



Data Mining for Prevention of Crimes

Neetu Singh^(✉), Chengappa Bellathanda Kaverappa,
and Jehan D. Joshi

University of Illinois at Springfield, Springfield, IL 62703, USA
{nsing2, cbell136, jjosh2}@uis.edu

Abstract. Preemptive measures are of utmost importance for crime prevention. Law enforcement agencies need to have an agile approach to solve everchanging crimes. Data analytics has proven to be an effective deterrent in the field of crime data analysis. Various countries like the United States of America have benefitted by this approach. The Government of India has also taken an initiative to implement data analytics to facilitate crime prevention measures. In this research paper, we have used R Studio, an open source data mining tool to perform the data analysis on the crime dataset shared by the Gujarat Police Department. To develop predictive model and study crime patterns we used various supervised and unsupervised data mining techniques such as Multiple Linear Regression, K-Means Clustering and Association Rules Analysis. The scope of this research paper is to showcase the effectiveness of data mining in the domain of crime prevention. In addition, an effort has been put forth to help the Gujarat Police Department to analyze their crime records and provide meaningful insights for decision making to solve the cases recorded.

Keywords: Data mining · Predictive model · Process modelling
Crime analysis

1 Introduction

Technology is a double-edged sword. Criminals have been using technology for various destructive purposes [1]. Preemptive measures are of utmost importance for crime prevention. This makes it imperative that law enforcement agencies need advanced crime analysis tools which will aid them in the process of crime prevention [2–4]. Various countries like the United States of America have benefitted by this approach [5, 6].

One of the challenges faced by the police departments is to minimize threats to society by investigating large volumes of data [7]. Various initiative has been taken by researchers to analyze crime data using data mining techniques [1, 8–11]. The Government of India has also taken an initiative to implement data analytics to facilitate crime prevention measures [11]. The Gujarat Police Department shares the same vision. A proposal was submitted by one of the authors to the Gujarat Police Department to access confidential crime data for research purposes. They graciously granted access to monthly and annual crime records for the years 2012 to 2016, for the State of Gujarat in India. The Gujarat Police Department categorized crime into 9 different categories such as Home Break-In, Injuries, Kidnapping, Murder, Attempt to

Murder, Police Raids, Rioting, Robbery and Theft. The goal of this paper is to develop a predictive model to predict future solving rates of the police department making them more proactive in nature and identifying crime associations and crime hotspots. This will further help the Gujarat Police Department to analyze their crime records and provide meaningful insights for decision making to solve the cases recorded. A survey on crime analysis found that 10% of the criminals commit 50% of the crimes [10]. So, any research initiative taken will help the police department solve crimes at a faster rate.

This research paper is a blend of supervised and unsupervised data mining techniques to analyze crime data. Here we implemented Multiple Linear Regression, Association Rules Analysis and K-Means Clustering algorithms to conduct a comparative study of various crime patterns from the dataset. The above-mentioned methods will help us identify underlying crime patterns and generate valuable insights from the dataset.

The motivation for conducting this research is to aid Gujarat Police Department in crime prevention and further help them understand the benefits of data mining in this domain. In this paper, different types of data mining methods have been applied and their subsequent results have been discussed which can be used by multiple enforcement agencies and as a point of reference for future research initiatives in the domain of crime analysis.

The paper is organized into six different sections. Section 1 is comprised of introduction. The background and information for crime analysis and prevention is discussed in literature review (Sect. 2). The methodology is discussed in Sect. 3. The different phases of data analysis are presented in Sect. 4. In Sect. 5 we discuss the results of data analysis which includes data visualization, graphs, thresholds and performance evaluation. The conclusion of the paper is presented in Sect. 6 which also includes the limitations and future scope of the research.

