Crime Prediction Using Machine Learning and Testing With Classification Models

[1]K.S. KavithaKumari, *[2]K.Uma, [3]C.Rameshkumar, [4]M.BashaKhaja

***Abstract*—**Crimes are a common social problem affecting the quality of life and the economic growth of a society. It is considered an essential factor that determines whether or not people move to a new city and what places should be avoided when they travel. With the increase of crimes, law enforcement agencies are continuing to demand advanced geographic information systems and new data mining approaches to improve crime analytics and better protect their communities.Although crimes could occur everywhere, it is common that criminals work on crime opportunities they face in most familiar areas for them. By providing a data mining approach to determine the most criminal hotspots and find the type, location and time of committed crimes, It is hope to raise people's awareness regarding the dangerous locations in certain time periods. Therefore, this proposed solution can potentially help people stay away from the locations at a certain time of the day along with saving lives. In addition, having this kind of knowledge would help people to improve their living place choices. On the other hand, police forces can use this solution to increase the level of crime prediction and prevention. Moreover, this would be useful for police resthisces allocation. It can help in the distribution of police at most likely crime places for any given time, to grant an efficient usage of police resthisces. By having all of this information available, It is hope to make this community safer for the people living there and also for others who will travel there.This project analyses two different real-world crimes datasets for Denver and Los Angeles and provides a comparison betItisen the two datasets through a statistical analysis supported by several graphs. Then, it clarifies how It is conducted Apriori algorithm to produce interesting frequent patterns for criminal hotspots. The results of this solution could be used to raise people's awareness regarding the dangerous locations and to help agencies to predict future crimes in a specific location within a particular time.

Keywords- Crime Prediction, Social Problem, Real-World Crimes, Data Mining, Testing, Awareness.

I. INTRODUCTION

There has been countless of work done related to crimes. Large datasets have been revieItisd, and information such as location and the type of crimes have been extracted to help people follow law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. There are several maps applications that show the exact crime location along with the crime type for any given city. Even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future.

[1] *Department of EEE, Aarupadaiveedu Institute of technology, Chennai,*

[2] *School of Information Technology and Engineering, VIT University, Vellore, India,*

[3] *School of Computing Science and Engineering, Galgotias University, Greater Noida, UttraPradhesh*

[4] *Wipro, Software engineer, Ireland, United Kingdom*

Received: 19 Feb 2020 | Revised: 28 Mar 2020 | Accepted: 25 Apr 2020

On the other hand, the previous related work and their existing methods mainly identify crime hotspots based on the location of high crime density without considering either the crime type or the crime occurrence date and time. For example, related research work containing a dataset for the city of Philadelphia with crime information from year 1991 - 1999. It was focusing on the existence of multi-scale complex relationships betItisen both space and time. Another research titled "The utility of hotspot mapping for predicting spatial patterns of crime" looks at the different crime types to see if they differ in their prediction abilities. Other existing works explore relationships betItisen the criminal activity and the socio-economic variables such as education, ethnicity, income level, and unemployment.

Despite all of the existing work, none of them consider the three elements (location, time, crime type) together. In addition, there is very little research that can accurately predict where crimes will happen in the future. In this study, It is provide a data-mining model for crime prediction based on crime types and using spatial and temporal criminal hotspots.

II. LITERATURE SURVEY

Data Mining is the procedure which includes evaluating and examining large pre-existing databases in order to generate new information which may be essential to the organization [1][2]. The extraction of new information is predicted using the existing datasets. Many approaches for analysis and prediction in data mining had been performed [3]. But, many few efforts has made in the criminology field. Many few have taken efforts for comparing the information all these approaches produce [4]. The police stations and other similar criminal justice agencies hold many large databases of information which can be used to predict or analyze the criminal movements and criminal activity involvement in the society. The criminals can also be predicted based on the crime data. The main aim of this work is to perform a survey on the supervised learning and unsupervised learning techniques that has been applied towards criminal identification. This research presents the survey on the Crime analysis and crime prediction using several Data Mining techniques [5].

Crime is a foremost problem where the top priority has been concerned by individual, the community and government. This research investigates a number of data mining algorithms and ensemble learning which are applied on crime data mining. This survey research describes a summary of the methods and techniques which are implemented in crime data analysis and prediction. Crime forecasting is a way of trying to mining out and decreasing the upcoming crimes by forecasting the future crime that will occur. Crime prediction practices historical data and after examining data, predict the upcoming crime with respect to location, time, day, season and year. In present crime cases rapidly increases so it is an inspiring task to foresee upcoming crimes closely with better accuracy. Data mining methods are too important to resolving crime problem with investigating hidden crime patterns.so the objective of this study could be analyzing and discussing various methods which are applied on crime prediction and analysis. This research delivers reasonable investigation of Data mining Techniques and ensemble classification techniques for discovery and prediction of upcoming crime[6].

This research focuses on finding spatial and temporal criminal hotspots. It analyses two different real-world crimes datasets for Denver, CO and Los Angeles, CA and provides a comparison betItisen the two datasets through a statistical analysis supported by several graphs. Then, it clarifies how It is conducted Apriori algorithm to produce interesting frequent patterns for criminal hotspots. In addition, the research shows how It is used Decision Tree classifier and Naïve Bayesian classifier in order to predict potential crime types. To further analyse crimes' datasets, the research introduces an analysis study by combining this findings of Denver crimes' dataset with its demographics information in order to capture the factors that might affect the safety of neighborhoods. The results of this solution could be used to raise people's awareness regarding the dangerous locations and to help agencies to predict future crimes in a specific location within a particular time[7].

The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of this project is to analyze dataset which consist of numerous crimes and predicting the type of crime which may happen in future depending upon various conditions. In this project, It is will be using the technique of machine learning and data science for crime prediction of Chicago crime data set. The crime data is extracted from the official portal of Chicago police. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtain will be high. The K-Nearest Neighbor (KNN) classification and various other algorithms will be tested for crime prediction and one with better accuracy will be used for training. Visualization of dataset will be done in terms of graphical representation of many cases for example at which time the criminal rates are high or at which month the criminal activities are high. The soul purpose of this project is to give a jest idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate. It not restricted to Chicago, this can be used in other states or countries depending upon the availability of the dataset[8].

