Criminal Identification, Crime Pattern Detection, and Prediction in India using Data Mining

Chithra Shaji Thomas
Assistant Professor
Department of Computer Science & Engineering
Mount Zion Institute of Science and Technology, Alappuzha, India

Abstract— Crimes in India are increasing at an alarming rate, and criminals are opting for queer activities to commit them. Newspapers, blogs, etc. are day to day filled with various crime incidents. So an approach is proposed for the design and implementation of three different applications: Criminal Identification, Crime Pattern Detection, and Prediction for Indian cities, using data mining. The approach is divided into eight modules— data extraction, data preprocessing, classification and accuracy verification, clustering methods comparison, clustering, map representation, WEKA implementation, and correlation and regression using Rtool. Criminal identification and prediction is done using KNN classification. Clustering comparison is done for three methods: k-means, k-medoid, and agglomerative hierarchical clustering, to find the best method. This proves k-means to be the best method for crime data clustering. Crime detection is done by using k-means clustering, which iteratively generates crime clusters that are based on similar crime attributes. Plotting of crime hotspots using open street map provides visualization to crime hotspots. Verification of k-means results is done using WEKA. WEKA verifies an accuracy of 96% and 97% in the formation of two crime clusters using selected crime attributes. WEKA is also used for other data mining operations such as association rule mining, decision tree, naive bayes prediction, etc. Regression and correlation steps will help to find relations among different attributes. This is done using R Studio. R and WEKA thus helps in prediction. The approach is useful in helping the investigating agencies in crime detection and criminals' identification, and for predicting crime trends.

Keywords: Criminal Identification, Crime Pattern, Prediction, Regression, Correlation, WEKA, Apriori, Decision Tree

I. INTRODUCTION

A method is proposed for the design and implementation of three different applications: Criminal identification, Crime pattern detection, and Prediction for Indian cities using data mining techniques. The approach is divided into three different applications consisting of total eight modules - 1. data extraction , 2.data preprocessing , (a) Criminal identification : 3. classification and accuracy verification, (b) Crime pattern detection : 4. clustering methods comparison, 5. clustering, 6. map representation, (c) Crime pattern Prediction and result verification : 7. WEKA implementation, and 8. Correlation and Regression using R tool. First module, DE extracts the unstructured crime dataset from various crime Web sources, for a particular time period. Second module, DP cleans, integrates and reduces the extracted crime data into structured crime instances. These instances are represented using predefined crime attributes. Next five modules are useful for crime detection, criminal identification and prediction, and crime verification, respectively. Criminal identification is done using KNN classification. Initially crime clustering comparison is done by comparing three methods: k-means, k-medoid, and hierarchical clustering. Comparison is done based on V-measure calculation. Crime pattern detection is by using k-means clustering, which iteratively generates crime clusters that are based on similar crime attributes. Plotting of hotspots and shortest distance representation using open street maps in GraphHopper API provides visualization to crime hotspots. Crime verification of results is done using WEKA. WEKA verifies an accuracy in the formation of two crime clusters using selected crime attributes. Regression and correlation steps will help to find relations among different attributes and thus in prediction. This is done using R tool, RStudio. The whole approach contributes in the betterment of the society by helping the investigating agencies in crime detection and criminals' identification, and thus reducing the crime rates.

II. RELATED WORK

Malathi and Baboo [8] used a classification technique (decision trees) to predict crime trends (out of four options) for the following year. They also describe the prediction of the numbers of crimes for a particular year using data from the previous 8 years, although it is not clear what method was used for this numeric prediction. In terms of the data used, no number of records is given; they mention that the data covers 9 years of crime information. Saltos used numerical prediction models and the data used is described in detail, both in terms of features and number of records. Another approach by Oatley and Ewart focused on the prediction of likelihood of repeated burglary for a particular property. For this purpose, they used a Bayesian belief network, using the following features or attributes: offender features; modus operandi features; property stolen; premise crime history; prevalence, incidence and concentration, which are numeric indicators of the distribution of crimes over an area. They used 70,000 records of burglary-related crimes, including motor vehicle theft, street robbery and burglary from dwelling houses. The focus of this research was the development of the software and the paper does not describe any evaluation of the proposed approach. In terms of the Bayesian belief network, the focus is on the interpretability of the output rather than the performance of the method, which is mentioned as part of the future work. In contrast, Saltos [11] approach focuses on the evaluation of prediction models, both in terms of their predictive performance, as well as their complexity, as an important practical aspect that is relevant for large volumes of data. Xue and Brown developed an approach for the prediction of future crime locations based on discrete choice theory and clustering. In terms of data, they used over 1,200 crime records. They compared their proposed