2 Literature Review

India is second most populous country in the world with a population of 1.3 billion people [11, 12]. According to the “7th Schedule Article 246 of the Indian Constitution”, Law and Order is a subject of the state [13]. This has created enormous pressure on state government to prevent crimes in their respective states. In this research paper, the authors have analyzed crime records for the state of Gujarat using various data mining techniques. The state of Gujarat has a population of 60.3 million people according to the latest census data [14]. The recorded number of actual strength of police personal in the year 2011 for the state of Gujarat was 72,838. This makes the task of maintaining law and order extremely challenging owing to the sparse police to population ratio which is 129.89 per 100,000 people [15].

Understanding the administrative structure of the Gujarat State Police Department was important for the authors to effectively collaborate with them. For a state jurisdiction the Head of the police force is Director General of Police (DGP). In a state there are many districts. For a Range that is a cluster of neighboring districts are under the control of Deputy Inspector General of Police (DIGP). There are some states where it

has two or more ranges based on geographical or population figure such a state is headed by the Inspector General of Police (IGP). A district is headed by Senior Superintendent of Police (SSP)/Superintendent of Police (SP) under whom the Assistant Superintendent of Police (ASP) and Deputy Superintendents of Police (DSP) act [11].

Researchers used various data mining techniques to help law enforcement agencies prevent crimes [16–19]. During our literature review we observed that to conduct crime hotspot analysis and crime pattern analysis various researchers preferred Density Based Clustering and K-Means Clustering which resulted in clusters based on the number of crime incidents recorded [1, 10, 20]. Based on the crime type and data type the method used for analysis greatly varies. Method such as social network analysis is used to analyze data from various social media platform, which is extremely important to prevent crimes [4]. Certain classification techniques such as K-NN (K Nearest Neighbors) classification help law enforcement agencies to classify crime types based on independent attributes [2].

Studying the work done by various other researchers we learnt that most of the researchers used one of the following above-mentioned data mining techniques. We identified this as our research gap and aimed to develop a holistic research paper which uses multiple unsupervised and supervised data mining methods along with statistical tests to test their performance and accuracy.

Market Basket Analysis using the Apriori algorithm has been a widely used data mining technique to study consumer behavior in the retail sector [21]. We wanted to study if market basket analysis can be used in other domains to study the underlying relationships in the data. We came across research carried out in the domain of transportation safety using Market Basket analysis to study the associations between fatal car crashes and various underlying factors [22]. We extrapolated this idea in the domain of crime data mining and after initial data preparation we used Market Basket analysis to study the associations between different crime types based on the number of crime incidence recorded.

3 Methodology

To effectively analyze the data using multiple data mining techniques and avoid any ambiguities we followed the SEMMA data mining implementation steps developed by the SAS institute [23]. SEMMA stands for “Sample”, “Explore”, “Modify”, “Model”, “Assess” [24].

- **Sample:** Various sampling strategies are used when the data is too large or complex to be analyzed. These sampling strategies try to replicate the properties of the population. In our dataset we analyzed the complete dataset to derive important insights and inferences.
- **Explore:** We conducted exploratory analysis by visualizing the data in R Studio to study descriptive statistics and to identify relationships between the variables in the dataset which helped us understand the data in a concise manner.

- **Modify:** Feature engineering is an important element which enriches the data for effective analysis and model development [10]. We added variables such as “Population”, “Crime Density” and “Crime Code” for calculating the Crime Density we wanted to know the population in each city as population is an attribute present in crime density formula. Secondly, the Crime Code was added to apply Multiple Linear Regression in which the different Crime Types were encoded to numerical data for gaining statistical parameters. These variables improved model accuracy.
- **Model:** We developed predictive models using Multiple Linear Regression and used unsupervised learning methods such as K-Means Clustering algorithm and Apriori algorithm to study underlying patterns and associations in the data.
- **Assess:** To evaluate the performance of the model it was tested on unseen data. Various statistical parameters such as ‘p-value’, ‘t-value’, ‘R-Squared’ and ANNOVA tests were performed to select the best model.

