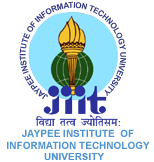
**CRIMINAL DATA**

**INVESTIGATION**

Crime pattern detection for investigation



From :

Nikita Jain (09104735)

Anushree Pai (09104737)

Yatharth Sharma(09104743)

**INTRODUCTION**

**GENERAL**

Crimes are a social nuisance and cost our society dearly in several ways. About 10% of the criminals commit about 50% of the crimes. This project helps in keeping a check over the patterns in crime scenario i.e. whether it’s increasing or decreasing and if increasing then what are the problem areas.

The first phase gave the brief overview of the project and its objective. The main objective is to find/ predict the probable suspects for the unsolved cases or test cases (in other words) from the past criminal record, crime record, and crime over past years records. The second phase concentrates mainly over shaping the project towards its ultimate goal. Herein we concentrate on predicting crime rates and crime trends over the past 50 years. Their growth, depreciation and other associated results concerning the crime scenario. Also we would predict the futuristic crime rates with proper visualizations.

**INTRODUCTION**

**PROBLEM STATEMENT**

1. Crime Pattern Detection

* Hot Spots/ high density crime areas
* Crime type clustering over those clusters
* Crime pattern detection and analysis

1. Criminal Pattern Detection

* Solving unsolved cases using cases of the past
* Criminal Pattern Detection and analysis over an area

**INTRODUCTION**

**FEATURES (other than the main objectives)**

1. Background checking of any maid servant: checking for any criminal background
2. Queries to find maximum crime at a place, in a year, place with max crime, percentage arrest, percentage domestic violence etc.
3. **Time series analysis of crime data** over the last 50 years with statistics
4. **Visualizations** with graphs

1. GUI – frontend
2. Prediction of crime over the next years with plotting of graph using **5- year moving average algorithm**
3. **Background check of any suspicious person** especially for the purpose of maid.

**BACKGROUND MATERIAL**

**LITERATURE STUDIED**

1) Crime Pattern Detection Using Data Mining

***-Shyam Varan Nath(Oracle Corporation)***

2) Crime Data Mining for Indian Police Information System

***-Manish Gupta, B. Chandraand M. P. Gupta***

3) An Optimal KD Model for Crime Pattern Detection Based on

Semantic Link Analysis-A Data Mining Tool

* ***C. Ram Singla, Deepak Dembla, Yogesh Chaba1 and Kevika Singla***

4) Data mining for intelligence led policing

5) Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model

***-A. Malathi and Dr. S. Santhosh Baboo***

6) Spatio-Temporal Crime Prediction Model Based On Analysis

Of Crime Clusters

- ***ESRA POLAT***

**BACKGROUND MATERIAL**

**COMPARATATIVE ANALYSIS OF RESEARCH PAPER STUDIED**

1. Crime Pattern Detection Using Data Mining: This paper talks about hotspots and use of k means clustering for crime pattern detection. It first identifies significant attributes in the database. Unlike other papers, it then gives weights to attributes in the data set. The most important attribute (e.g. type of crime) is given the highest priority (weight) as compared to other attributes. We are using this feature of the research paper in our project. The data with missing values are made as test cases.
2. Crime Data Mining for Indian Police Information System : This paper is all about India’s crime analysis system. It gives ways to enhance the currently existing system in Indian Police System called Crime Criminal Information System (CCIS). It suggests to divide the database according to respective states, using classification, to make the data more easier to analyze. In our project, we are subdividing the data into different types of crime allowing the user to get information of those crimes easily (e.g. percentage of the particular crime in a particular year, the hotspot of that particular crime)
3. An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool : The system finds the critical path of the serial killers who are striking over again and again and determines links between their crime activities locations occurred in the past, travel record, and background history, etc. These findings increase the chance of finding these repeat offenders*.* Criminal intelligence analysis therefore requires the ability to integrate information from multiple crime incidents or even multiple sources and discover regular patterns about the structure, organization, operation, and information flow in

criminal networks. If a particular criminal uses a pattern of path to commit consecutive crimes, then the next crime location of this serial killer can be predicted from the pattern observed. Eg: in dhoom 2 , the last location of crime of criminal Hrithik Roshan was predicted by the pattern he formed from his previous crime locations.

1. Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. Malathi and Dr. S. Santhosh Baboo :The crime data is divided into days of the week, to observe Spatio temporal distribution of crime. To the clustered results, a classification algorithm was applied to predict the future crime

pattern. The classification was performed to find in which category a cluster would be in the next year. This allows us to build a predictive model on predicting next year’s records using this year’s data.

1. Spatio-Temporal Crime Prediction Model Based On Analysis

Of Crime Clusters: Objective is Development of a spatio-temporal crime prediction model based on geographical information systems coupled with spatial statistical methods. In this paper clustering analyses are used to identify hot spots. Cluster analysis aims to collect data into groups according to several algorithms which are Kmeans,Nnh hierarchical, spatio-temporal analysis of crime (STAC), fuzzy, ISODATA, and geographical analysis machine (GAM) clustering techniques. Clusters of STAC do include more homogenous areas than the other methods.STAC is not restricted to include all the observations hence STAC is able to indicate denser crime areas than other methods.This is important in crime prevention for allocating resources effectively. If all the area is going to be searched, there is no meaning to form crime prediction models.

**APPENDIX B**

**IEEE REFERENCE**

1. Crime Pattern Detection Using Data Mining: Hsinchun Chen, Wingyan hung, Yi Qin, Michael Chau, Jennifer , Rong Zheng, Homa Atabakhsh, “Crime Data Mining: A General Framework and Some Examples”, IEEE Computer Society April 2004.
2. An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool: Agrawal R., Gupta A., Sarawagi A., *“Modeling Multidimensional Database*s”, ICDE’97 pages 232-243. IEEE Press, 1997.
3. An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool : Fayyad, U. Data Mining and Knowledge Discovery: Making Sense Out of Data. IEEE Expert, v. 11, no. 5, pp. 20-25,October 1996.
4. Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. Malathi and Dr. S. Santhosh Baboo : Abraham, T. O. (2006) Investigative profiling with computer forensic log data and association rules," in Proceedings of the IEEE International Conference on Data Mining (ICDM'02), Pp. 11– 18.
5. Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. Malathi and Dr. S. Santhosh Baboo : Brown, D.E. (1998) The regional crime analysis program: A frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics,

**BACKGROUND MATERIAL**

**COMPARATATIVE ANALYSIS OF RESEARCH PAPER STUDIED**

|  |  |  |  |
| --- | --- | --- | --- |
| **PAPER** | **ALGORITHM USED** | **ALGO WE USED** | **REASON** |
| Crime Pattern Detection Using Data Mining-Shyam Varan Nath | K-means Clustering | Spacio-Temporal Analysis Of crime Algorithm  (STAC) | By comparative analysis it was found that STAC gave a better result and was hence implement. |
| Crime Data Mining for Indian Police Information System | Classification | Clustering | Classes not predefined |
| An Optimal KD Model for Crime Pattern Detection Based on  Semantic Link Analysis-A Data Mining Tool | Association | Association | For link analysis of crime patterns |
| Data mining for intelligence led policing | Spatio-temporal clustering using : When and Where | Spatio-temporal clustering using : When and Where |  |
| Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. Malathi | K-means clustering, HYB algorithm | None | Better approach and clustering algorithm found in other research papers |
| Spatio-Temporal Crime Prediction Model Based On Analysis  Of Crime Clusters | Kmeans,Nnh hierarchical, spatio-temporal analysis of crime (STAC), fuzzy,  ISODATA, and geographical analysis machine (GAM) clustering techniques. | Spacio Temporal Analysis Of crime  (STAC) | Out of all the algos mentioned in the paper, STAC is a weighted average of all the mentioned algos |

