



CRIMES IN KARNATAKA A DECADAL ANALYSIS AND PREDICTIVE MODELING THROUGH DATA MINING TECHNIQUES (2011-2021)

Prajvalarani Sushil Kumar Bogar

Dr. Amit Singhal

DECLARATION: I AS AN AUTHOR OF THIS PAPER /ARTICLE, HERE BY DECLARE THAT THE PAPER SUBMITTED BY ME FOR PUBLICATION IN THE JOURNAL IS COMPLETELY MY OWN GENUINE PAPER. IF ANY ISSUE REGARDING COPYRIGHT/PATENT/OTHER REAL AUTHOR ARISES, THE PUBLISHER WILL NOT BE LEGALLY RESPONSIBLE. IF ANY OF SUCH MATTERS OCCUR, PUBLISHER MAY REMOVE MY CONTENT FROM THE JOURNAL WEBSITE. FOR THE REASON OF CONTENT AMENDMENT /OR ANY TECHNICAL ISSUE WITH NO VISIBILITY ON WEBSITE /UPDATES, I HAVE RESUBMITTED THIS PAPER FOR THE PUBLICATION.FOR ANY PUBLICATION MATTERS OR ANY INFORMATION INTENTIONALLY HIDDEN BY ME OR OTHERWISE, I SHALL BE LEGALLY RESPONSIBLE. (COMPLETE DECLARATION OF THE AUTHOR AT THE LAST PAGE OF THIS PAPER/ARTICLE

Abstract

Crime record bureaus and police stations have begun using data mining techniques in response to the demands placed on them by the large amounts of data held by police departments, as well as the variety and severity of criminal actions. With a reasonable amount of data, an experienced police officer may accurately assess crime patterns; nevertheless, as the number of instances and the severity of crimes rise, the quantity of facts and statistics also increases. Instead of tracking down specific offenders, the emphasis is on criminal locations. The police will be the primary users of the system, and they will occasionally be able to forecast when crimes are likely to occur and the likelihood that they will occur in the near future. This paper primarily examines the kmeans clustering method as it pertains to data mining. It explores how this algorithm might be employed to identify areas prone to criminal activity and to forecast the likelihood of future criminal incidents. To better allocate resources, such as police officers, to areas with a higher crime rate, this study might be useful to the country's law enforcement.

Keywords: Crime Analysis, Data mining tools, Crime bureau., , classification, correlation, Visualization, prediction, K means algorithm, modus operand.



1. INTRODUCTION

The traditional and old-style methods of operation in most public and commercial organizations have been utterly replaced by new techniques and advancements in computer technology. In most western countries, police stations, record bureaus, and related units of the crime investigation department have been computerized. This has led to the use of advanced software to analyze, detect, and predict crime. Other features, such as time series data with general packet radio service and geographic information systems, can provide information about hotspots and related details. Such divisions are linked with Internet, intranet, wifi, and blue tooth technologies.

1.1.

Crime investigation is becoming more complicated at a rate that matches the rate of change in technology, the nature of crimes, and society at large. As a result, crime record bureaus frequently make efforts to adapt. When it comes to doing their jobs, the Indian police have likewise opted for innovative processes and cutting-edge technology.

1.2. Crime analysis

From the vast amounts of data kept by the department, the primary goal of crime analysis is to equip investigators and police with the knowledge they need to help control and prevent future criminal actions. Investigating a crime requires a complicated set of factors to be considered, including but not limited to: emotional instability, jealousy, vengeance, inferiority complex, ego, gain, harm to others' property or reputation, and so on. Getting money quick, lusting, etc., The following section discusses the various frameworks suggested and developed based on the literature review, which are necessary for a machine to perform this kind of intelligent task. Therefore, it is important to have knowledge about external factors in addition to reports in order to analyze the crime. In order to reach a conclusion, the following steps are taken in the crime analysis process: first, all types of crimes are included in the data set. Then, only economic crimes related to the Indian Penal Code (IPC) are considered. Then, the analysis is narrowed down to a specific IPC section and one subgroup, such as cheating. Finally, any crimes involving family members are taken into account.