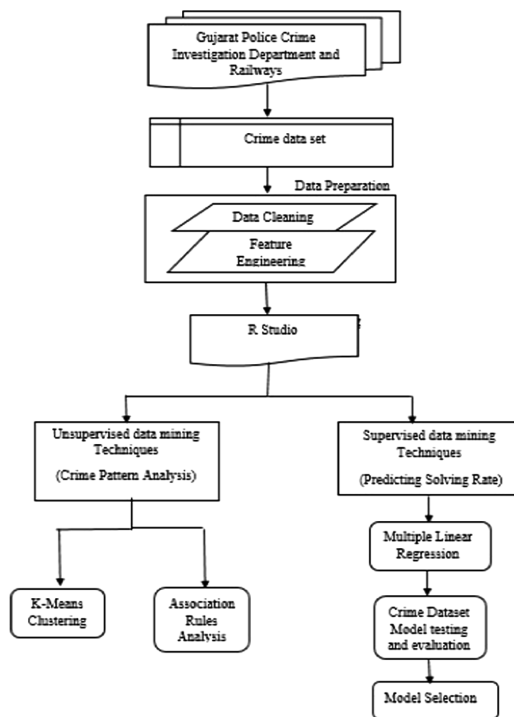


Fig. 1. Data analysis – conceptual model

A conceptual model is developed which would help us perform the data analysis activities in a structured manner. Figure 1 depicts this conceptual model. It illustrates the data analysis processes and techniques used for developing the predictive and classification model using SEMMA. First the data was obtained from the police

department after which the annual crime data was extracted from the main dataset. This crime dataset is used for detailed analysis. To prepare the data various irrelevant fields and null values were omitted and the variable solving rate was checked for correctness and errors were modified. To further enrich the data, we added new features such as “Crime Density”, “Crime Code” and “Population”. Average imputation method was used to impute population values for districts whose populations weren’t available [1]. The cleaned data was loaded to R Studio for further analysis.

We decided to use two approaches for our analysis “Crime pattern analysis” to study patterns and associations in the dataset using K-Means Clustering and Association Rules mining [4]. Predictive models were also developed using Multiple Linear Regression to predict the departments solving rate in subsequent years [25].

Various model testing techniques such as ANNOVA, variable importance tests and p-values were used for finalizing the variables and model selection. The model was validated by testing it on unseen test data. The model’s stability was further tested by performing Ten-Fold cross validation [26].

4 Data Analysis

Data analysis uses various data mining techniques to analyze the data and uncover information previously unknown to us [26, 27]. Here selecting the best algorithms that best fits our research needs and help the Gujarat State Police Department was of utmost importance. Various steps were undertaken to successfully analyze the data and extract meaningful information from the data.

4.1 Data Preprocessing and Preparation

The dataset comprised of variables such as “Year”, “City”, “Crime type”, “Public” (Cases recorded), “Found” (Cases resolved), and “Percentage” (Crime solving percentage). Data cleaning and transformation was carried out to derive better insights from the data. Three additional variables namely “Population”, “Crime Density” and “Crime Code” were added. The population for the years 2012 to 2016 were derived based on the annual population growth rate of India and census data for the year of 2011 [1, 28]. In addition, the Crime density per 100,000 of the general population was calculated by mathematical formula:

$$\text{Crime density} = \left(\frac{\text{Public(Cases Recorded)}}{\text{Population}} \right) * 100,000$$

The variable “Crime Code” was derived by encoding variable “Crime types” as sequential numbers from 1 to 9.

4.2 Data Mining Techniques

Data mining techniques are used to extract valuable information from the data [27]. These techniques are mainly of two types supervised and unsupervised.

Supervised learning: In this method the dataset is divided into two parts namely Training dataset and Test dataset. The models are developed using the training data and validated by comparing the model's prediction with that of the unseen test data [5, 27].

Unsupervised learning: In this class of data mining technique, the dataset is analyzed for underlying patterns and relations between different variables of a dataset [5, 27].

We have applied both of these data mining techniques to analyze the data, such as Multiple Linear Regression (supervised learning) to develop predictive models, K-Means Clustering and Association Rules mining (unsupervised learning) for studying the underlying patterns and the associations between different attributes in the dataset.

