Lost Prop	Lost Prop	Open or Active			Out of SF									
4	######	######	3:13	2020	Sunday	######	1E+11	953196	2E+08	2E+08	II	Initial		4083	Assault	Aggravat	Firearm, [	Open or /	23RD ST	2E+07	Bayview	Potrero H	10	37.755	-122.4	(37.7548	54	2	9	
5	######	######	3:38	2020	Sunday	######	1E+11	953262	2E+08	2E+08	II	Initial		28100	Malicious	Vandalisi	Malicious	Open or /	VALENCI	2E+07	Mission	Mission	9	37.767	-122.4	(37.7665	53	3	2	
6	######	######	9:40	2020	Saturday	######	1E+11	953227	2E+08		II	Coplogic	TRUE	6244	Larceny	Larceny ·	Theft, Frc	Open or Active			Park									
7	######	######	13:40	2020	Sunday	######	1E+11	953362	2E+08	2E+08	II	Initial		64020	Non-Crim	Other	Mental H	Open or /	04TH ST	2E+07	Southerr	Financial	6	37.784	-122.4	(37.7840	32	1	10	
8	######	######	16:18	2020	Sunday	######	1E+11	953350	2E+08	2E+08	II	Initial		12010	Weapons	Weapons	Weapon,	Cite or Ar	ORTEGA	3E+07	Taraval	Sunset/F	4	37.751	-122.5	(37.7510	39	10	7	
9	######	######	22:00	2020	Wednesc	######	1E+11	953006	2E+08	2E+08	II	Initial		74000	Missing F	Missing A	Missing A	Open or /	FILLMOF	3E+07	Northern	Western ,	5	37.78	-122.4	(37.7804	97	4	11	
10	######	######	14:00	2020	Friday	######	1E+11	953214	2E+08		II	Coplogic	TRUE	6244	Larceny	Larceny ·	Theft, Frc	Open or /	HEMLOC	3E+07	Northern	Japantov	5	37.786	-122.4	(37.7862	101	4	11	
11	######	######	11:13	2020	Sunday	######	1E+11	953296	2E+08	2E+08	II	Initial		61030	Other	Other	Death Re	Open or /	HYDE ST	3E+07	Central	Nob Hill	3	37.791	-122.4	(37.7909	16	6	3	
12	######	######	9:00	2020	Saturday	######	1E+11	953241	2E+08		II	Coplogic	TRUE	6372	Larceny	Larceny	Theft, Otl	Open or /	05TH ST	2E+07	Southerr	South of	6	37.781	-122.4	(37.7806	32	1	10	
13	######	######	12:00	2020	Saturday	######	1E+11	953334	2E+08	2E+08	II	Initial		71000	Lost Prop	Lost Prop	Lost Prop	Open or /	PACIFIC ,	3E+07	Central	Chinatow	3	37.796	-122.4	(37.7964	107	6	3	
14	######	######	15:26	2020	Sunday	######	1E+11	953389	2E+08	2E+08	II	Initial		64020	Non-Crim	Other	Mental H	Open or /	ARBALLI	2E+07	Taraval	Lakeshoi	7	37.719	-122.5	(37.7191(	42	10	8	
15	######	######	3:26	2020	Thursday	######	1E+11	953246	2E+08		IS	Coplogic	TRUE	5073	Burglary	Burglary ·	Burglary,	Open or /	WALLER	3E+07	Park	Haight As	8	37.771	-122.4	(37.7706	28	7	5	
16	######	######	10:00	2020	Sunday	######	1E+11	953360	2E+08	2E+08	II	Initial		71000	Lost Prop	Lost Prop	Lost Prop	Open or /	FULTON	3E+07	Richmon	Outer Ric	1	37.771	-122.5	(37.7713	8	8	4	
17	######	######	13:45	2020	Saturday	######	1E+11	953382	2E+08	2E+08	II	Initial		15200	Offences	Other	Domestic	Open or /	OCEAN A	2E+07	Taraval	West of T	7	37.723	-122.5	(37.7234	64	10	8	
18	######	######	12:53	2020	Sunday	######	1E+11	953321	2E+08	2E+08	II	Initial		28160	Malicious	Vandalisi	Malicious	Open or /	LOMBAF	3E+07	Northern	Marina	2	37.8	-122.4	(37.8002	15	4	6	
19	######	######	2:57	2020	Saturday	######	1E+11	953129	2E+08	2E+08	II	Initial		4134	Assault	Simple A:	Battery	Open or /	ARKANS	2E+07	Bayview	Potrero H	10	37.764	-122.4	(37.7637	54	2	9	
20	######	######	22:04	2020	Saturday	######	1E+11	953185	2E+08	2E+08	II	Initial		68020	Miscellar	Miscellar	Miscellar	Open or /	WILLIE B	3E+07	Bayview	Bayview I	10	37.733	-122.4	(37.7333	86	2	9	
21	######	######	7:56	2020	Sunday	######	1E+11	953266	2E+08	2E+08	VS	Vehicle Supplemer		7045	Recovere	Recovere	Vehicle, I	Open or /	YOSEMI	2E+07	Bayview	Bayview I	10	37.725	-122.4	(37.7247	78	2	9	
22	######	######	8:30	2020	Sunday	######	1E+11	953278	2E+08	2E+08	II	Initial		72000	Non-Crim	Non-Crim	Found P	Open or /	STEINER	3E+07	Central	Marina	2	37.799	-122.4	(37.7987	15	4	6	
23	######	######	12:20	2020	Sunday	######	1E+11	953299	2E+08	2E+08	II	Initial		27170	Other Mis	Other	Resisting	Cite or Ar	16TH ST '	2E+07	Mission	Mission	10	37.766	-122.4	(37.7656	53	3	9	
24	######	######	0:00	2020	Saturday	######	1E+11	953306	2E+08	2E+08	IS	Initial Supplement		5071	Burglary	Burglary ·	Burglary,	Cite or Ar	BOARDN	2E+07	Out of SF	South of	6	37.775	-122.4	(37.7751	32	1	10	
25	######	######	21:00	2020	Saturday	######	1E+11	953174	2E+08	2E+08	II	Initial		19057	Disorderl	Intimidati	Terrorist '	Open or /	PERSIA /	2E+07	Ingleside	Excelsior	11	37.723	-122.4	(37.7231	80	9	1	
26	######	######	20:30	2020	Saturday	######	1E+11	953159	2E+08	2E+08	II	Initial		64070	Suspicioi	Suspicioi	Suspicioi	Open or /	ORBEN F	3E+07	Northern	Pacific H	5	37.788	-122.4	(37.7880	102	4	11	
27	######	######	19:55	2020	Sunday	######	1E+11	953403	2E+08	2E+08	II	Initial		64070	Suspicioi	Suspicioi	Suspicioi	Open or /	08TH ST	2E+07	Southerr	South of	6	37.776	-122.4	(37.7762	32	1	10	
28	######	######	22:10	2020	Saturday	######	1E+11	953169	2E+08	2E+08	VS	Vehicle Supplemer		71012	Other Off	Other Off	License I	Open or /	TIOGA A'	2E+07	Ingleside	Visitacior	10	37.717	-122.4	(37.7165	75	9	9	
29	######	######	20:30	2020	Saturday	######	1E+11	953160	2E+08	2E+08	IS	Initial Supplement		3014	Robbery	Robbery	Robbery,	Cite or Ar	STEUAR	2E+07	Central	Financial	3	37.793	-122.4	(37.7932	108	1	3	

Figure: Dataset

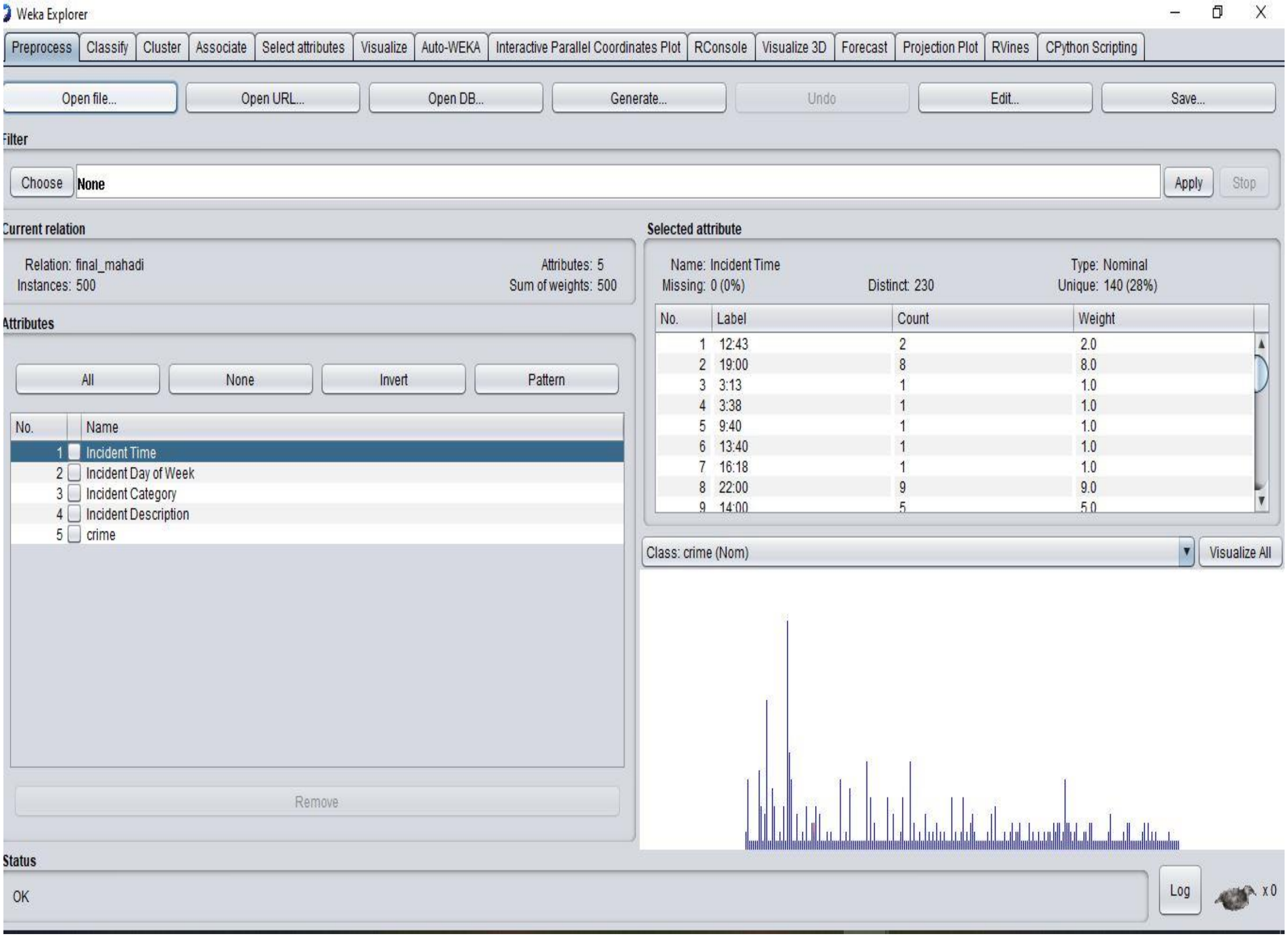
For reaching our destiny, we do not need all of the attributes. From 41 attributes we have selected only 4 attributes for our dataset. The attributes are Incident time, Incident day, Incident category, Incident description. We have also added an attribute which indicate the crime is occur or not. Finally, we select 5 attributes for our final datasets.



	A	B	C	D	E
1	Incident Time	Incident Day of Wee	Incident Category	Incident Description	crime
2	12:43	Saturday	Assault	Battery	yes
3	19:00	Thursday	Lost Property	Lost Property	yes
4	3:13	Sunday	Assault	Firearm, Discharging in Grossl	yes
5	3:38	Sunday	Malicious Mischief	Malicious Mischief, Breaking \	yes
6	9:40	Saturday	Larceny Theft	Theft, From Locked Vehicle, >:	yes
7	13:40	Sunday	Non-Criminal	Mental Health Detention	yes
8	16:18	Sunday	Weapons Offense	Weapon, Carrying Concealed \	yes
9	22:00	Wednesday	Missing Person	Missing Adult	yes
10	14:00	Friday	Larceny Theft	Theft, From Locked Vehicle, >:	yes
11	11:13	Sunday	Other	Death Report, Cause Unknowr	yes
12	9:00	Saturday	Larceny Theft	Theft, Other Property, $50-$2(	yes
13	12:00	Saturday	Lost Property	Lost Property	yes
14	15:26	Sunday	Non-Criminal	Mental Health Detention	yes
15	3:26	Thursday	Burglary	Burglary, Other Bldg., Unlawfu	yes
16	10:00	Sunday	Lost Property	Lost Property	yes
17	13:45	Saturday	Offences Against Th	Domestic Violence (secondar)	yes
18	12:53	Sunday	Malicious Mischief	Malicious Mischief, Vandalism	yes
19	2:57	Saturday	Assault	Battery	yes
20	22:04	Saturday	Miscellaneous Inves	Miscellaneous Investigation	yes

Figure: Modified Dataset

We need data pre-processing for more data utilization. We have a large dataset of 0.4 million rows and we cannot load the whole dataset in Weka because the system does not support. After that we select 39 thousand rows for our analysis. But there also have some problem that we faced. The problem is when we load the dataset in weka there are so many noisy data which interrupt our analysis. For this problem the prediction result is indescribable. So, we select 500 rows manually for our research. As Apriori algorithm does not support numerical values, so we have decided to take nominal attributes. There are five attributes we have selected such as Incident time, Incident day, Incident category, Incident description, crime.



