Applications of Data Mining Techniques and Algorithms in Crime Analysis and Digital Forensics: A Review

Narmadha Devi. A. S[1], Dr. K. Sivakumar[2]

Research Scholar, Saveetha School of Engineering, SIMATS, Chennai-602105, India
https://orcid.org/0000-0001-5718-5139
narmadhadevias9010.sse[at]saveetha.com

Professor, Department of Mathematics, Saveetha School of Engineering, SIMATS, Chennai-602105, India
sivakumarkaliappan.sse[at]saveetha.com

Abstract: *Crime analysis is an essential method for figuring out why certain crimes keep happening over and over again. Data analytics applied to criminal offenses have the potential to aid law enforcement in the age of ever-more-pervasive digital technologies by allowing them to more rapidly apprehend suspects and apprehend crimes. Data mining is a technique for extracting meaningful information from large amounts of unstructured data. Using analytic and predictive techniques, as in "predictive policing," has been proven to be quite useful in identifying criminals. With the rise in crime over the years comes a corresponding rise in the volume and complexity of crime data that must be stored and analyzed in massive data warehouses. Furthermore, modern criminals are becoming increasingly technologically sophisticated, necessitating the use of cutting-edge tools to stay one step ahead of the law enforcement establishment. The primary purpose of this article is to provide a comprehensive overview of the algorithms and methods currently in use to detect and track down criminals.*

Keywords: Data Mining, crime analysis, Digital forensics, Naïve Bayes Classifiers, Predictive approach

1. Introduction

The ever-increasing crime rate is a huge cause for worry since it will inevitably slow down the spread of effective government. Criminal activity cannot be analyzed since it is neither predictable nor random. While burglaries, arson attacks, and the like have decreased, more serious crimes like homicide, sex abuse, gang rape, and so on have grown. While we can't examine the victims themselves, we may examine the scene of the crime. Intelligence agencies and local police departments have a hard time analyzing the data to spot trends in criminal activity or anticipate where crimes could occur in the future. Therefore, a powerful analysis tool is required to easily and rapidly examine crime data and provide actionable criminal trends.

The term "predictive policing" refers to a method of criminal identification that uses statistical analysis and forecasting. We need to use cutting-edge technology to stay one step ahead of criminals who are becoming increasingly sophisticated in their methods of operation as a direct result of the rising crime rate over the years, which will necessitate the management of vast quantities of crime data stored in warehouses that would be extremely challenging to analyze manually. What follows is a brief outline of the remainder of this paper's structure: In the Literature Survey section, we take a quick look at some similar studies that have used data mining methods and algorithms to analyze criminal behaviour.

Types of Crime Analysis
The term "crime analysis" is used to describe a set of regular, analytical procedures that, regularly, compile information regarding crime trends and patterns. The field of crime analysis may be broken down even further into subfields depending on the scope, methodology, and data used in each study [2]:

1) Intelligent Analysis: Intelligent analysis's primary goal is to assist law enforcement agencies to locate and apprehend criminal networks engaged in illegal behavior. Intelligence studies rely on data collected by law enforcement through means such as surveillance, participant observation, eavesdropping, etc. Communications, itineraries, and bank records of the suspects in question are all examples of the kinds of data that may be obtained during an inquiry.
2) Investigative Analysis: "criminal profiling" also describes this practice. Profiling criminals who have committed significant crimes are a part of this procedure. The primary goal of this sort of study is to aid law enforcement in the identification of offenders by revealing distinguishing traits, social behaviors, and the like.
3) Tactical Analysis: To uncover patterns, identify potential suspects, and close ongoing cases by connecting the dots between previous investigations, criminologists study the intricacies of criminal investigation and analysis. Information gathered by patrol officers in the field concerning possible criminal activities is also analyzed.
4) Strategic Analysis: A strategic crime analyst use statistical tools to sift through voluminous electronic information. These experts deal with event variables including date, place, time, and category.
5) Administrative Analysis: Its focus is on disseminating information to the public and those working in law enforcement, government, and other related fields on crimes committed and the results of related research. It's a method of selecting relevant results from previous analyses and presenting them in the most effective way possible. Its major goal is to educate listeners.