R programming provides us with various packages to perform analysis on dataset. We used R programming to clean the data, developed data mining models and test models. The following data mining methods were used to analyze our data.

Multiple Linear Regression: This method has more than one independent variables which predicts and establishes a relationship between our dependent variable. The difference between the predicted value \mathbf{P}_i and actual value \mathbf{A}_i is termed as error rate. The Regression can be expressed as \mathbf{Y}_i [5, 27]: $\mathbf{Y}_i = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \epsilon_i$

Where:

| | |
|-------------------------|--|
| Y | denotes the measured value of dependent variable |
| X | denotes the value of the independent variables |
| β_0 | denotes a constant |
| $\beta_1 \dots \beta_n$ | denotes estimated regression co-efficient |
| ϵ_i | denotes residual |

K-Means Clustering: This method groups similar objects together to analyze crime patterns in the dataset. The algorithm uses distance computation functions to assign an observation to a cluster. An object is assigned to the cluster that it has the closest distance. The user decides the number of clusters in such a way that it minimizes the within sum of squares distance between the clusters [27, 29].

Association Rule Mining: The Association Rules analysis is an unsupervised learning technique which is used to extrapolate the rules associated with each other. In this case we have mined the associations of different types of crime committed with each other type [27, 30, 31]. The Association Rules are generated as a Transaction ID when applied. Each rule is represented by a transaction ID. The Association Rules are evaluated using the metrics - Support, Confidence and Lift.

Support is the measurement in which the crime type is repeated in a set of transactions given in the dataset. Confidence is the ratio of the support of two different crime types together to the support measure of individual crime committed regardless of associated crime is repeated in other transaction sets with other crime type. It also measures the accuracy of the associated rule [32]. Lift is measured as ratio of the

support of two crimes together to the support of two crimes individually committed [30, 31].

The algorithm used for mining Association Rules is Apriori algorithm using R programming in R Studio. The different crime types are mined and crime types which are associated to each other based on the above three measures are extracted carefully. This shows the crime type that are committed by the criminals based on the previous committed crimes. The police department can use this prediction to analyze the next crime type of the suspect and be watchful for prevention of crime.

5 Results and Discussion

After this preliminary analysis of dataset, we would like to help the Gujarat Police Department to be more proactive and introduce counter measures in due course of time. We developed a predictive model using regression [5, 27] in R which would predict the cases resolved. These will be the cases which need to be solved on high priority to minimize the crime in Gujarat. The regression model provided an accuracy of 85.5%. We observed in Table 3 the “Crime Type Thieves” is not significant. This leads to an interesting finding that theft related cases are the least resolved ones which needs to be considered by Gujarat Police. Theft being the least resolved crime type might be one of the reasons that we found theft as one of the most happening crime in Gujarat as represented in Table 1.

Table 1. Regression model results

| Model variables | Coefficient estimate | Std. error | t value | P value |
|---|----------------------|------------|---------|-------------|
| (Intercept) | -18.216 | 3.677 | -4.953 | 8.18e-07*** |
| Crime type injuries | 129.338 | 4.795 | 26.973 | <2e-16*** |
| Crime type kidnapping | 41.597 | 4.820 | 8.630 | <2e-16*** |
| Crime type murder | 21.021 | 4.818 | 4.363 | 1.38e-05*** |
| Crime type murder attempt | 18.607 | 4.741 | 3.924 | 9.11e-05*** |
| Crime type raid | 11.112 | 4.856 | 2.288 | 0.02226* |
| Crime type riot | 28.269 | 4.842 | 5.838 | 6.54e-09*** |
| Crime type robbery | 14.714 | 4.734 | 3.108 | 0.00192** |
| Crime type thieves | -7.058 | 5.120 | -1.378 | 0.16827 |
| Cases recorded | 0.405 | 0.006 | 61.996 | <2e-16*** |
| Population | 7.718 | 0.957 | 8.066 | 1.53e-15*** |
| Signif. codes: | ***0.001 | **0.01 | *0.05 | |
| <i>Model statistics</i> | | | | |
| Adjusted R-squared: 0.8556 F-statistic: 853.4 p-value: <2.2e-16 | | | | |

To test and validate the Multiple Linear Regression model we used the analysis of variance test and the model prediction function on unseen test data. Table 2 below

displays the result of the analysis of variance test where the F-value of the variables of our choice are of significance.