**BACKGROUND MATERIAL**

**COMPARATATIVE ANALYSIS OF RESEARCH PAPER STUDIED**

|  |  |  |  |
| --- | --- | --- | --- |
| **PAPER** | **FEATURES** | **FEATURES WE USED** | **REASON** |
| Crime Pattern Detection Using Data Mining-Shyam Varan Nath | Crime and Location  HotSpots | Crime and Location  HotSpots | Feature used but not the algorithm |
| Crime Data Mining for Indian Police Information System | Classes through classification | Clusters through clustering | Makes work of police faster and efficient |
| An Optimal KD Model for Crime Pattern Detection Based on  Semantic Link Analysis-A Data Mining Tool | Serial killer patterns detection viz. next location of crime, time, etc | Serial killer pattern detection and prediction | To ensure safety and security of the next crime spot at that time |
| Data mining for intelligence led policing | 1. Hot spot maps of the recent period.  2. Temporal hot spot maps to show what changed.  3. Prediction maps of the upcoming period.  4. A where/when analysis with description of the clusters found.  5. A predicted week-distribution of crime over the upcoming period: on what days and times is the highest crime rate to be expected?  6. Crime rate graphs with basic statistics.  7. Hot spot lists of the most frequent offenders. | 1. Hot spot maps of the recent period.  2. Temporal hot spot maps to show what changed.  3. Prediction maps of the upcoming period.  4. A where/when analysis with description of the clusters found.  5. A predicted week-distribution of crime over the upcoming period: on what days and times is the highest crime rate to be expected?  6. Crime rate graphs with basic statistics.  7. Hot spot lists of the most frequent offenders. | All features used but algorithm used of some other paper |
| Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. | 1) To develop a data cleaning algorithm.  2) Crime prediction in both space and time to help police departments in tactical and planning operations. | Develop a data cleaning algorithm: using KNN | Used a better algorithm for the second feature (found in other research papers) |
| Spatio-Temporal Crime Prediction Model Based On Analysis  Of Crime Clusters | identify hot spots and forecast | identify hot spots and forecast | Feature used with this paper algorithm |

**BACKGROUND MATERIAL**

**COMPARATATIVE ANALYSIS OF RESEARCH PAPER STUDIED**

|  |  |  |
| --- | --- | --- |
| **PAPER** | **LOOPHOLES** | **TROUBLESHOOTING** |
| Crime Pattern Detection Using Data Mining-Shyam Varan Nath | Quality of the input data may be inaccurate , have missing information or it may be error prone. | Proper preprocessing of data |
| Crime Data Mining for Indian Police Information System | No loopholes mentioned | No troubleshooting |
| An Optimal KD Model for Crime Pattern Detection Based on  Semantic Link Analysis-A Data Mining Tool | No loopholes mentioned | No troubleshooting |
| Data mining for intelligence led policing | **Based on selecting variables :**  Traditional analytical tools require the analyst to look at variables one by one. This way of working is not viable for rich data sets containing many variables.  **2. Static results :**  Existing systems usually generate static reports that do not allow interaction. They cannot be used to find the explanations behind the numbers they present.  **3. Difficult extraction:** It is typically hard to extract data from police source systems because of old and diverse database systems with data models based on transactions instead of analysis. | 1.**Ready database**  providing a single all-embracing database in which all the data has been extracted, linked, cleaned and augmented. The aim is that the database covers 99% of the information need. The remaining 1% needs dedicated data extraction and preparation.  2. **Interactive analysis**  The system works like an interactive analytical instrument in which every part of the results is clickable to 'zoom in'. In this way, the user can simply embark on an analytical journey without the need to first design the process  3. Although the user is working interactively, it is important to keep track of the steps that were taken to reach a result, especially because such documentation can be required in court. Therefore, the system keeps track of the history of each result. |
| Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. | As the dataset is very large, this method is very time consuming | Enhancing the KNN approach by using a new distance metric  methods combined with generalized relevance learning to perform  the classification and missing value treatment simultaneously. |
| Spatio-Temporal Crime Prediction Model Based On Analysis  Of Crime Clusters | All points are included in clusters is one of the limitations of these approach. Spatial outliers are forced to be included to clusters, hence cluster orientation and sizes are deviating from the optimal. | Either exclude the outliers or include them but get satisfied with a non-optimal answer |

**BACKGROUND MATERIAL**

**COMPARATATIVE ANALYSIS OF RESEARCH PAPER STUDIED**

|  |  |
| --- | --- |
| **PAPER** | **ADVANTAGES/CONTRIBUTIONS** |
| Crime Pattern Detection Using Data Mining-Shyam Varan Nath | It clusters the crimes based on the weighing techniques, to come up with crime groups which contain the possible crime patterns. |
| Crime Data Mining for Indian Police Information System | Provides and efficient technique to form classes |
| An Optimal KD Model for Crime Pattern Detection Based on  Semantic Link Analysis-A Data Mining Tool | 1. To formulate crime pattern detection as machine learning task and to thereby use data mining to support police detectives in solving unsolved crimes. 2. Link based modeling technique is able to identify the crime location patterns from a large number of crime locations making the job for crime detectives easier. 3. A model for predicting the crime hot-spots that will help in the deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources |
| Data mining for intelligence led policing | Many case studies that give and in-depth analysis of case |
| Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. | Clustering techniques were analyzed in their efficiency in forming accurate clusters, speed of creating clusters, efficiency in identifying crime trend, identifying crime s, crime density of a state and efficiency of a state in controlling crime rate. |
| Spatio-Temporal Crime Prediction Model Based On Analysis  Of Crime Clusters | By comparative analysis it was found that STAC gave a better result and was hence implement  Partitioning based clustering algorithms are preferably used to allocate resources  effectively |

**DESIGN ANALYSIS AND MODELLING**

**REQUIREMENT ANALYSIS**

* Data sets of crime types having attributes such as : year of crime, X and Y coordinates of crime, location, place, time, date, day of the week ,arrests, domestic violence etc.
* Data sets of criminals i.e. the solved cases having attributes such as: name of criminal, place , XY coordinates, crime committed, etc .
* Data sets of unsolved cases (treated as test cases) containing some attributes matching the tables we have and then we predict the criminal using DM techniques.
* Crime records for the last 50 years for time series analysis of data.
* Pre processing of data sets to remove redundant or missing attributes.

**DESIGN ANALYSIS AND MODELLING**

**DESIGN AND MODELLING**

**USER**

**POLICE**

**ADMIN**

**DESIGN ANALYSIS AND MODELLING**

**ALGORITHM TO BE USED**

* **Time Series analysis** of data and forecasting using **5-year moving average**
* **Clustering** for Crime pattern Detection
* **Clustering and NN** for Criminal Pattern Detection
* **Association** for Criminal Pattern Detection i.e. to find out the serial killer link etc.
* **K-Nearest Neighbor** Algorithm for pre-processing of data

**APPENDIX A**

**THE DATA FORMAT AND SAMPLE DATA**

* All data sets are in the form of CSV files loaded in MySql of WAMP server.
* Csv files of all clusters
* Csv files of crime type clusters
* Csv files of association clusters
* Csv files of criminal prediction with error

**WEKA TOOL ANALYSIS**

**DBSCAN**

=== Run information ===

Relation: murderers

Instances: 19453

Attributes: 24

1 , Last Name , First Name ,Middle Name ,Street Address ,City , State , Zip Code , Residence County , X , Y ,Height , Weight, Race, Gender , Date of Birth , Victim , Status , Classification , Conviction County , Conviction State , Age of Victim , Age of Offender at Time of Offense

Crimes

Test mode: evaluate on training data

=== Clustering model (full training set) ===

MakeDensityBasedClusterer:

Wrapped clusterer:

kMeans

======

Number of iterations: 8

Within cluster sum of squared errors: 194684.65456352133

Missing values globally replaced with mean/mode

Cluster centroids:

Cluster#

Attribute Full Data 0 1

(19453) (7181) (12272)

==========================================================

1 9728 10662.1287 9181.3916

Last Name JOHNSON JOHNSON BROWN

First Name MICHAEL MICHAEL MICHAEL

Middle Name A A A

Street Address Out of State Illinois Department of Corrections Out of State

City CHICAGO CHICAGO CHICAGO

State IL IL

Zip Code

Residence County IDOC IDOC COOK

X -88.6195 -88.6375 -88.609

Y 40.9639 40.9499 40.9721

Height 533.3028 531.254 534.5016

Weight 186.3227 183.9981 187.683

Race W B W

Gender M M M

Date of Birth 5/14/1982 4/19/1979 5/14/1982

Victim Victim under the age of 18 Victim under the age of 18 Victim under the age of 18

Status Compliant Compliant Compliant

Classification Sexual Predator Sexual Predator Sexual Predator

Conviction County COOK COOK COOK

Conviction State IL IL IL

Age of Victim 13.0

Age of Offender at Time of Offense 18.0

Crimes AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16 AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16 AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16

Fitted estimators (with ML estimates of variance):

Cluster: 0 Prior probability: 0.3692

Time taken to build model (full training data) : 6.97 seconds

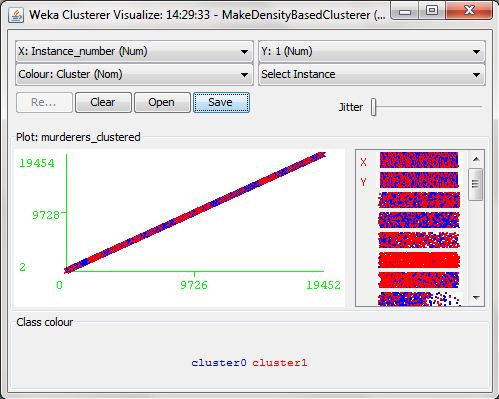
=== Model and evaluation on training set ===