2. Literature Survey

Mugdha Sharma et al. The classification fault of selecting attributions with higher values was solved by using the suggested improved ID3 method to provide importance-attribute significance on the at-attributes which have fewer values but greater importance. The experimental data reveal that the improved ID3 algorithm obtains more logical and efficient categorization rules. E-mail-based criminal activity analysis was also suggested using the Z-crime technology.

Sushant Bharti et al. To uncover the potential future criminal partner and alternative network outside the actual one, a hidden link algorithm was devised. The significance of the node is also examined in this work. Using this research, we can learn what role each node plays in the network. This helps identify influential members of the network and their relative strengths. In addition to analyzing the co-offender network in India and projecting its likely future expansion, this research provided a predictive technique for crime analysis that may be used to prevent crimes before they had happened.

Shiju Sathyadevan et al. Criminal pattern analysis using an Apriori-proposed method. Association rules that reveal overarching database tendencies may also be derived using this approach. The report also suggests using the naive Bayes approach to train the model using crime statistics. It was found that the Naive Bayes algorithm achieved 90% accuracy in testing.

Prashant K. Khobragade et al.A new version of the Forensic Tool Kit, version 4.0, is offered; it allows for remote data inquiry and graphical analysis. When looking at distant data, it's important to look at things like process data, service data, driver data, network device data, and network data. File and data analysis are both performed by this program. Furthermore, this instrument is utilized to investigate the attacked system. Memory data (both physical and digital) is examined as part of a criminal inquiry.

K.Zakir Hussain etal. used data mining methods to study of criminal activity. In this study, we present a new method for analyzing criminal investigations (CIA). The police community used this resource to aid in the solving of homicidal cases. Evidence gathered from interviews with victims and witnesses formed the basis for this conclusion. Both a forensic and behavioral lens were used in the investigation and analysis. It revealed information on the unidentified perpetrator and offered investigation tips and techniques for interviews and court proceedings.

Architecture of Basic Crime Prediction System

The following steps make up the engineering of the primary framework:
1) Stacking Crime Data: The client starts by obtaining the criminal activity dataset from the National Crime Records Bureau (NCRB) of India. Information on all types of infractions committed in India between 2011 and 2015 is included in this collection. This dataset allows researchers to examine a wide range of factors. The information is being recorded using. CSV
2) Information Pre-Processing: Following the accumulation of evidence of wrongdoing, the next step in this approach is the preliminary processing of the collected data. It's a process for transforming data from its raw form into a more digestible version by adding missing attributes, streamlining the data, and simplifying the layout.
3) Utilization of AI Calculation: After initial processing is done, usable data is collected and processed. Methods of sorting and grouping are applied to the computations based on these criteria. Characterization calculations like Credulous Bayes undermine the controlled learning concept by requiring random inspections, such as splitting the data into test and train testing (say, 80% train tests and 20% test tests) so that the classifier may be made ready to recognize a previously undetected crime. K-mean and other methods of grouping data rely on independent calculations on how to divide up the wrongdoing records into manageable chunks before creating the groups.
4) Information Visualization: Data visualization is the practice of representing large amounts of data in the form of pictures or graphs to aid in the viewer's quick and easy interpretation of that data. The effects may be seen with the help of appropriate diagrams or guides revealing vulnerable areas with a high probability of violations. Bar graphs, boxplots, heat maps, scatterplots, and similar graphs are often used to show perception.
5) Design Recognition: Next, the system uses a process called design ID to keep tabs on how often crimes of a similar sort and classification are clustered together. Police may use evidence from high-impact cases to inform the development of effective strategies for countering criminal activity and reducing crime rates.
6) Forecast: Models that may anticipate future misconduct are driven by the endless instances collected.
7) Chief: The information gathered in this way is useful to law enforcement agencies in their efforts to perfect their mediation processes via the use of practical preparedness.