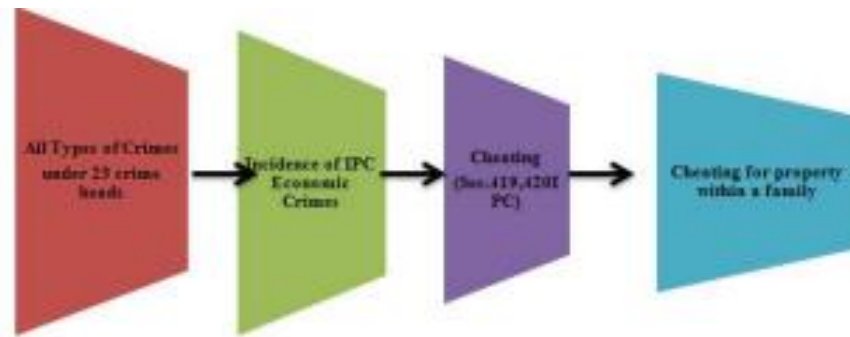


Figure 1: displays the crime analysis technique.

- **Types of criminal analysis:**

One of the best resources still available to assist police departments is crime analysis.

- The primary goal of administrative crime analysis is to give police supervisors summary facts, statistics, and broad trend information.
- Investigative crime analysis entails creating victim and suspect profiles for investigators based on information analysis. Occasionally, it is referred to as "criminal investigative analysis."
- Currently, intelligence analysis happens more frequently at the state and federal levels. It focuses on terrorism, organized crime, and providing information analysis and presentation support for particular investigations.

1.3. Crime survey

Based on the review of the literature. In India, one IPC crime is registered every seventeen seconds. There is a crime against women every three minutes, a molestation case every fifteen minutes, and a rape case every twenty-nine minutes. Every nine minutes, a case of abuse by a spouse or family member is reported; every twenty-nine minutes, an attempt at murder is made. Every two minutes there is a theft case, every minute there is a crime against property, and every eight minutes there is an economic crime. The idea of sophisticated and intelligent police is



rapidly gaining traction in today's global society. This includes placing a strong focus on predicting and problem-solving skills in law enforcement.

1.4. Data mining concepts used

Data mining involves sifting through massive amounts of data storage for specific data based on predetermined criteria. Information science, statistics, data visualization, machine learning, HPC, AI, and visualization are all interrelated in data mining.

Data mining involves sifting through massive amounts of data storage for specific data based on predetermined criteria. Information science, statistics, data visualization, machine learning, HPC, AI, and visualization are all interrelated in data mining.

Extraction of crime files

The data needs to be collected and combined from many sources; for example, criminal data could be in the form of PDFs, tables, Excel spreadsheets, web pages, static or dynamic photos, or recorded speech. The data then needs to be transformed into a common format for processing, like CSV or ASCII. Transforming data into a usable format and retrieving it are the two main steps in the data extraction process.

Preprocessing From

Data needs to be cleaned up by eliminating noise. Data can be transformed by aggregating, standardizing, and smoothing it. Patterns need to be reviewed to prevent redundant patterns. Cleaning the data involves ignoring the entire record or attribute if it causes confusion or inconsistency, filling missing values manually or automatically by assigning a default value (usually the mean or mode of the attribute) is one method used to preprocess data. To handle noisy data, a regression or clustering procedure can be used, or a neighborhood or binning approach.



Classification of crime data

The two phases of supervised learning that make up classification are training and testing. As a result, there is a training set and a testing set within the dataset. The classifier's architecture allows it to "learn" the membership criteria for data items from the training set. In order to determine the classifier's accuracy, we use the testing set. Some common types of classifiers include decision trees, artificial neural networks, support vector machines, Bayesian models, and discriminant analysis.

2. LITERATURE REVIEW

Swain, P. K., et.al., (2023) In recent decades, the serious global problem of an ever-increasing mortality toll from vehicle accidents has gained significant attention. This worldwide disaster is nothing new to India, thanks to her rapidly increasing motorization. This research forecasts the number of deaths in India due to road accidents from 2022 to 2031 using two very simple but effective and versatile techniques for time series data: the autoregressive integrated moving average method (ARIMA) and the exponential smoothing method. After comparing the outcomes from the two approaches, we see that they are consistent with one another and the previous research. There has never been an effort like this one to compare and contrast two time series analysis methods using the same data. The information was culled from two sources: the National Crime Record Bureau's Accidental Deaths and Suicides in India (ADSI) Report (2021) and the Ministry of Road Transport and Highways, India's annual report (2020). Among the many possible models tested, the ARIMA (2, 2, 2) and exponential smoothing (M, A, N) models proved to be the most appropriate for the provided data. The ARIMA (2, 2, 2) model outperforms the other one in terms of AIC and BIC. According to our model selection criterion, this model is the best. Additionally, the report shows that the number of road accident deaths in India is expected to rise over the next decade.