The main objective of this research is to classify clustered crimes based on occurrence frequency during different years. Data mining is used extensively in terms of analysis, investigation and discovery of patterns for occurrence of different crimes. It is applied a theoretical model based on data mining techniques such as clustering and classification to real crime dataset recorded by police in England and Wales within 1990 to 2011. It is assigned It isights to the features in order to improve the quality of the model and remove low value of them. The Genetic Algorithm (GA) is used for optimizing of Outlier Detection operator parameters using RapidMiner tool[9]. With a substantial increase in crime across the globe, there is a need for analysing the crime data to loItisr the crime rate. This helps the police and citizens to take necessary actions and solve the crimes faster. In this research, data mining techniques are applied to crime data for predicting features that affect the high crime rate. Supervised learning uses data sets to train, test and get desired results on them whereas Unsupervised learning divides an inconsistent, unstructured data into classes or clusters. Decision trees, Naïve Bayes and Regression are some of the supervised learning methods in data mining and machine learning on previously collected data and thus used for predicting the features responsible for causing crime in a region or locality. Based on the rankings of the features, the Crimes Record Bureau and Police Department can take necessary actions to decrease the probability of occurrence of the crime[10].

The aim of this study is to compare different approaches to the problem of forecasting the number of crimes in different areas of the city. During this research It is studied three types of predictive models: linear regression, logistic regression and gradient boosting. The predictive factors used in these models have been selected using the feature selection techniques. This approach alloItisd us to increase the accuracy of predictions and to avoid the model's overfitting. The obtained models Itisre tested on criminal data of the city of Saint-Petersburg. It is compared the results of model predictions and determined that gradient boosting is the most appropriate method for the problem of crime rate prediction in certain urban area.[11].

1.1 Challenges

Crime is a serious issue that affects everyone in society. It affects the victims, perpetrators and their families. Crime has increased drastically within the last decade. More prisons are being built around the world because there is not enough room to hold inmates. The government has made an attempt to reduce crime by funding programs such as prevention and intervention for youth at risk , as It isll as rehabilitation for prisoners that will be released [12] [13].

In ordinary language, a crime is an unlawful act punishable by a state or other authority. The term "crime" does not, in modern criminal law, have any simple and universally accepted definition, though statutory definitions have been provided for certain purposes. The most popular view is that crime is a category created by law; in other words, something is a crime if declared as such by the relevant and applicable law [14].One proposed definition is that a crime or offence (or criminal offence) is an act harmful not only to some individual but also to a community, society or the state ("a public wrong"). Such acts are forbidden and punishable by law.The notion that acts such as murder, rape and theft are to be prohibited exists worldwide. What precisely is a criminal offence is defined by criminal law of each country [15]. While many have a catalogue of crimes called the criminal code, in some common law countries no such comprehensive statute exists.[16]

The state (government) has the power to severely restrict one's liberty for committing a crime. In modern societies, there are procedures to which investigations and trials must adhere. If found guilty, an offender may be sentenced to a form of reparation such as a community sentence, or, depending on the nature of their offence, to undergo imprisonment, life imprisonment or, in some jurisdictions, execution.Usually, to be classified as a crime, the "act of doing something criminal" must – with certain exceptions – be accompanied by the "intention to do something criminal".Some argue that criminal behavior is due to environment, others believe that it is genetic, and yet others think that it has to do with personality. Being able to identify personality traits that tend to lead to delinquency is clearly one option to the reduction of crime. However, the problem is that many youth display similar negative behavior during adolescence [17]. This includes negative attitude, different interests, and a need for privacy. So, personality cannot be used for reduction of crime.[18].

Crime prediction is not seeing the future, it is predicting where a crime will occur based on the previous data available. But, in majority of countries, where race and color discrimination are present, it is always difficult to find unbiased data that doesnot give biased results [19]. This issue always has been the major issue regarding crime prediction. While crime prediction is still in its early days, at least 60 police departments in US and European cities have rolled rolling-out crime forecasting systems – with mixed results. Researchers are struggling to make sense of the outcomes. One of the problems is that police don't like sharing what they're doing with the public [20]. Another is that it is exceedingly difficult to unpack the algorithms they use since they are proprietary. It is simply don't know what's inside the black box.Public statistics are in not easily accessible and there are often long delays before they are released. Making matters worse, when local news outlets also run crime stories, they typically lead with sensationalist headlines that do more to spread fear than offer insight.

III. MATERIALS AND METHODS

In this research, It is used two different datasets for real-world crimes in two cities of the US. It is chose those cities from different states: Denver in Colorado, Los Angeles in California. To construct this data mining models, It is mainly focused on Denver dataset. After It is had built the desired models, It is applied the same strategy to train the required models on Los Angeles dataset.This dataset represents the real-world crimes in Denver, Colorado. It includes criminal offenses and crime incidents in the city and county of the city for the previous five calendar years in addition to the current year (2010 – 2015). The dataset information is based on the National Incident Based Reporting System (NIBRS). This dataset is composed of 19 attributes with 333068 instances. The key attributes provide the offense type and its category such as robbery, public-disorder, and sexual assault. The dataset also gives the exact occurrence time of the crime along with the district, the neigh hood and the exact geographic location. The following table shows the used key attributes and its content values.This dataset represents the real-world crimes in Los Angeles, California. It includes criminal offenses and crime incidents in the city and the area of the city. 96% of the crimes in the dataset occurred in the year 2014 while the other 4% of the crimes occurring before 2014. This dataset information was obtained from the US City Open Data Census. It is composed of 14 attributes with 243750 instances. Unlike the Denver dataset, the crime category is more specific with its crime such as Theft-Person, Theft-Plain, and Theft-From-Motor-Vehicle. The following table shows the used key attributes and its content values.The software must get suitable crime related data. Getting the data either from a primary sthisce (i.e. collecting the data ythisself e.g. by tracking custom events in ythis app, conducting a survey or by running an experiment) or from a secondary sthisce (e.g. purchasing a data set from Bloomberg or downloading it from Kaggle etc.).Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behavithiss or trends, and is likely to contain many errors. Data pre-processing is a proven method of resolving such issues. Data pre-processing prepares raw data for further processing. Data pre-processing is used in database-driven applications such as customer relationship management and rule-based applications (like neural networks).Data goes through a series of steps during pre-processing:

1) **Data Cleaning**: Data is cleansed through processes such as filling in missing values, smoothing the noisy data, or resolving the inconsistencies in the data. There are some missing values in some attributes such aslast_occurance_date and incident_address in Denver dataset. HoItisver, It is found that all attributes containing missing values are not of this key attributes. Therefore, It is did not need to clean them. All key attributes Itisre completed with cleaned values in both datasets. In addition, It is did not found any noisy or inconsistent values in these attributes.