Table 2. Analysis of variance

| | Degrees of freedom | Sum square | Mean square | F- value | P-value |
|------------|--------------------|------------|-------------|----------|--------------|
| Crime type | 8 | 5887627 | 735953 | 414.363 | <2.2e-16*** |
| Public | 1 | 9153575 | 9153575 | 5153.733 | <2.2e-16*** |
| Population | 1 | 115544 | 115544 | 65.055 | 1.528e-15*** |
| Residuals | 1429 | 2538055 | 1776 | | |

Table 3 displays how well our predictive model performed on unseen test data we got an R-Squared value of 77% which signifies that our model did perform well on unseen test data as well.

Table 3. Model prediction statistics

| RMSE | R squared | Mean absolute error |
|------------|-----------|---------------------|
| 56.6043764 | 0.7701536 | 20.6575616 |

To check the impact and importance of the variable in our model we ran a variable importance test. Here a higher overall value signifies higher importance and relevance. From the Table 4 it is evidently clear that the variable ‘Cases Recorded’ has the highest significance towards predicting our crime solving rate.

Table 4. Variable importance table

| Variable name | Overall |
|---------------------------|-----------|
| Crime type injuries | 26.973354 |
| Crime type kidnapping | 8.630258 |
| Crime type murder | 4.362551 |
| Crime type murder attempt | 3.924314 |
| Crime type raid | 2.288467 |
| Crime type riot | 5.837583 |
| Crime type robbery | 3.108232 |
| Crime type thieves | 1.378487 |
| Cases recorded | 61.995915 |
| Population | 8.065636 |

We decided to perform K-Means Clustering and Association Rules Analysis to identify patterns and associations in the dataset which will help us to have deeper

analysis of the crime data. To explore the underlying patterns in our data we used the K-Means clustering technique [9]. The Elbow method in Fig. 2 shows six clusters are optimal number of clusters as adding another doesn't improve the total within sum of squares. Using this as our K-value we performed K-Means clustering in R. Figure 3 represents the six homogenous clusters were created based on variables "Cases recorded" and "Cases resolved".

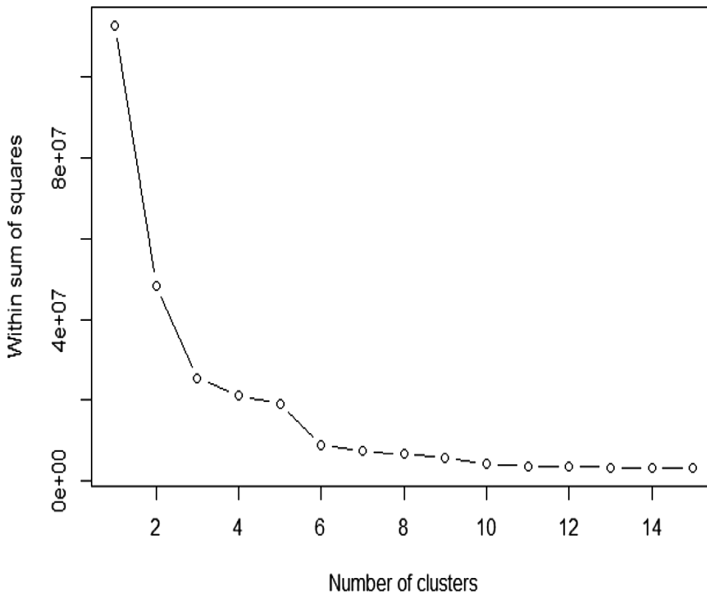


Fig. 2. A plot describing the within sum of squares (WSS) by Elbow method

As observed in Fig. 3, clusters can be segregated into different crime zones based on the number of cases recorded. In addition, Cluster 5 is a high-risk crime zone where the number of cases recorded is over 2500 crime instances which is extremely high as compared to the average.