approach with a traditional hot spot identification method and found that their models outperform the traditional ones. A classification approach has been used by Yu et al.[7] to classify areas into hot spots and cold spots, and to predict if an area will be a hot spot for residential burglary. They defined a hot spot as an area with at least 1 crime. They experimented with different levels of aggregation of historical data, and a variety of classification techniques: k-Nearest Neighbor (k-NN), Decision trees, Support Vector Machines (SVM), Neural Network, Naive Bayes and ensemble learning. They found that the best results were obtained with the 1-nearest neighbor and the neural network algorithms. Unlike previous research, they focused on the prediction of crime frequency as a numeric value rather than as a label, because the definition of a hot spot may vary according to: (1) area-1 crime in a low-crime area may constitute a hotspot, while 10 or more crimes may be considered as a hotspot in a high crime area; (2) crime type - some crimes, e.g., anti-social behavior, are much more frequent than others such as armed robbery, and thus, hotspots for different type of crimes need to be defined proportionately to their frequency. The proposed approach also used a large number of records and discusses the time required to build and test prediction models based on such large volumes of data - an aspect that has not been addressed in previous research, but is very important in today's context of large amounts of data available and the practical issues involved in their analysis.

Some of the authors have discussed primary clustering (Chen et al.[6]; Kulis and Jordan[14]; Malathi et al.[9]) and classification (Okonkwo and Enem [17]) techniques for crime detection, criminal identification theoretically; however, none of them provides a sound implementation for the same. Although some papers (Nath [4]; Malathi and Baboo [8]; Malathi et al. [9]) discuss application of k-means for crime detection, but these and other works (Ehlers [12]; Chen et al.[6]; Hussain et al. 2012) are deficient in integration among crime detection, criminal identification and prediction, and crime verification. Malathi et al. (2011) work with crime attributes—number of crimes of a particular crime type, e.g., murder and burglary, versus years. The proposed method, in addition to Malathi et al. Works with attributes number_ of_crimes_committed_in_year versus crime_year. These crime attributes are considered as follows: (1) independent of attributes crime_location and crime_type; (2) dependent on attribute crime_location, but independent of attribute crime_type, etc. Nath [4] tries to detect crime suspects based on their races, age and sex. On the other hand, proposed classification scheme speculates suspects based on the finegrained attributes—suspect_name, suspect_age,suspect_sex, suspect_facial_feature, and suspect_nationality. Mande et al.[16] states that criminal identification is based on Gaussian mixture models. They solely depend on the eye-witness information. They confine to only one state of India, i.e., Andhra Pradesh for criminal records. Proposed method works on crime data of several Indian cities that are selected based on their crime rates. Jin et al. in their proposal define the position of crime events with longitude and latitude using k-means. Proposed scheme defines the formation of clusters firstly using kmeans and then using map. Okonkwo and Enem [17] confer about terrorism attack as a type of crime. They recommend government to set up data mining agencies within the law enforcement agencies where various criminal data should be consolidated and mined. They focus on KNN's theoretical details, but there is no implementation provided. Li and Juhola[15] says that crime research is an area that can benefit from better visualization and DMT. The new proposal provides a consolidated and visualized approach for crime detection, criminal identification and prediction, and crime verification to shield India from crimes.

III. METHODOLOGY

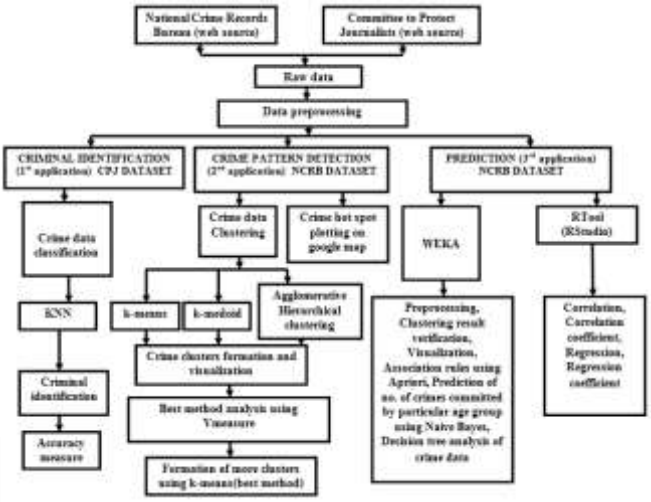

Fig. 1: Work-Flow Diagram

The work flow starts with DE step followed by DP step, which generated final crime database. This database is then supplied to other modules—clustering and classification. Clustering implementation uses k-means. k-means groups crime instances iteratively into clusters with similar attributes for crime detection. Clustering is then followed by implementation of maps for improved visual aid to kmeans. The classification step uses KNN which discovers similarities among different crimes and organizes them into predefined classes for criminal identification and prediction. Clustering and classification methods are implemented in python. The method then employs WEKA for crime verification of the k-means results. WEKA verifies high accuracy in the formation of two crime clusters using selected crime attributes. The proposed method being an integration of various data mining modules such as DE, data preprocessing, clustering, visualization and classification, it gains insight into the crimes and facilitates into the detection of prime crime suspects by filtering out the huge crime data. This method can help the police and justice departments to narrow down the identification of criminals. This in turn will reduce the cost and time of crime investigation. The method also uses techniques of Correlation and Regression. Correlation is a statistical technique used to determine the degree to which two variables are related. If the result of correlation is 1 that means there is perfect relation between the two attributes, if result is 0 then there is no relation between the two attributes, hence there must be strong relation between the attributes to get significant result.