Filter: None

Current relation

Relation: final_mahadi
Instances: 500
Attributes: 5
Sum of weights: 500

Attributes

No.	Name
1	Incident Time
2	Incident Day of Week
3	Incident Category
4	Incident Description
5	crime

Selected attribute

Name: Incident Time
Missing: 0 (0%)
Distinct: 230
Type: Nominal
Unique: 140 (28%)

No.	Label	Count	Weight
1	12:43	2	2.0
2	19:00	8	8.0
3	3:13	1	1.0
4	3:38	1	1.0
5	9:40	1	1.0
6	13:40	1	1.0
7	16:18	1	1.0
8	22:00	9	9.0
9	14:00	5	5.0

Class: crime (Nom)

Status

OK

Figure: Data Pre-processing

3.3. Data Mining Techniques

Data mining techniques are used to obtain useful data from valuable knowledge. These techniques are mostly supervised and unsupervised by two types.

Supervised learning: In this learning method there are two parts which is named by training Dataset and Test Dataset. The models are created using training data and validated by comparing prediction of the model with that of the unseen test data.

Unsupervised learning: Data mining of unlabeled data is known as unsupervised learning. The dataset is analyzed in this data mining technique class, for underlying patterns and relationships between a dataset's various variables.

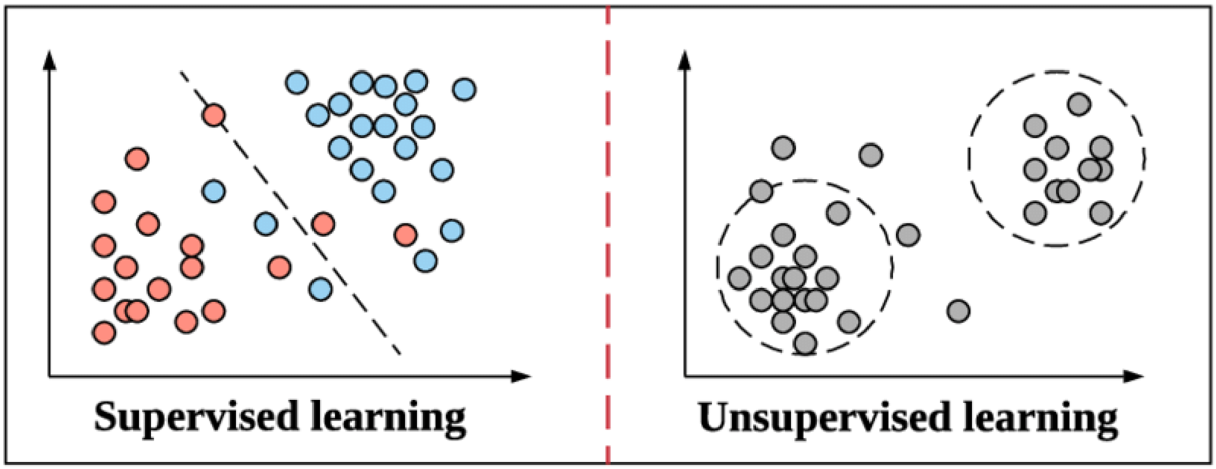


Figure: Supervised and Unsupervised learning

From this types of methods, we use Unsupervised learning method. From Unsupervised learning method we use Apriori algorithm and k-means clustering algorithm for generating some patterns of crime and make association among various attributes in the dataset.

Apriori Algorithm:

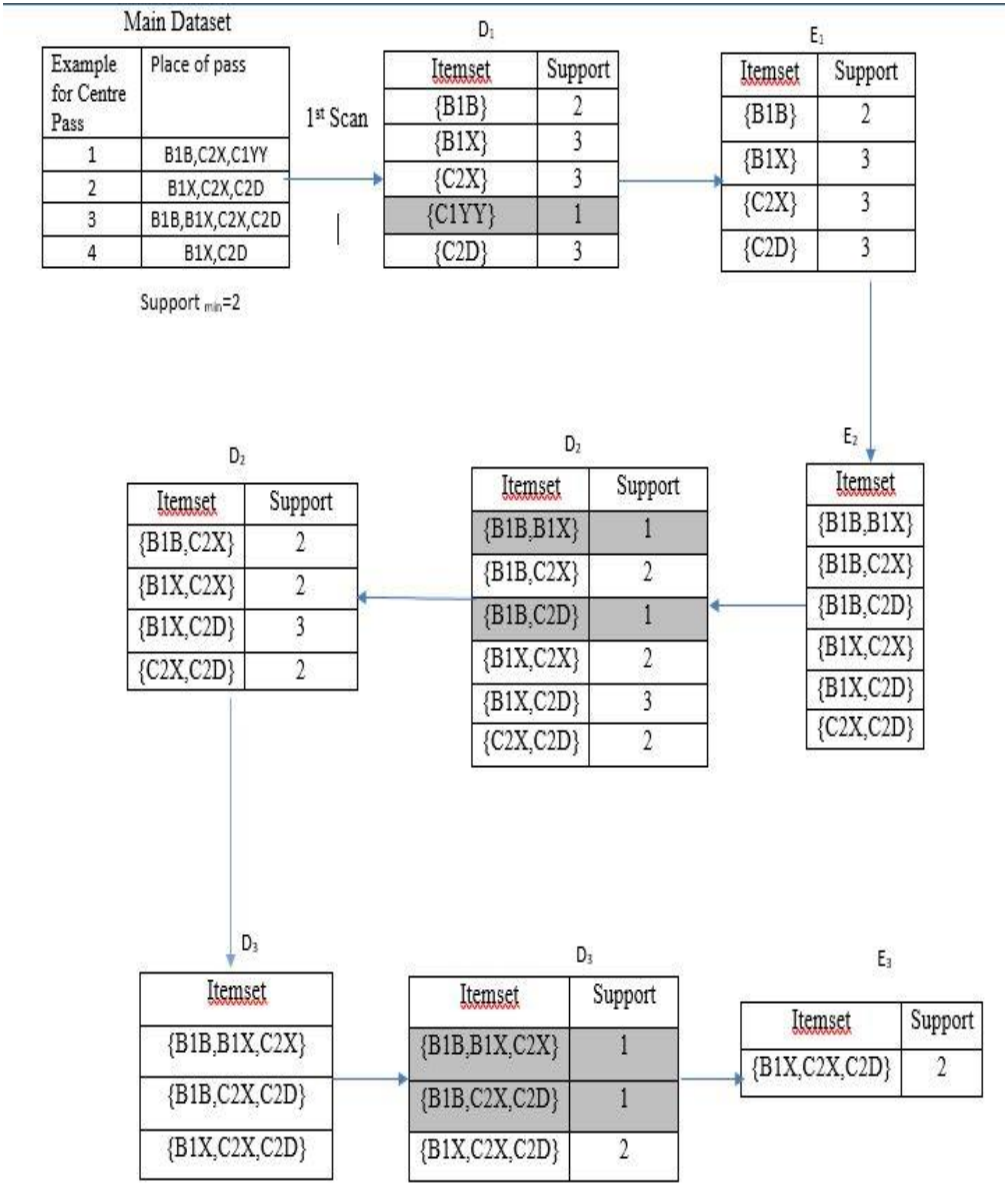
Step 1: Start with reading each item in a dataset.

Step 2: Calculate the support of every item.

Step 3: If support is greater or equal to minimum support then insert items to frequent itemset otherwise remove the item.

Step 4: Now find confidence for each non empty subsets.

Step 5: If confidence greater or equals to minimum confidence then insert to association rules otherwise remove subsets.



Main Dataset

Example for Centre Pass	Place of pass
1	B1B,C2X,C1YY
2	B1X,C2X,C2D
3	B1B,B1X,C2X,C2D
4	B1X,C2D

Support $_{min}$=2

1st Scan

D_1

Itemset	Support
{B1B}	2
{B1X}	3
{C2X}	3
{C1YY}	1
{C2D}	3

E_1

Itemset	Support
{B1B}	2
{B1X}	3
{C2X}	3
{C2D}	3

E_2

Itemset
{B1B,B1X}
{B1B,C2X}
{B1B,C2D}
{B1X,C2X}
{B1X,C2D}
{C2X,C2D}

D_2

Itemset	Support
{B1B,B1X}	1
{B1B,C2X}	2
{B1B,C2D}	1
{B1X,C2X}	2
{B1X,C2D}	3
{C2X,C2D}	2

D_2

Itemset	Support
{B1B,C2X}	2
{B1X,C2X}	2
{B1X,C2D}	3
{C2X,C2D}	2

D_3

Itemset
{B1B,B1X,C2X}
{B1B,C2X,C2D}
{B1X,C2X,C2D}

D_3

Itemset	Support
{B1B,B1X,C2X}	1
{B1B,C2X,C2D}	1
{B1X,C2X,C2D}	2

E_3

Itemset	Support
{B1X,C2X,C2D}	2

Procedure of Apriori algorithm

K-means Clustering:

Step 1: Start with selecting the number of cluster center.
Step 2: In these step set initial cluster center randomly.
Step 3: Putting object to nearest cluster center.
Step 4: Now recalculate the new cluster center.
Step 5: Creating cluster based on smallest distance.
Step 6: Objects move to cluster and get output.