Clustered Instances

0 7805 ( 40%)

1 11648 ( 60%)

Log likelihood: -82.79991



**kMeans**

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -S 10

Relation: murderers

Instances: 19453

Attributes: 24

1 , Last Name , First Name ,Middle Name ,Street Address ,City , State , Zip Code , Residence County , X , Y ,Height , Weight, Race, Gender , Date of Birth , Victim , Status , Classification , Conviction County , Conviction State , Age of Victim , Age of Offender at Time of Offense

Test mode: evaluate on training data

=== Clustering model (full training set) ===

kMeans

======

Number of iterations: 8

Within cluster sum of squared errors: 194684.65456352133

Missing values globally replaced with mean/mode

Cluster centroids:

Cluster#

Attribute Full Data 0 1

(19453) (7181) (12272)

==========================================================

1 9728 10662.1287 9181.3916

Last Name JOHNSON JOHNSON BROWN

First Name MICHAEL MICHAEL MICHAEL

Middle Name A A A

Street Address Out of State Illinois Department of Corrections Out of State

City CHICAGO CHICAGO CHICAGO

State IL IL

Zip Code

Residence County IDOC IDOC COOK

X -88.6195 -88.6375 -88.609

Y 40.9639 40.9499 40.9721

Height 533.3028 531.254 534.5016

Weight 186.3227 183.9981 187.683

Race W B W

Gender M M M

Date of Birth 5/14/1982 4/19/1979 5/14/1982

Victim Victim under the age of 18 Victim under the age of 18 Victim under the age of 18

Status Compliant Compliant Compliant

Classification Sexual Predator Sexual Predator Sexual Predator

Conviction County COOK COOK COOK

Conviction State IL IL IL

Age of Victim 13.0

Age of Offender at Time of Offense 18.0

Crimes AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16 AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16 AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16

Time taken to build model (full training data) : 4.29 seconds

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

Clustered Instances

0 7181 ( 37%)

1 12272 ( 63%)

**FARTHEST FIRST**

=== Run information ===

Scheme: weka.clusterers.FarthestFirst -N 2 -S 1

Relation: murderers

Instances: 19453

Attributes: 24

1 , Last Name , First Name ,Middle Name ,Street Address ,City , State , Zip Code , Residence County , X , Y ,Height , Weight, Race, Gender , Date of Birth , Victim , Status , Classification , Conviction County , Conviction State , Age of Victim , Age of Offender at Time of Offense

Test mode: evaluate on training data

=== Clustering model (full training set) ===

FarthestFirst

==============

Cluster centroids:

Cluster 0

3135.0 BUTTON RICHARD L 1222 N FRINK PEORIA IL 61606.0 PEOR -89.60907998 40.70238845 600.0 215.0 W M 5/16/1939 Victim under the age of 18 Compliant Sexual Predator AZ AZ 8.0 56.0 AGGRAVATED CRIMINAL SEXUAL ABUSE/VICTIM 13-16

Cluster 1

19015.0 ROBINSON JEROME A Out of State ISOLD MS 38754.0 OOFS -88.6195128024768 40.9638723671553 500.0 130.0 B M 7/4/1980 Victim over the age of 18 Non-Compliant Sexual Predator COOK IL 22.0 AGGRAVATED CRIMINAL SEXUAL ABUSE

Time taken to build model (full training data) : 0.25 seconds

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

Clustered Instances

0 11223 ( 58%)

1 8230 ( 42%)

**DESIGN , ANALYSIS AND MODELLING**

**ALGORITHMS USED**

* **Time Series Analysis** of Crime Data using **5 year Moving Average Algorithm**

**CRIME ANALYSIS :**

* **DBScan Algorithm** to do clustering of hot spots i.e. density based clustering based on X and Y coordinates
* **DBScan Algorithm** to create clusters inside the clustering performed on above clusters based on type of crime
* **PAM :** To find out the most dangerous crime over the created cluster based on costs
* **Apriori Association Algorithm** to detect the pattern of crime over the area and its prediction of risk.

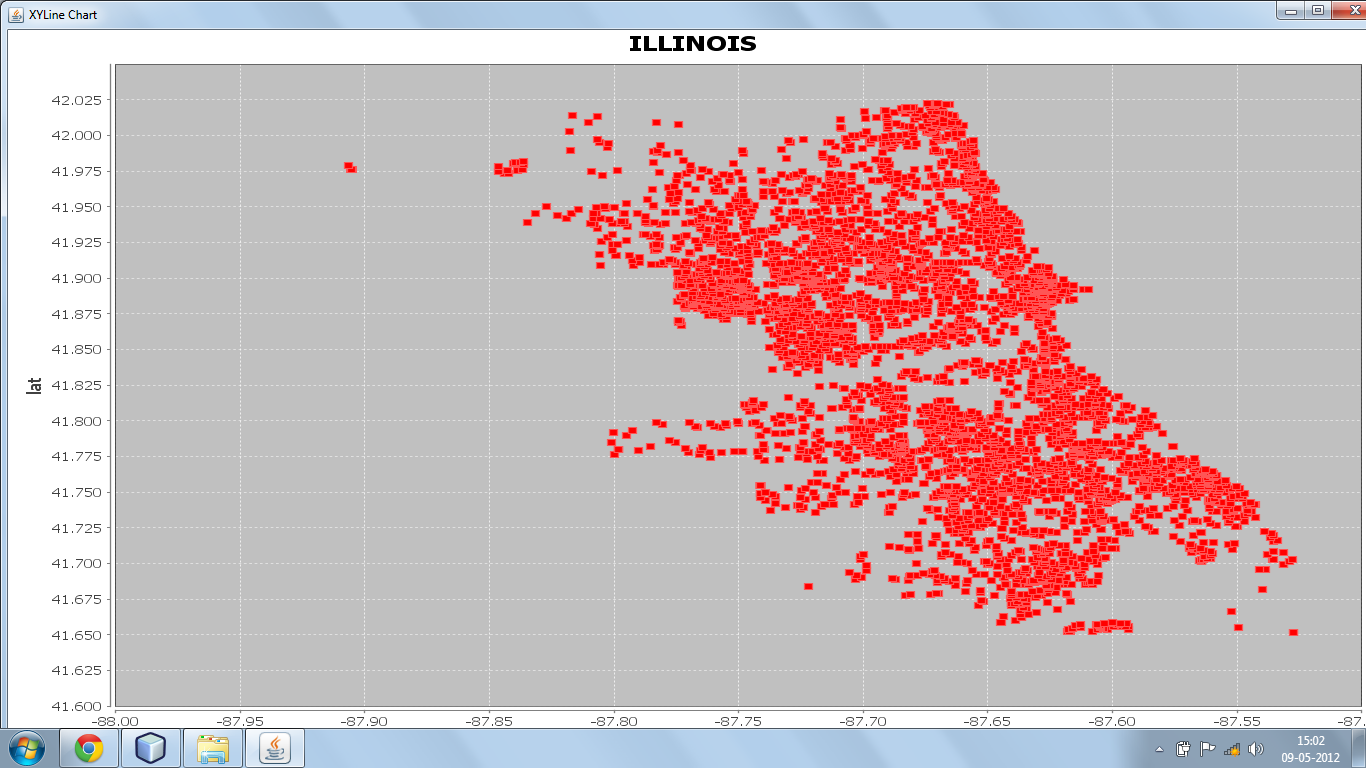
**CRIMINAL ANALYSIS:**

* **Neural Network Algorithm**  to predict the possible criminals based on inputs to the network
* **Associative Criminal Trend Analysis using APRIORI algorithm**  to detect the pattern of criminals age wise, location wise , street wise etc.

**EXPERIMENTAL RESULTS**

**PROCESSES AND ALGORITHM DISCUSSION IN DETAIL(CRIME)**

1. **PRE-PROCESSING:**
2. Cleaning of the data sets files (csv files) by removing the records having missing attributes or faulty attributes through the use of JAVA code.