3. Review Analysis

The evaluation is based solely on a data mining examination of criminal activity. The primary goal is to reduce crime by utilising cutting-edge technology so that law enforcement can stay one step ahead of criminals. One method involves using data on cases filed and arrests made over the course of five years to determine crime rates. A large number of reported incidents over the last five years is readily apparent. Arrest rates are modest, but lowering crime is the objective.
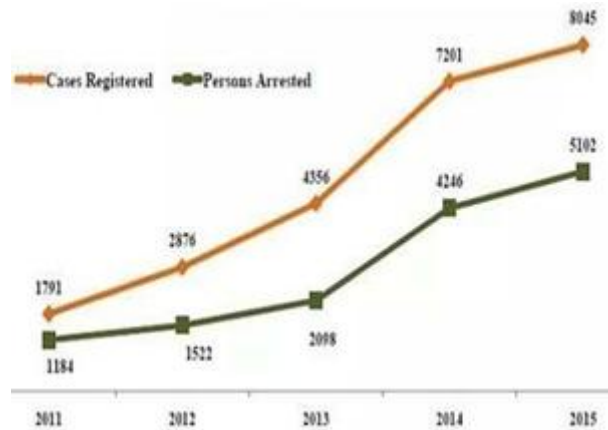
Figure 1: Cases Registered Versus Persons Arrest

Table I: Different Data Mining Approaches Implemented for Crime Analysis

P. No.	Focus	Methods/ Tool	Advantages	Futurework
[1]	The goal is to foresee and reduce criminal activity.	Z-Crime Tool ID3Algorithm	Data mining as a tool for anticipating criminal behavior deserves further attention.	Two distinct strategies for identifying salient features and assigning them to distinct classes. E-mails will soon include an automated reply option integrated specifically for mobile use cases.
[2]	The structure, functioning, and eventual breakup of organized crime networks are the subject of intense study. To analyse crime data statistically and discover important trends, hidden linkages, and link predictions.	The hidden link detection algorithm	Predictive Methodology Preventing further criminal activity is another potential benefit of crime analysis. To better understand municipal crime trends, we may build visual and intuitive criminal and intelligence investigative approaches.	In the future, we will apply a predictive method of crime analysis to assist prevent crimes before they happen.
[3]	Finding and analyzing classification accuracy using various test sets is the main focus here.	Naïve Bayes Classifiers Apriori algorithm	Bayes's theorem, used in classification, yields an accuracy of more than 90%.	Future study is needed on the issue.
[4]	Information gathering from various online sources is the main focus here.	ForensicToolKit4.0tool	Memory forensic analysis and remote system forensics are two common applications discussed by the tool's users.	to resort to other methods of forensic memory investigation.
[5]	The goal is to use data mining tools to identify illicit activity.	A technique known as Criminal Investigative Analysis (CIA) was applied.	Datamining is used to investigate and predict criminal behavior, namely homicides. Data-mining methods improve both the rate and depth of analysis.	
[6]	The goal is to discover the elusive cybercriminal by following the trail of his or her questionable remarks.	There was the usage of a pattern-detection technique which may be considered suspicious.	The proposed method was dynamic in nature, difficult to construct, ontology-aware, database- and data-mining-friendly, and only slightly more efficient than the state-of-the-art.	To address Big Data issues, it will be necessary to integrate with HA-DOOP.

4. Conclusion

The study of experimental data led us to the conclusion that the state-of-the-art ID3 algorithm is the most logical and efficient set of classification criteria. The goal of the hidden link algorithm was to uncover previously undetected connections between criminals, illuminating new pathways of potential collaboration and alternative networks outside the actual one. Using Bayes's theorem, classification methods achieved an accuracy of above 90%. The file and data are created by the forensic kit tool, which is also used to evaluate the data and the victim system itself while an assault is in progress. The accuracy of the Criminal Investigation Analysis (CIA) tool, which is intended to aid in the solving of violent crimes, was poor.