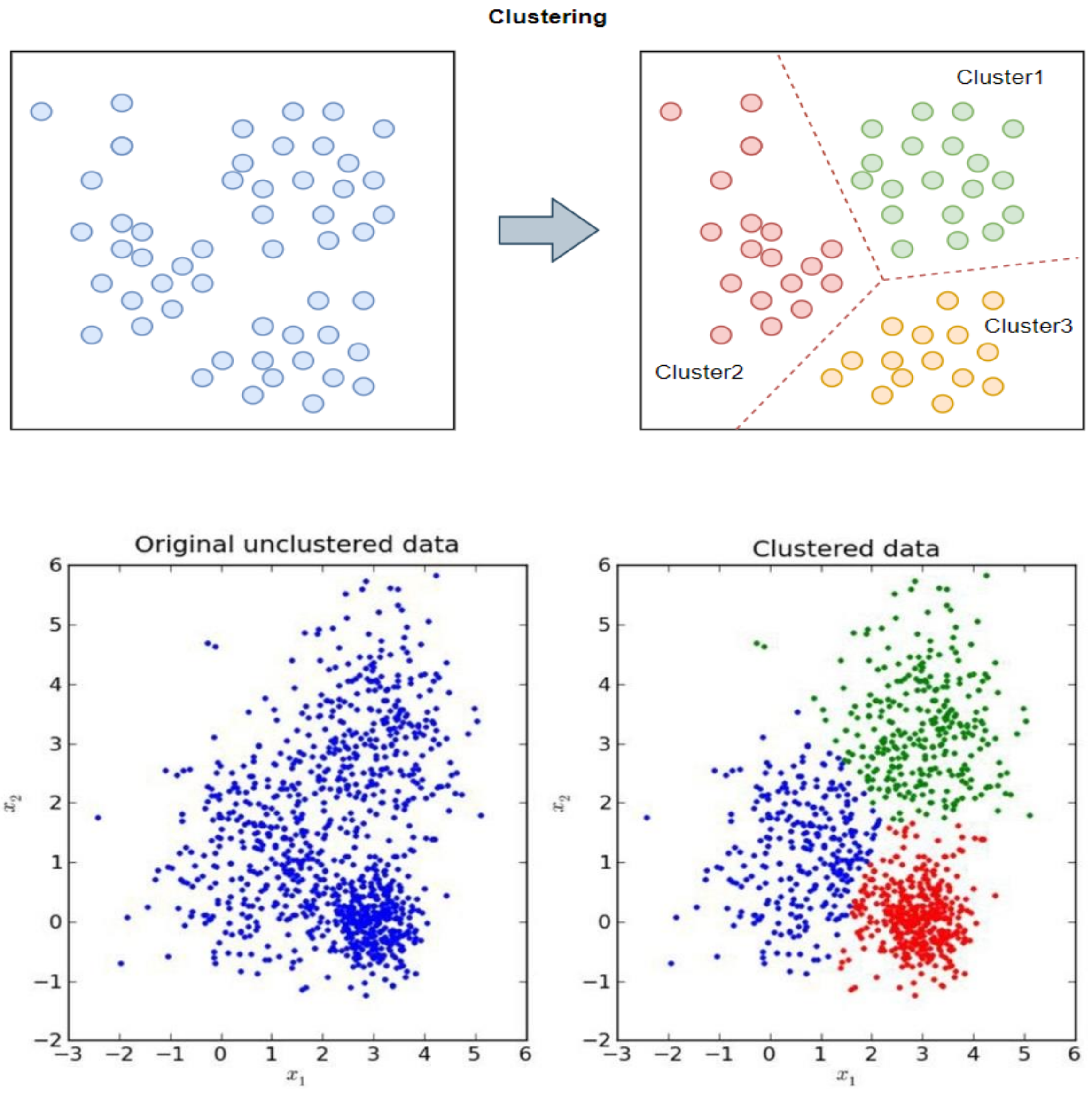


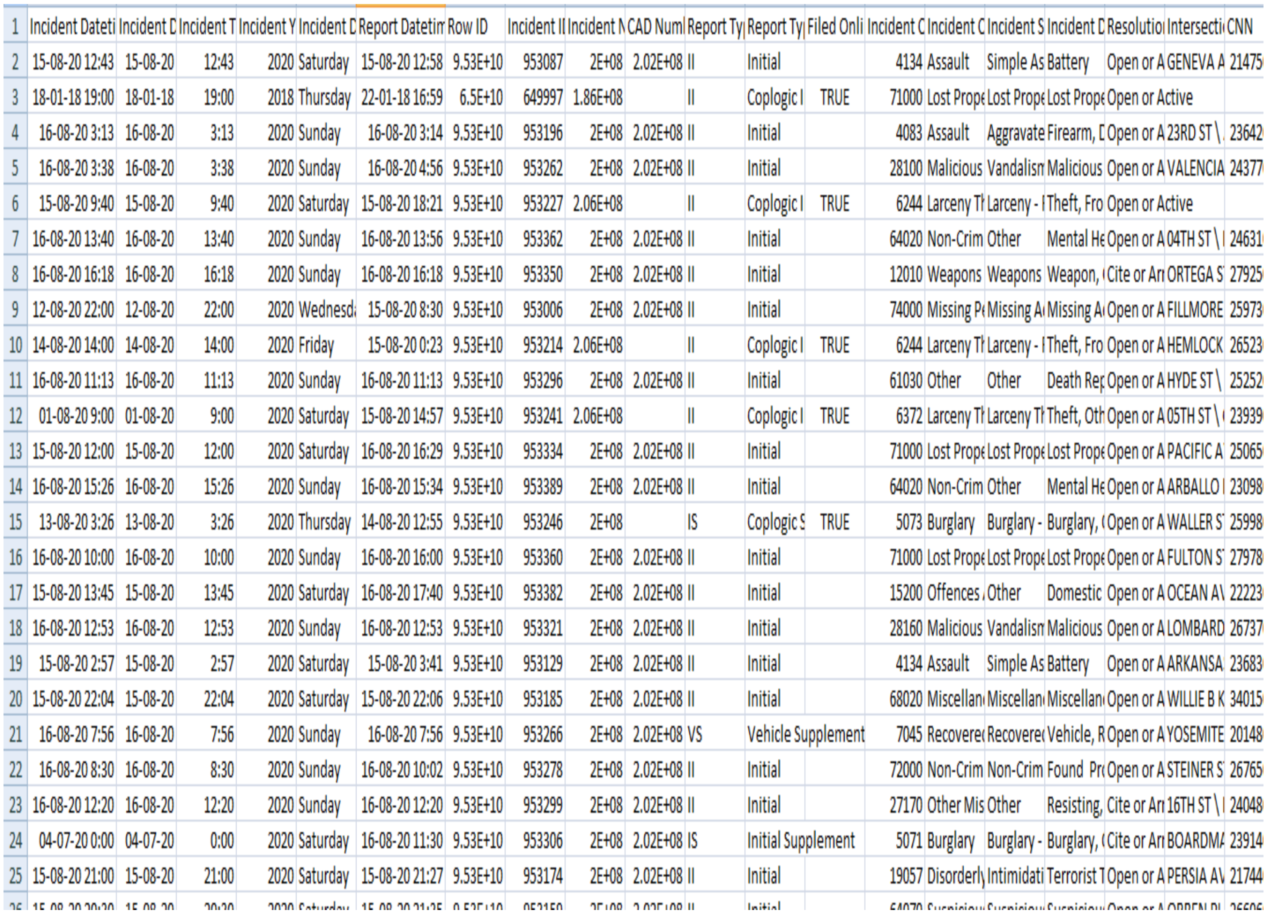
Figure: Clustering algorithm

Chapter 4

4. Result and Discussion

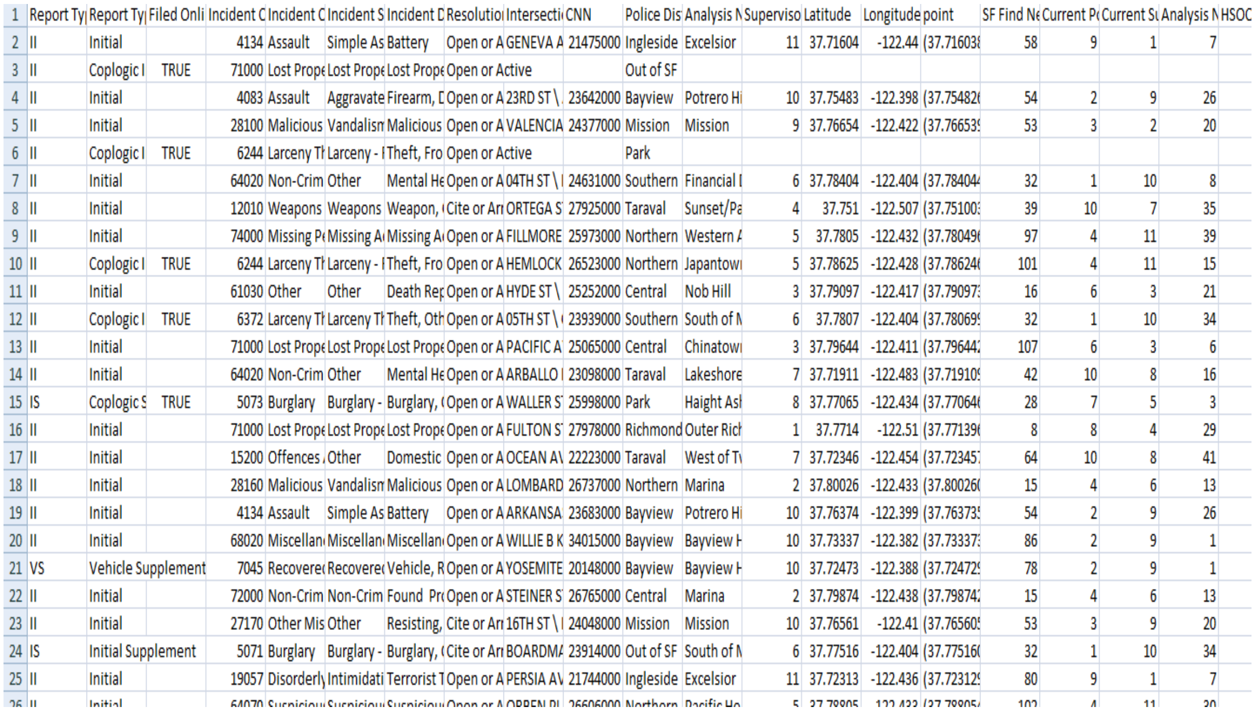
4.1 Result Analysis

Using the given below dataset, some valuable crime detection patterns have been formed. In our dataset there are 40 thousand instances and more than 15 attributes. For our research, we narrow down our focus into a specific point to figure out the best out come to help police department. So we worked with 5 special attributes (Incident time, Incident day of week, Incident category, Incident description, crime) and 500 instances.



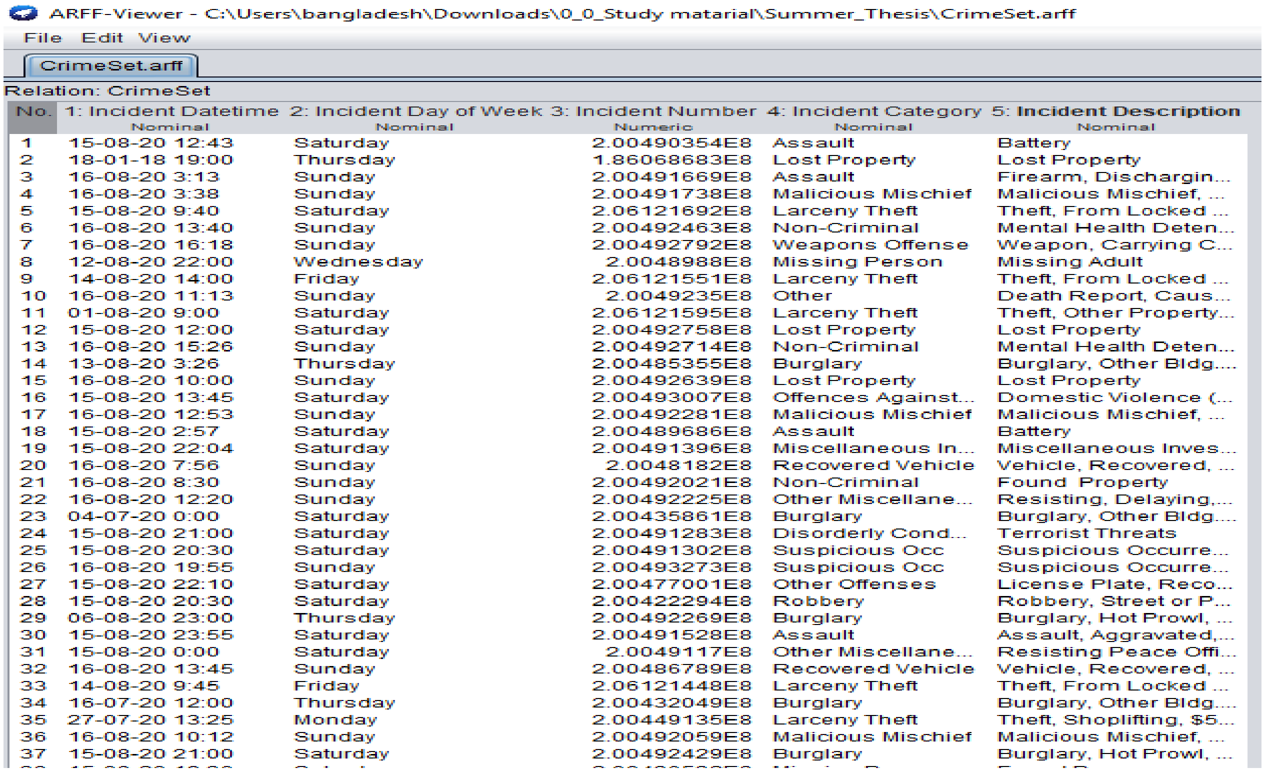
1	Incident Dateti	Incident D	Incident T	Incident Y	Incident D	Report Datetim	Row ID	Incident I	Incident N	CAD Num	Report Ty	Report Ty	Filed Onli	Incident C	Incident C	Incident S	Incident D	Resolutio	Intersecti	CNN
2	15-08-20 12:43	15-08-20	12:43	2020	Saturday	15-08-20 12:58	9.53E+10	953087	2E+08	2.02E+08	II	Initial		4134	Assault	Simple As	Battery	Open or A	GENEVA A	21475
3	18-01-18 19:00	18-01-18	19:00	2018	Thursday	22-01-18 16:59	6.5E+10	649997	1.86E+08		II	Coplogic I	TRUE	71000	Lost Prope	Lost Prope	Lost Prope	Open or Active		
4	16-08-20 3:13	16-08-20	3:13	2020	Sunday	16-08-20 3:14	9.53E+10	953196	2E+08	2.02E+08	II	Initial		4083	Assault	Aggravate	Firearm, D	Open or A	23RD ST \	23642
5	16-08-20 3:38	16-08-20	3:38	2020	Sunday	16-08-20 4:56	9.53E+10	953262	2E+08	2.02E+08	II	Initial		28100	Malicious	Vandalism	Malicious	Open or A	VALENCIA	24377
6	15-08-20 9:40	15-08-20	9:40	2020	Saturday	15-08-20 18:21	9.53E+10	953227	2.06E+08		II	Coplogic I	TRUE	6244	Larceny Th	Larceny -	Theft, Fro	Open or Active		
7	16-08-20 13:40	16-08-20	13:40	2020	Sunday	16-08-20 13:56	9.53E+10	953362	2E+08	2.02E+08	II	Initial		64020	Non-Crim	Other	Mental He	Open or A	04TH ST \	24631
8	16-08-20 16:18	16-08-20	16:18	2020	Sunday	16-08-20 16:18	9.53E+10	953350	2E+08	2.02E+08	II	Initial		12010	Weapons	Weapons	Weapon,	Cite or Ar	ORTEGA S	27925
9	12-08-20 22:00	12-08-20	22:00	2020	Wednesd	15-08-20 8:30	9.53E+10	953006	2E+08	2.02E+08	II	Initial		74000	Missing P	Missing A	Missing A	Open or A	FILLMORE	25973
10	14-08-20 14:00	14-08-20	14:00	2020	Friday	15-08-20 0:23	9.53E+10	953214	2.06E+08		II	Coplogic I	TRUE	6244	Larceny Th	Larceny -	Theft, Fro	Open or A	HEMLOCK	26523
11	16-08-20 11:13	16-08-20	11:13	2020	Sunday	16-08-20 11:13	9.53E+10	953296	2E+08	2.02E+08	II	Initial		61030	Other	Other	Death Rep	Open or A	HYDE ST \	25252
12	01-08-20 9:00	01-08-20	9:00	2020	Saturday	15-08-20 14:57	9.53E+10	953241	2.06E+08		II	Coplogic I	TRUE	6372	Larceny Th	Larceny Th	Theft, Oth	Open or A	05TH ST \	23939
13	15-08-20 12:00	15-08-20	12:00	2020	Saturday	16-08-20 16:29	9.53E+10	953334	2E+08	2.02E+08	II	Initial		71000	Lost Prope	Lost Prope	Lost Prope	Open or A	PACIFIC A	25065
14	16-08-20 15:26	16-08-20	15:26	2020	Sunday	16-08-20 15:34	9.53E+10	953389	2E+08	2.02E+08	II	Initial		64020	Non-Crim	Other	Mental He	Open or A	ARBALLO	23098
15	13-08-20 3:26	13-08-20	3:26	2020	Thursday	14-08-20 12:55	9.53E+10	953246	2E+08		IS	Coplogic S	TRUE	5073	Burglary	Burglary -	Burglary,	Open or A	WALLER S	25998
16	16-08-20 10:00	16-08-20	10:00	2020	Sunday	16-08-20 16:00	9.53E+10	953360	2E+08	2.02E+08	II	Initial		71000	Lost Prope	Lost Prope	Lost Prope	Open or A	FULTON S	27978
17	15-08-20 13:45	15-08-20	13:45	2020	Saturday	16-08-20 17:40	9.53E+10	953382	2E+08	2.02E+08	II	Initial		15200	Offences	Other	Domestic	Open or A	OCEAN AV	22223
18	16-08-20 12:53	16-08-20	12:53	2020	Sunday	16-08-20 12:53	9.53E+10	953321	2E+08	2.02E+08	II	Initial		28160	Malicious	Vandalism	Malicious	Open or A	LOMBARD	26737
19	15-08-20 2:57	15-08-20	2:57	2020	Saturday	15-08-20 3:41	9.53E+10	953129	2E+08	2.02E+08	II	Initial		4134	Assault	Simple As	Battery	Open or A	ARKANSA	23683
20	15-08-20 22:04	15-08-20	22:04	2020	Saturday	15-08-20 22:06	9.53E+10	953185	2E+08	2.02E+08	II	Initial		68020	Miscellan	Miscellan	Miscellan	Open or A	WILLIE B K	34015
21	16-08-20 7:56	16-08-20	7:56	2020	Sunday	16-08-20 7:56	9.53E+10	953266	2E+08	2.02E+08	VS	Vehicle Supplement		7045	Recovere	Recovere	Vehicle, R	Open or A	YOSEMITE	20148
22	16-08-20 8:30	16-08-20	8:30	2020	Sunday	16-08-20 10:02	9.53E+10	953278	2E+08	2.02E+08	II	Initial		72000	Non-Crim	Non-Crim	Found Pr	Open or A	STEINER S	26765
23	16-08-20 12:20	16-08-20	12:20	2020	Sunday	16-08-20 12:20	9.53E+10	953299	2E+08	2.02E+08	II	Initial		27170	Other Mis	Other	Resisting,	Cite or Ar	16TH ST \	24048
24	04-07-20 0:00	04-07-20	0:00	2020	Saturday	16-08-20 11:30	9.53E+10	953306	2E+08	2.02E+08	IS	Initial Supplement		5071	Burglary	Burglary -	Burglary,	Cite or Ar	BOARDMA	23914
25	15-08-20 21:00	15-08-20	21:00	2020	Saturday	15-08-20 21:27	9.53E+10	953174	2E+08	2.02E+08	II	Initial		19057	Disorderl	Intimidati	Terrorist T	Open or A	PERSIA AV	21744
26	15-08-20 20:20	15-08-20	20:20	2020	Saturday	15-08-20 21:25	9.53E+10	953150	2E+08	2.02E+08	II	Initial		64070	Suspiciou	Suspiciou	Suspiciou	Open or A	ORBEN PL	26606