****

1. **PART A: HOT SPOT DETECTION USING X AND Y COORDINATES**

**Use of DBScan Algorithm to find out the hot spots i.e. the places with the maximum density of crime activities (any crime) using the X and Y coordinates of the crime activities noted.**

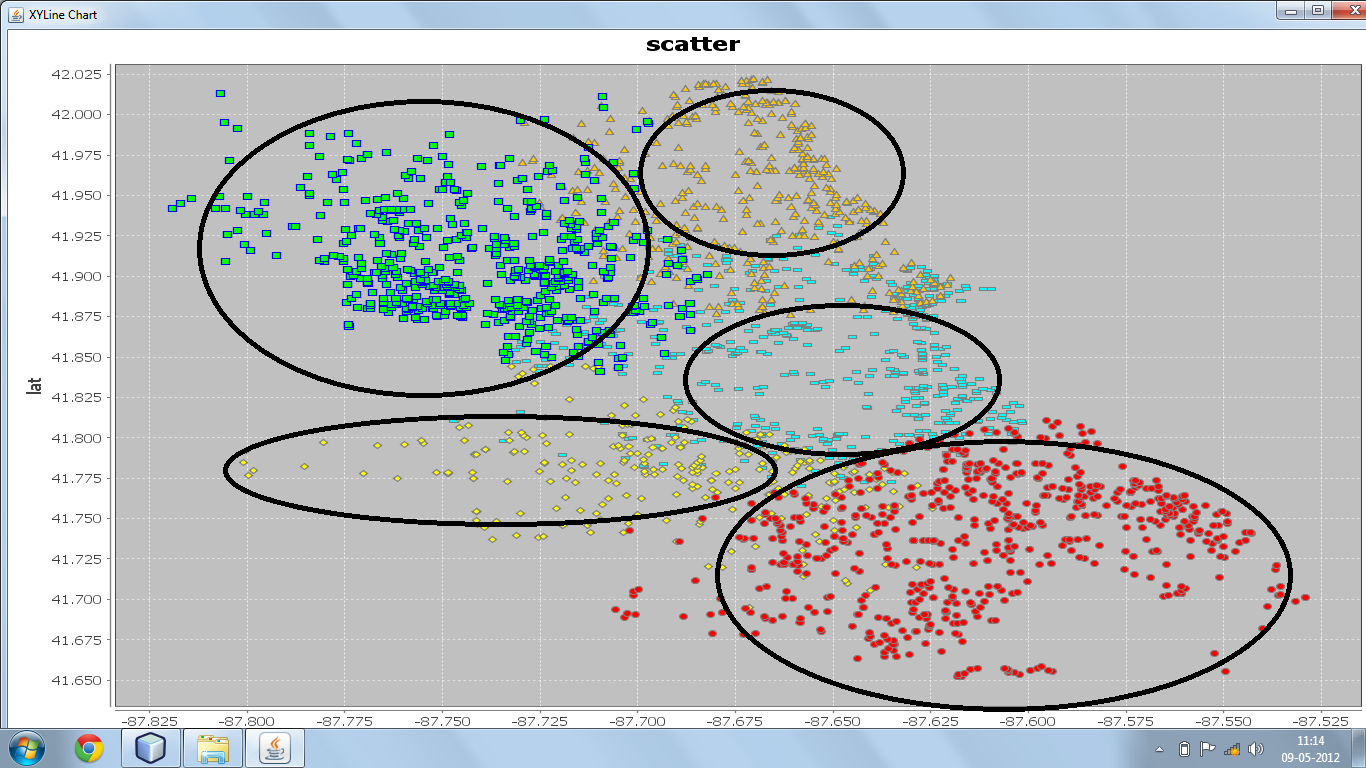
* Algorithm used: DBScan
* Threshold used :

To make cores : 0.0926

To combine cores : 0.0927

* Sparcity problem : Yes , pruned out sparsed data
* Outliers : Yes, core points not satisfying the minimum number of points required to form a core (here the minimum number of points criteria has been set to be 500)
* Complexity of algorithm used (space and time)
* Reason for threshold taken :
* Reason for algorithm used :

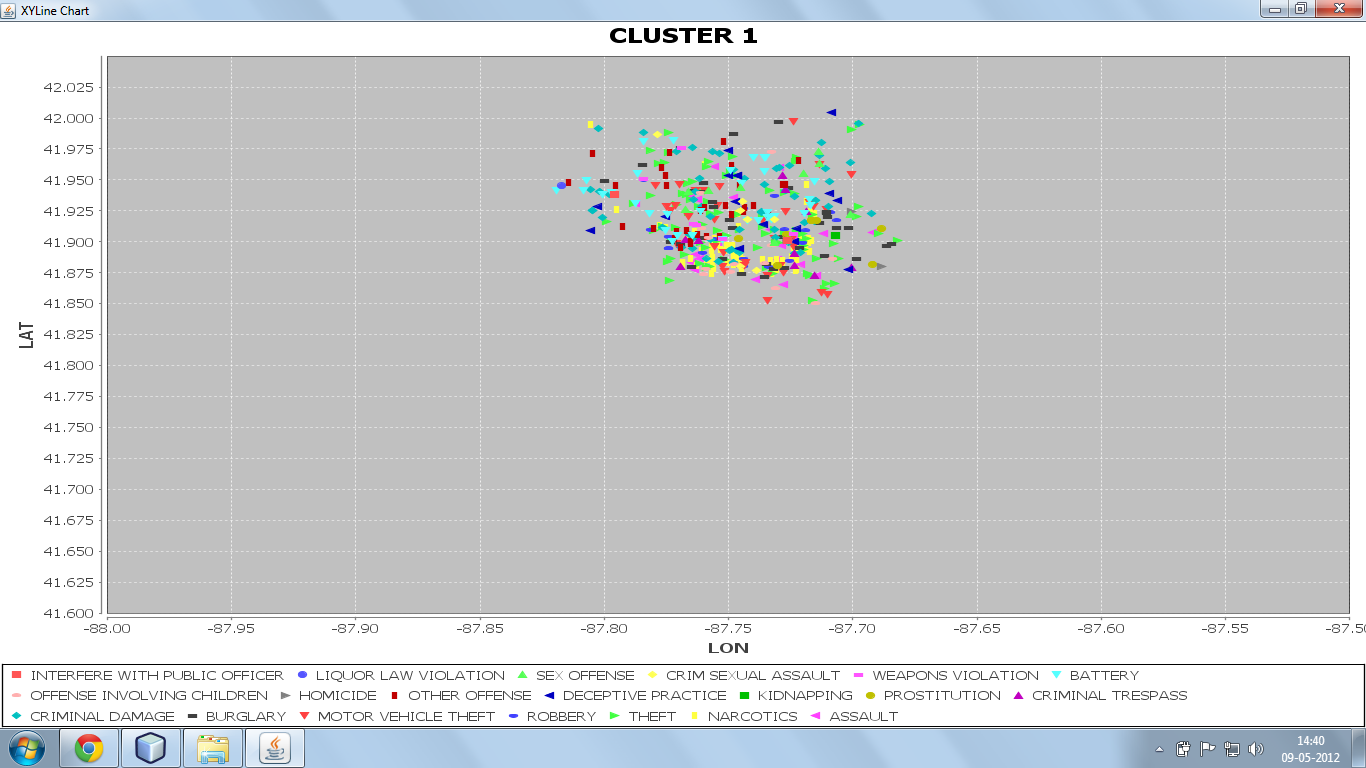
Space time complexity : O(n) and O(n2)