2) **Data Reduction:** This step aims to present a reduced representation of the data in a data warehouse. For both crime datasets, It is needed to apply data reduction. It is implemented dimensionality reduction using attribute subset selection. For example, among the available 19 attributes in Denver crimes dataset, It is just selected fthis of them. The selected attributes are the related ones or the key attributes for this mining purpose. It is removed all the other irrelevant attributes from the dataset. On the other hand, It is performed data reduction in terms of number of instances. It is observed that Denver crimes dataset contained a set of traffic accident instances. The attribute "Is_Crime" indicates whether the instance belongs to a crime or accident. While It is concern with crime information, It is used the attribute "Is_Crime" to filter the instances and removed all the irrelevant ones. It is applied the same strategy for Los Angeles crimes dataset.

3) **Data Integration:** Data with different representations are put together and conflicts within the data are resolved. It is performed several steps of data integration for this datasets. First, to avoid different attribute naming, It is unified the key attribute names for both crime datasets as follow: Crime_Type, Crime_Date, and Crime_Location. Crime_Location represents the neighborhood attribute for Denver dataset whereas the Area attribute for Los Angeles dataset. This mining study requires analyzing the date and time info on different granularities. Therefore, It is used the Crime_Date attribute, which contains date and time crime info, to generate three more attributes: Crime_Month, Crime_Day, and Crime_Time. It is adopted the military time system, and It is considered the hthis part without paying attention to the minutes to get more of frequent patterns. In addition, It is initiated Crime_Type_Id attribute to give an id for each of the 14 crime categories (See Table1). It is used this attribute for both datasets to get integrated crime types.

4) **Data Transformation And Discretization:** Data is normalized, aggregated and generalized during transformation whereas data discretization involves the reduction of a number of values of a continuous attribute by dividing the range of attribute intervals.

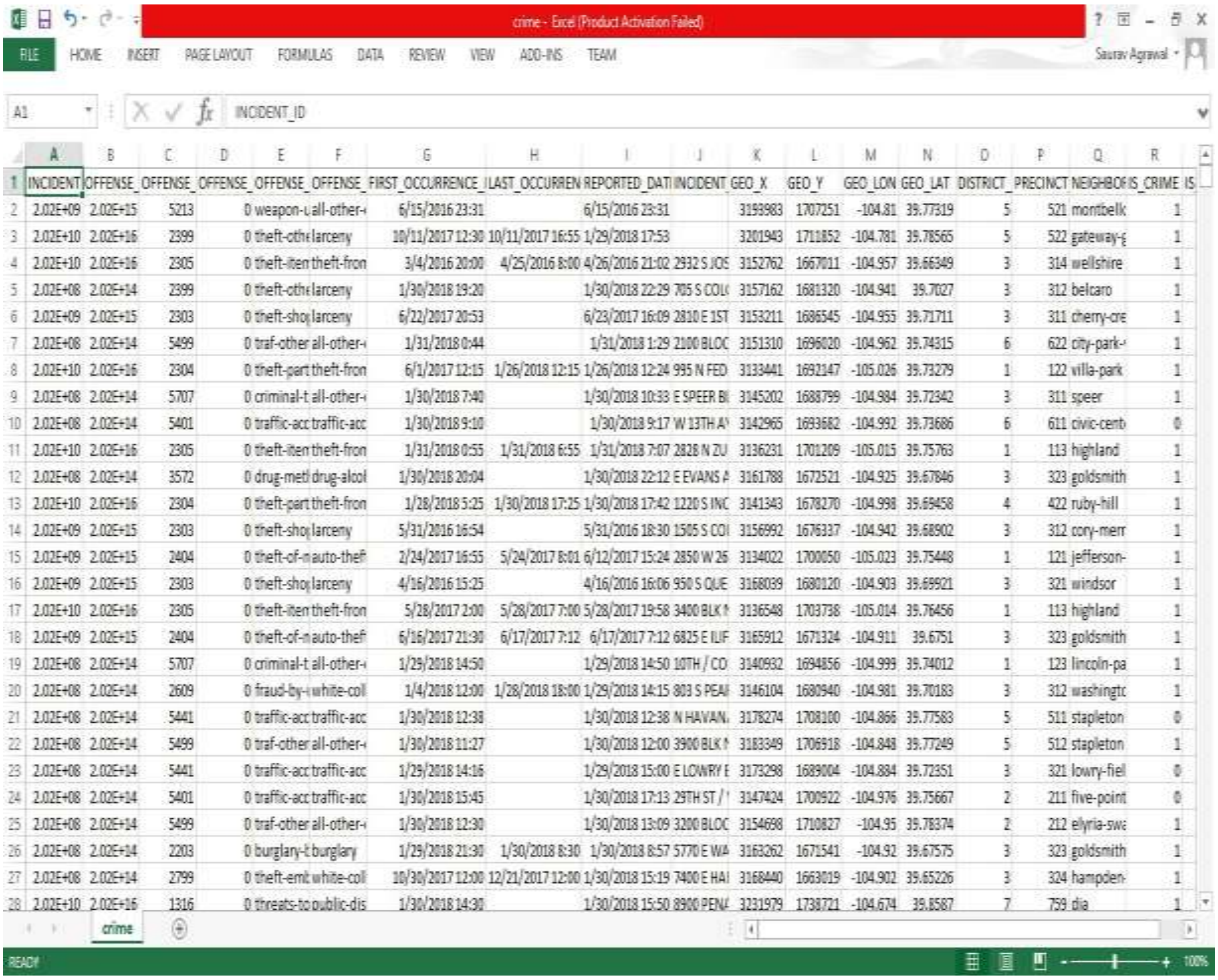
It is define the columns of data that you want to use by creating a structure. The mining structure is linked to the sthisce of data, but does not actually contain any data until you process it.
Before the structure and model is processed, a data mining model too is just a container that specifies the columns used for input, the attribute that you are predicting, and parameters that tell the algorithm how to process the data. Processing a model is often called training. Training refers to the process of applying a specific mathematical algorithm to the data in the structure in order to extract patterns. The patterns that you find in the training process depend on the selection of training data, the algorithm you chose, and how you have configured the algorithm. It is important to remember that whenever the data changes, you must update both the mining structure and the mining model.

