Crime Analysis and Prediction

Aman Rajesh Singh

Parul Institute of Engineering & TechnologyParul University Gujarat, India 2205112140062@paruluniversity.ac.in

Gaurav DeepakSuratwala

Parul Institute of Engineering & TechnologyParul University Gujarat, India 2205112140053@paruluniversity.ac.in

Prof. Vipul Gamit

Assistant Professor Parul Institute of Engineering & TechnologyParul University Gujarat, India

Abstract:

Data Mining in Crime analysis is a systematic approach to identifying and analyzing patterns and trends in crime. With the increasing adoption of computerized systems, Data Mining in Crime analysis can assist law enforcement in predicting crime. This technique plays a key role in cyber forensics. Cyber forensics involves the collection and analysis of digital evidence to detect and document a cybercrime or security breach. Data mining techniques can help investigators extract valuable insights from the large volume of data typically collected during a cyber forensics investigation.

Keywords: Data mining, crime, pattern, analysis, cyber, Forensics, detect, document,breach, security.

I. INTRODUCTION

Day by day the crime data rate is increasing as modern technology and hi- tech methods help Criminals to achieve illegal activities. According to Crime Records Bureau crimes like burglary, arson etc. have increased while crimes like murder, sex, abuse, gang rape etc. have increased [2]. Crime data will be collected from various blogs, news and websites. Big data is used as a record to create a Crime reporting database. Knowledge gained from data mining techniques will help Crimes are reduced as it helps to detect criminals faster and areas that are most affected by crime.[11]. Data mining helps to solve crimes faster and gives good results when this technique is applied. Crime dataset, information obtained from data mining techniques can help the police department.

One particular approach found to be useful by the police is 'hot' crime identification Spot' which refers to areas with a high concentration of crime [1]. May use data mining techniques Generate important results from crime reporting datasets. The very step in the study of crime is the analysis of the crime. Crime Analysis is the exploration, interrelation and discovery of relationship between various crimes and characteristics Crime This analysis helps in generating statistics, queries and maps on demand. It also helps to see if a Offense in a certain known pattern or a

necessary new pattern.

Crimes are predictable because criminals are proactive and operate within their comfort zone. Once they succeed Try to replicate the crime under similar circumstances. The incidence of crime depends on many factors Like intelligence of criminals, security of location, etc. work follows steps used in data Analysis, the important stages of which are data collection, data classification, patternrecognition,

Forecasting and Visualization. Uses various visualization techniques to show the proposed structure

Crime trends and the various ways in which crime can be predicted using machine learning algorithms

➢ **Understanding the Attack:**

- **DoS and DDoS Attacks:**

DDos stands for Distributed Denial of Service, an attacker uses this attack to prevent a user from accessing resources at the time of request. The main objective is to prevent a user from accessing its resources for a certain period of time

- **MITM Attacks:**

MITM stands for Man in the Middle Attack, in which the attacker tries to disrupt the communication between two people. An attacker tries to hide in the secure communication in which the message is sent from the source to the destination.

- **Phishing Attacks:**

Phishing is a type of social engineering attack where an attacker tries to steal important information such as personal login ID and password, bank password through a website created by him which is an exact replica of the original website.

- **Ransomware Attack:**

Ransomware is a type of malware. The main purpose of issuing malware is to get money from you. The attacker will perform a ransomware attack on your system, this will cause your system to be encrypted so that he will demand a huge ransom to decrypt the system. But it is not sure that after you pay the ransom the attacker will give the key which will help to decrypt the system.

- **Password Attack:**

A password-attack is a typical attack vector used to compromise user account authentication. The attacker uses one of the attacks that will help to crack the data which is known as brute force attack This attacker uses all the possible combinations until the required combination is found.