Figure 4.1.1: Actual Dataset



1	Report Ty	Report Ty	Filed Onli	Incident C	Incident C	Incident S	Incident D	Resolutio	Intersecti	CNN	Police Dis	Analysis N	Superviso	Latitude	Longitude	point	SF Find Ne	Current P	Current S	Analysis N	HSOC
2	II	Initial		4134	Assault	Simple As	Battery	Open or A	GENEVA A	21475000	Ingleside	Excelsior	11	37.71604	-122.44	(37.71603	58	9	1	7	
3	II	Coplogic I	TRUE	71000	Lost Prope	Lost Prope	Lost Prope	Open or Active			Out of SF										
4	II	Initial		4083	Assault	Aggravate	Firearm, D	Open or A	23RD ST \	23642000	Bayview	Potrero Hi	10	37.75483	-122.398	(37.75482	54	2	9	26	
5	II	Initial		28100	Malicious	Vandalism	Malicious	Open or A	VALENCIA	24377000	Mission	Mission	9	37.76654	-122.422	(37.76653	53	3	2	20	
6	II	Coplogic I	TRUE	6244	Larceny Th	Larceny -	Theft, Fro	Open or Active			Park										
7	II	Initial		64020	Non-Crim	Other	Mental He	Open or A	04TH ST \	24631000	Southern	Financial	6	37.78404	-122.404	(37.78404	32	1	10	8	
8	II	Initial		12010	Weapons	Weapons	Weapon,	Cite or Ar	ORTEGA S	27925000	Taraval	Sunset/Pa	4	37.751	-122.507	(37.75100	39	10	7	35	
9	II	Initial		74000	Missing Pe	Missing A	Missing A	Open or A	FILLMORE	25973000	Northern	Western A	5	37.7805	-122.432	(37.78049	97	4	11	39	
10	II	Coplogic I	TRUE	6244	Larceny Th	Larceny -	Theft, Fro	Open or A	HEMLOCK	26523000	Northern	Japantow	5	37.78625	-122.428	(37.78624	101	4	11	15	
11	II	Initial		61030	Other	Other	Death Rep	Open or A	HYDE ST \	25252000	Central	Nob Hill	3	37.79097	-122.417	(37.79097	16	6	3	21	
12	II	Coplogic I	TRUE	6372	Larceny Th	Larceny Th	Theft, Oth	Open or A	05TH ST \	23939000	Southern	South of M	6	37.7807	-122.404	(37.78069	32	1	10	34	
13	II	Initial		71000	Lost Prope	Lost Prope	Lost Prope	Open or A	PACIFIC A	25065000	Central	Chinatow	3	37.79644	-122.411	(37.79644	107	6	3	6	
14	II	Initial		64020	Non-Crim	Other	Mental He	Open or A	ARBALLO	23098000	Taraval	Lakeshore	7	37.71911	-122.483	(37.71910	42	10	8	16	
15	IS	Coplogic S	TRUE	5073	Burglary	Burglary -	Burglary,	Open or A	WALLER S	25998000	Park	Haight Ash	8	37.77065	-122.434	(37.77064	28	7	5	3	
16	II	Initial		71000	Lost Prope	Lost Prope	Lost Prope	Open or A	FULTON S	27978000	Richmond	Outer Rich	1	37.7714	-122.51	(37.77139	8	8	4	29	
17	II	Initial		15200	Offences	Other	Domestic	Open or A	OCEAN AV	22223000	Taraval	West of Tw	7	37.72346	-122.454	(37.72345	64	10	8	41	
18	II	Initial		28160	Malicious	Vandalism	Malicious	Open or A	LOMBARD	26737000	Northern	Marina	2	37.80026	-122.433	(37.80026	15	4	6	13	
19	II	Initial		4134	Assault	Simple As	Battery	Open or A	ARKANSA	23683000	Bayview	Potrero Hi	10	37.76374	-122.399	(37.76373	54	2	9	26	
20	II	Initial		68020	Miscellan	Miscellan	Miscellan	Open or A	WILLIE B K	34015000	Bayview	Bayview H	10	37.73337	-122.382	(37.73337	86	2	9	1	
21	VS	Vehicle Supplement		7045	Recovere	Recovere	Vehicle, R	Open or A	YOSEMITE	20148000	Bayview	Bayview H	10	37.72473	-122.388	(37.72472	78	2	9	1	
22	II	Initial		72000	Non-Crim	Non-Crim	Found Pr	Open or A	STEINER S	26765000	Central	Marina	2	37.79874	-122.438	(37.79874	15	4	6	13	
23	II	Initial		27170	Other Mis	Other	Resisting,	Cite or Ar	16TH ST \	24048000	Mission	Mission	10	37.76561	-122.41	(37.76560	53	3	9	20	
24	IS	Initial Supplement		5071	Burglary	Burglary -	Burglary,	Cite or Ar	BOARDMA	23914000	Out of SF	South of M	6	37.77516	-122.404	(37.77516	32	1	10	34	
25	II	Initial		19057	Disorderly	Intimidati	Terrorist T	Open or A	PERSIA AV	21744000	Ingleside	Excelsior	11	37.72313	-122.436	(37.72312	80	9	1	7	
26	II	Initial		64070	Suspiciou	Suspiciou	Suspiciou	Open or A	ORBEN PL	26606000	Northern	Pacific He	5	37.78805	-122.433	(37.78805	102	4	11	30	

Figure 4.1.2: Actual Dataset



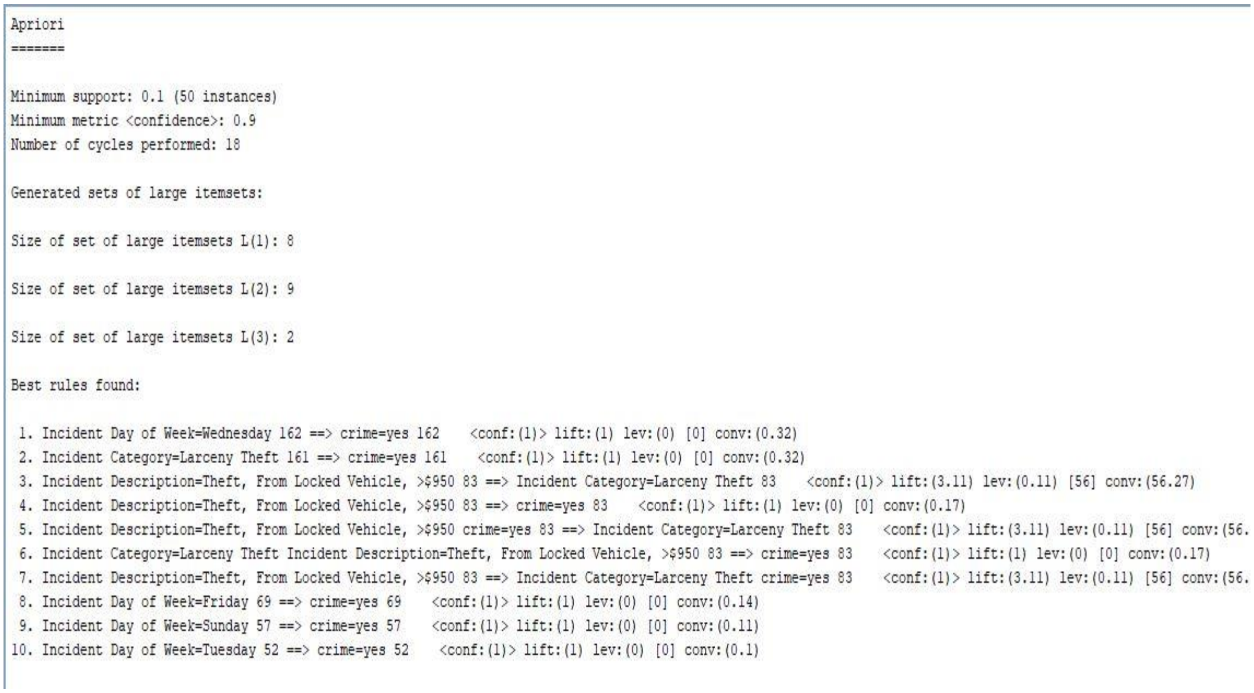
ARFF-Viewer - C:\Users\bangladesh\Downloads\0_0_Study matarial\Summer_Thesis\CrimeSet.arff