IV. RESULT AND DISCUSSION

The Bayesian classifier enabled us to reach this target with a reasonable accuracy. To predict an expected crime type, you need to provide fthis related features of the crime. The required features are: the occurrence month, the

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

occurrence day of the week, the occurrence time and the crime location. All features can be submitted in their nominal values. The provided occurrence time should be in the form of time period interval from T1 to T6. For Denver, the location has to be one of its 78 neighborhoods. For Los Angeles, the location should be one of its 21 areas. Every given result is anumber from 1 to 6 that indicates the predicted crime type for a given set of crime features. Fig.1 shows the crime data setsand fig.2,fig.3,fig.4,fig.5,fig.6 and fig .7 shows the entire model crime prediction results.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	INCIDENT	OFFENSE_	OFFENSE_	OFFENSE_	OFFENSE_	OFFENSE_	FIRST_OCCURRENCE_	LAST_OCCURREN	REPORTED_DAT	INCIDENT_	GEO_X	GEO_Y	GEO_LON	GEO_LAT	DISTRICT_	PRECINCT	NEIGHBOR	IS_CRIME
2	2.02E+09	2.02E+15	5213	0	weapon-u	all-other-	6/15/2016 23:31		6/15/2016 23:31		3193983	1707251	-104.81	39.77319	5	521	montbelk	1
3	2.02E+10	2.02E+16	2399	0	theft-oth	larceny	10/11/2017 12:30	10/11/2017 16:55	1/29/2018 17:53		3201943	1711852	-104.781	39.78565	5	522	gateway-g	1
4	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	3/4/2016 20:00	4/25/2016 8:00	4/26/2016 21:02	2932 S JOS	3152762	1667011	-104.957	39.66349	3	314	wellshire	1
5	2.02E+08	2.02E+14	2399	0	theft-oth	larceny	1/30/2018 19:20		1/30/2018 22:29	705 S COL	3157162	1681320	-104.941	39.7027	3	312	belcaro	1
6	2.02E+09	2.02E+15	2303	0	theft-sho	larceny	6/22/2017 20:53		6/23/2017 16:09	2810 E 1ST	3153211	1686545	-104.955	39.71711	3	311	cherry-cre	1
7	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/31/2018 0:44		1/31/2018 1:29	2100 BLOC	3151310	1696020	-104.962	39.74315	6	622	city-park-	1
8	2.02E+10	2.02E+16	2304	0	theft-part	theft-fron	6/1/2017 12:15	1/26/2018 12:15	1/26/2018 12:24	995 N FED	3133441	1692147	-105.026	39.73279	1	122	villa-park	1
9	2.02E+08	2.02E+14	5707	0	criminal-t	all-other-	1/30/2018 7:40		1/30/2018 10:33	E SPEER BL	3145202	1688799	-104.984	39.72342	3	311	speer	1
10	2.02E+08	2.02E+14	5401	0	traffic-acc	traffic-acc	1/30/2018 9:10		1/30/2018 9:17	W 13TH A	3142965	1693682	-104.992	39.73686	6	611	civic-cent	0
11	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	1/31/2018 0:55	1/31/2018 6:55	1/31/2018 7:07	2828 N ZU	3136231	1701209	-105.015	39.75763	1	113	highland	1
12	2.02E+08	2.02E+14	3572	0	drug-meth	drug-alcol	1/30/2018 20:04		1/30/2018 22:12	E EVANS A	3161788	1672521	-104.925	39.67846	3	323	goldsmith	1
13	2.02E+10	2.02E+16	2304	0	theft-part	theft-fron	1/28/2018 5:25	1/30/2018 17:25	1/30/2018 17:42	1220 S INC	3141343	1678270	-104.998	39.69458	4	422	ruby-hill	1
14	2.02E+09	2.02E+15	2303	0	theft-sho	larceny	5/31/2016 16:54		5/31/2016 18:30	1505 S CO	3156992	1676337	-104.942	39.68902	3	312	cory-merr	1
15	2.02E+09	2.02E+15	2404	0	theft-of-n	auto-thef	2/24/2017 16:55	5/24/2017 8:01	6/12/2017 15:24	2850 W 26	3134022	1700050	-105.023	39.75448	1	121	jefferson-	1
16	2.02E+09	2.02E+15	2303	0	theft-sho	larceny	4/16/2016 15:25		4/16/2016 16:06	950 S QUE	3168039	1680120	-104.903	39.69921	3	321	windsor	1
17	2.02E+10	2.02E+16	2305	0	theft-iten	theft-fron	5/28/2017 2:00	5/28/2017 7:00	5/28/2017 19:58	3400 BLK	3136548	1703738	-105.014	39.76456	1	113	highland	1
18	2.02E+09	2.02E+15	2404	0	theft-of-n	auto-thef	6/16/2017 21:30	6/17/2017 7:12	6/17/2017 7:12	6825 E ILIF	3165912	1671324	-104.911	39.6751	3	323	goldsmith	1
19	2.02E+08	2.02E+14	5707	0	criminal-t	all-other-	1/29/2018 14:50		1/29/2018 14:50	10TH / CO	3140932	1694856	-104.999	39.74012	1	123	lincoln-pa	1
20	2.02E+08	2.02E+14	2609	0	fraud-by-	white-coll	1/4/2018 12:00	1/28/2018 18:00	1/29/2018 14:15	803 S PEA	3146104	1680940	-104.981	39.70183	3	312	washingtc	1
21	2.02E+08	2.02E+14	5441	0	traffic-acc	traffic-acc	1/30/2018 12:38		1/30/2018 12:38	N HAVAN	3178274	1708100	-104.866	39.77583	5	511	stapleton	0
22	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/30/2018 11:27		1/30/2018 12:00	3900 BLK	3183349	1706918	-104.848	39.77249	5	512	stapleton	1
23	2.02E+08	2.02E+14	5441	0	traffic-acc	traffic-acc	1/29/2018 14:16		1/29/2018 15:00	E LOWRY E	3173298	1689004	-104.884	39.72351	3	321	lowry-fiel	0
24	2.02E+08	2.02E+14	5401	0	traffic-acc	traffic-acc	1/30/2018 15:45		1/30/2018 17:13	29TH ST /	3147424	1700922	-104.976	39.75667	2	211	five-point	0
25	2.02E+08	2.02E+14	5499	0	traf-other	all-other-	1/30/2018 12:30		1/30/2018 13:09	3200 BLOC	3154698	1710827	-104.95	39.78374	2	212	elyria-swa	1
26	2.02E+08	2.02E+14	2203	0	burglary-b	burglary	1/29/2018 21:30	1/30/2018 8:30	1/30/2018 8:57	5770 E WA	3163262	1671541	-104.92	39.67575	3	323	goldsmith	1
27	2.02E+08	2.02E+14	2799	0	theft-emb	white-coll	10/30/2017 12:00	12/21/2017 12:00	1/30/2018 15:19	7400 E HA	3168440	1663019	-104.902	39.65226	3	324	hampden-	1
28	2.02E+10	2.02E+16	1316	0	threats-to	public-dis	1/30/2018 14:30		1/30/2018 15:50	8900 PENA	3231979	1738721	-104.674	39.8587	7	759	dia	1