Singh, S., et.al., (2023) during the past few decades, on a worldwide scale, and even the most recent climate predictions (CMIP6) show that problems will only become worse as a result of intensified global warming. Understanding the geographical trend and variability associated with HWs is necessary for accurately analyzing the repercussions of these hazards, which have



negative effects on biophysical and human systems. In order to have a better idea of how vulnerable certain regions in India are, we ran a joint exceedance probabilistic analysis of HW features using a bivariate copula. We chose as HW-vulnerable areas those that had somewhat higher HW frequencies, intensities, durations, and joint return periods of both. The Indian states of Odisha, Karnataka, Madhya Pradesh, Uttar Pradesh, Punjab, Bihar, Chhattisgarh, Gujrat, Telangana, Andhra Pradesh, and Uttar Pradesh have all been identified as susceptible, and according to recent climate projections, they are expected to become even more so in the future. We used nonstationary models based on extreme value theory (EVT) to examine the relative impact of three important large-scale climatic oscillations (LSCOs)—ENSO, IOD, and AMO—on HW in order to reveal the spatiotemporal variability in HW characteristics. Statistical examination of the generalized extreme value distribution of temperature extremes using LSCOs shows that in April, 25.9% of the overall region was impacted by ENSO, while in May and June, 23.1% was impacted by IOD.

R. W. a. M. Peter (2001) Identified and clarified the difficulties encountered by law enforcement in recording and investigating crimes. This article delves into topics including data accuracy and the effectiveness of investigation methods, offering insights that can greatly enhance police strategies and administration. The technology aspect of crime recording systems is one of the main focuses of the study. In a time when IT improvements are vital to police work, Peter assesses how tech has altered the reliability and effectiveness of crime recording. The investigation will most certainly include discussions of database integration, digital evidence gathering methods, and the impact of technology on the investigative landscape as a whole. To discuss the larger picture of efficient crime recording, which includes its function in reducing crime, strengthening community policing, and boosting public confidence in the police. By evaluating the obstacles and accomplishments in crime recording and investigation systems, Peter's work undoubtedly provides useful insights to the continuing dialog surrounding the optimization of police techniques. This important part of law enforcement is complicated, but it provides a foundation for understanding it by looking at technology developments, problems, and wider ramifications. The findings offered here can help researchers and practitioners improve crime management practices and shape the field's future.



3. RESEARCH METHODOLOGY

We were fortunate enough to interview four important Karnataka police stations—the Central station, the Siddapura headquarters, the Muddebihal division, and the Kampegowda Nagar division—during our data collecting. A database is used to store and process the acquired data. Due to document-to-document variation in field count, content, and size, crime data is inherently unstructured. Here are some more advantages of using an unstructured database:

- Significant amounts of data that are organized, semi-organized, and unstructured.
- Simple and adaptable object-oriented programming.

In order to conduct this study, interview guides were utilized. In addition, there were audio recorders and writing implements. The interview guide explained the study's goals.

Table 1: variables found in the gathered data

Variables	Description	Examples
ID	Each crime's unique identification number	1,2,3
Area	Place of the crime	Karnataka
Day of week	Day of the week that there was a crime	Tuesday, Wednesday
Day hour	The time of day the crime was committed	11,9,10
Incident month	month in which the incident occurred	June ,July
Incident time tag	Time frame during which the incident took place	Evening ,Morning
Crime category	Type of crime	Rape, Theft
Longitude	Location of the crime's longitude	32,6000
Latitude	Latitude at which the crime occurred	0.4003
Incident date time	The timing and date of the crime	6/11/2010;10:10

- **Data clustering/classification**



The kmeans algorithm will be employed for the purpose of prediction analysis. Partitioning the clusters according to their means is the primary use case of this approach. The first stage is to group a certain number of objects into what are called K clusters. At first, we choose K objects to serve as the cluster centers. Once again, the cluster center determines the assignment of these objects. Next, the cluster means are revised once again. The majority of other clustering algorithms use this one as their foundation.