Relation: CrimeSet

No.	1: Incident Datetime	2: Incident Day of Week	3: Incident Number	4: Incident Category	5: Incident Description
	Nominal	Nominal	Numeric	Nominal	Nominal
1	15-08-20 12:43	Saturday	2.00490354E8	Assault	Battery
2	18-01-18 19:00	Thursday	1.86068683E8	Lost Property	Lost Property
3	16-08-20 3:13	Sunday	2.00491669E8	Assault	Firearm, Dischargin...
4	16-08-20 3:38	Sunday	2.00491738E8	Malicious Mischief	Malicious Mischief, ...
5	15-08-20 9:40	Saturday	2.06121692E8	Larceny Theft	Theft, From Locked ...
6	16-08-20 13:40	Sunday	2.00492463E8	Non-Criminal	Mental Health Deten...
7	16-08-20 16:18	Sunday	2.00492792E8	Weapons Offense	Weapon, Carrying C...
8	12-08-20 22:00	Wednesday	2.0048988E8	Missing Person	Missing Adult
9	14-08-20 14:00	Friday	2.06121551E8	Larceny Theft	Theft, From Locked ...
10	16-08-20 11:13	Sunday	2.0049235E8	Other	Death Report, Caus...
11	01-08-20 9:00	Saturday	2.06121595E8	Larceny Theft	Theft, Other Property...
12	15-08-20 12:00	Saturday	2.00492758E8	Lost Property	Lost Property
13	16-08-20 15:26	Sunday	2.00492714E8	Non-Criminal	Mental Health Deten...
14	13-08-20 3:26	Thursday	2.00485355E8	Burglary	Burglary, Other Bldg....
15	16-08-20 10:00	Sunday	2.00492639E8	Lost Property	Lost Property
16	15-08-20 13:45	Saturday	2.00493007E8	Offences Against...	Domestic Violence (...
17	16-08-20 12:53	Sunday	2.00492281E8	Malicious Mischief	Malicious Mischief, ...
18	15-08-20 2:57	Saturday	2.00489686E8	Assault	Battery
19	15-08-20 22:04	Saturday	2.00491396E8	Miscellaneous In...	Miscellaneous Inves...
20	16-08-20 7:56	Sunday	2.0048182E8	Recovered Vehicle	Vehicle, Recovered, ...
21	16-08-20 8:30	Sunday	2.00492021E8	Non-Criminal	Found Property
22	16-08-20 12:20	Sunday	2.00492225E8	Other Miscellane...	Resisting, Delaying,...
23	04-07-20 0:00	Saturday	2.00435861E8	Burglary	Burglary, Other Bldg....
24	15-08-20 21:00	Saturday	2.00491283E8	Disorderly Cond...	Terrorist Threats
25	15-08-20 20:30	Saturday	2.00491302E8	Suspicious Occ	Suspicious Occurre...
26	16-08-20 19:55	Sunday	2.00493273E8	Suspicious Occ	Suspicious Occurre...
27	15-08-20 22:10	Saturday	2.00477001E8	Other Offenses	License Plate, Reco...
28	15-08-20 20:30	Saturday	2.00422294E8	Robbery	Robbery, Street or P...
29	06-08-20 23:00	Thursday	2.00492269E8	Burglary	Burglary, Hot Prowl, ...
30	15-08-20 23:55	Saturday	2.00491528E8	Assault	Assault, Aggravated,...
31	15-08-20 0:00	Saturday	2.0049117E8	Other Miscellane...	Resisting Peace Offi...
32	16-08-20 13:45	Sunday	2.00486789E8	Recovered Vehicle	Vehicle, Recovered, ...
33	14-08-20 9:45	Friday	2.06121448E8	Larceny Theft	Theft, From Locked ...
34	16-07-20 12:00	Thursday	2.00432049E8	Burglary	Burglary, Other Bldg....
35	27-07-20 13:25	Monday	2.00449135E8	Larceny Theft	Theft, Shoplifting, $5...
36	16-08-20 10:12	Sunday	2.00492059E8	Malicious Mischief	Malicious Mischief, ...
37	15-08-20 21:00	Saturday	2.00492429E8	Burglary	Burglary, Hot Prowl, ...

Figure 4.1.2: Selected Dataset

Association Rule Mining: We used Association Rule mining Apriori algorithm to generate some pattern. As you can see (Figure 4.1.3) shows some generated patterns.



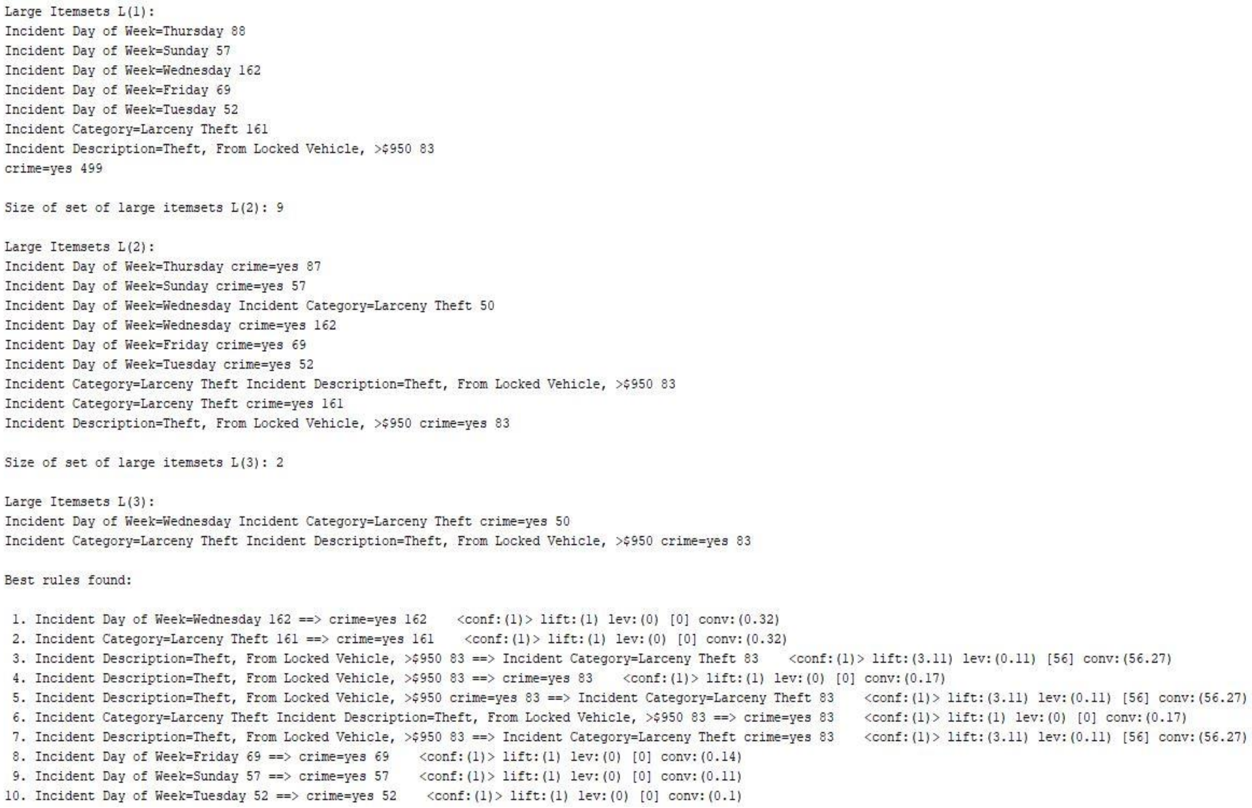
```
Apriori
=======

Minimum support: 0.1 (50 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 9

Size of set of large itemsets L(3): 2

Best rules found:

 1. Incident Day of Week=Wednesday 162 ==> crime=yes 162    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.32)
 2. Incident Category=Larceny Theft 161 ==> crime=yes 161    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.32)
 3. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.27)
 4. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> crime=yes 83    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.17)
 5. Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83 ==> Incident Category=Larceny Theft 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.
 6. Incident Category=Larceny Theft Incident Description=Theft, From Locked Vehicle, >$950 83 ==> crime=yes 83    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.17)
 7. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft crime=yes 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.
 8. Incident Day of Week=Friday 69 ==> crime=yes 69    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.14)
 9. Incident Day of Week=Sunday 57 ==> crime=yes 57    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.11)
10. Incident Day of Week=Tuesday 52 ==> crime=yes 52    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.1)
```

Figure 4.1.3: Apriori Algorithm metric type confidence

For five hundred instances, the number of cycle performed is 18. The minimum support founds 0.1 and the maximum confidence founds 0.9. The number of best rules N was selected 10. Total 10 best rules are found from this analysis. The result is based on confidence metric type and top best rules are:

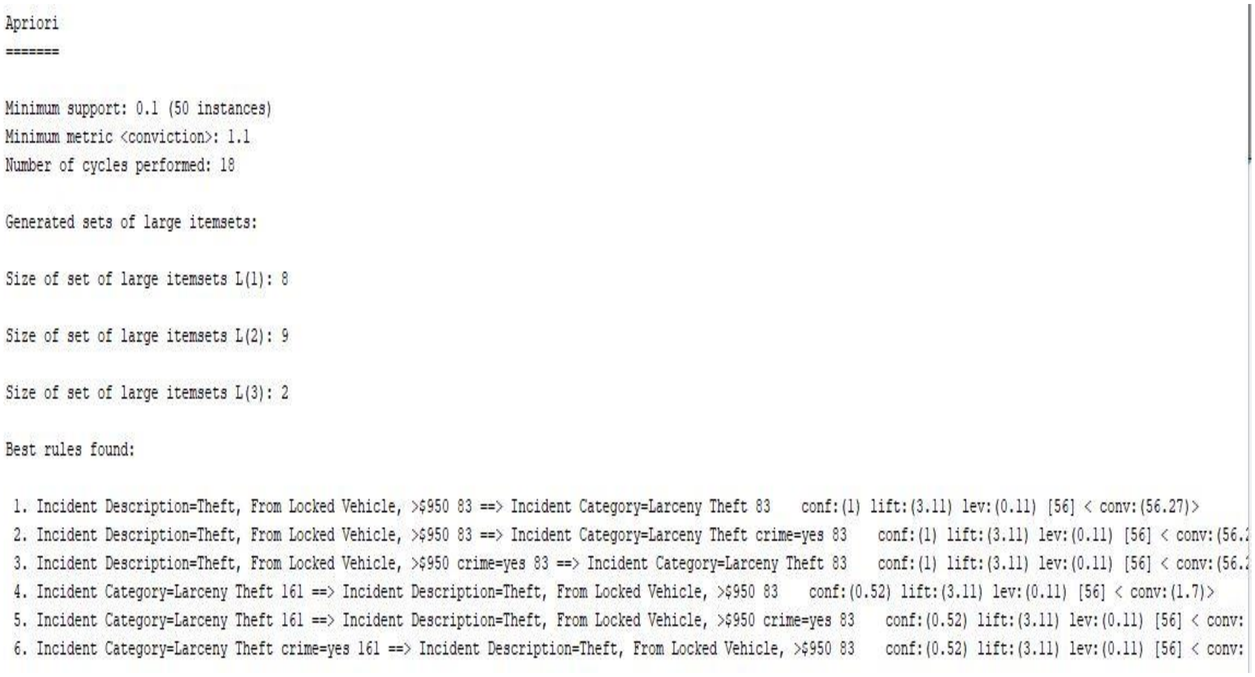
- The most crime happened on Wednesday about 162 offences.
- The most crime category is larceny theft it’s about 161 cases
- The most crime description found is theft from locked vehicles

This association rule mining suggests us a possible pattern for detecting crime before it happens that is when the day is Wednesday there is some possibility of happing crime like theft from locked vehicles. If we narrow down the pattern, we can say there must be high change of happing theft offense in this particular area. The confidence of each rules found approximately 1. For showing large item sets in details we turned on the output item set “True”. In below (figure 4.1.4) large item set are described accordingly.



```
Large Itemsets L(1):
Incident Day of Week=Thursday 88
Incident Day of Week=Sunday 57
Incident Day of Week=Wednesday 162
Incident Day of Week=Friday 69
Incident Day of Week=Tuesday 52
Incident Category=Larceny Theft 161
Incident Description=Theft, From Locked Vehicle, >$950 83
crime=yes 499

Size of set of large itemsets L(2): 9

Large Itemsets L(2):
Incident Day of Week=Thursday crime=yes 87
Incident Day of Week=Sunday crime=yes 57
Incident Day of Week=Wednesday Incident Category=Larceny Theft 50
Incident Day of Week=Wednesday crime=yes 162
Incident Day of Week=Friday crime=yes 69
Incident Day of Week=Tuesday crime=yes 52
Incident Category=Larceny Theft Incident Description=Theft, From Locked Vehicle, >$950 83
Incident Category=Larceny Theft crime=yes 161
Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83

Size of set of large itemsets L(3): 2

Large Itemsets L(3):
Incident Day of Week=Wednesday Incident Category=Larceny Theft crime=yes 50
Incident Category=Larceny Theft Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83

Best rules found:

 1. Incident Day of Week=Wednesday 162 ==> crime=yes 162    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.32)
 2. Incident Category=Larceny Theft 161 ==> crime=yes 161    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.32)
 3. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.27)
 4. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> crime=yes 83    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.17)
 5. Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83 ==> Incident Category=Larceny Theft 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.27)
 6. Incident Category=Larceny Theft Incident Description=Theft, From Locked Vehicle, >$950 83 ==> crime=yes 83    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.17)
 7. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft crime=yes 83    <conf:(1)> lift:(3.11) lev:(0.11) [56] conv:(56.27)
 8. Incident Day of Week=Friday 69 ==> crime=yes 69    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.14)
 9. Incident Day of Week=Sunday 57 ==> crime=yes 57    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.11)
10. Incident Day of Week=Tuesday 52 ==> crime=yes 52    <conf:(1)> lift:(1) lev:(0) [0] conv:(0.1)
```