After detailed analysis of cluster 5 it is evident that all the recorded instances of crime are related to theft (Table 5). Hence, it is recommended for the Gujarat Police Department to allocate more resources and prioritize theft related cases in Ahmedabad city. In Table 5 we found an interesting inference as the cases of theft are not properly resolved they do not have a high significance in helping us predict the cases resolved.

Our predictive model helped us to identify that theft is the most happening crime in Gujarat (Table 1). Crime pattern analysis using K-means clustering further confirms our results by providing the details that over 2500 recorded instances of crime are related to theft (Fig. 3, Table 5). So, planned to extrapolate different association between crime types using a novel approach of performing market basket analysis. To conduct this analysis using the annual crime dataset we split the dataset based on the crime type and grouped it based on cases recorded to create our baskets [32]. We used

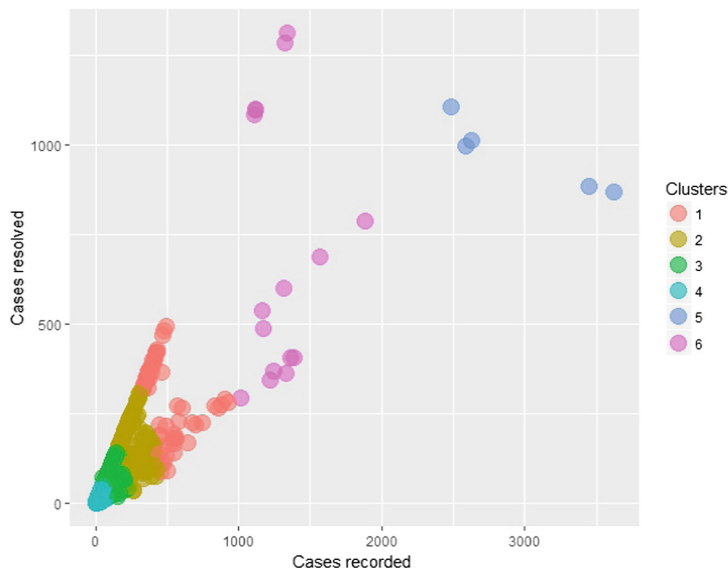


Fig. 3. K-means clustering result

Table 5. Contents of K-means cluster 5

| City | Year | Crime code | Crime type | Cases recorded | Cases resolved | Cases resolved (%) |
|-----------|------|------------|------------|----------------|----------------|--------------------|
| Ahmedabad | 2012 | 6 | Thieves | 3622 | 871 | 24.04 |
| Ahmedabad | 2013 | 6 | Thieves | 3449 | 886 | 25.69 |
| Ahmedabad | 2014 | 6 | Thieves | 2589 | 999 | 38.59 |
| Ahmedabad | 2015 | 6 | Thieves | 2487 | 1108 | 44.55 |
| Ahmedabad | 2016 | 6 | Thieves | 2623 | 1014 | 38.66 |

different Support and Lift values to analyze these baskets and mine relevant Association Rules as displayed in Table 6.

Based on the lift values we can infer that suspects committing Murder and Riot are most likely to commit Kidnapping. Here a set of interesting rules that are relevant and will help the Gujarat State Police department are displayed in Table 6. Owing to these rules the department can narrow down their search to relevant suspects. In addition, the understanding of the association of crime types can help police department to focus on resolving those crime types simultaneously. For example, if police department strategize to resolve the crime types of murder, riot and robbery it can help them to resolve the kidnapping crime type.