IV. EXPERIMENTATION AND RESULTS

A. Preprocessed Dataset:

Preprocessing is done on the extracted data by selecting the required attributes and integrating the datasets from NCRB and CPJ. This dataset is used for KNN implementation. Datasets needed for clustering comparison, k-means implementation, WEKA, map, and Rtool implementations are then made using the data collected from NCRB site.

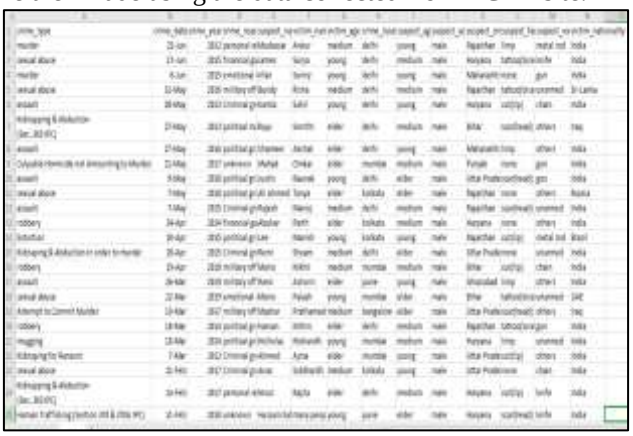

Fig. 2: Dataset for KNN

B. KNN Classification:

KNN is used to filter the crime dataset to get the details of crimes which are similar to a presently happened crime. This will help to get the details of criminals who follow similar crime patterns. As KNN algorithm is used, effective filtration is done and all of its nearest neighbors are analysed.

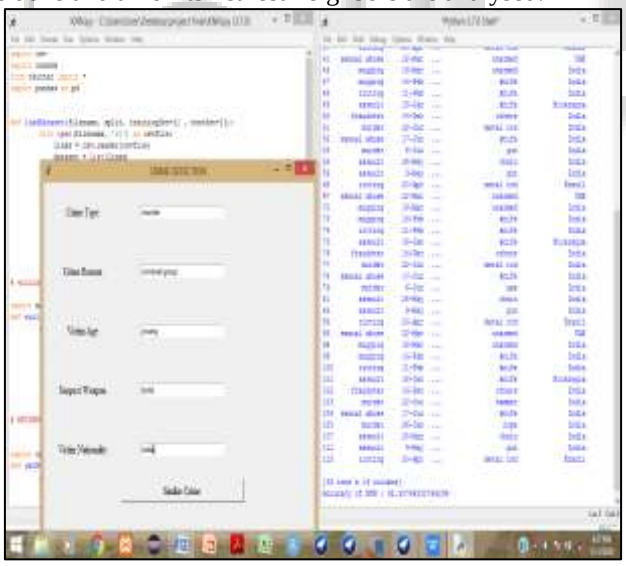

Fig. 3: KNN Output

C. Clustering methods comparison and k-means clustering:

1) Clustering methods comparison:

Initially three methods are used for crime data clustering:

1) k-means
2) k-medoid clustering
3) Agglomerative Hierarchical clustering.

Clustering is done on a dataset using three of these methods and these three outputs are compared to find the best method. Further clustering operations are done using that best method.

