



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 1, January 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijrset@gmail.com

www.ijrset.com

A Review on Analysis and Prediction of Crime Patterns using Data Mining

Sanika Nimkar, Nikita Kamnani, Tejal Gohatre, Unnati Gadling, Niyati Wadhwa, Prof. Sumit Sagane

U. G. Student, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

U. G. Student, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

U. G. Student, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

U. G. Student, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

U. G. Student, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

Assistant Professor, Department of Computer Science Engineering, PRPCEM, Amravati, Maharashtra, India

ABSTRACT: Crime analysis and prevention is a methodical way of identifying and studying crime patterns and trends. This system can predict areas where crime is likely to happen and can show where crime is more common. With the growing use of computer systems, crime data analysts can assist law enforcement in solving crimes more quickly. A method to uncover useful information from disorganized data, to bridge computer science and criminal justice, and to speed up crime solving is Data Mining. Rather than focusing on why crimes happen, like the criminal's background or personal issues, we concentrate on understanding the factors contributing to daily crimes. In recent years, data mining techniques have been used to examine crime data from various sources, seeking patterns and trends in criminal activities. This can enhance crime detection efficiency and even automate crime alerts. However, there are many data mining techniques, and selecting the right one is crucial for effective crime detection. It represents the use of data mining in solving crimes highlights research gaps and challenges and serves as a guide for newcomers to the field of crime data mining.

KEYWORDS: Data Mining, Crime Detection, Crime Patterns and trends, Navie Bayes, Apriori Algorithm, Mongo DB

I. INTRODUCTION

The crime rate is steadily increasing, and it is a concerning trend. Crime is difficult to predict because it doesn't follow a set pattern and criminals often use modern technology to their advantage. Crime Records Bureau statistics show that certain crimes like burglary and arson have decreased, while more serious crimes like murder, sexual abuse, and gang rape have increased. While we can't predict who will be victims of crime, we can predict areas with a higher likelihood of crime occurring. Predictions aren't 100% accurate, but they do help reduce crime to some extent, especially in sensitive areas. Building a powerful crime analytics tool is crucial. This involves collecting and evaluating crime records. Crime prevention and detection are challenging but essential to solve crimes more efficiently. Various studies have developed techniques for crime solving, which can speed up the process and help automated systems detect criminals. Technology is advancing rapidly and can aid in addressing these challenges. Crime patterns are always changing and growing, and analyzing previously stored crime data is increasingly complex. To overcome these challenges, data mining techniques are used. Data mining involves various learning algorithms to extract hidden knowledge from large volumes of crime data. Data mining helps identify patterns and trends in crimes, making it easier to solve crimes quickly and even automate criminal detection.

The reason for choosing this method is to have only data about the known crimes and get the crime pattern for a particular place. Therefore, classification techniques that will rely on the existing and known solved crimes, will not give good predictive quality for future crimes.



II. OBJECTIVE

1. To collect large volumes of data for analyzing and extracting hidden knowledge.
2. To predict and detect criminal behaviors.

III. PROBLEM STATEMENT

- a) **Increase in Crime Data:** We're dealing with more and more crime information that needs to be stored and analyzed. This is due to the growing volume of data related to criminal activities.
- b) **Incomplete and Inconsistent Data:** Analyzing this data is challenging because it often lacks important information and is inconsistent. In other words, the data may be missing key details or may not be reliable.
- c) **Difficulty Obtaining Crime Data:** There's a limitation in accessing crime data records from law enforcement departments. This means that getting the necessary data from these sources can be difficult or restricted.
- d) **Accuracy Depends on Training Set:** The accuracy of any program used for analyzing crime data relies heavily on the quality and accuracy of the data used to train it. In other words, if the data used to teach the program is not accurate, the results it produces may also be less reliable.

IV. METHODOLOGY

A. Data Collection: In the data collection step, we are collecting data from different websites like news sites, blogs, social media, RSS feeds, etc. The collected information is retained in the database for subsequent analysis or manipulation. Since the collected data is unstructured, we use Mongo DB. Crime data is unstructured data since for documents with varying fields, content, and sizes, a schema-less database is preferable.

B. Classification: For classification, we are using an algorithm called Naive Bayes is both a supervised learning and statistical classification method. The Naive Bayes classifier, a probabilistic model, assigns classifications based on an input that gives a probability distribution of a set of all classes rather than providing a single output. The algorithm categorizes a news article into a specific crime type that fits the best.

C. Pattern Identification: The third phase is the pattern identification phase where we have to identify trends and patterns in crime. For finding crime pattern that occurs frequently we are using the Apriori algorithm. Apriori can be used to identify patterns in the database through association rules to reveal general trends. The result of this stage is the specific crime trends in a given location.

D. Prediction: For prediction, we are using the decision tree concept. A decision tree is similar to a graph in which the internal node represents a test on an attribute, and each branch represents the outcome of a test. The main advantage of using a decision tree is that it is simple to understand and interpret. This feature helps the algorithms to make better decisions about variables.