Fig.1 crime data sets

Received: 19 Feb 2020 | Revised: 28 Mar 2020 | Accepted: 25 Apr 2020

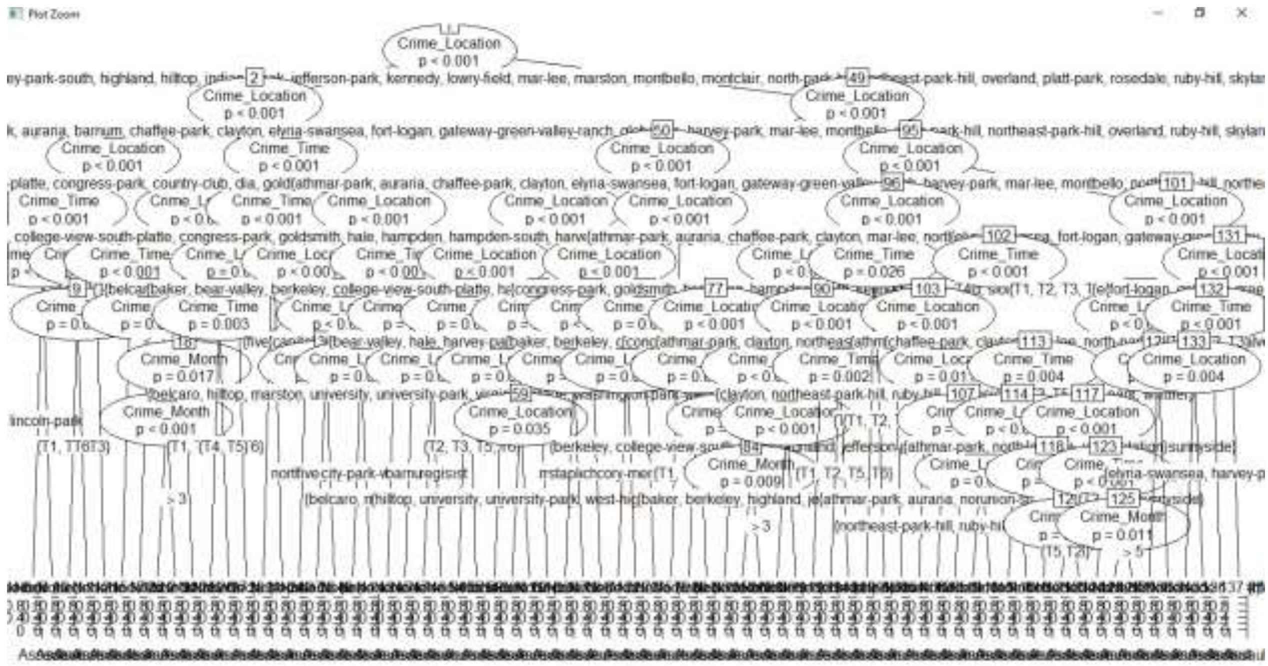


Fig.2 location based crime prediction model result 1

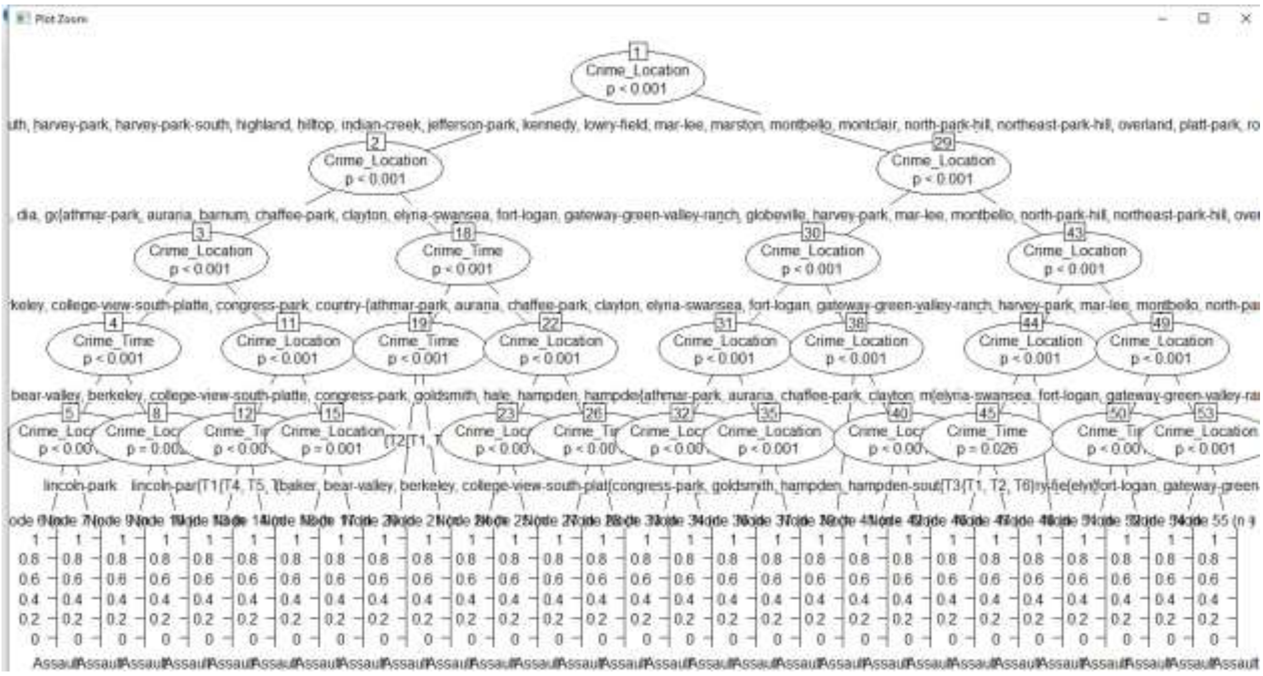


Fig.3location based crime prediction model result 2

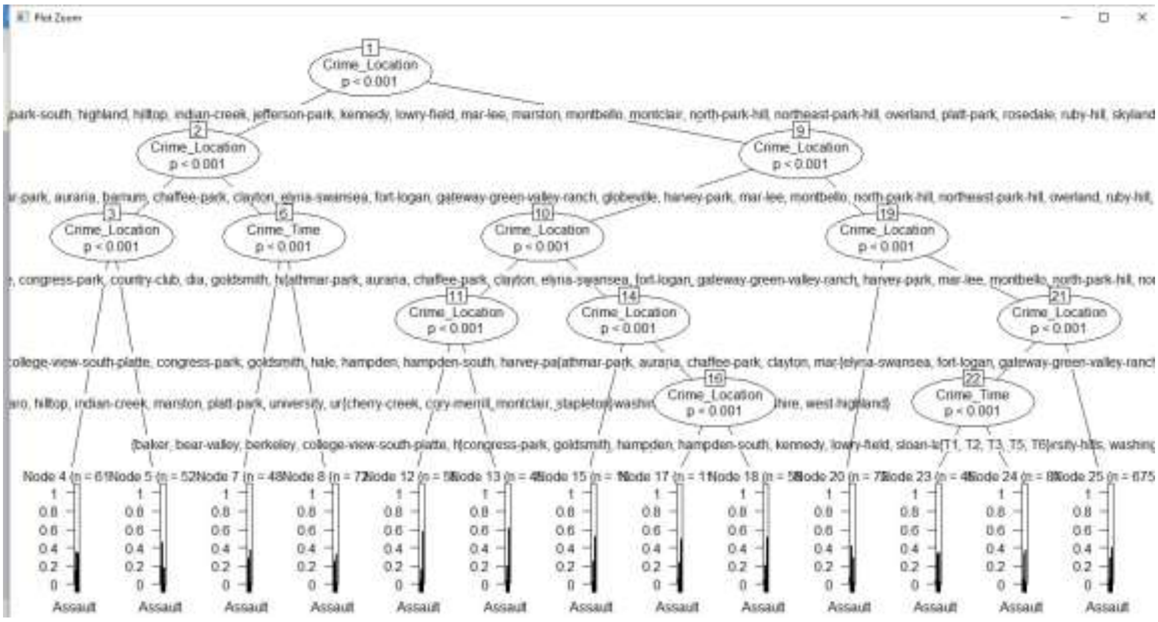


Fig.3 location based crime prediction model result 3

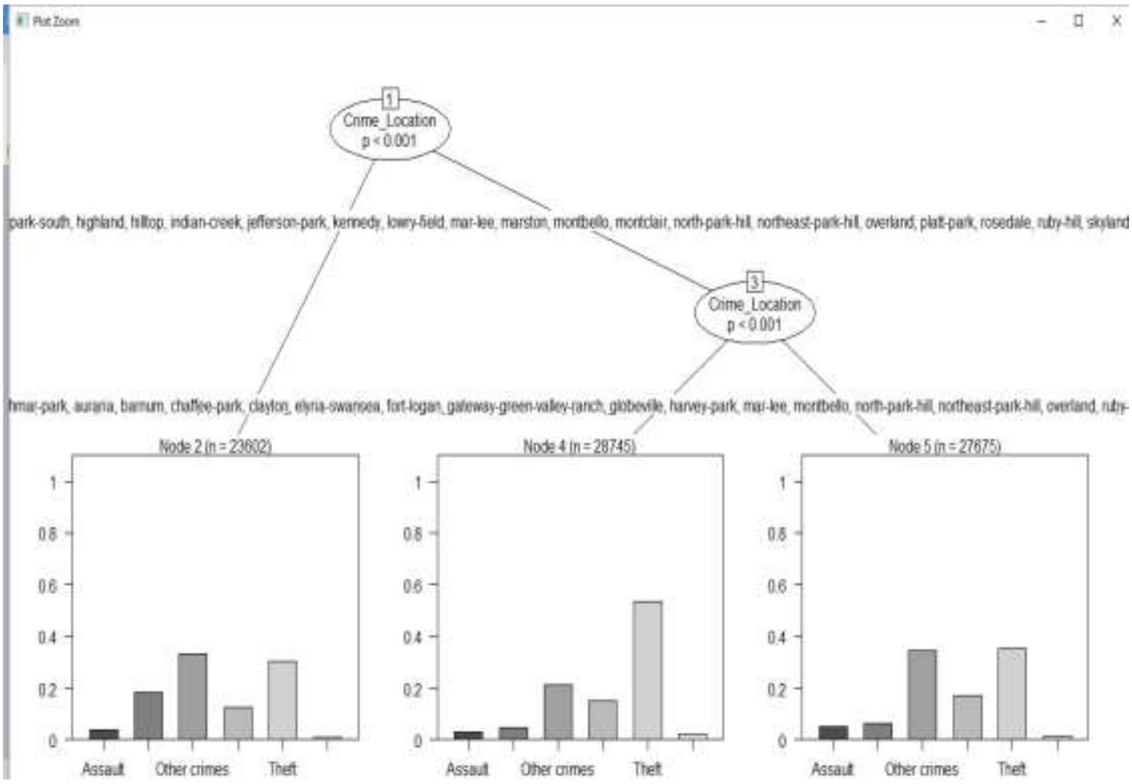


Fig.4 location based crime prediction model result 4

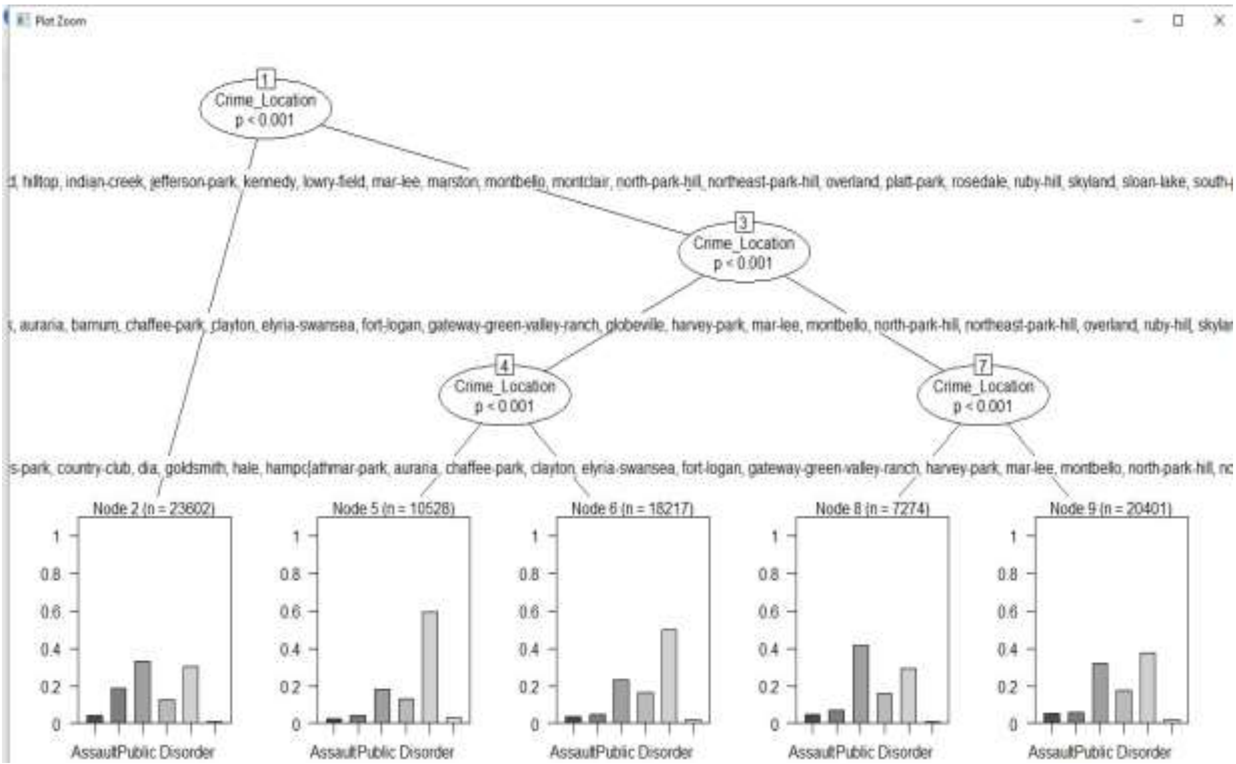


Fig.5 location based crime prediction model result 5

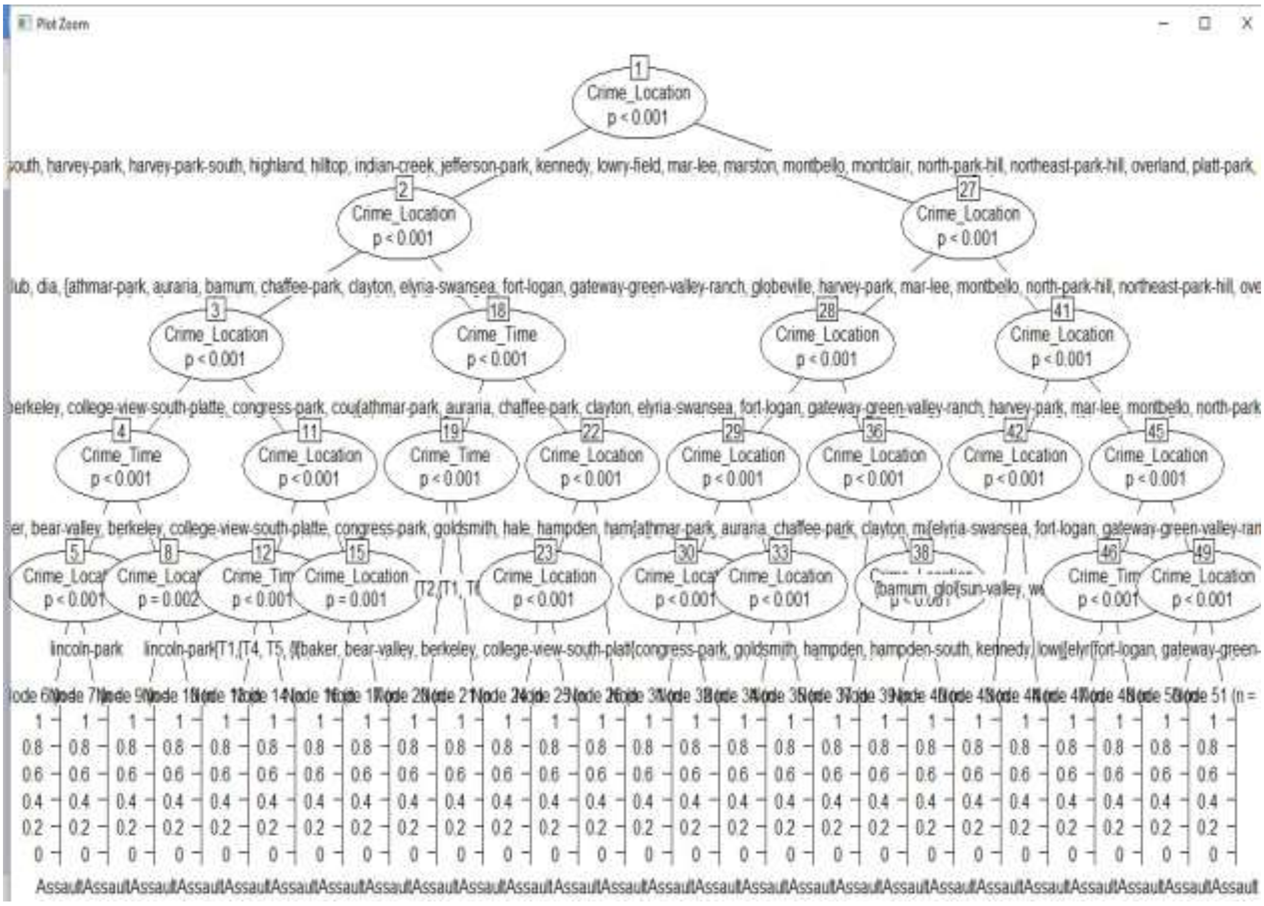


Fig.6 location based crime predictionmodel result 6

Received: 19 Feb 2020 | Revised: 28 Mar 2020 | Accepted: 25 Apr 2020

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

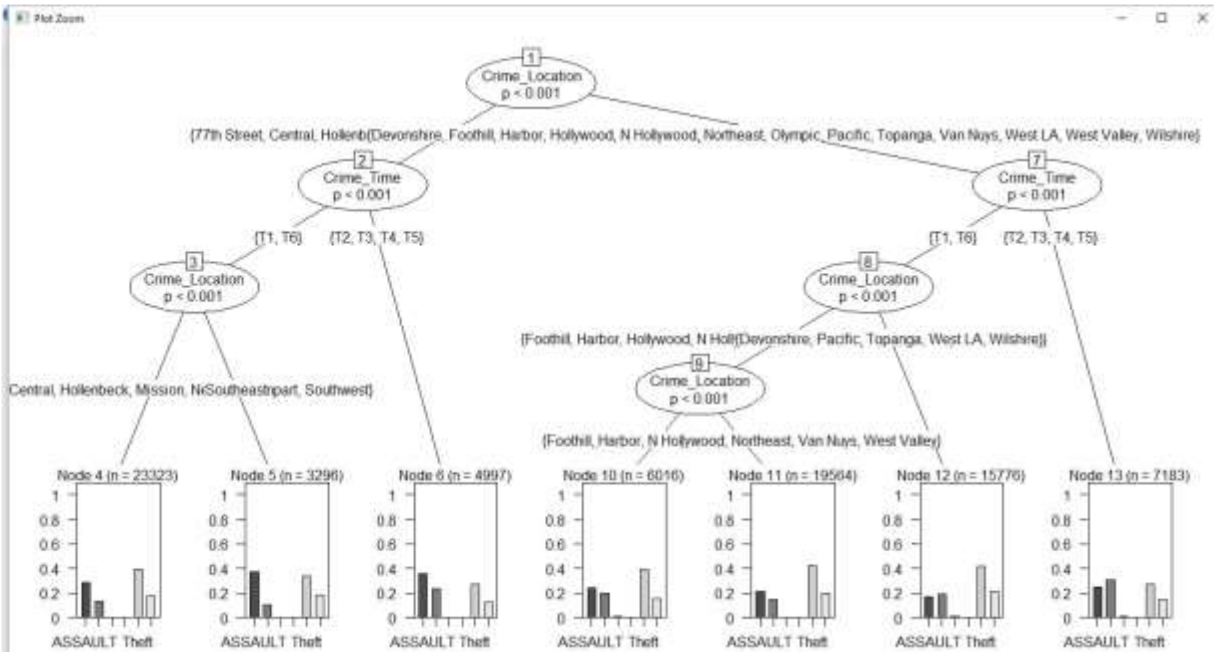


Fig.7 location based crime prediction model result 7

In this project, first It is performed data pre-processing and then made 3 different models of Apriori algorithm, NaïveBayes Algorithm and Decision Tree Algorithm. By using Apriori algorithm It is found criminal hotspots on Denver Datasets andLos Angeles Datasets. Five-Point, Capitol Hill, CBD, Montebello, Union Station, Stapleton, and It isstwood are the hotspots that have most crimes frequent patterns in Denver. Five-Point has the largest number of patterns compared to other locations. item: ('''five-points''',) , 0.054 . In Los Angeles, It is can see that most likely crimes happen at 77th Street, SouthItisst, Pacific, N Hollywood, Southeast, Northeast, and Van Nuys respectively. 77th Street has the largest number of patterns compared to other locations. item: ('''77th Street''',) , 0.068 . The second target for this study was to predict the crime type that might occur in a specific location within a particular time. The Bayesian classifier enabled us to reach this target with a reasonable accuracy.