- **Prediction**

The decision tree concept is being utilized for the purpose of prediction. A decision tree resembles a graph in which each branch indicates the result of a test on an attribute and each internal node reflects the test itself. One major perk of decision trees is how easy they are to learn and use. Also, it's great with big data sets and has a robust character, which are additional benefits. When it comes to variables, this characteristic aids the algorithms in making better selections. A model is constructed in accordance with each location. Therefore, we feed the prediction engine up-to-date information and attributes in order to get the regions that are prone to crime. Several visualisation techniques are employed to display the outcome..

4. DATA ANALYSIS

The names of the crimes are shortened, and the number of reported instances falls into each category according to the crime analysis. It is worth mentioning that the data indicates a wide variety of illegal behaviors, each with its own unique incidence. With 37,512 reports, theft is clearly the most common crime, highlighting how common it is and the damage it could do to the communities it hits. The next most common type of drug-related violation is 'DRU..,' which has 13,500 documented cases. This shows that substance abuse and illegal drug activities are a big concern. The varied nature of criminal incidents is reflected in other notable categories such as Assault (ASS...), Robbery (ROB..), and Non-Specified Crimes (NON..), each of which had 11,630 instances. Additional dimensions of criminal activity include kidnapping (9,000 incidents), fraud (10,000 episodes), and sexual offenses (5,500 cases). These numbers show how many different types of crimes are happening in the community and how serious they are, so law enforcement needs to take proactive steps to stop them. 'Other' appears to be a grouping of



several infractions not specifically mentioned in the table, with 12,000 instances falling into it. This may include a wide variety of illegal acts that may not fit neatly into the defined categories, but which do add to the general crime scene.

Table 2: Distribution of crimes by time of day.

Crime name	Crime analysis
ARR	11,000
Assault	11,630
Drug-related offenses	13,500
Fraud	10,003
Kidnapping	9000
Non-Specified Crimes	12,650
Other	12,000
Robbery	11,600
Sexual Offenses	5500
Theft	37,512

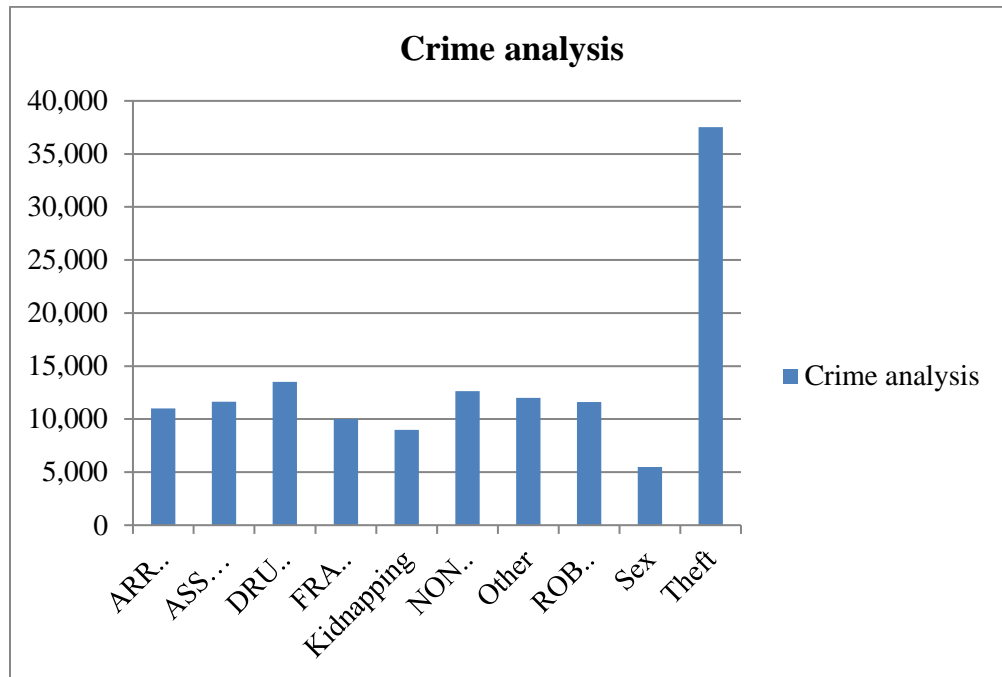


Figure 2: Distribution of crimes by time of day.