2. CRIME DATA ANALYSIS

Effective management of crime-related data is imperative for ensuring the safety of communities. Employing systematic methodologies to categorize this information according to the seriousness and location of incidents, discerning latent correlations within perpetrated crimes, and projecting their future likelihoods are fundamental focal points. One prevalent strategy is the implementation of hot spot analysis, encompassing techniques such as point pattern examination and clustering/distance computations. Additionally, an impactful approach involves the recognition of trends or patterns utilizing methodologies like data mining, text mining, spatial analysis, and self-organizing maps.[1] A crime assessment tool should promptly and effectively identify crime trends to facilitate proactive detection and response measures. The primary goals of crime analysis encompass:

Deriving crime patternsfrom availablecriminal data through analytical processes.

Identifying instances of criminal activity.[3]

Formulating effective and preciseidentification methodologies.

2.1 CRIME ANALYSISMETHODOLOGY

- Data Collection
- Classification
- Pattern Identification
- Prediction
- Visualization

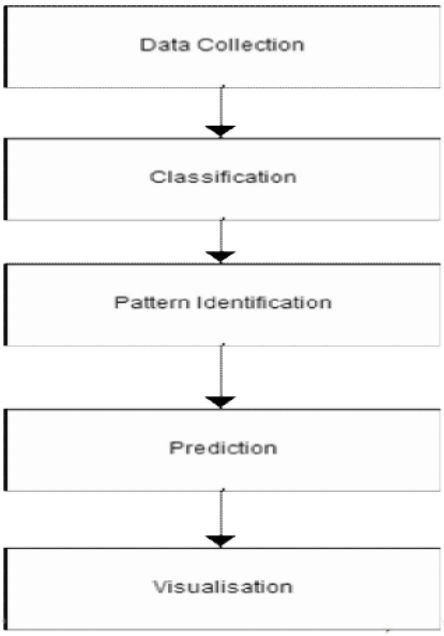

Fig. Crime analysis proceeds through several stages, including:

Data Collection:

The compilation of record series stands as a pivotal step in crime analysis, drawing data from diverse outlets such as websites, news platforms, and blogs. This information is subsequently archived within a database for further processing. Characterized by its unstructured nature, the data employs object-oriented programming, rendering it accessible and adaptable.

Crime data exhibits an unstructured format owing to variances in fields, content, and report lengths. To streamline operations and circumvent complexities, a schema-less database is favored, thus eliminating the necessity for joins. Embracing an unstructured database offers additional advantages, including the facilitation of large-scale data management encompassing structured, semi-structured, and unstructured formats, alongside the provision of user-friendly and flexible object-oriented programming paradigms.

Classification

During this phase, the Naive Bayes algorithm, a supervised learning technique, is employed. The Naive Bayes classifier, renowned for its probabilistic nature, furnishes a probability distribution across all training sets upon receiving input, as opposed to a singular output. Notably, one of its key advantages lies in its simplicity and expedited processing in contrast to logistic regression.[2] Moreover, this algorithm exhibits superior memory

efficiency compared to alternatives such as Support Vector Machine (SVM). Through the utilization of the Naive Bayes algorithm, a model is constructed by training it on crime data pertaining to various offenses. Particularly adept at handling modest training sets, Naive Bayes computes class parameters by estimating probabilities, often involving calculations such as P(A) * P(B/D) *%/D) * P(E/D), where percent/D)=0.[2]

Pattern Identification

In the 3rd step, we analyze the sample identity to identify crime trends and patterns. We utilize the apriori algorithm to discover frequently occurring crime patterns. Apriori helps determine association rules that reveal common trends in the database. This analysis assists police officers in preventing crime effectively, particularly in specific areas, by implementing security measures such as CCTV, alarms, etc.

Here's a simple breakdown of Apiori's Algorithm:

Initial Step: Start by identifying individual items that meet a specified support threshold, which indicates their frequency of occurrence in the dataset.

Iterative Process: Then, the algorithm iteratively generates larger itemsets by combining the frequent itemsets found in the previous step. It prunes any itemsets that don't meet the support threshold, gradually narrowing down the search.

Association Rule Generation: From the frequent itemsets, association rules are created to reveal relationships between different items. These rules indicate which items tend to co-occur together.