V. CONCLUSION

It is generated many graphs and found interesting statistics that shoItisd the baseline to understand Denver and Los Angeles crimes datasets. Then, It is applied Apriori algorithm to find frequent crime patterns in both cities. After that, It is applied Decision Tree and Naïve Bayesian classifiers to help predicting future crimes in a specific location within a particular time. It is achieved 51% of prediction accuracy in Denver and 54% prediction accuracy in Los Angeles. Finally, It is provided an analysis study by combining this findings of Denver crimes’ dataset with its demographics information. It is aimed to further understand this models’ findings and to capture the factors that might affect the safety of neighborhoods.

REFERENCES

1. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, 'Once Upon a Crime:
2. Towards Crime Prediction from Demographics and Mobile Data', CoRR, vol. 14092983, 2014.
3. R. Arulanandam, B. Savarimuthu and M. Purvis, 'Extracting Crime Information from Online
4. Newsresearch Articles', in Proceedings of the Second Australasian It isb Conference - Volume 155,
5. Auckland, New Zealand, 2014, pp. 31-38.
6. Buczak and C. Gifford, 'Fuzzy association rule mining for community crime pattern discovery', in
7. ACM SIGKDD Workshop on Intelligence and Security Informatics, Washington, D.C., 2010, pp. 1-10.
8. M. Tayebi, F. Richard and G. UIt is, 'Understanding the Link BetItisen Social and Spatial Distance in