Another interesting finding of the association rules analysis is that theft is usually not associated with any another crime type. Overall, we have observed theft is the most happening and recorded crime type in the state of Gujarat but is usually not associated with any other crime type. On the other hand, robbery is associated with most of the

Table 6. Association rules for crime types

| Rules | Support | Confidence | Lift* | Count |
|---|---------|------------|-------|-------|
| [Murder, Riot] → [Kidnapping] | 0.1 | 0.8 | 2.9 | 39 |
| [Riot, Robbery] → [Kidnapping] | 0.1 | 0.8 | 2.8 | 41 |
| [Murder, Robbery] → Murder attempt] | 0.1 | 0.8 | 5.8 | 38 |
| [Kidnapping, Murder Attempt] → [Murder] | 0.1 | 0.8 | 4.9 | 37 |
| [Home Break-In, Murder Attempt, Robbery] → [Riot] | 0.08 | 0.9 | 4.5 | 30 |
| [Home Break-In, Murder Attempt, Raid] → [Robbery] | 0.05 | 1.0 | 5.5 | 21 |

*Lift > 1 explains the significance of an association rule.

crime types. It would be interesting for us to further explore the difference between robbery and theft and focus on the cases reported and resolved for both crime types. We plan to extend our analysis on these lines in future.

6 Conclusion

As the population keeps growing with urbanization taking place new types of crime are committed making the cities vulnerable for the public to be safe. A smart and robust approach would help the law enforcement agencies to maintain public safety and peace in the cities by preventing crimes to be committed. The data mining techniques have been proven to be effective in analyzing the dataset and gather insights which are useful in many domains. This research paper used few data mining techniques to predict and identify the pattern of crime types committed in the state of Gujarat. This will help the Gujarat Police Department to identify the suspects and prevent them to further committing crime. The Association Rules suggest the Crime Types that are associated based on the cases recorded and the Multiple Linear Regression model has helped us to develop a predictive model which predicts the resolved cases with accuracy of 85.5%. The K-Means Clustering has shown the high crime risk zone city amongst the other cities.

In this research we have used yearly data (2012–2016) to predict the solving rate and to analyze the patterns of crime. The results of this research will help Gujarat police department to analyze their crime records and provide meaningful insights for decision making to solve the cases recorded. Currently we are working on monthly crime dataset to develop more robust models that could be implemented with the department. To successfully implement these methodologies, we are studying the As-Is crime solving process of the Gujarat Police Department and subsequently modelling a To-Be process which would integrate the data mining approach in the crime solving process.

Acknowledgement. The authors sincerely appreciate and are thankful for the cooperation of The Gujarat Police Department, India for providing us access to the dataset used in this study.