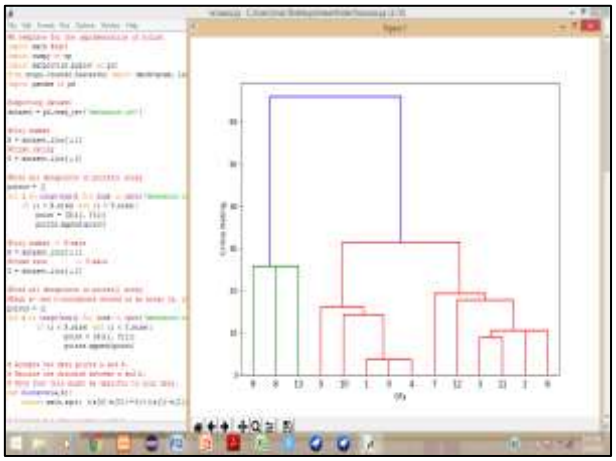

Fig. 4: Hierarchical clustering output 1

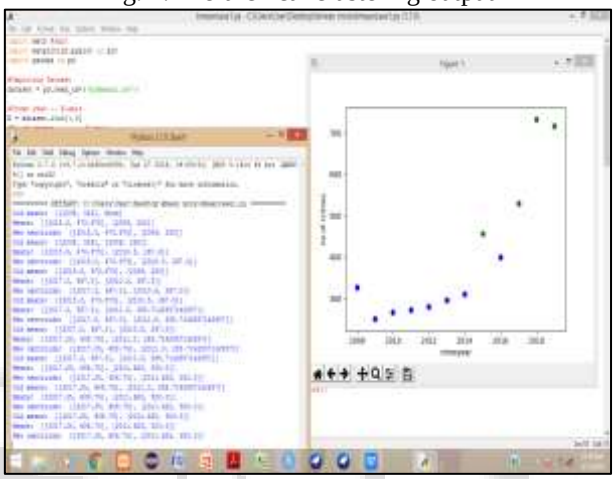

Fig. 5: k-means output 1

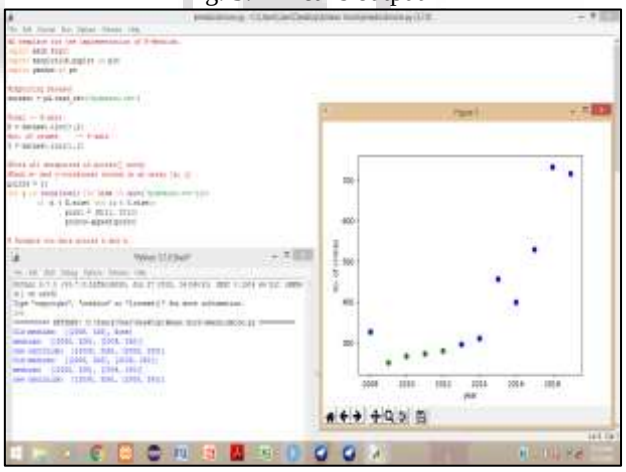

Fig. 6: k-medoid output 1

2) Conclusions derived are:

1) On analyzing the clustering results (output 1) based on V-measure, it is found that homogeneity and completeness parameters are satisfied by k-means and k-medoid results.
2) Each cluster of k-means and k-medoid has data-points belonging to same class label. Also in k-means and k-medoid, all data-points belonging to same class are clustered into same cluster. Thus the accuracy of hierarchical clustering is less compared to other two methods. So it is not chosen for crime dataset clustering.

3) Also, Hierarchical clustering results are very difficult to analyse in case of large datasets. Therefore this method is not selected.
4) K-means and k-medoid showed similar results. So comparison is done for another dataset.

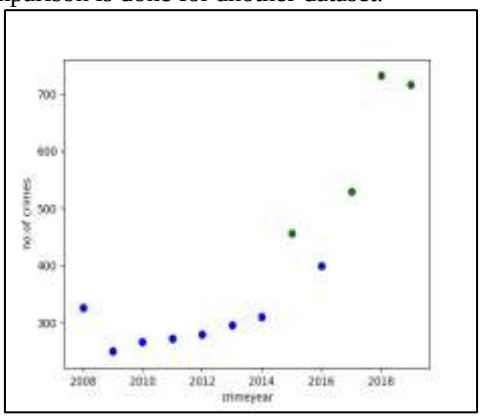

Fig. 7: k-means output 2

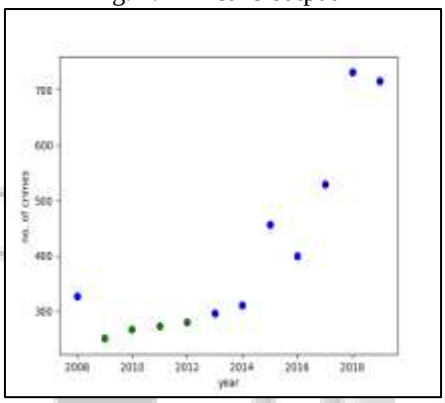

Fig. 8: k-medoid output 2

3) Results based on output 2:

K-Means is more accurate in terms of finding and defining clusters, whereas K-Medoids is faster but less accurate. K-Medoids is less accurate because it splits the first cluster (in blue) down its left half, causing for those data points to be included in the second cluster (in green). Visually, this makes sense, but this type of clustering visualization did not reflect the aim of depicting the two dense collections of points from x=2008 to x=2014 and x=2015 to x=2019 as separate clusters. On the contrary, K-Means did well to cluster the points from x=2008 to x=42014 and x=2015 to x=2019 separately, as K-Means is more mathematically suitable for cluster visualization based on density. Thus, for crime data clustering k-means is chosen as the best method.

Year	No. of crimes	Class label	k-mean cluster	k-medoid cluster
2008	326	Low	Cluster 1	2
2009	250	Low	1	1
2010	266	Low	1	1
2011	272	Low	1	1
2012	280	Low	1	1
2013	296	Low	1	2
2014	310	Low	1	2
2015	456	High	2	2
2016	400	Low	1	2
2017	530	High	2	2
2018	732	High	2	2
2019	717	High	2	2

Table I: Results of Clustering Comparison of k-means and k-medoid (output 2)

D. Final conclusion of output 2, based on V-measure:

The accuracy comparison is made based on V-measure. The calculation of V-measure first requires the calculation of two terms:

1) Homogeneity: Perfect homogeneous clustering is one where each cluster has data-points belonging to the same class label.
2) Completeness: Perfectly complete clustering is one where all data-points belonging to the same class are clustered into the same cluster.