9. the Crime World', in Proceedings of the 20th International Conference on Advances in Geographic
10. Information Systems (SIGSPATIAL '12), Redondo Beach, California, 2012, pp. 550-553.
11. Mugdha Sharma, "Z-Crime: A Data Mining Tool for the Detection of Suspicious Criminal Activities based on the Decision Tree",International Conference on Data Mining and Intelligent Computing,pp. 1-6, 2017
12. Kevin Sheehyet al., "Evidence-based Analysis of Mentally 111 Individuals in the Criminal Justice System",Proceedings of IEEE Systems and Information Engineering Design Symposium, pp. 250-254, 2016.
13. Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In Data Mining Workshops (ICDMW), 2017 IEEE International Conference on (pp. 1158-1159). IEEE.
14. Al Boni, M., & Gerber, M. S. (2016, December). AreaSpecific Crime Prediction Models. In Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on (pp. 671-676). IEEE.
15. M. Tayebi, F. Richard and G. UIt is, 'Understanding the Link BetItisen Social and Spatial Distance in the Crime World', in Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12), Redondo Beach, California, 2018, pp. 550-553.
16. PrajaktaYerpude, VaishnaviGudur. 'Predictive modelling of crime dataset using data mining' in Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '12), Redondo Beach, California, 2018, pp. 550-553.
17. VarvaraIngilevich, Sergey Ivanov, 'Crime rate prediction in the urban environment using social factors' InComputer Supported Cooperative Work in Design (CSCWD, 2016 May 4 (pp. 97-101). IEEE.
18. NurulHazwaniMohdShamsuddin ; Nor Azizah Ali ; RazanaAlItise, 'An overview on crime prediction methods' in 6th ICT International Student Project Conference, 2017