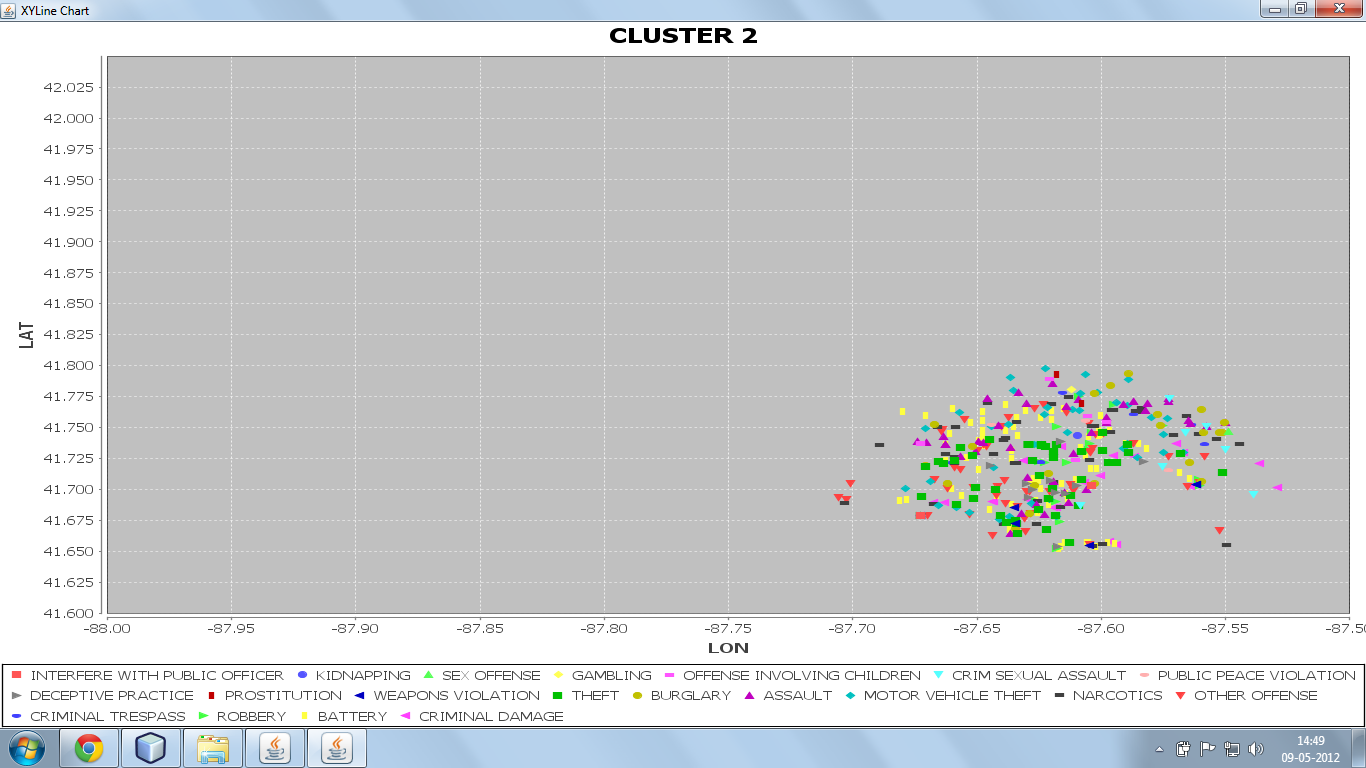


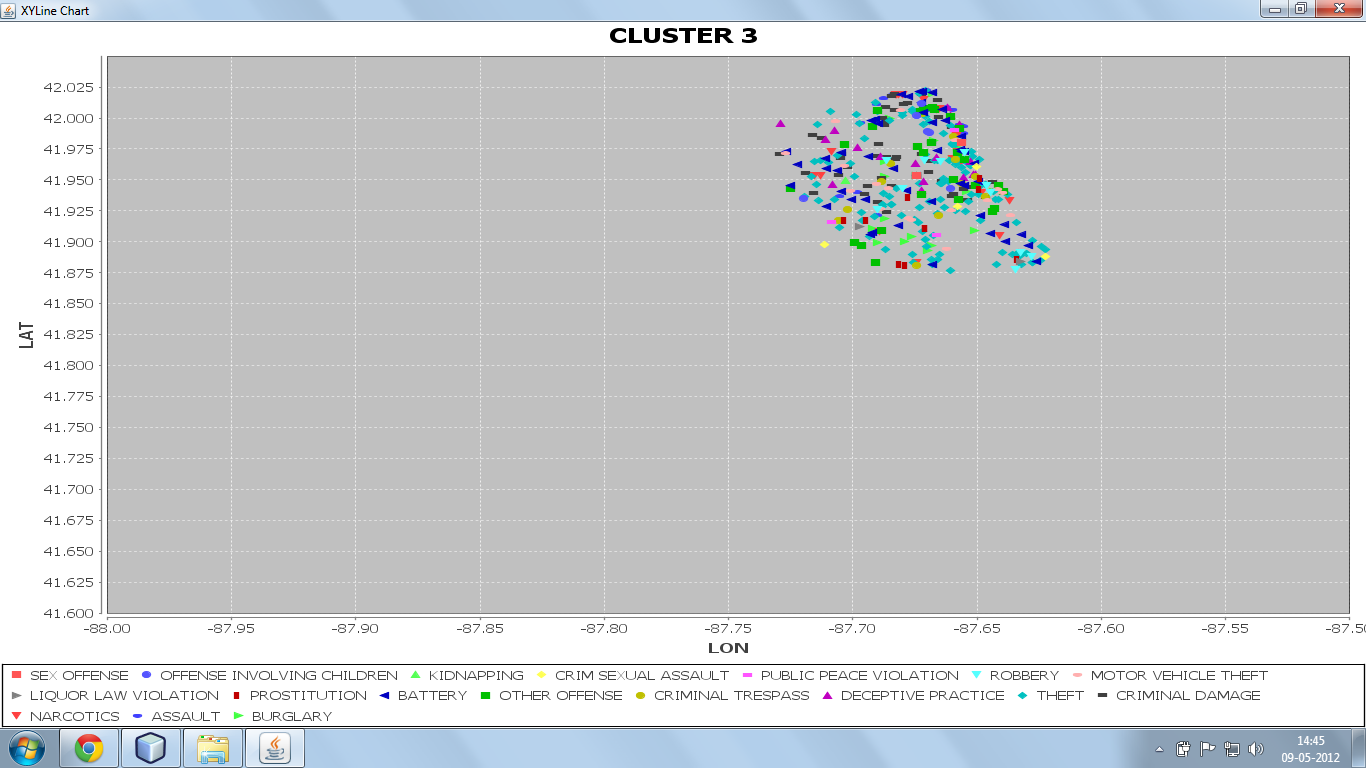
1. **PART B : CLUSTERING ON THE BASIS OF CRIME TYPE ON THE CREATED CLUSTERS**

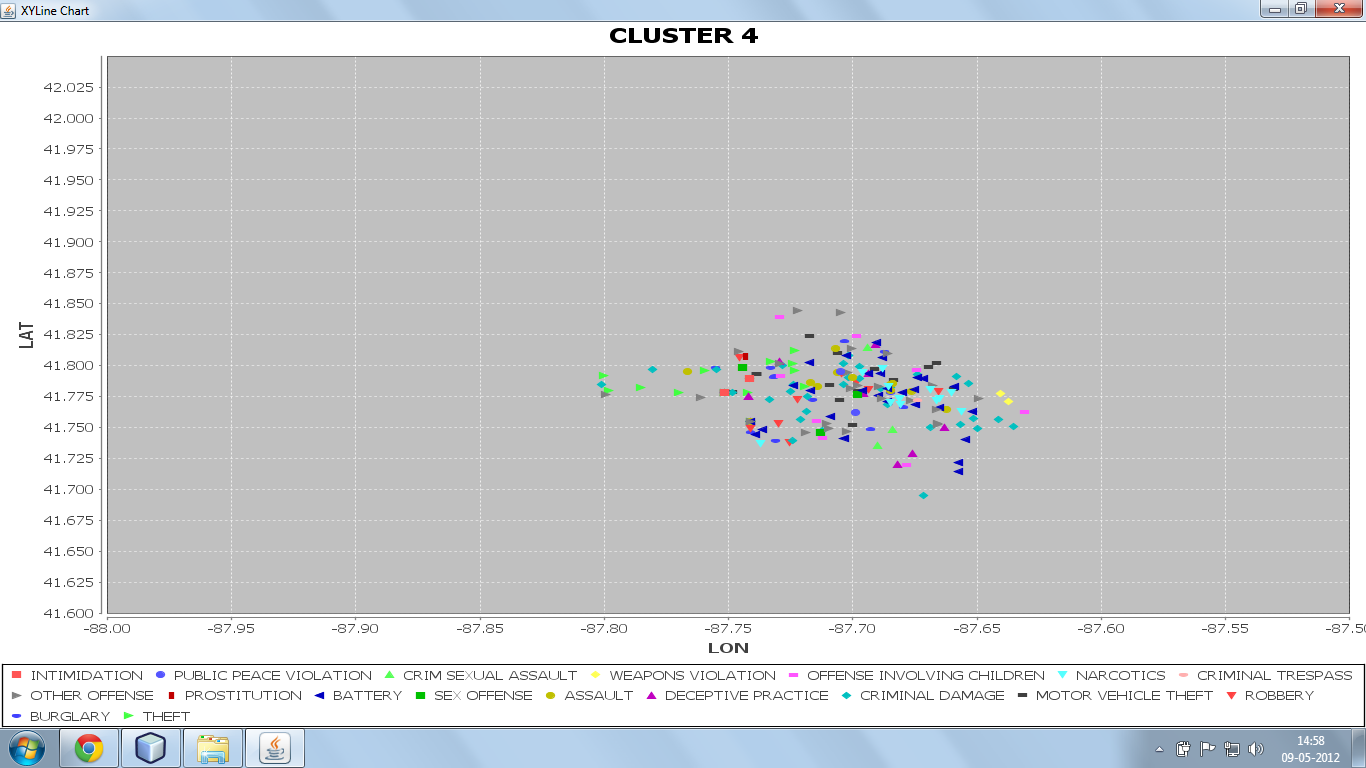
**Use of DBScan Algorithm to find out particular crime type hot spots in the clusters defined.**