On analyzing k-means and k-medoid results, each cluster of k-means has data-points belonging to same class label. Also in k-means result, all data-points belonging to same class are clustered into same cluster. Thus homogeneity and completeness parameters are satisfied for k-means clustering. Thus v-measure is high for k-means, and k-means is chosen as the best method for crime dataset clustering.

1) k-means clustering:

K-means is thus proved to be the best method for crime data clustering by accuracy comparison, and the accuracy of k-means result is also verified and proved using WEKA. So more clusters are formed using k-means.

2) Case 1 - Crime detection in India during 1990–2019:

k-Means aims to group objects (crimes in India during 1990–2019) as—A number of crimes in 1990, B number of crimes in 1991, C number of crimes in 1992 etc. into precise clusters. Clusters are based on the two crime attributes crime_year and number_of_crimes_committed_in_year.

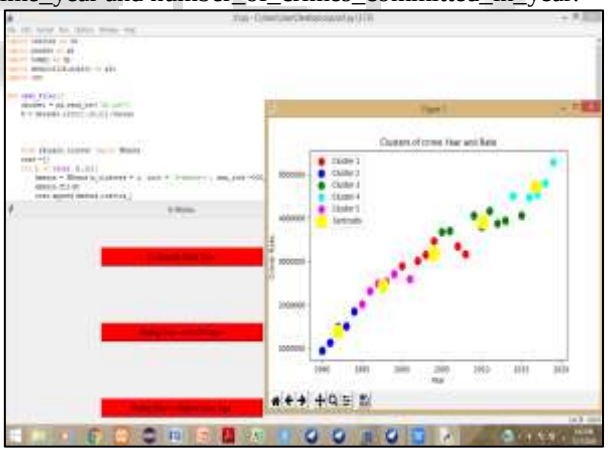

Fig. 9: k-means case 1 output

By analyzing result (fig.9), it is found that the highest crime rate in India is seen in the past five years. Also crimes show a huge increase in rate from 1990 to 2019. The highest crime rate cluster is the cluster 2014 to 2019(light blue, in fig.9).

3) Case 2 - Crime detection in Delhi during 1990–2019:

Clusters are generated to detect the number of crimes in a specified location (say Delhi) during 1990–2019. Here again, attributes crime_year and number_of_crimes_committed_in_year are used to generate clusters, but they depend on the attribute crime_location = Delhi and independent of the attribute crime_type. In the

same way, for other crime locations in India, similar clusters can be generated.

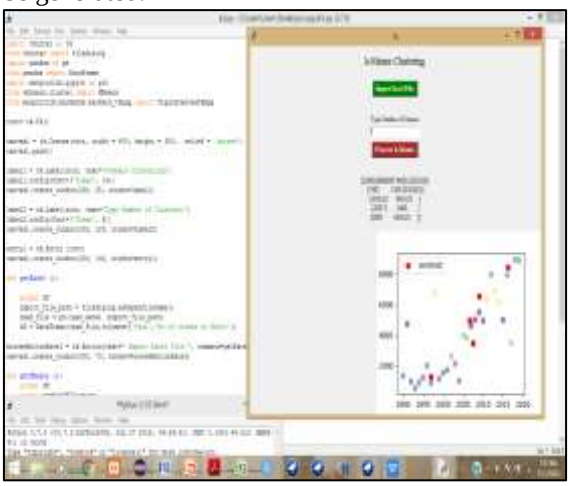

Fig. 10: k-means case 2 output

In case 2, output can be obtained for required cluster number, by giving no. of cluster as input. The output (in fig.10) for case 2 shows that Delhi showed highest crime rate during the years 2011, 2016, 2018, and 2019.

4) Case 3 – Detection of highest crime type in India during 2019:

Clusters are generated for this case where attributes crime_type and crime_location are specified. This case helps to detect which crime type is at peak in a given location in a year. For example, results like Manipur has high crime rate for kidnapping during 2019.

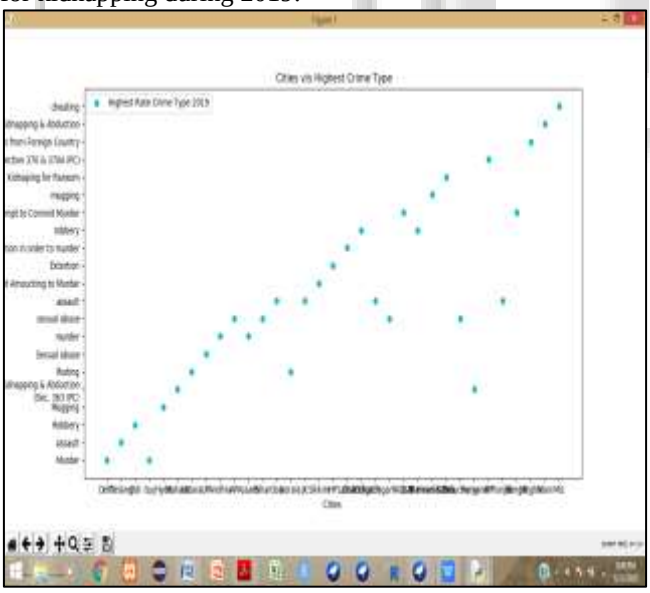

Fig. 11: k-means case 3 output

E. Map Representation:

The hotspots of crime is plotted on map which will help to ensure security in those areas. The shortest distance between those areas is also marked using the application. This will help to create a security network. This is done using OpenStreetMap which is available online. The application is supported by Graph Hopper API. Three map representations are available: map, satellite, and hybrid.