5. CONCLUSION

We examined data mining's potential for spotting trends in criminal activity by use of the clustering method. In this case, we were able to help investigators by using data mining to formalize criminal pattern recognition as a machine learning job. The current scope of our study involves predicting locations prone to crime. As for future scope, we may also estimate when the crime is likely to occur. This goes hand in hand with trying to guess where the crime took place. We will use various test sets to evaluate the efficacy of frequent-item sets and predictions. As a result, the system will pick up on shifting crime trends just by looking at them. Additionally, the elements that contribute to criminal activity evolve with moments. Moving through the available crime data will allow us to discover new causes of crime. Complete precision is impossible to attain because we are just taking a small number of elements into account. We need to identify more criminal attributes if we want to improve our prediction outcomes. An individual criminal's likely crime is predicted by our software. The Apriori Algorithm, in conjunction with association rule mining, will be utilized to achieve this goal. This is what a criminal needs to decide what crime to commit next.



REFERENCES

1. A. E. a. U. N, "Geographic Information Systems Technologies In Crime Analysis And Mapping," International Journal of Data Mining Techniques ans Applicatins, vol. 1, no. 2, pp. 117-120, 2012
2. A. Murray, "Explanatory spatial data analysis techniques for examining urban crimeimplications for evaluating treatment," British Journal of Criminolgy, vol. 41, no. 2, pp. 309-329, 2001.
3. a. W. D. M. Morabito, "Crime forecasting using data mining techniques," 11th IEEE International Journal, vol. 29, no. 10, pp. 779-786, 2011.
4. B. a. B. .L, "learning organisations in the public sector," A study of police agencies Employing Information And Technology To Advance Knowledge., vol. 1, no. Public Administration Review 63, pp. 30-43, 2003.
5. B. M. M. a. B. J. L, Learning organisations in the Public Sector, 2003.
6. D. E. Brown, "The Regional crime analysis program," IEEE internaltional journal on systems and man, vol. 3, no. 3, pp. 2848-2853, 2005.
7. F. U. M, "Knowledge Discovery and Data Mining: Towards a unifying framework," in AAAI Press, Menlo Park, Carliforni, Portland, Orgeon, Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining..
8. F.-T. a. M. Green, "Crime and society," International Journal of Social Economics, vol. 29, no. 6, pp. 781-795, 2002.
9. Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng,Homa Atabakhsh Crime Data Mining: An Overview and Case Studies, Artificial Intelligence Lab, Department of Management Information Systems, University of Arizona, Tucson, AZ 85721, USA 2010.
10. Large Data Clustering And Classification Schemes For Data Mining by Babu, T Ravindra, Computer Science and Automation (csa)Dec-2006.
11. Longbing Cao, Huaifeng Zhang, Yanchang Zhao, Dan Luo, and Chengqi Zhang, Combined Mining: Discovering Informative Knowledge in Complex Data, IEEE Transactions on systems, man,and cybernetics-part B: Cybernetics, Vol-41, no.3, June 2011.



12. Predicting Crime: The review of Research Department of Justice, www.justice.gc.ca. [9]. Liu Ying , University crime data mining, ACM,2004.
13. Singh, S., Yadav, A., & Goyal, M. K. (2023). Univariate and bivariate spatiotemporal characteristics of heat waves and relative influence of large-scale climate oscillations over India. *Journal of Hydrology*, 130596.
14. Swain, P. K., Tripathy, M. R., & Agrawal, K. (2023). Forecasting road accidental deaths in India: an explicit comparison between ARIMA and exponential smoothing method. *International Journal of Injury Control and Safety Promotion*, 1-14.
15. W. C. H. Chen, "A general framework and some examples," in IEEE international conference, 2004.

Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentricontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper maybe removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Prajvalarani Sushil Kumar Bogar

Dr. Amit Singhal