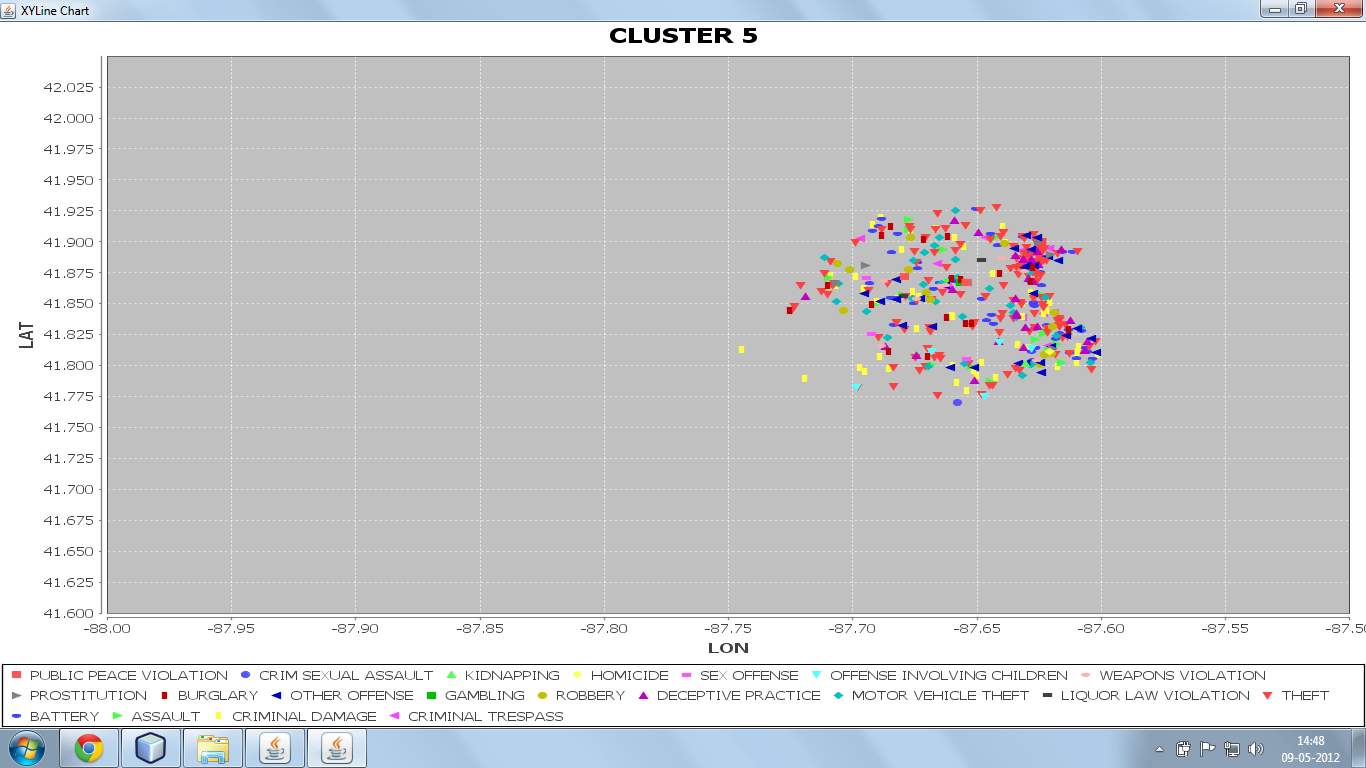
* Algorithm used : DBScan
* Reason of the algorithm used :
* Space time complexity : O(n) and O(n2)
* Threshold used : 0.053
* Reason of threshold :
* Outliers : Yes, pruned out those values not satisfying threshold.
* Sparcity problem : Yes, scatterd data. Pruning performed to deal.





****

****

****

1. **PART C : APPLYING ASSOCIATION TO DETECT THE PATTERN OF CRIME OVER ILLINOIS LOCATION, TIME, CRIME TYPE**

* **Use of Apriori Association Algorithm to detect the support and confidence of sets including time of crime, location of crime and type of crime. Also to tell the percentage possibility of crime in a particular area and particular street at a particular time.**
* Algorithm used : Apriori
* Reason of algorithm :
* Threshold : average of all supports
* Sparcity : no
* Outliers : yes , pruned out
* Association performed on groups i.e.

At times :

* 6p.m to 8p.m
* 8 p.m to 10p.m
* 10 p.m to 12p.m

On every distinct location like street, restaurant, park of distinct types of crime.

**EXPERIMENTAL RESULTS**

**PROCESSES AND ALGORITHM DISCUSSION IN DETAIL(CRIMINAL)**

1. **NEURAL NETWORK WITH ONE HIDDEN LAYER TO PREDICT CRIMINAL BASED ON HEIGHT, WEIGHT , GENDER AND RACE**

**As the activation function used is hyperbolic tangent function, the values of the inputs(height, weight, gender and race) has been converted into values between 0-1 as follows**

Ranges taken as inputs :

Height(feetinch)

0.1 <=500

0.2 [501,503]

0.3 [504,506]

0.4 [507,509]

0.5 [510,600]

0.6 [601,603]

0.7 >=604

Weight(in pounds)

0.1 <=150

0.2 [151,170]

0.3 [171,190]

0.4 [191,210]

0.5 [211,230]

0.6 >=231

Gender

0.1 Male

0.5 Female

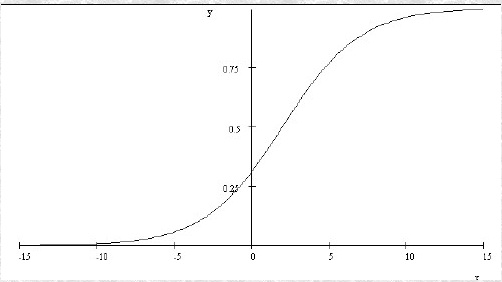
Race

0.1 Black

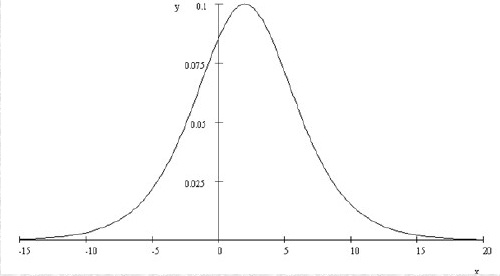
0.2 White

0.3 Other/Unknown

**HYPERBOLIC TANGENT FUNCTION(ACTIVATION FUNCTION FOR NEURAL NETWORK)**



**DIFFERENTIATION OF HYPERBOLIC TANGENT FUNCTION**



1. **ASSOCIATION ALGORITHM USED TO PREDICT THE POSSIBLITY OF AGE OF CRIMINAL COMMITTING WHAT TYPE OF CRIME , WHEN AND WHERE**

ASSOCIATION: apriori to detect criminal trends based on age of offender, most dangerous location , street and state.

Divided into groups based on the age of the offender at the time of crime performed :

* Below 18
* 18-35
* 35 and above

**CONCLUSION**

* **We successfully determine the hot spots / high density crime areas as shown in the plots.**
* **Association gives us the percentage risk of a particular crime occurring at a particular area at a particular location at a particular time.**
* **The neural network used is trained over the data sets using supervised learning, then the weights fixed are used to predict the 4th unknown parameter given three known parameter.**
* **The association used gives us a visualization of the age group of the offenders seen mostly at what time what location and what type of crime**

**FUTURE WORK**

* **Currently done only on the state of Illinois. The future aspect of the project is to extend these over the entire state of U.S**
* **Application of biometrics to predict criminals based on 3D facial recognition and fingerprints matching using data mining algorithms.**

**EXPERIMENTAL RESULTS**

**PSEUDO CODE**

**DBSCAN :**

**Input:**

D ={t1,t2….tn} //Set of Elements

MinPts //Number of points in cluster

Eps //Maximum distance for density measure

**Output:**

K={K1, K2…..Kk}

**DBSCAN Algorithm**

K=0;

for i=1 to n do

if ti is not in a cluster, then

X={tj|tj is density-reachable from ti};

if X is a valid cluster, then

k=k+1;

Kk=X;

**APRIORI ALGORITHM :**

**Input :**

I //itemsets

D //Database of transactions

s //Support

**Output:**

L //Large itemsets

**Aprioiri algorithm**

k=0; //k is used as the scan number

L={};

For each Ii€ Ck do

Ci =0;//Initial counts for each itemset are 0

For each tj€ CK do

If Ii € tj then

Ci=ci+1;

For each Ii € Ck do

If ci>=(s X |D|) do

Lk = Lk U Ii; L = L U Lk;

Ck+1 =Apriori-gen(Lk)

Until Ck+1={};

**NEURAL NETWORK :**

**Input:**

N //neural network

X={x1,…xh} //input tuple consisting of values for input attributes only

**Output:**

Y={yi,…ym};

Propagation Algorithm:

For each node i in the input layer do

Output xi on each output arc from i;

For each hidden layer do

For each node i do

Si=summation(j=1 to k) wji\*xji;

For each output arc from I do

Output (1-e-Si)/(1+e-CSi);

For each node i in the output payer do

Si=summation(j=1 to k) wji\*xji;

Output yi =1/(1+e-CSi);

**NN**

N //Startig neural network

X={x1,…xh} //Input tuple from training set

D={d1,…dm} // Output tuple desired

Output:

N //Impoved neural network

Backpropagation algorithm

Propagation(N,X);

E=1/2\*[summation(i=1 to m) (di-yi)2]

Gradient (N,E);

**APPENDIX C**

**CODE**

**NEURAL NETOWRK**

class Main {

double err1=0,err2=0,err3=0,err11=0,err22=0,err33=0;

double[] inputlayer = new double[3];

double[] hiddenlayer = new double[3];

double outputlayer;

double output;

double inputweights[][] = new double[hiddenlayer.length][inputlayer.length];

double hiddenweights[] = new double[hiddenlayer.length];

double hiddenoutput[] = new double[hiddenlayer.length];

void disp(){

System.out.println("input weights:");

for(int i=0;i<3;i++){

for(int j=0;j<3;j++){

System.out.println(inputweights[i][j]);

}

}

System.out.println("hidden weights weights:");

for(int i=0;i<3;i++){

System.out.println(hiddenweights[i]);

}

}

void getInputValues(double wt,double r,double g) {

inputlayer[0]=wt;

inputlayer[1]=r;

inputlayer[2]=g;

}

void getHiddenWeights() {

System.out.println("enter the values of weights of HIDDEN LAYER to OUTPUT LAYER");

for (int i = 0; i < hiddenlayer.length; i++) {

hiddenweights[i]=Math.random() \* 1;

System.out.println(hiddenweights[i]);

}

}

double activationFunction(final double d) {

return((1-Math.exp(-d))/(1+Math.exp(-d)));

}

double activationFunctionDiff(final double d){

return((2\*Math.exp(-d))/(Math.pow(1+Math.exp(-d),2)));

}

void getInputWeights() {

System.out.println("enter the values of weights of INPUT LAYER to HIDDEN LAYER");

for (int i = 0; i < inputlayer.length; i++) {

for (int j = 0; j < hiddenlayer.length; j++) {

inputweights[j][i]=Math.random()\*1;

System.out.println(inputweights[j][i]);

}

}

}

void calculateHiddenNodes() {

int j = 0;

int k=0;

hiddenlayer[0]=0;

hiddenlayer[1]=0;

hiddenlayer[2]=0;

while (j < inputlayer.length){

for (int i = 0; i < inputweights.length; i++) {

hiddenlayer[k] += inputlayer[i] \* inputweights[i][j];

}

j = j + 1;

k=k+1;

}

for (int i = 0; i < inputweights.length; i++) {

hiddenoutput[i] = activationFunction(hiddenlayer[i]);

}

}

void setNewWeights(double err){

err1=hiddenweights[0]\*err;

err2=hiddenweights[1]\*err;

err3=hiddenweights[2]\*err;

err11=inputweights[0][0]\*err1+inputweights[0][1]\*err2+inputweights[0][2]\*err3;

err22=inputweights[1][0]\*err1+inputweights[1][1]\*err2+inputweights[1][2]\*err3;

err33=inputweights[2][0]\*err1+inputweights[2][1]\*err2+inputweights[2][2]\*err3;

inputweights[0][0]+= err1\*activationFunctionDiff(hiddenlayer[0])\*inputlayer[0];

inputweights[1][0]+= err1\*activationFunctionDiff(hiddenlayer[0])\*inputlayer[1];

inputweights[2][0]+= err1\*activationFunctionDiff(hiddenlayer[0])\*inputlayer[2];

inputweights[0][1]+= err2\*activationFunctionDiff(hiddenlayer[1])\*inputlayer[0];

inputweights[1][1]+= err2\*activationFunctionDiff(hiddenlayer[1])\*inputlayer[1];

inputweights[2][1]+= err2\*activationFunctionDiff(hiddenlayer[1])\*inputlayer[2];

inputweights[0][2]+= err3\*activationFunctionDiff(hiddenlayer[2])\*inputlayer[0];

inputweights[1][2]+= err3\*activationFunctionDiff(hiddenlayer[2])\*inputlayer[1];

inputweights[2][2]+= err3\*activationFunctionDiff(hiddenlayer[2])\*inputlayer[2];

hiddenweights[0]+= err\*activationFunctionDiff(outputlayer)\*hiddenlayer[0];

hiddenweights[1]+= err\*activationFunctionDiff(outputlayer)\*hiddenlayer[1];

hiddenweights[2]+= err\*activationFunctionDiff(outputlayer)\*hiddenlayer[2];

}

double calculateOutput() {

outputlayer=0;

for (int i = 0; i < hiddenweights.length; i++){

outputlayer += hiddenoutput[i] \* hiddenweights[i];

}

output = activationFunction(outputlayer);

return output;

}

}

%>

//backpropagation

double prob=0;

if(Math.abs(0.1-o)<Math.abs(0.5-o))

{ch="m";}

else

{

ch="f";}

err=(g-o);

errp=Math.abs(err)\*100;

out.println("<tr><td>"+i+"</td><td>"+wt+"</td><td>"+r+"</td><td>"+h+"</td><td>"+gender+"</td><td>"+ch+"</td><td>"+err+"</td><td>"+errp+"</td></tr>");

count=0;

while(Math.abs(err)>0.1 && count!=20)

{

count=count+1;

n.setNewWeights(err);

n.calculateHiddenNodes();

o= n.calculateOutput();

err=(g-o);

}

}

**ASSOCIATION**

BufferedReader CSVFile = new BufferedReader(new FileReader(f));

String dataRow = CSVFile.readLine();

FileWriter fw= new FileWriter("D:/minor2/PROJECTCSV/CriminalSupport.csv");

BufferedWriter o = new BufferedWriter(fw);

double count=0;

while(dataRow !=null)

{

count++;

dataRow = CSVFile.readLine();

}

CSVFile.close();

BufferedReader CountFile = new BufferedReader(new FileReader(f1));

String dataRow1 = CountFile.readLine();

int sum=0;

while(dataRow1!=null)

{

String[] k = dataRow1.split(",");

double i = Double.parseDouble(k[3]);

d=(double)i/count;

o.write(k[0]);

o.write(",");

o.write(k[1]);

o.write(",");

o.write(k[2]);

o.write(",");

String h=new Double(d).toString();

o.write(h);

o.write("\n");

dataRow1=CountFile.readLine();

**CLUSTERING USING DBSCAN:**

package pkgfinal;

import java.io.\*;

import java.util.\*;

class p

{

String ccode,date,addres,crime,crimed,area,arrest,domes,beat;

int x,y,year;

float lat;

float lon;

public p() {

this.ccode = "";

this.date = "";

this.addres = "";

this.crime ="";

this.crimed = "";

this.area = "";

this.arrest = "";

this.domes = "";

this.beat = "";

this.x = 0;

this.y = 0;

this.year = 0;

this.lat = (float)0.0;

this.lon = (float)0.0;

}

}

public class hdgh

{

class cen

{

float cx,cy;

int size;

String initial;

}

public void db(){

try{

FileInputStream fstream = new FileInputStream("e:/small.csv");

DataInputStream in = new DataInputStream(fstream);

BufferedReader br = new BufferedReader(new

InputStreamReader(in));

FileWriter fw = new FileWriter("e:/test2/cen.csv");

FileWriter fwinner = new FileWriter("e:/test2/core.csv");

int minpnts=50,x,y;

String []temp;

String input;

int i=0,k=0,flag=0,ctr=0;

float rad = (float) 0.14;

String str;

List inner=new LinkedList();

List outer=new LinkedList();

p p5=new p();

p p3=new p();

p p2=new p();

int m=0;

while ((str = br.readLine()) != null)

{

if(ctr%10000==0)

{

System.out.println(ctr);

}

try

{

p p1=new p();

int innerctr=0;

ctr++;

temp=str.split(",");

p1.ccode=temp[0];

p1.date =temp[1];

p1.addres= temp[2];

p1.beat =temp[3];

p1.crime= temp[4];

p1.crimed= temp[5];

p1.area= temp[6];

p1.arrest=temp[7];

p1.domes=temp[8];

p1.x=Integer.parseInt(temp[9]);

p1.y=Integer.parseInt(temp[10]);

p1.year= Integer.parseInt(temp[11]);

String lat =temp[12];

String lon =temp[13];

p1.lat=Float.parseFloat(lat);

p1.lon=Float.parseFloat(lon);