Fig. 12: Map representation

F. Regression and Correlation:

Correlation is a statistical technique used to determine the degree to which two variables are related. Regression models are used for prediction. Regression model is only a model, it is an indication of what may be. Correlation and Regression are done using R tool (Rstudio).

1) Regression graph:

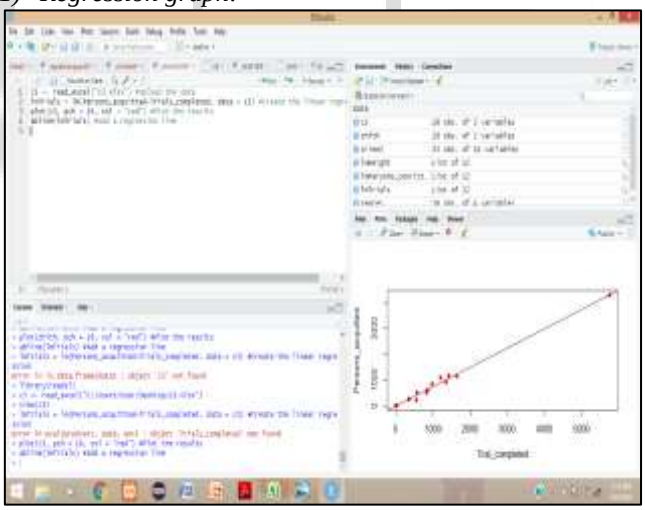

Fig. 13: Regression graph

2) Regression models:

Output 1:

```
> lmHeight = lm(Persons_convicted~Trial_completed,
data = c3) #Create the linear regression
> summary(lmHeight) #Review the results
Call:
lm(formula = Persons_convicted ~ Trial_completed,
data = c3)
Residuals:
     Min       1Q   Median       3Q      Max
-211.528  -30.455   -8.923   70.920  193.177
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     15.36894   32.49414   0.473    0.643
```

```
Trial_completed  0.24680    0.02032  12.149 1.72e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 109.9 on 16 degrees of freedom
Multiple R-squared:  0.9022,    Adjusted R-squared:  0.8961
F-statistic: 147.6 on 1 and 16 DF,  p-value:1.721e-09
```

From this output, β1= 15.36 and β2=0.24, are obtained. X is assumed as 100. These are given in equation, Y=β1+β2*X+∑ , β1=slope, β2=intercept of regression line. This gives regression coefficient,

Y= β1+β2*X+∑ ----------(1)
= 0.24*100+15.36 = 39.36 ~ 39

This makes regression model. From this, assumption can be made that, for every 100 rape case trials completed, only 39 are convicted of the rape charges. Thus by regression, prediction can be done for the number of criminals who got convicted of the crime charges against the no. of trials completed during the year(X value can be given according to the no.of trials completed). Prediction depends on the dataset taken as model.

Output 2 :

```
> chith <- read_excel("chith.xls") #Upload the data
> lmHeight = lm(Persons_acquitted~Trial_completed,
data = chith) #Create the linear regression
> summary(lmHeight) #Review the results
Call:
lm(formula = Persons_acquitted ~ Trial_completed,
data = chith)
Residuals:
    Min      1Q  Median      3Q     Max
-193.177  -70.920    8.923   30.455  211.528
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -15.36894   32.49414  -0.473    0.643
Trial_completed   0.75320    0.02032  37.076   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 109.9 on 16 degrees of freedom
Multiple R-squared:  0.9885,    Adjusted R-squared:  0.9878
F-statistic:  1375 on 1 and 16 DF,  p-value:< 2.2e-16
```

From this, β1=-15.36,β2=0.75, let X=100, which gives:
Y= 59.64 ~ 60, which means that about 60 people gets acquitted when 100 rape trials are done.

From output 1 and output 2, the results can be concluded as of 100 rape case trials completed in an year, about 60 people getacquitted of the rape charges and only 40 people gets convicted Thus the no. of people convicted is less than the number of people getting acquitted. Also it results that, as no. of trials are increased, more people get acquitted. This regression model can be used again for finding relations by changing value of X (no._of_trials_completed). Prediction is done based on the model dataset entered.

3) Correlation:

In the case of crime data, regression graph shows a correlation of 0.99(obtained by analyzing the graph), which is a strong correlation. It means that there is a strong correlation between persons_acquitted and trials_completed. This means that if no. of trials are high, the no. of persons acquitted may be more.

G. WEKA:

1) Clustering result verification using WEKA:

Clustering is done using WEKA on the same dataset on which initial k-means was performed. Results are then compared to verify the accuracy of initial k-means clustering results.

