Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Forensic Breakthrough for Crime Scene Analysis via Mobile Data and NLP Applications

B. Krishna Veni[1], Khateeb Syed Muskan[2], Bingi Pranitha Raj[2], Jenne Sumithra[2], Gogula Sai Ganesh[2], Subramanyam Sujeeth Kumar[2]

[1]Assistant Professor, Computer Science and Engineering, Gates Institute of Technology, Gooty, AP, India
[2]UG student, Computer Science and Engineering, Gates Institute of Technology, Gooty, AP, India

Correspondence

B. Krishna Veni

Assistant Professor, Department of CSE, Gates Institute of Technology, Gooty, AP, India

- Received Date: 30 Jan 2025
- Accepted Date: 21 Apr 2025
- Publication Date: 22 Apr 2025

Abstract

With the rise of digital communication, chat logs have become a crucial source of evidence in forensic investigations. The crime scene analysis by integrating Natural Language Processing (NLP) techniques to analyze chat data for identifying key information, suspect behavior, and potential criminal intent. Using advanced NLP methods such as sentiment analysis, named entity recognition, and text classification, it can detect hidden connections, and contextual meanings within conversations. This automated analysis enhances the efficiency and accuracy of forensic investigations, reducing manual effort and enabling law enforcement to make data-driven decisions. By leveraging chat-based forensic analysis, this contributes to modern crime-solving techniques, offering a powerful tool for digital evidence examination. Advancements in Natural Language Processing (NLP) and digital forensics have revolutionized crime scene investigations, particularly in analyzing digital communication for crucial evidence. forensic breakthrough that leverages chat logs and NLP applications to extract, analyze, and interpret conversations relevant to criminal cases. By providing law enforcement agencies with an intelligent and scalable tool for chat-based forensic analysis, this significantly improves digital evidence handling, accelerates crime resolution, and supports judicial proceedings with data-driven insights. It ultimately contributes to modern forensic science by bridging the gap between technology and criminal investigations, making digital communication a key asset in solving crimes efficiently.

Introduction

It enhances crime scene investigation by leveraging Natural Language Processing (NLP) to analyze chat logs and other textual data. It seeks to identify crucial evidence, detect patterns, and support law enforcement in solving crimes more efficiently. By automating the extraction of insights from digital communications, this can help forensic experts detect suspicious behavior, establish connections between suspects, and reconstruct events leading to a crime.

Chat Log Analysis: Processing and analyzing digital conversations (e.g., social media, emails, messaging apps) to identify threats, criminal intent, or hidden patterns.

NLP Techniques: Using AI-driven NLP models for entity recognition, sentiment analysis, topic modeling, and behavioral analysis to extract meaningful insights.

Crime Scene Correlation: Integrating chat log analysis with other forensic evidence to strengthen investigations.

Automation & Efficiency: Reducing manual effort by automating data extraction and interpretation, allowing faster decision-making.

Legal & Ethical Considerations: Ensuring compliance with data privacy laws and forensic standards while maintaining the admissibility of evidence in court.

Related Work

Several studies have proposed approaches for forensic analysis of mobile data. However, most of these approaches focus on specific aspects of mobile data. Our approach integrates the source of mobile data and uses machine learning and NLP techniques to analyse and identify potential leads and suspects.

Present technologies used in crime detection and prediction

The term "crime forecasting" denotes the fundamental procedure of anticipating crimes. Predicting a crime in advance requires the use of tools. Currently, police utilize technologies to help them do particular duties, such as hearing in on an accused's phone conversation or recording odd criminal behaviour using a body cam. Stakeouts are the most commonly used surveillance method by police personnel and collect information on all kinds of suspects.

Such illegal activities are seen during

Citation: Krishna Veni B, Muskan KS, Bingi PR, Jenne S, Gogula SG, Subramanyam SK. Forensic breakthrough for crime scene analysis via mobile data and NLP applications. GJEIIR. 2025;5(2):45.

stakeouts or patrols, guns, narcotics, and other suggestions discovered during home explorations. Additionally, surveillance techniques like facial recognition, license plate recognition, and body cameras are used. While each of these techniques is efficient, they all operate independently.

ML techniques used in crime prediction

In machine learning-based crime prediction. Predictions were made using crime statistics. This machine-learning-based criminal investigation technique includes data collection, categorization, pattern recognition, prediction, and visualization.

Methodology

Our approach consists of the following stages:

Upload mobile data

In this module investigators upload mobile data obtained from suspects or victims. This includes chat logs a secure interface ensures data integrity and confidentiality during the upload process.

Extract data

This module focuses on extracting relevant text-based communication from the uploaded files. It supports multiple file types and messaging platforms. The extracted data is structured into a consistent format suitable for Natural Language Processing (NLP), removing noise like metadata or irrelevant system messages.