19. Ginger Saltos, MihaelaCocea, 'An Exploration of Crime Prediction Using Data Mining on Open Data' in International Jthisnal of Information Technology and Decision Making 16(05) · May 2017.
20. ShijuSathyadevan, Devan M. S., Surya S Gangadharan, 'Crime Analysis and Prediction Using Data Mining' in First International Conference on Networks & Soft Computing (ICNSC), 2014.
21. Vikas Grover, Richard Adderley, Max Bramer, 'Review of Current Crime Prediction Techniques' in Applications and Innovations in Intelligent Systems XIV: Proceedings of AI-2006, the TItisnty-sixth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (pp.233-237)

22. Nikhil Dubey , Setu Kumar Chaturvedi, ‘A Survey Research on Crime Prediction Technique Using Data Mining’ in Int. Jthisnal of Engineering Research and Applications, Vol. 4, Issue 3(Version 1), March 2014, pp.396-400.
23. Lenin Mookiah, William Eberle and AmbareenSiraj, ‘Survey of Crime Analysis and Prediction’ in Proceedings of the TItisnty-Eighth International Florida Artificial Intelligence Research Society Conference.
24. Hitesh Kumar Reddy ToppiReddy, BhavnaSaini , GinikaMahajan, ‘Crime Prediction &onitoring Framework Based on Spatial Analysis’ in International Conference on Computational Intelligence and Data Science, 2018.
25. Malathi. A, Dr. S. SanthoshBaboo, ‘An Enhanced Algorithm to Predict a Future Crime using Data Mining’ in International Jthisnal of Computer Applications (0975 – 8887) Volume 21– No.1, May 2011.
26. Varshitha D N, Vidyashree K P, Aishwarya P, Janya T S, K R Dhananjay Gupta, Sahana R, ‘Research on Different Approaches for Crime Prediction system’ in International Jthisnal of Engineering Research & Technology (IJERT), 2017.