E. Visualization: High-crime areas can be visually depicted using a heat map which indicates the level of activity, usually darker colors to indicate low activity and brighter colors to indicate high activity. The below figure is an example of a heat map.

Area sensitivity	Notable event	VIP presence	Criminal group	crime
Yes	Yes	Yes	No	yes
Yes	Yes	No	Yes	no
No	No	No	Yes	no
Yes	No	No	No	no
Yes	Yes	Yes	yes	yes
No	Yes	No	No	no

Table 1.1 Table for Decision Tree for Delhi

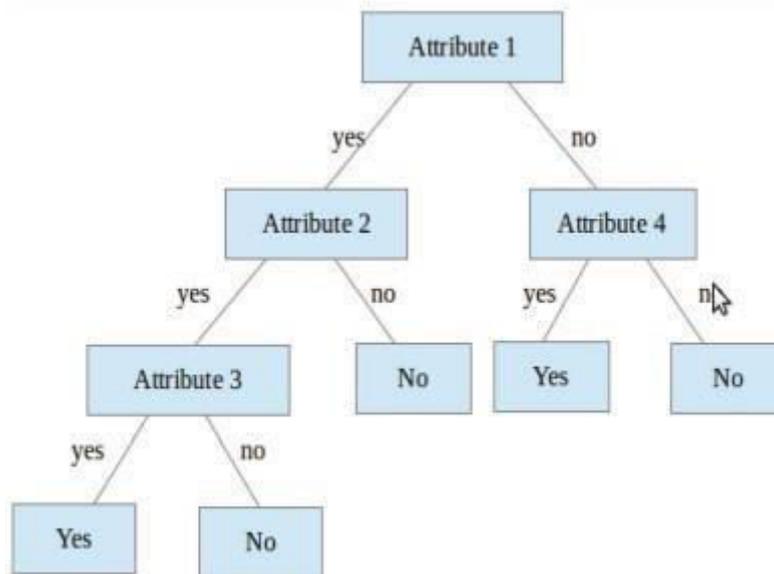


Fig. 1.1 Example of Decision Tree

V. CONCLUSION

Crimes are characterized which changes over time and increase continuously. The changing and increasing of crime lead to the issues of understanding crime behavior, crime prediction, precise detection, and managing large volumes of data obtained from various sources. Research interests have tried to solve these issues. However, this research still has gaps in crime detection accuracy. This leads to challenges in the field of crime detection. The challenges include modeling crimes to identify effective algorithms for crime detection, achieving accurate results, preparing the data and transformation, and processing time. It has been tested the accuracy of classification and prediction based on different test sets.

Classification is done based on the Bayes theorem which showed more than 90% accuracy. By using this algorithm, numerous news articles were trained and a model was built. For testing, test data is being input into the model, which shows better results.

REFERENCES

1. S. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in Networks Soft Computing (ICNSC), 2014 First International Conference on, Aug 2014, pp. 406–412.
2. Thammano and P. Kesisung, "Enhancing k-means algorithm for solving classification problems," in Mechatronics and Automation (ICMA), 2013 IEEE International Conference on, Aug 2013, pp. 1652–1656.
3. C.-H. Yu, M. Ward, M. Morabito, and W. Ding, "Crime forecasting using data mining techniques," in Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, Dec 2011, pp. 779–786.
4. Jayaweera, C. Sajeewa, S. Liyanage, T. Wijewardane, I. Perera, and A. Wijayasiri, "Crime analytics: Analysis of crimes through newspaper articles," in Moratuwa Engineering Research Conference (MERCon), 2015, April 2015, pp. 277–282.
5. Malathi. A and Dr. S. Santhosh Baboo. Article:an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1-6, May 2011. Published by Foundation of Computer Science.
6. M. Sharma, "Z - crime: A data mining tool for the detection of suspicious criminal activities based on decision tree," in Data Mining and Intelligent Computing (ICDMIC), 2014 International Conference on, Sept 2014, pp. 1–6.
7. V. Vaithiyanathan, K. Rajeswari, R. Phalnikar, and S. Tonge, "Improved apriori algorithm based on selection criterion," in Computational Intelligence Computing Research (ICCIC), 2012 IEEE International Conference on, Dec 2012, pp. 1–4.
8. Thammano and P. Kesisung, "Enhancing k-means algorithm for solving classification problems," in Mechatronics and Automation (ICMA), 2013 IEEE International Conference on, Aug 2013, pp. 1652–1656.



9. Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in Proceedings of the Eleventh International Conference on Information and Knowledge Management, ser. CIKM '02. New York, NY, USA: ACM, 2002, pp. 515–524. [Online]. Available: <http://doi.acm.org/10.1145/584792.584877>.
10. S. Kaza, D. Hu, H. Atabakhsh, and H. Chen, "Predicting criminal relationships using multivariate survival analysis," in Proceedings of the 8th Annual International Conference on Digital Government Research: Bridging Disciplines & Domains, ser. dg.o '07. Digital Government Society of North America, 2007, pp. 290–291. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1248460.1248524>.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details