Figure 4.1.4: Apriori algorithm with detail item set



```
Apriori
=======

Minimum support: 0.1 (50 instances)
Minimum metric <conviction>: 1.1
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 9

Size of set of large itemsets L(3): 2

Best rules found:

 1. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft 83    conf:(1) lift:(3.11) lev:(0.11) [56] < conv:(56.27)>
 2. Incident Description=Theft, From Locked Vehicle, >$950 83 ==> Incident Category=Larceny Theft crime=yes 83    conf:(1) lift:(3.11) lev:(0.11) [56] < conv:(56.2
 3. Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83 ==> Incident Category=Larceny Theft 83    conf:(1) lift:(3.11) lev:(0.11) [56] < conv:(56.2
 4. Incident Category=Larceny Theft 161 ==> Incident Description=Theft, From Locked Vehicle, >$950 83    conf:(0.52) lift:(3.11) lev:(0.11) [56] < conv:(1.7)>
 5. Incident Category=Larceny Theft 161 ==> Incident Description=Theft, From Locked Vehicle, >$950 crime=yes 83    conf:(0.52) lift:(3.11) lev:(0.11) [56] < conv:
 6. Incident Category=Larceny Theft crime=yes 161 ==> Incident Description=Theft, From Locked Vehicle, >$950 83    conf:(0.52) lift:(3.11) lev:(0.11) [56] < conv:
```

Figure 4.1.5: Apriori algorithm metric type conviction

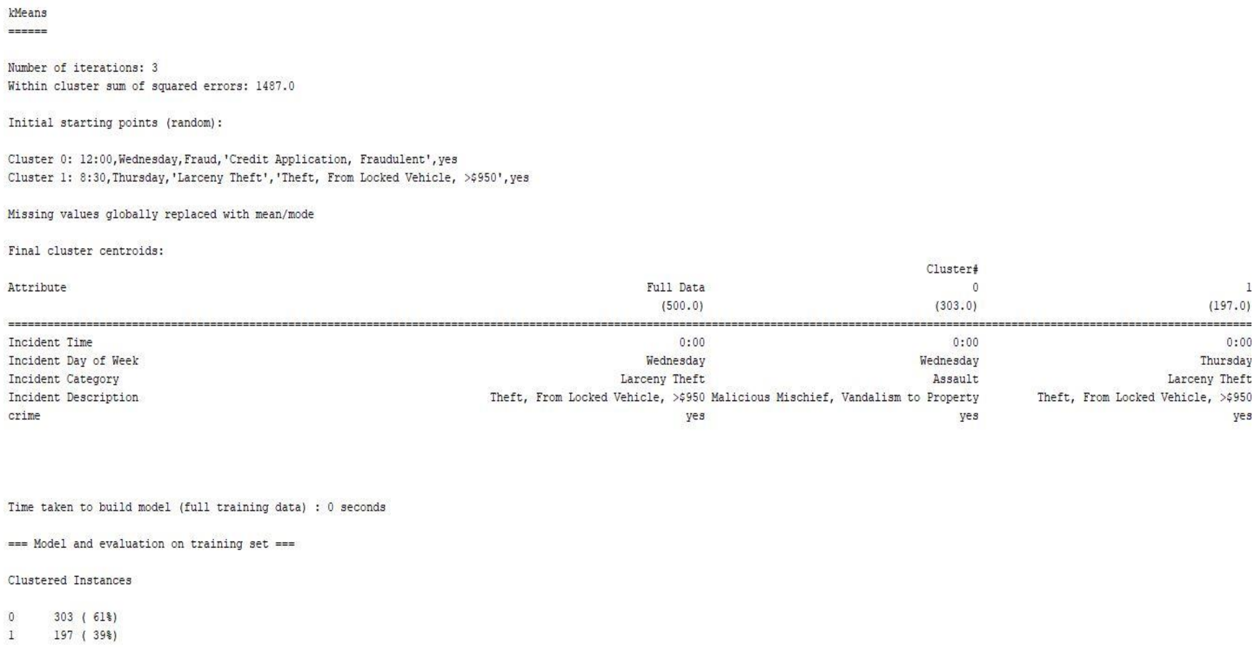
In figure 4.1.3, we found pattern with the metric type confidence. The pattern showed Wednesday is the most probable day for happening crimes related to theft from locked vehicles. In the conviction metric type (figure 4.1.5) same kind of pattern we found. Most probable incident is larceny theft and description is theft from locked vehicles. And surprisingly all the rule found is about theft. So from this kind of pattern we can make a decision that is whatever the day is there must be high chance of happing incident like theft from locked vehicles in this particular area.

Clustering: K-Means clustering is one the method that is used in clustering. For future prediction in the field of data mining and statistics clustering is widely used algorithm. In our dataset after using K-Means clustering we found some pattern that is more specific than the apriori algorithm. In figure 4.1.6, Total 3 iteration has occurred to generate the desire pattern. No of clustering is 2, so two clustering patterns have found here. We ignored the missing or noisy value for this algorithm. Apart from the apriori algorithm this K-Means clustering is indicating the possible time of crime along with the day and crime type.

From the K-Means clustering centroids, there must be a high possible of chance of credit card fraud if the day is Wednesday and the time is 12.00PM. This pattern found from 303(61%) instances. This pattern can make alert the police department to be more cautious about this day and time. Another clustering pattern is if the day is Thursday and the time is 8.PM there must be a high possible of chance for happing incident like theft from locked vehicles. This pattern found from 197(39%) instances.

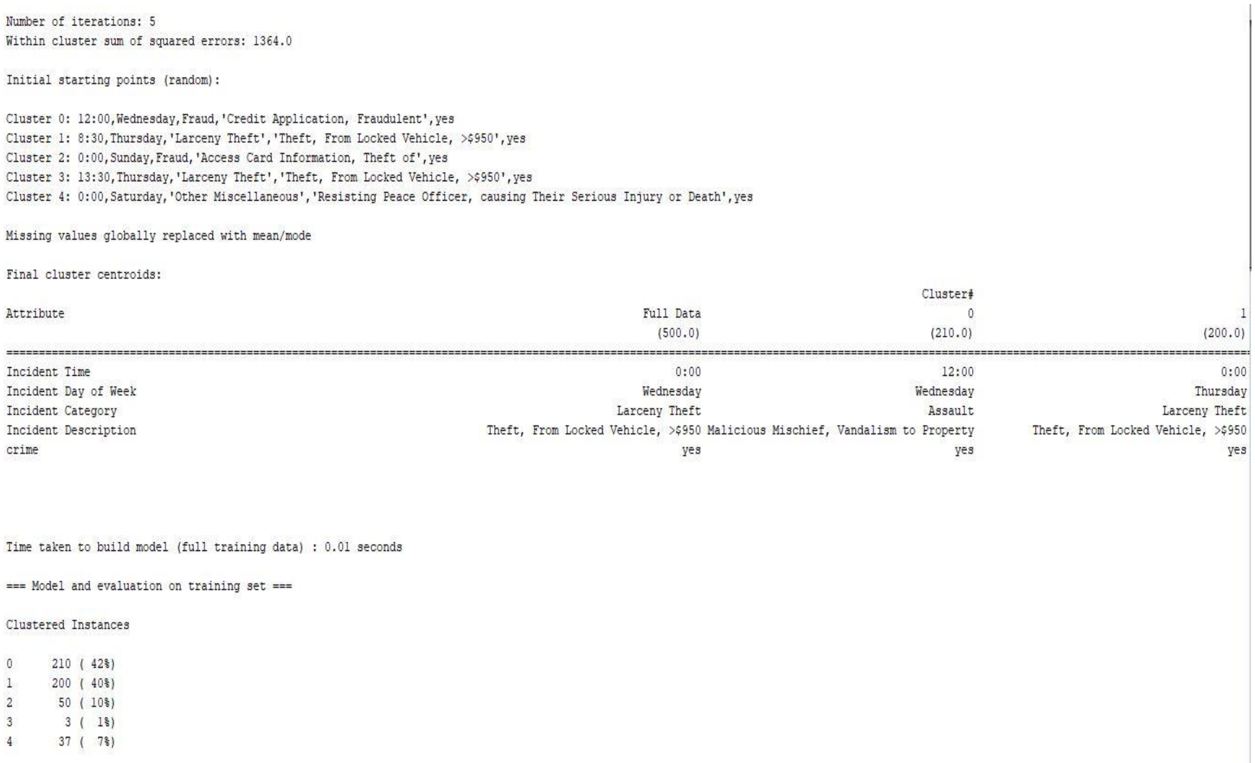
For finding more patterns with K-Means clustering we set number of clustering to five (figure 4.1.7). We found more patterns that are generated from the dataset. Apart from the two clusters we had before, new three patterns also suggest incident day and time with incident type. This time 42% instances suggest that high possibility for happing crime in Wednesday at 12.00PM is credit card fraud. The second cluster with 40% instances made pattern that Thursday at 8.30AM incident like theft from locked vehicles can be happened. Rest of the clustering pattern has low % of instances. So we can ignore them for this time being.

From these two K-Means clustering result we can make a decision that whatever the number of cluster is (2 or 5) the best pattern can never be changed. Though the instance % goes down but the internal result remains same.



```
kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 1487.0

Initial starting points (random):

Cluster 0: 12:00,Wednesday,Fraud,'Credit Application, Fraudulent',yes
Cluster 1: 8:30,Thursday,'Larceny Theft','Theft, From Locked Vehicle, >$950',yes

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                                                                            Cluster#
Attribute                                                          Full Data                                       0                                    1
                                                                     (500.0)                                 (303.0)                              (197.0)
=========================================================================================================================================================
Incident Time                                                           0:00                                    0:00                                 0:00
Incident Day of Week                                               Wednesday                               Wednesday                             Thursday
Incident Category                                              Larceny Theft                                 Assault                        Larceny Theft
Incident Description                      Theft, From Locked Vehicle, >$950 Malicious Mischief, Vandalism to Property    Theft, From Locked Vehicle, >$950
crime                                                                    yes                                     yes                                  yes



Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      303 ( 61%)
1      197 ( 39%)
```

Figure 4.1.6: K-Means clustering with 2 clusters



```
Number of iterations: 5
Within cluster sum of squared errors: 1364.0

Initial starting points (random):

Cluster 0: 12:00,Wednesday,Fraud,'Credit Application, Fraudulent',yes
Cluster 1: 8:30,Thursday,'Larceny Theft','Theft, From Locked Vehicle, >$950',yes
Cluster 2: 0:00,Sunday,Fraud,'Access Card Information, Theft of',yes
Cluster 3: 13:30,Thursday,'Larceny Theft','Theft, From Locked Vehicle, >$950',yes
Cluster 4: 0:00,Saturday,'Other Miscellaneous','Resisting Peace Officer, causing Their Serious Injury or Death',yes

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                                                                            Cluster#
Attribute                                                          Full Data                                       0                                    1
                                                                     (500.0)                                 (210.0)                              (200.0)
=========================================================================================================================================================
Incident Time                                                           0:00                                   12:00                                 0:00
Incident Day of Week                                               Wednesday                               Wednesday                             Thursday
Incident Category                                              Larceny Theft                                 Assault                        Larceny Theft
Incident Description                      Theft, From Locked Vehicle, >$950 Malicious Mischief, Vandalism to Property    Theft, From Locked Vehicle, >$950
crime                                                                    yes                                     yes                                  yes



Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      210 ( 42%)
1      200 ( 40%)
2       50 ( 10%)
3        3 (  1%)
4       37 (  7%)
```

Figure 4.1.7: K-Means clustering with 5 clusters

No. Cluster	Time	Day	Incident type	Incident description	Clustered(%)
0	**12.00**	**Wednesday**	**fraud**	**credit card fraud**	**61%**
1	**8.30**	**Thursday**	**theft**	**Theft from locked vehicle**	**39%**

Table 4.1.1: K-Means cluster with 2 number of clustering

No. Cluster	Time	Day	Incident type	Incident description	Clustered(%)
0	**12.00**	**Wednesday**	**fraud**	**credit card fraud**	**61%**
1	**8.30**	**Thursday**	**theft**	**Theft from locked vehicle**	**39%**
2	**00.00**	**Sunday**	**fraud**	**Access card info**	**10%**
3	**13.30**	**Thursday**	**theft**	**Theft from locked vehicle**	**1%**
4	**00.00**	**Saturday**	**other**	**Miscellaneous**	**7%**

Table 4.1.2: K-Means cluster with 5 number of clustering

4.2 Visualization:

For better understanding the outcome of our research, visualization is more convenient than other approaches. There are two visualization tools. First one is basic 2D visualization and second one is 3D visualization. We used both types of visualization for our dataset. And the outcome is more understandable than others.

Figure 4.1.8, X axis indicates the incident day of week and the Y axis indicates the incident description. Like previous two algorithms we have shown above, most probable crime happens on Wednesday. In visualization we also saw the same pattern. Wednesday is relatively the most offensive day of the week. Thursday and Friday is also in this competition. From this pattern anyone can easily get idea about crime percentage of the day.