Cluster ID	k-means (Code)	k-means(WEKA)
G1	321.5	333.4
G2	196	201

Table II: k-means clusters

Error calculation:
G1 Cluster= mod (333.4-321.5)/333.4= 0.0356
G2 Cluster= mod (201-196)/201= 0.0248
So, accuracy measure:
G1 cluster= 100-3.56= 96.44% ; G2 cluster= 100-2.48 = 97.52%, which proves the high accuracy of k-means.

2) Association rules using Apriori :

Association rules are derived in WEKA using attributes crime_type, suspect_age and suspect_weapon_type of the dataset used previosly for KNN. Preprocessing is done for selection of 3 attributes, for apriori implementation, from the large dataset.

Frequent item sets are found out and then association rules are derived as:

Crime_type = robbery ➔ Suspect_age = medium

This is the best rule found. This denotes that most of the robberies are done by medium aged people. The association rule is formed for the dataset entered as input.

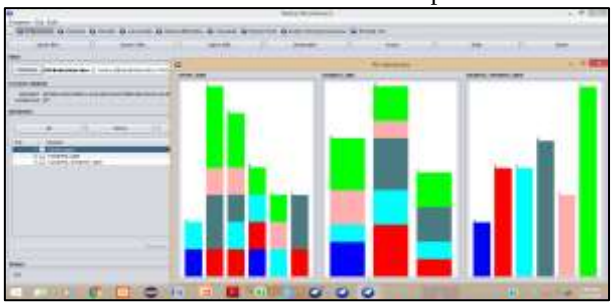

Fig. 14: Visualization of Apriori dataset

Fig. 15: Apriori result

3) Naive Bayes method for Prediction:

Naive Bayes classification method in WEKA will help to predict particular values. Here, the case is used for predicting the no. of crimes done by a particular age group. Test case used is of the state Andhra Pradesh where crime type is murder, gender is male, and age group is between 7-8 (s-ei) years. Thus, the most probabilistic answer for the given test case is obtained as, number of murder committed is 0-3 (z-th).

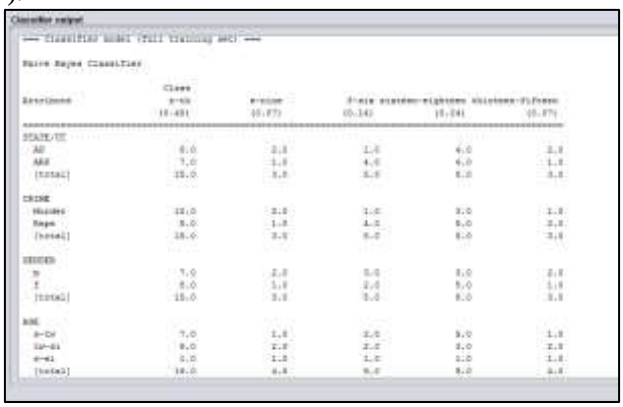

Fig. 16: Naive Bayes output

4) Decision tree analysis of crime data :

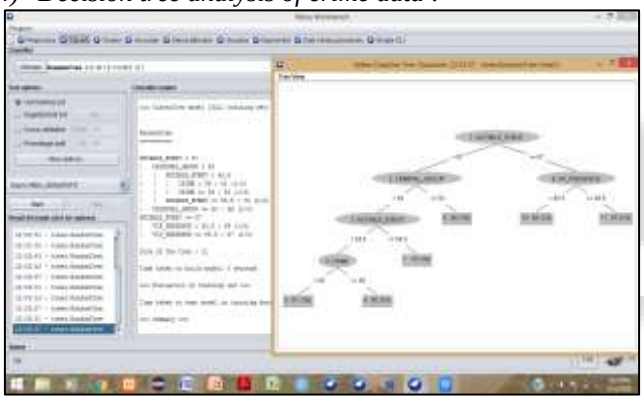

Fig. 17: Decision tree analysis

A data giving crime occurrence based on different factors can be easily analysed and crime ouurence conditions can be found out using decision tree concept, by tree visualization. Classification will help to analyse which of the factors lead to high crime occurrence. This analysis helps to ensure more security when such factors and situations occur.