Apply forensic activity

Here, forensic techniques are applied to authenticate and preserve the extracted data. This step strengthens the credibility of findings in legal contexts.

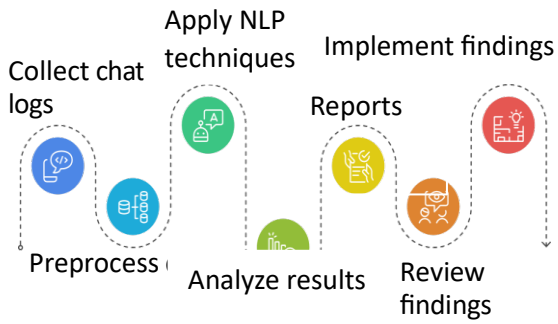
Filter data

The system filters the large volume of communication data based on user-defined parameters such as keywords, time ranges, user IDs, or emotional tone. NLP-based classification helps narrow down conversations that are contextually relevant to the investigation.

Analyze suspicious chats

This module uses advanced NLP methods such as sentiment analysis, named entity recognition (NER), to identify suspicious or criminal conversations. The system flags potential threats, identifies patterns of behaviour, and visualizes connections between individuals, helping investigators focus on critical leads.

Flow diagram:



Result & Discussion

Mobile Data Analysis

- Successfully extracted and analysed chat logs.
- Identified patterns and anomalies that can be used to identify potential leads and suspects.

NLP Analysis

- Applied NLP techniques, including sentiment analysis, entity recognition, and topic modelling, to analyse call logs from mobile.
- Successfully identified sentiment and emotional tone in text data, as well as entities and topics related to crime scene investigation.

Integration of Mobile Data Analysis and NLP Analysis

- Successfully integrated mobile data analysis and NLP analysis to provide a comprehensive view of crime scene investigation.
- Identified potential leads and suspects using the integrated approach.

The results of the project demonstrate the potential of using chat log and NLP applications for forensic analysis in crime scene investigation. The integrated approach provides a comprehensive view of the crime scene and can be used to identify potential leads and suspects.

Conclusion

The integration of Natural Language Processing (NLP) in forensic investigations, particularly for analyzing chat logs, represents a significant breakthrough in crime scene analysis. By leveraging NLP techniques such as sentiment analysis, named entity recognition, and topic modeling, investigators can extract crucial insights from digital conversations, uncover hidden relationships, and detect suspicious behavior in real time. It enhances traditional forensic methods by enabling automated, scalable, and efficient processing of vast amounts of text data, reducing human error and investigation time. Additionally, machine learning models can continuously improve, adapting to evolving criminal tactics and linguistic patterns.

Despite its potential, challenges such as data privacy, ethical considerations, and adversarial manipulations must be addressed to ensure the responsible and effective deployment of NLP-based forensic tools. With continued advancements, NLP will play a pivotal role in modern crime scene analysis, enabling law enforcement agencies to solve cases more effectively and uphold justice in the digital age.

References

1. Kim S, Joshi P, Kalsi PS, Taheri P (2018) Crime analysis through machine learning. In: Paper presented at the IEEE 9th annual information technology electronics and mobile communication conference. IEEE, Vancouver 1-3 November 2018.
2. Tabedzki C, Thirumalaiswamy A, van Vliet P (2018) Yo home to Bel-Air: predicting crime on the streets of Philadelphia. In: the University of Pennsylvania, CIS 520: machine learning.
3. Bharati A, Sarvanaguru RAK (2018) Crime prediction and analysis using machine learning. Int Res J Eng Technol 5(9):1037–1042.
4. Bandekar SR, Vijayalakshmi C (2020) Design and analysis of machine learning algorithms for the reduction

of crime rates in India.

5. Hossain S, Abtahee A, Kashem I, Hoque M, Sarker IH (2020) Crime prediction using Spatio-temporal data. arXiv preprint arXiv:2003.09322.
6. Jha P, Jha R, Sharma A (2019) Behaviour analysis and crime prediction using big data and machine learning. Int J Recent Technol Eng 8(1):461–468.
7. Mohammed Inayathulla and Karthikeyan C, "Image Caption Generation using
8. Deep Learning for Video Summarization Applications" International Journal of Advanced Computer Science and Applications (IJACSA), 15(1),2024. https://dx.doi.org/10.14569/IJACSA.2024.0150155.
9. Mohammed, I., Chalichalamala, S. (2015). TERA: A Test Effort Reduction Approach by Using Fault Prediction Models. In: Satapathy, S., Govardhan, A., Raju, K., Mandal, J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham. https://doi.org/10.1007/978-3-319-13728-5_25.
10. Inayathulla, M., Karthikeyan, C. (2022). Supervised Deep Learning Approach for
11. Generating Dynamic Summary of the Video. In: Suma, V., Baig, Z., Kolandapalayam Shanmugam, S., Lorenz, P. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19-1012-8_18.