Figure 4.1.9, visualization with jitter effect helps to identify the particular period in the interval among two or more effect of a data signal in better way. Each crime description is now more visible and identical than before.

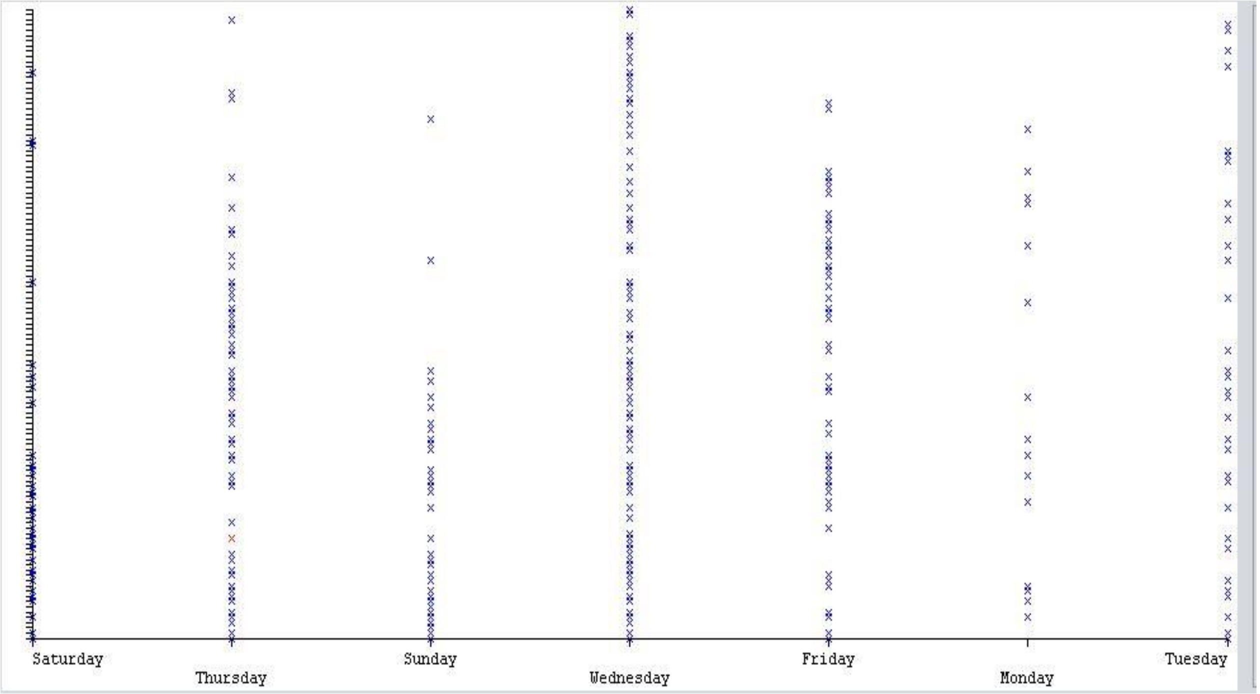


Figure 4.1.8: Visualization

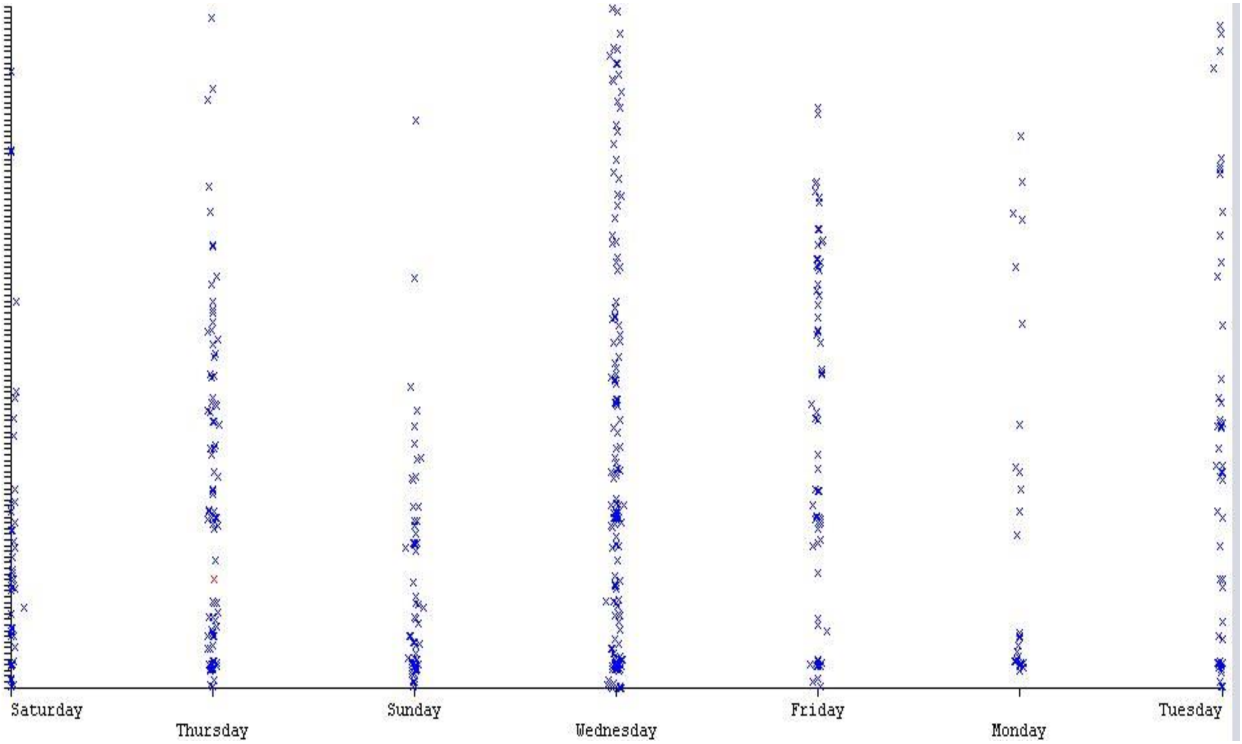
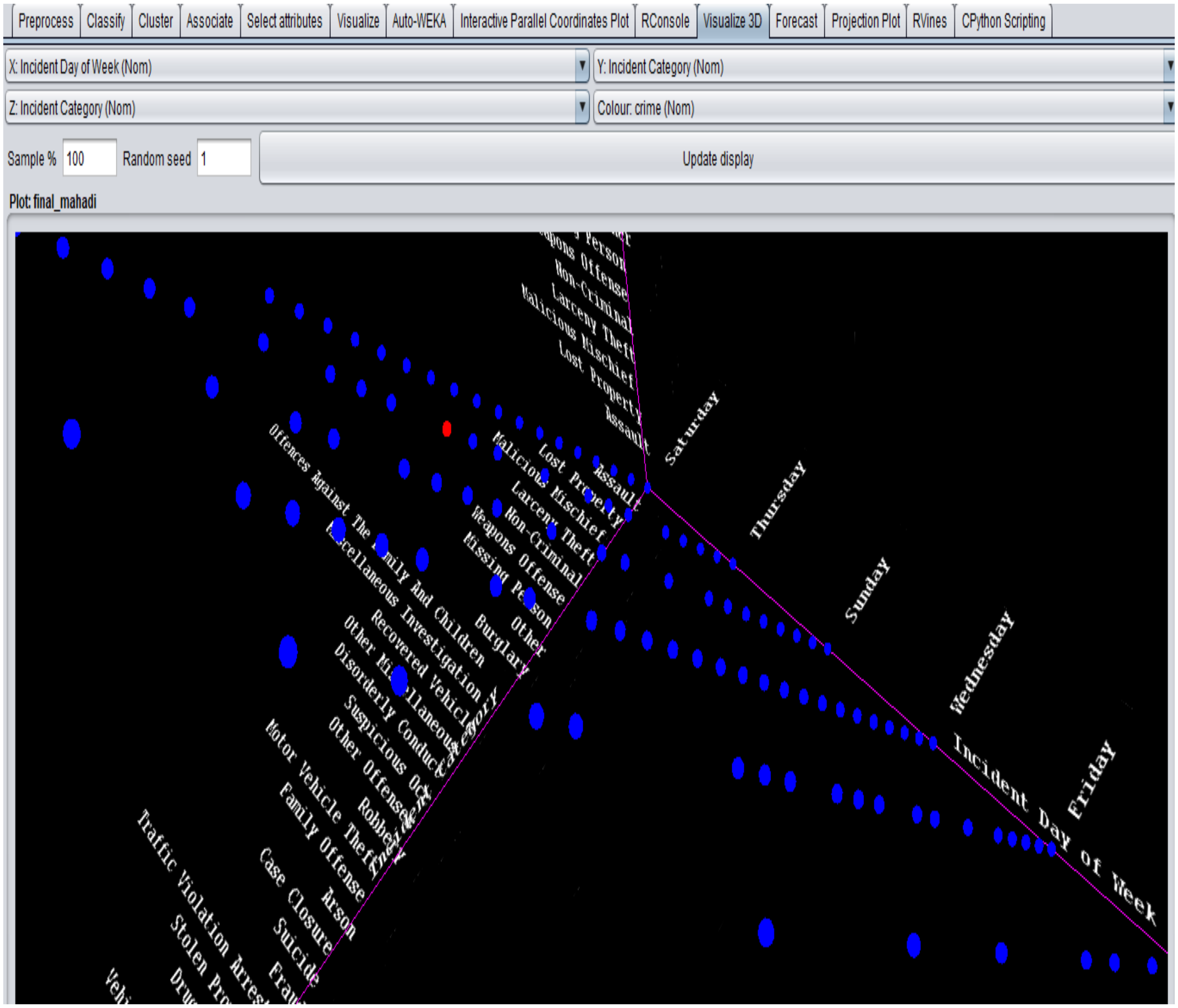


Figure 4.1.8: Visualization with jitter effect

Visualization in 3D, Days of week lies in X axis and incident category lies in both Y and Z axis. Wednesday is more continuous than other days of the week. This refers crimes in Wednesday are more than other days of the week.



X: Incident Day of Week (Nom)

Y: Incident Category (Nom)

Z: Incident Category (Nom)

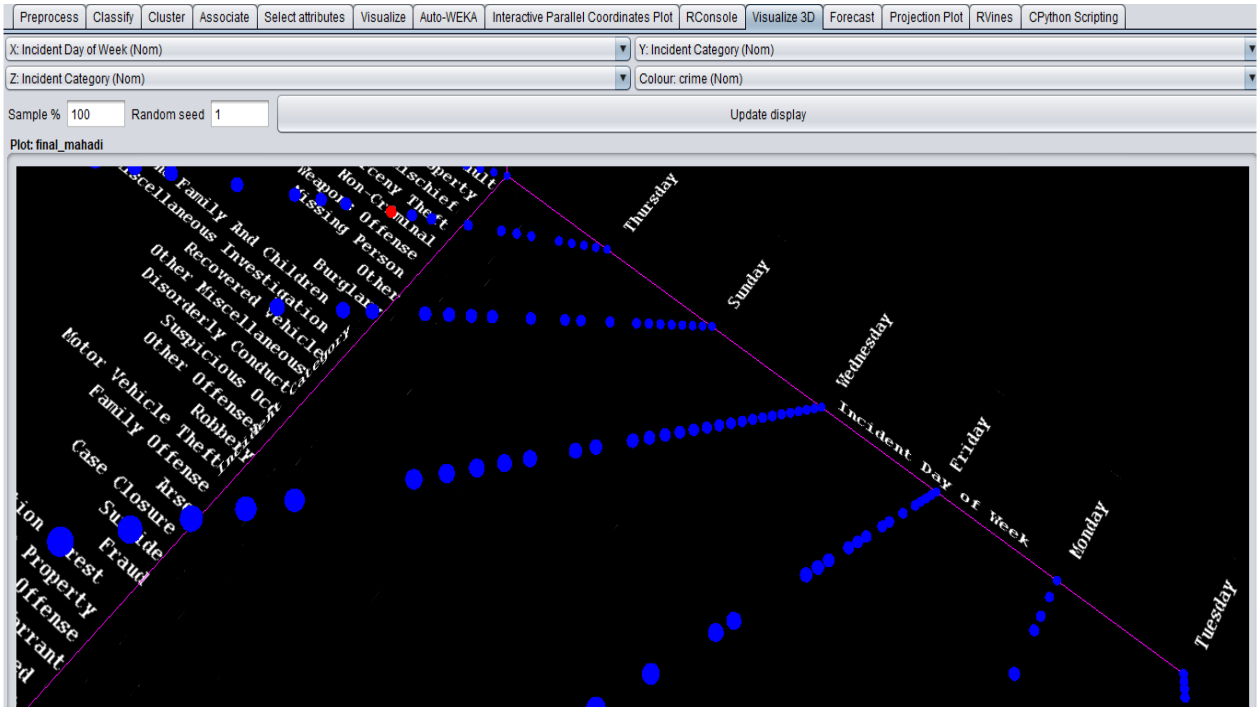
Colour: crime (Nom)

Sample % 100

Random seed 1

Plot: final_mahadi

Figure 4.1.10: Visualization in 3D



X: Incident Day of Week (Nom)

Y: Incident Category (Nom)