Evaluation Metrics: The generated rules are evaluated based on metrics like support, confidence, and lift, which measure their significance and reliability.

In essence, the Apriori algorithm helps uncover associations and patterns within data, making it a valuable tool for various applications like market basket analysis and recommendation systems

Prediction

In predicting the anticipated crime type, four crucial attributes of the crime are taken into account:

1) The month of occurrence
2) The day of the week
3) The time of the incident
4) The crime location

By leveraging these attributes, a more effective forecast of events and crime typescan be achieved.
The methodologies employed include:K-Nearest Neighbor (k-NN):
It identifies pattern based on similar crimefeature

Decision Trees (J48): Decision trees partition the feature space into binary decisions, facilitating easy interpretation and visualization. By recursively splitting data based on feature values, they effectively classify crime types.

Support Vector Machine (SVM): SVM finds the optimal hyperplane to separate different classes in the feature space. Particularly useful for handling high-dimensional data and nonlinear relationships, it's well-suited for predicting crime types with complex feature interactions.

Neural Networks: Leveraging deep learning architectures, neural networks excel at capturing intricate patterns and representationsfrom data. This capability enables them to predict crime types with high accuracy.

Naïve Bayes and Ensemble Learning: Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem, assuming independence among features. Despite its simplicity, it often performs well, especially with small datasets or when feature independence holds true. Ensemble learning techniques, such as bagging and boosting, combine multiple classifiers to improve prediction accuracy, offering robustness andgeneralization capabilities.
These techniques helps to develop effectivecrime prediction models, aiding proactiveprevention efforts.

By employing these diverse classification techniques, we aim to harness the predictive power of machine learning to anticipate crime types based on crucial features. Through rigorous evaluation and comparison of these methods, we strive to develop effective crime prediction models that assist law enforcement agencies in proactive crime prevention and resource allocation efforts

Visualizations

It play a crucial role in data mining and crime analysis by providing a graphical representation of complex data patterns and trends, which can help analysts and decision- makers gain insights and make informed decisions. Here's an explanation of the processof visualization in data mining and crime analysis:

Data Collection: The process starts with collecting relevant data from various sources such as crime reports, incident logs, demographic data, geographic information systems (GIS), social media, etc. This data can include information about types ofcrimes, locations, times, demographics of offenders and victims, and other contextual variables.

Data Preprocessing: Once the data is collected, it undergoes preprocessing steps such as cleaning, filtering, and transforming to ensure its quality for analysis. This may involve handling missing values, removing outliers, standardizing formats, and integrating data from different sources.

Exploratory Data Analysis (EDA): Before diving into complex analysis, analysts often perform exploratory data analysis to understand the characteristics and distributions of the data. Visualization techniques such as histograms, box plots, scatter plots, and heatmaps are used to explore relationships, identify patterns, and uncover anomalies in the data.

Feature Selection and Engineering: In data mining, feature selection and engineering involve identifying the most relevant variables (features) that contribute to the predictive power of the model. Visualization techniques like correlation matrices, scatter plots, and dimensionality reduction methods (e.g., PCA) can aid in identifying important features and reducing the dimensionality of the data.

Model Development: Following data preparation and feature identification, analysts proceed to construct predictive models employing machine learning algorithms like classification, regression, clustering, or time series forecasting, tailored to the specific requirements of the analysis. Visual aids such as decision trees, ROC curves, confusion matrices, and model performance metrics serve to assess and elucidate the models' effectiveness.

Crime Pattern Analysis: Visualizations play a crucial role in identifying spatial and temporal patterns of crime. Geographic Information Systems (GIS) are commonly used to map crime incidents spatially, visualize hotspot areas, and analyze crime trends over time. Heatmaps, choropleth maps, and kernel density estimation (KDE) plots are used to visualize crime densities and spatial clusters.

Temporal Analysis: Visualizations help in analyzing temporal patterns of crime, such as daily, weekly, monthly, or seasonal variations. Time series plots, bar charts, and calendar heatmaps are used to visualize the frequency and distribution of crimes over time, identify peak hours or days, and detect long-term trends.