System.out.println("new value is : "+p1.lat+p1.lon);

int size=inner.size();

if(i>0)

{ flag=0;

int flag3=0;

for(int q=0;q<outer.size();q++)

{

inner=(List)outer.get(q);

innerctr=0;

for(int e=0;e<inner.size();e++)

{

p2=(p)inner.get(e);

float diff= (float) Math.sqrt(Math.pow((p1.lat-p2.lat),2)+Math.pow((p1.lon-p2.lon),2));

if(diff<0.0926)

{

innerctr++;

}

else{

break;

}

}

if(innerctr==inner.size())

{

flag3++;

inner.add(p1);

outer.remove(q);

outer.add(inner);

flag=1;

break;

fwinner.write(temp[0]+",");

}

}

if(flag==0)

{

inner=new LinkedList();

inner.add(p1);

outer.add(inner);

}

}

else

{

inner.add(p1);

outer.add(inner);

i++;

}

}

catch(Exception e)

{System.out.println(e);

}

}

System.out.println("core");

System.out.println("outer"+(outer.size()));

float cenx=0,ceny=0;

for(int q=0;q<outer.size();q++)

{

inner=(List)outer.get(q);

System.out.println(inner.size());

p3=(p)inner.get(0);

fw.write(p3.ccode+","+inner.size()+",");

cenx=0;

ceny=0;

for(int e=0;e<inner.size();e++)

{

p2=(p)inner.get(e);

fwinner.write(p2.ccode+",");

cenx=cenx+p2.lat;

ceny=ceny+p2.lon;

}

cenx=(float)(cenx/inner.size());

ceny=(float)(ceny/inner.size());

fw.write(cenx+","+ceny+"\n");

fwinner.write("\n");

}

fw.close();

fwinner.close();

}

**CLUSTERING USING DBSCAN-2:**

package pkgfinal;

import java.io.\*;

import java.util.Iterator;

import java.util.LinkedList;

import java.util.List;

public class Final {

void plotl()

{

try{

FileInputStream fstream = new FileInputStream("e:/test/outf.csv");

DataInputStream in = new DataInputStream(fstream);

BufferedReader br = new BufferedReader(new InputStreamReader(in));

FileWriter fw[]=new FileWriter[8];

fw[0]=new FileWriter("e:/test/outclus.csv");

String []temp;

String str;

String []temp2;

String str2;

int flag=0,flag2=0;

String str3;

List inner=new LinkedList();

List outer=new LinkedList();

int ctr=0;

int cnt=0;

int sizearr;

int wctr=0;

int j=0,k=0;

int len;

int fctr=0;

while ((str = br.readLine()) != null)

{

try{

System.out.println(fctr);

j=0;

temp=str.split(",");

ctr=0;

{

while(j<temp.length)

{

FileInputStream fstream1 = new FileInputStream("e:/small.csv");

DataInputStream in1 = new DataInputStream(fstream1);

BufferedReader br1 = new BufferedReader(new InputStreamReader(in1));

String a=temp[j];

while ((str2 = br1.readLine()) != null)

{

k=0;

temp2=str2.split(",");

if(a.equals(temp2[0]))

{

while(k<temp2.length)

{

fw[fctr].write(temp2[k]+",");

k++;

}

ctr++;

fw[fctr].write("\n");

int r=1;

while(r>0)

{

str3=br1.readLine();

temp2=str3.split(",");

if(temp[j+r].equals(temp2[0]))

{int z=0;

while(z<temp2.length)

{

fw[fctr].write(temp2[z]+",");

z++;

}

fw[fctr].write("\n");

ctr++;

}

else{

j=j+r;

break;}

r++;

}

}

}

br1.close();

j++;

} }

}catch(Exception e)

{System.out.println(e);} fctr++;

}for(int i=0;i<8;i++)

{

fw[i].close();

}

}

catch(Exception e)

{

System.out.println(e);

} }

void read()

{

try{

FileInputStream fstream = new FileInputStream("e:/test/finalcluster.csv");

DataInputStream in = new DataInputStream(fstream);

BufferedReader br = new BufferedReader(new InputStreamReader(in));

FileInputStream fstream3 = new FileInputStream("e:/test/outlier.csv");

DataInputStream in3 = new DataInputStream(fstream3);

BufferedReader br3 = new BufferedReader(new InputStreamReader(in3));

FileWriter fw = new FileWriter("e:/test/clustf.csv");

String []temp;

String str;

String []temp2;

String str2;

int flag=0,flag2=0;

cen c1 =new cen();

cen c2=new cen();

cen c3=new cen();

cen c4=new cen();

List inner=new LinkedList();

List outer=new LinkedList();

int ctr=0;

int cnt=0;

int sizearr;

int wctr=0;

int j=0,k=0;

int len;

while ((str = br.readLine()) != null)

{

j=0;

temp=str.split(",");

while(j<temp.length)

{

FileInputStream fstream1 = new FileInputStream("e:/test/core.csv");

DataInputStream in1 = new DataInputStream(fstream1);

BufferedReader br1 = new BufferedReader(new InputStreamReader(in1));

String a=temp[j];

while ((str2 = br1.readLine()) != null)

{ cnt=0;

k=0;

temp2=str2.split(",");

if(a.equals(temp2[0]))

{

while(k<temp2.length)

{

fw.write(temp2[k]+",");

inner.add(temp2[k]);

cnt++;

k++;

}

break;

}

}

br1.close();

j++;

}

fw.write("\n");

outer.add(inner);

inner= new LinkedList();

}

fw.close();

System.out.println("outer"+outer.size());

for(int y=0;y<outer.size();y++)

{ List a=(List)outer.get(y);

System.out.println("inner"+a.size());

}}

catch(Exception e)

{

System.out.println(e);

} }

void fc()

{

try

{

List plz=new LinkedList();

List cluster=new LinkedList();

int i=0;

FileInputStream fstream = new FileInputStream("e:/test/cen.csv");

DataInputStream in = new DataInputStream(fstream);

BufferedReader br = new BufferedReader(new InputStreamReader(in));

FileInputStream fstream1 = new FileInputStream("e:/test/core.csv");

DataInputStream in1 = new DataInputStream(fstream1);

BufferedReader br1 = new BufferedReader(new InputStreamReader(in1));

FileWriter fwout = new FileWriter("e:/test/outlier.csv");

FileWriter fw = new FileWriter("e:/test/finalcluster.csv");

String []temp;

String str;

int flag=0,flag2=0;

cen c1 =new cen();

cen c2=new cen();

cen c3=new cen();

cen c4=new cen();

List inner=new LinkedList();

List outer=new LinkedList();

int ctr=0;

int sizearr;

int wctr=0;

int j=0;

List out=new LinkedList();

while ((str = br.readLine()) != null)

{

wctr++;

System.out.println(str);

try

{

cen c1 =new cen();

temp=str.split(",");

c1.cx=Float.parseFloat(temp[2]);

c1.cy=Float.parseFloat(temp[3]);

c1.size=Integer.parseInt(temp[1]);

c1.initial=temp[0];

if(c1.size<50)

{

out.add(c1);

fwout.write(c1.initial+",");

} else

{ cluster.add(c1); }

} catch(Exception e)

{ System.out.println("yatharth"); }

}

List inner2=new LinkedList();

cen c5=new cen();

for( i=0;i<cluster.size();i++)

{

flag2=0;

if(i==0)

{

c2=(cen)cluster.get(i);

inner.add(c2);

outer.add(inner);

System.out.println("iniout"+outer.size());

}

else

{

c2=(cen)cluster.get(i);

flag2=0;

for(int q=0;q<outer.size();q++)

{

int innerc=0;

inner2 =(List)outer.get(q);

for(int w=0;w<inner2.size();w++)

{

c5=(cen)inner2.get(w);

float diff2= (float) Math.sqrt(Math.pow((c5.cx-c2.cx),2)+Math.pow((c5.cy-c2.cy),2));

if(diff2<(float)0.09277)

{ innerc++;

} }

if(innerc==inner2.size())

{

inner2.add(c2);

outer.remove(q);

outer.add(inner2);

flag2=1;

break;

} } if(flag2==0)

{

System.out.println("f0");

else{

inner2=new LinkedList();

inner2.add(c2);

outer.add(inner2);

System.out.println("out"+outer.size());

} }

} }

System.out.println("out"+outer.size());

for(int q=0;q<outer.size();q++)

{ inner2 =(List)outer.get(q);

for(int w=0;w<inner2.size();w++)

{ c5=(cen)inner2.get(w);

fw.write(c5.initial+","); }

fw.write("\n"); }

fwout.close();

fw.close();

} catch(Exception e)

{ System.out.println("yatharth2");

System.out.println(e); }

System.out.println("fcover"); }