5. Future Work

The task of the future will include using this new technology to better understand and control criminal behavior. Since criminal investigation analysis tools don't rely on crystal balls or substitute in-depth research, they have their limitations in terms of precision and specificity. To find the distant node, conveniently analyze the data for the report presentation, and use additional tools for memory forensic analysis, we will utilise a forensic toolkit. Future studies will include analyzing the network of Co- offenders in India and making predictions about the likely future network of offenders, as well as using a predictive method in crime analysis to prevent crimes before they happen.

References

[1] Mugdha Sharma. "Z-CRIME: A Data Mining Tool for the Detection of Suspicious Criminal Activities Based onDecision Tree", International Conference on Data Mining and Intelligent Computing, 5-6 September2014.

[2] Alkesh Bharati, Dr. Sarvanaguru RA.K ,"Crime Prediction and Analysis using Machine Learning" In International Research Journal of Engineering and Technology (IRJET), Vol no.05, pp. 2395-0072, 2018

[3] Sushant Bharti, Ashutosh Mishra. "Prediction of Future possible offender's network and role of offender's", Fifth International Conference on Advances in Computing andCommunications,2015.

[4] Shiju Sathyadevan, Devan M.S and Surya Gangadharan. S. "Crime Analysis and Prediction Using Data Mining", First International Conference onNetworks&SoftComputing,2014.

[5] Prashant K. Khobragade and Latest G. Malik."DataGeneration and Analysis for Digital Forensic Application using Data mining", Fourth International Conference onCommunicationSystemsandNetwork-Technologies,2014

[6] K.Zakir Hussain, M. Durairaj and G. RabialahaniFarzana. "Criminal Behavior Analysis By Using Data Mining Techniques", IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012), March 30-31, 2012.

[7] Mahmood Ali,KhajaMoizuddin Mohammed, and Lakshmi Rajamani ."Framework for Surveillance of Instant Messages in Instant messengers and Social networking sites using Data Mining and Ontology", IEEE-Students' Technology Symposium, 2014.

[8] Q. Ansari ,TapasyaPatki , A. B. Patki and V. Kumar. "Integrating Fuzzy Logic and Data Mining: Impact on Cyber Security ", IEEE-Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.

[9] Chandra, Manish Gupta, and M. P. Gupta. "A Multivariate Time Series Clustering Approach for Crime Trends Prediction", IEEE International Conference on Systems, Man and Cybernetics, 2008.

[10] Rajesh Kumar, Nikhil R. Pal, BhabatoshChanda and J.D. Sharma. "Detection of Fraudulent Alterations in Ball-Point Pen Strokes using Support Vector Machines", IEEE India Conference (INDICON), 2009.

[11] VeenaH. Bhat, Prasanth G Rao, Abhilash RV, Deepa Shenoy, Venugopal K R and L M Patnaik. "A Novel Data Generation Approach for Digital Forensic Application in DataMining", IEEE Second International Conference on Machine Learning and Computing, 2010.

[12] T. Subbulakshmi, Dr. S. Mercy Shalinie V. Ganapathi Subramanian, K. Balakrishnan, D. Anand Kumar, and K. Kannathal."Detection of DDoS Attacks using Enhanced Support Vector Machines with Real Time Generated Dataset", Third International Conference on Advanced Computing (ICAC), 2011.

[13] Mayank Pandey and Vadlamani Ravi. "Detecting phishing e-mails using Text and Data mining", IEEE International Conference on Computational Intelligence and Computing Research, 2012.

[14] Vinit Kumar Gunjan, Amit Kumar, and Sharda Avdhanam."A Survey of Cyber Crime in India",15th International Conference on Advanced Computing Technologies (ICACT), 2013.

[15] Malathi. A and Dr. S. Santhosh Baboo. "An enhanced algorithm to predict a future crime using data mining", International Journal of Computer Applications, 1-6, May2011.

[16] Vinod Gendre, Nitish Kumar Chandrakar, Lalit Kumar P. Bhaiya, "A Survey on Crime Detection And Prediction Techniques" ijraset, 2022.