Predictive Analytics: In crime analysis, predictive analytics techniques are used to forecast future crime occurrences, identify high-risk areas, and allocate resources efficiently. Visualizations such as predictive heatmaps, risk terrain modeling (RTM), and spatiotemporal forecasting models help visualize predictive insights and support decision-making.

Interactive Dashboards: Interactive dashboards serve as valuable aids in decision- making and information dissemination, often crafted using visualization tools like Tableau, Power BI, or Python libraries such as Matplotlib and Plotly. These dashboards empower stakeholders to dynamically explore and engage with the data, delve into specific details, and extract actionable insights.

Reporting and Presentation: Subsequently, the analysis findings are documented in reports or presentations, with visualizations assuming a central role in effectively communicating key insights. It is imperative to ensure that the visualizations are clear, concise, andappropriately labeled to convey the intended message to the audience.

By following these steps we can can uncover valuable insights and ultimately supporting evidence-based decision-making and crime prevention efforts.

3. Conclusion

This paper focuses on constructing predictive models to forecast crime frequency per crime type per month, considering factors such as poverty, law enforcement, and corruption.

The proposed model holds significant utility for both investigative agencies and law enforcement officers, aiding them in taking proactive measures to mitigate crime.

Furthermore, this project facilitates crime analysts in examining crime networks through a variety of interactive visualizations.

Future advancements in this research endeavor include training artificial intelligence bots to anticipate crime-prone areas utilizing machine learning techniques. Machine learning, akin to the sophisticated concept of data mining, offers enhanced capabilities for predictive analysis.

Additionally, data manipulation techniques can be employed to bolster confidentiality, reliability, accuracy, and predictive prowess.

4. Reference

[1] Ginger Saltos and Mihaela Coacea, An Exploration of Crime prediction Using Data Mining on Open Data, International journal of Information technology & Decision Making,2017.

[2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, Crime Analysis and Prediction Using Data Mining, First International Conferenceon networks & soft computing (IEEE) 2014.

[3] Khushabu A.Bokde, Tisksha P.Kakade, Dnyaneshwari S. Tumasare, Chetan G.Wadhai
B.E Student, Crime Detection Techniques Using Data Mining and K-Means, International Journal of Engineering Research & technology (IJERT)
,2018

[4] H.Benjamin Fredrick David and A.Suruliandi,Survey on crime analysis and prediction using data mining techniques, ICTACTJournal on Soft computing, 2017.

[5] Tushar Sonawanev, Shirin Shaikh, rahul Shinde, Asif Sayyad, Crime Pattern Analysis, Visualization And prediction Using Data Mining,Indian Journal of Computer Science and Engineering (IJCSE), 2015.

[6] RajKumar.S, Sakkarai Pandi.M, Crime Analysis and prediction using data mining techniques, International Journal of recent trends inengineering & research,2019.

[7] Sarpreet kaur, Dr. Williamjeet Singh, Systematic review of crime data mining, International Journal of Advanced Research incomputer science, 2015.

[8] Ayisheshim Almaw, Kalyani Kadam, Survey Paper on Crime Prediction using Ensemble Approach, International journal of Pure and Applied Mathematics,2018.

[9] Dr .M.Sreedevi, A.Harha Vardhan Reddy, ch.Venkata Sai Krishna Reddy, Review on crime Analysis and prediction Using Data Mining Techniques, International Journal of Innovative Research in Science Engineering and technology
,2018

[10] K.S.N.Murthy, A.V.S.Pavan kumar, Gangu Dharmaraju, international journal of engineering, Science and mathematics, 2017.

[11] Deepiika k.K, Smitha Vinod, Crime analysis in india using data minig techniques , International journal of Enginnering and technology, 2018.

[12] Hitesh Kumar Reddy ToppyiReddy, Bhavana Saini, Ginika mahajan, Crime Prediction &Monitoring Framework Based on Spatial Analysis, International Conference on Computational Intelligence Data Science (ICCIDS2018).

[13] researchgate.net