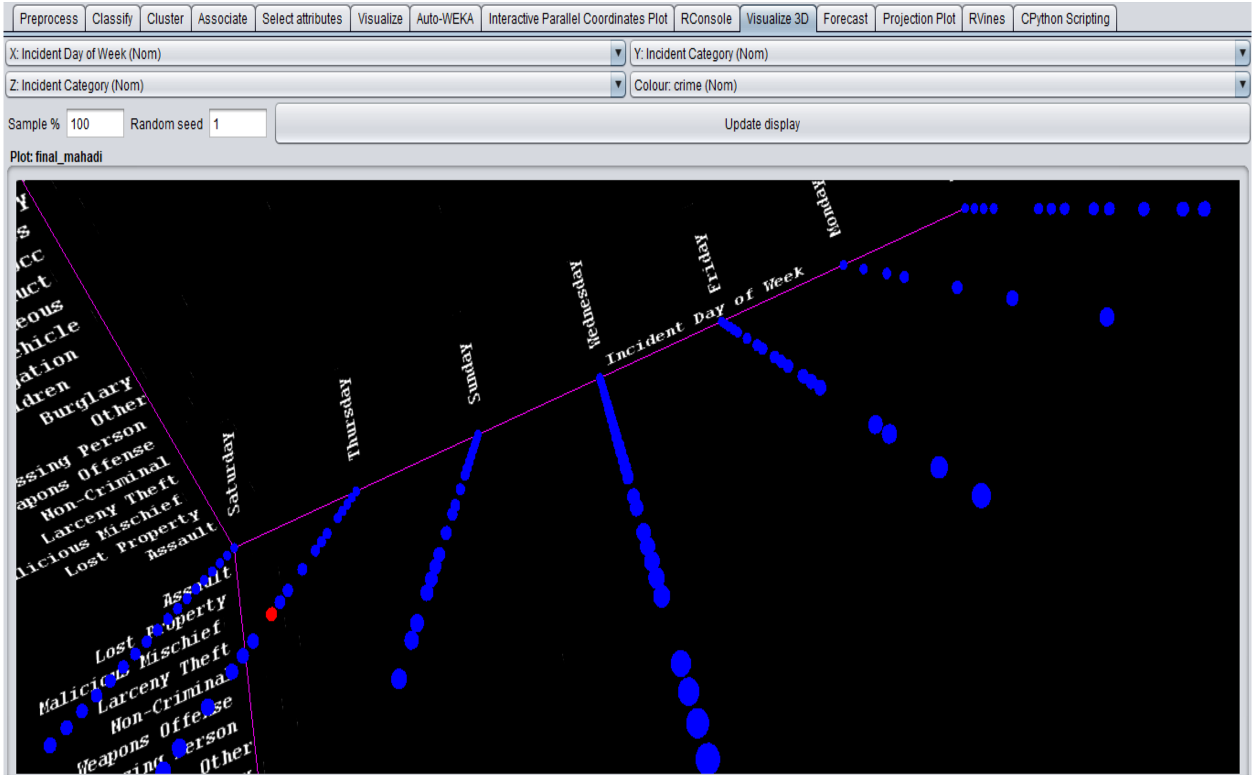
Z: Incident Category (Nom)

Colour: crime (Nom)

Sample % 100 Random seed 1

Plot: final_mahadi

Figure 4.1.11: Visualization in 3D



X: Incident Day of Week (Nom)

Y: Incident Category (Nom)

Z: Incident Category (Nom)

Colour: crime (Nom)

Sample % 100 Random seed 1

Plot: final_mahadi

Figure 4.1.12: Visualization in 3D

4.3 Discussion

In this research our dataset is collected from San Francisco police department website as mentioned earlier. We worked with SFPD website data because it’s more reliable and original than other datasets. We had more than forty thousand instances and 15 attributes. For our research we used five hundred instances and 5 attributes that already mentioned before.

After applying two algorithms we got close result in both analyses. In apriori algorithm 10 best rules were found, among them Wednesday is the most probable day for occurring crimes. Crime category and crime description are also identified that are larceny theft, theft from locked car. In K-Means clustering same Wednesday is the most probable day for occurring crime. In apriori algorithm, crime occurring time was not mentioned but after using K-Means clustering the probable time has been found along with the day. From visualization Saturday and Thursday are also in the competition of crime. Several cases like theft from locked car and fraud access card incident has been found from these two days also. So this analysis is another achievement of this research. After analyzing both algorithms our research got the desire output that was expected. Police department can take help from this analysis and prevent crime from occurring.

In the end, Visualization in two ways helps users to understand the actual outcome. For predicting crime using these two data mining techniques can be useful for further analysis.

Chapter 5

5. Conclusion

As the population continues to expand with urbanization, new forms of crime are dedicated to make communities vulnerable to public safety. The fast growth in cases of offence which contribute to the creation of new techniques of criminal investigations. A few of the one's popular methods of criminal investigation often used enforcement agencies today, use of such data mining methods is being used by agencies to build a description of an accused, a description of the suspect and unidentified Characteristics of a criminal case that is newly taking place, which is dedicated by the very same perpetrator(s) with a similar criminal history Instances. Strong association rules about crime in this study Cases are created by mining using the association rule, which is One of the Methods of Data Mining. Police data varies from the standard databases, as previously mentioned, widely used for applications of data mining, including temporal, Along with unstructured free text, spatial and geographical data Lands. Fields These fields include data used by qualified people. In order to connect and identify crimes, analysts. It is critical to identify which day crime occurred to much that's why law enforcement agencies cannot take proper action earlier. For that this research generate a pattern using previous criminals' dataset and using the data mining technique. This pattern will help police department or others law enforcement agencies to take necessary action earlier. It will be beneficial for the country and also people. To analyze crime data from a database, data mining methods are used. The outcomes of this data mining could theoretically be used in the next few years to mitigate and even deter crime. We assume there is a bright future for crime data mining to enhance the efficiency and effectiveness of investigative and intelligence research.

Chapter 6

6. Future Work

Crime data is a vital field in which successful data processing tools for crime detection play an integral role for investigators and law enforcers to continue with investigations and help resolve criminal cases. New forms of crime are committed to making neighborhoods vulnerable to public safety as the population continues to increase with urbanization. The exponential rise in criminal investigations leading to the introduction of modern criminal investigation techniques. A few of the common criminal investigation techniques commonly used by law enforcement agencies today, agencies are using such data mining methods to construct a picture of an accused, a description of the alleged and unidentified features of a newly occurring criminal case, dedicated by the same perpetrator(s) with a similar criminal background instances. Additional work is needed to determine if it was possible to attribute the crime to a network and offer a trust level for each person inside.

In this research we work a small size of dataset there has some attribute because of some technical limitation but in future we will work with all attributes of the dataset. In this research we work with only 500 rows of data but in the next time we will enlarge the rows of data. By analyzing all rows of dataset we will be able to predict the whole scenario of the city crime. We will try to take all attributes because we can make relation between all the attributes and will generate more patterns. That will be more helpful for the law enforcement agencies of the city. Through working on criminal investigation data sets such as FBI and crime detection of counter-terrorism initiatives, the scope of this study can be further strengthened. Then we will try to work with the crime hotspot area's dataset.

To support law enforcement agencies, deter crimes, researchers used different data mining techniques. During our literature review, we found that different researchers favored Density based Clustering and K-Means Clustering to perform crime hotspot analysis and crime pattern analysis, resulting in clusters depending on the number of reported crime incidents. The approach used for investigation differs widely on the basis of the type of crime and data type. Another change to this approach is the introduction of any integrated ERP program.

References:

1) Chao Yang, Hongbo Liu, Yeqing Sun, Ajith Abraham 2012 Multi-Knowledge Extraction From Violent Crime Datasets Using Swarm Rough Algorithm 12th International Conference On Hybrid Intelligent Systems (His)

2) Jieling Jin, Yuanchang Deng 2017 A Comparative Study On Traffic Violation Level Prediction Using Different Models 4th International Conference On Transportation Information And Safety (Ictis)

3) Sachin Kumar and Durga Toshniwal 2015 A Data Mining Framework To Analyze Road Accident Data Journal Of Big Data,

4) Adesola Falade*, Ambrose Azeta, Aderonke Oni, Isaac Odun-ayo 2019 Systematic Literature Review of Crime Prediction and Data Mining https://doi.org/10.18280/rces.060302

5) Abhinav Srivastava, Amlan Kundu, Shamik Sural and Arun K. Majumdar 2008 Credit Card Fraud Detection Using Hidden Markov Model IEEE Transactions On Dependable And Secure Computing 5

6) G.C. Oatley, J. Zeleznikow, B.W. Ewart, “Matching and Predicting Crimes,” In Applications and Innovations in Intelligent Systems XII in Proceedings of AI2004, The Twenty-fourth SGAI International Conference on Knowledge Based Systems and Applications of Artificial Intelligence. Ann Macintosh, Richard Ellis and Tony Allen Ed. London: Springer, , pp. 19-32, 2004.

7) R. William Adderley, “The use of data mining techniques in crime trend analysis and offender profiling,” Ph.D. thesis, University of Wolverhampton, Wolverhampton, England , 2007.

8) 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.

9) Chen, W. Chung, Y. Qin, M. Chau, J. Xu, G. Wang, R. Zheng, and H. Atabakhsh, "Crime Data Mining: An Overview and Case Studies," in Proceedings of the 3rd National Conference for Digital Government Research (dg.o 2003), pp. 1-5, Boston, MA, May 18-21, 2003.

10) H. Chen, H. Atabakhsh, T. Petersen, J. Schroeder, T. Buetow, L. Chaboya, C. O'Toole, M. Chau, T. Cushna, D. Casey, Z. Huang, "COPLINK: Visualization for Crime Analysis," ACM International Conference Proceeding Series, Proceedings of the 2003 annual national conference on digital government research,, Vol. 130, pp 1-6, Boston, MA, 2003.

11) R.V. Hauck, H. Atabakhsh, P. Ongvasith, H. Gupta, H. Chen, "Using Coplink to Analyze Criminal-Justice Data," Computer, vol. 35, no. 3, pp. 30-37, Mar. 2002.

12) Neetu Singh(&), Chengappa Bellathanda Kaverappa, and Jehan D. Joshi June,2018 Data Mining for Prevention of Crimes DOI: 10.1007/978-3-319-92043-6_55

13) Shyam Varan Florida Atlantic University, "Crime Pattern Detection Using Data Mining", Conference Paper · January 2007 DOI: 10.1109/WI-IATW.2006.55 · Source: IEEE Xplore

14) Peng Chen,Justin Kurland "Time, Place, and Modus Operandi: A Simple Apriori Algorithm Experiment for Crime Pattern Detection" This work was supported by Natural Science Foundation project (71704183).

15) Mehmet Sevri, Hacer Karacan, and M. Ali Akcayol "Crime Analysis Based on Association Rules Using Apriori Algorithm" International Journal of Information and Electronics Engineering, Vol. 7, No. 3, May 2017. DOI: 10.18178/IJIEE.2017.7.3.669

16) Arvind Venkatesh1, Hemanth S2, Juyin Shafaq Imtiaz Inamdar3, Lokesh S4 "Detection and Analysis of Crime Patterns Using Apriori Algorithm" Volume: 06 Issue: 04 | Apr 2019 International Research Journal of Engineering and Technology (IRJET)

17) De Bruin ,J.S.,Cocx,T.K,Kosters,W.A.,Laros,J. and Kok,J.N(2006) Data mining approaches to criminal carrer analysis ,"in Proceedings of the Sixth International Conference on Data Mining (ICDM"06) ,Pp. 171-177

18) Manish Gupta1*, B.Chandra1 and M. P. Gupta1,2007 Crime Data Mining for Indian Police Information System

19) A.Malathi ,Dr.S.Santhosh Baboo. D.G. Vaishnav College,Chennai ,2011 Algorithmic Crime Prediction Model Based on the Analysis of Crime Clusters.

20) Malathi.A 1 ,Dr.S.Santhosh Baboo 2 and Anbarasi . A 31 Assistant professor ,Department of Computer Science ,Govt Arts College ,Coimbatore , India . 2 Readers , Department of Computer science , D.G. Vaishnav Collge ,Chennai , India , 2011 An intelligent Analysis of a city Crime Data Using Data Mining

21) Malathi , A; Santhosh Baboo , S, 2011 An Enhanced Algorithm to Predict a Future Crime using Data Mining

22) Kadhim B.Swadi al-Janabi . Department of Computer Science . Faculty of Mathematics and Computer Science .University of Kufa/Iraq , 2011 A Proposed Framework for Analyzing Crime DataSet using Decision Tree and Simple K-means Mining Algorithms.

23) Jyoti Agarwal, Renuka Nagpal, Rajni Sehgal "Crime Analysis using K-Means Clustering" International Journal of Computer Applications (0975 – 8887) Volume 83 – No4, December 2013

24) Prof.Neha Mishra, Prof.Pooja Shelke, Prof.Neha Mishra, Prof.Pooja Shelke, "Data Mining – A necessity for Crime Detection" nternational Journal on Recent and Innovation Trends in Computing and Communication Data Mining – A necessity for Crime Detection, Article March ,2015.