References

1. Malathi, A., Baboo, S.S.: An enhanced algorithm to predict a future crime using data mining. *Int. J. Comput. Appl.* **21**(1) (2011). ISSN 0975-8887
2. Hassani, H., et al.: A review of data mining applications in crime. *Stat. Anal. Data Min.: ASA Data Sci. J.* **9**(3), 139–154 (2016)
3. Bloomberg, J.: How the FBI Proves Agile Works for Government Agencies (2012). <https://www.cio.com/article/2392970/agile-development/how-the-fbi-proves-agile-works-for-government-agencies.html>
4. David, H., Suruliandi, A.: Survey on crime analysis and prediction using data mining techniques. *ICTACT J. Soft Comput.* **7**(3) (2017)
5. McClendon, L., Meghanathan, N.: Using machine learning algorithms to analyze crime data. *Mach. Learn. Appl.: Int. J. (MLAIJ)* **2**(1) (2015)
6. Li, X., et al.: GDP growth vs. criminal phenomena: data mining of Japan 1926–2013. *AI Soc.* **33**, 1–14 (2013)
7. Chen, H., et al.: Crime data mining: a general framework and some examples. *Computer* **37**(4), 50–56 (2004)
8. Tayal, D.K., et al.: Crime detection and criminal identification in India using data mining techniques. *AI Soc.* **30**(1), 117–127 (2015)
9. Thota, L.S., et al.: Cluster based zoning of crime info. In: *Proceedings of the 2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, pp. 87–92. IEEE (2017)
10. Nath, S.V.: Crime data mining. In: Elleithy, K. (ed.) *Advances and Innovations in Systems, Computing Sciences and Software Engineering*, pp. 405–409. Springer, Dordrecht (2007). https://doi.org/10.1007/978-1-4020-6264-3_70
11. Gupta, M., et al.: Crime data mining for Indian police information system. In: *Proceeding of the 2008 Computer Society of India* (2008)
12. Wikipedia: World Population Prospects: The 2017 Revision. https://en.wikipedia.org/wiki/United_Nations_Department_of_Economic_and_Social_Affairs
13. Government of India: The Constitution of India (2015). <http://lawmin.nic.in/olwing/coi/coi-english/coi-4March2016.pdf>
14. Government of Gujarat: Official Gujarat State Portal (2011). <http://www.gujaratindia.com/state-profile/demography.html>
15. Bureau of Police Research and Development: Data on Police Organizations in India (2015). <http://www.bprd.nic.in/WriteReadData/userfiles/file/201607121235174125303FinalDATABOOKSMALL2015.pdf>
16. Nissan, E.: An overview of data mining for combating crime. *Appl. Artif. Intell.* **26**(8), 760–786 (2012)
17. Caplan, J.M., et al.: Joint utility of event-dependent and environmental crime analysis techniques for violent crime forecasting. *Crime Delinq.* **59**(2), 243–270 (2013)
18. Dos Santos, M.J., Kassouf, A.L.: A cointegration analysis of crime, economic activity, and police performance in São Paulo city. *J. Appl. Stat.* **40**(10), 2087–2109 (2013)
19. Yu, C.-H., et al.: Crime forecasting using data mining techniques. In: *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW)*, pp. 779–786. IEEE (2011)
20. Birant, D., Kut, A.: ST-DBSCAN: an algorithm for clustering spatial–temporal data. *Data Knowl. Eng.* **60**(1), 208–221 (2007)
21. Chen, Y.-L., et al.: Market basket analysis in a multiple store environment. *Decis. Support Syst.* **40**(2), 339–354 (2005)

22. Pande, A., Abdel-Aty, M.: Market basket analysis of crash data from large jurisdictions and its potential as a decision support tool. *Saf. Sci.* **47**(1), 145–154 (2009)
23. Kurgan, L.A., Musilek, P.: A survey of knowledge discovery and data mining process models. *Knowl. Eng. Rev.* **21**(1), 1–24 (2006)
24. Azevedo, A.I., Santos, M.F.: KDD, SEMMA and CRISP-DM: a parallel overview. In: *Proceedings of the IADIS European Conference Data Mining 2008, DM 2008 Proceeding*, pp. 182–185 (2008)
25. João, P., et al.: Predictive model for criminality in lisbon (2010)
26. Adderley, R.: *The use of data mining techniques in crime trend analysis and offender profiling*. University of Wolverhampton, United Kingdom (2007)
27. Shmueli, G., et al.: *Data Mining for Business Analytics: Concepts, Techniques, and Applications* in R. Wiley, New York (2017)
28. Census of India: Ahmedabad City Census 2011 Data (2011). <http://www.census2011.co.in/census/city/314-ahmedabad.html>
29. Zubi, Z.S., Mahmud, A.A.: Using data mining techniques to analyze crime patterns in the libyan national crime data. *Recent Adv. Image Audio Sig. Process.* **8**, 79–85 (2014)
30. Englin, R.: *Indirect association rule mining for crime data analysis*. Eastern Washington University, Cheney, Washington (2015)
31. Sevri, M., et al.: Crime analysis based on association rules using apriori algorithm. *Int. J. Inf. Electron. Eng.* **7**(3), 99 (2017)
32. Tan, P.-N., et al.: Selecting the right objective measure for association analysis. *Inf. Syst.* **29**(4), 293–313 (2004)