V. Conclusion

Crimes in India are rising at an alarming rate because of the factors such as increase in poverty, migration, unemployment, frustration, illiteracy and corruption. Crime investigating agencies search the database of criminals manually or with some computer data analyst which is a tedious process and takes much more time. So to contribute toward combating crimes and to identify criminals, an integrated technology of is proposed using data mining techniques for Indian cities. Selection of Indian cities is based on their crime rates. Firstly, unstructured crime data is extracted from various crime Web sources and then preprocessed the crime data into structured instances that are represented using predefined crimes attributes. Initially a clustering comparison is done among three methods, which proves k-means as the best method. The method then applies k-means clustering for crime detection during 2008–2019 through three cases. Case 1 clusters crimes in India from 1990 to 2019 irrespective of crime location and crime type. It gives an overview of crime rate. Case 2 detects crimes in specific location, e.g., Delhi, irrespective of crime type. This helps in analyzing crime trends in previous years for a specific location. Case 3 detects crimes of specific type and in specific location, for easy analysis of highest crime type at a given location. To enhance k-means results, the method performs map implementation which helps to plot crime hotspots on map. The method also applies KNN classification for criminals' identification and prediction. KNN looks at the past crimes and finds similar ones that match the current crime based on the number of nearest neighbors' matched. KNN accuracy is also verified. The proposed method then uses WEKA to verify k-means, Case1 results. Accuracy in the formation of two crime clusters is measured using selected crime attributes. WEKA verifies an accuracy of about 96% and 97% in the formation of two crime clusters. WEKA also helps in prediction of crime patterns by using decision tree, Naive Bayes, and Apriori. Also, correlation regression methods are applied using R tool for making models and to analyse data. Investigating agencies can utilize the proposed data mining tool to ease their crime investigation process in three fields: criminal identification, crime pattern detection, and prediction. The proposed method can speed up the crime solving process by processing and filtering the voluminous crime data within a short span of time. Thus, this method can aid the law enforcement agencies to enforce the security of citizens of India. In future can enhance data privacy and other security measures of this crime-based data mining system and can also collaborate with security agencies in India. The best method now existing for crime data classification is KNN. In future this can be replaced by newly introduced methods with high accuracy.

References

[1] T. H. Grubesic, on the application of fuzzy clustering for crime hot spot detection, Journal of Quantitative Criminology 22(1) (2003) 77–105.

[2] K. Dahbur and T. Muscarello, Classification system for serial criminal patterns, Artificial Intelligence and Law 11(4) (2006) 251–269.

[3] Y. Peng, G. Kou, Y. Shi and Z. Chen, A descriptive framework for the field of data mining and knowledge discovery, International Journal of Information Technology & Decision Making 7(4) (2006) 639–682.

[4] S. V. Nath, Crime pattern detection using data mining, in Web Intelligence and Intelligent Agent Technology Workshops, 2006 IEEE/WIC/ACM International Conference on (IEEE, 2008), pp. 41–44.

[5] S. Lin and D. E. Brown, An outlier-based data association method for linking criminal incidents, Decision Support Systems 41(3) (2008),http://www.sciencedirect.com/science/article/pii/S0167923604001344.

[6] Chen H, Chung W, Xu JJ, Wang G, Qin Y, Chau M ,Crime data mining: a general framework and some examples. Comput IEEE 37(4):50–56, 2008.

[7] Yu G, Shao S, Luo B (2008) Mining crime data by using new similarity measure. In: 2nd International conference genetic and evolutionary computing (WGEC), IEEE, Washington, USA, pp 389–392.
[8] Malathi A, Baboo SS (2011) Evolving data mining algorithms on the prevailing crime trend—an intelligent crime prediction model. Int J Sci Eng Res 2(6).
[9] Malathi A, Babboo SS, Anbarasi A (2011) An intelligent analysis of a city crime data using data mining. In: International conference information electronic engineering, vol 6. IACSIT Press, Singapore, pp 130–134.
[10] Sayal R, Kumar VV (2011) A novel similarity measure for clustering categorical data sets, Int J Comput Apps 17(1):25–30.
[11] Saltos and Cocea, An exploration of crime prediction using data mining on open data, International journal of Information Technology and Decision making (2017), pp. 40–53.
[12] Ehlers D (1998) Predicting crime: a statistical glimpse of the future, Nedbank ISS Crime Index. Halfway House: Inst for Security Stud 2(2).
[13] Han J, Kamber M, Pei J (2012) Data mining: concepts and techniques, 3rd edn. Morgan Kaufmann Publishers Inc, San Francisco, CA.
[14] Kulis B, Jordan MI (2011) Revisiting k-means: new algorithms via bayesian nonparametrics. In: 29th International conference machine learning. Omnipress, Edinburgh, pp 513–520.
[15] Li X, Juhola M (2014) Country crime analysis using self organizing map, with special regard to demographic factors. AI Soc 29(1): 53–68.
[16] Mande U, Srinivas Y, Murthy JVR (2012a) Feature specific criminal mapping using data mining techniques and generalized Gaussian mixture model. Int J Comput Sci Commun Netw 2(3):375–379.
[17] Okonkwo RO, Enem FO (2011) Combating crime and terrorism using data mining techniques. In: 10th International conference IT people centred development, Nigeria Computer Society, Nigeria.
[18] Wang, G.; Chen, H.; and Atabakhsh, H. 2004. Automatically detecting deceptive criminal identities. *Communications of the ACM* 47(3):70–76.
[19] M. Al Boni and M. S. Gerber. Predicting crime with routine activity patterns inferred from social media. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2016.
[20] Ying-Lung Lin, Meng-Feng Yen, Liang-Chih Yu. Grid-Based Crime Prediction Using Geographical Features. International Journal of Geo-Information,2018.
[21] Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav. Crime Pattern Detection, Analysis & Prediction. International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2017.

