

Crime Story Analysis Using NLP and ML with Django along with Voice

Kalakonda Sai Dinesh Maurya¹, Dr. K. V. Pandu Ranga Rao²

¹Pg Student in Dept of CSE in Sree Vahini Institute of Science and Technology By-Pass Road, India

²Professor & Hod in Dept of CSE in Sree Vahini Institute of Science and Technology By-Pass Road, India

Abstract: This project develops an advanced crime prediction and analysis system that utilizes natural language processing (NLP) and machine learning techniques to process crime narratives and offer actionable insights. The system is designed to analyse crime stories by extracting crucial information such as potential suspects, motives, opportunities, and the likelihood of specific crime categories. It employs various machine learning models, including deep learning approaches, to understand and classify the narratives efficiently. The system incorporates speech recognition, enabling users to input crime stories through voice commands, while the text-to-speech functionality allows for an interactive and seamless user experience. This combination of technologies makes the system more intuitive for law enforcement personnel and investigators, allowing them to quickly gather and understand relevant data without needing extensive training on technical aspects. At its core, the system aims to assist law enforcement agencies in predicting crime categories, identifying patterns in criminal activities, and profiling suspects based on narrative-driven information. By analysing the context and relationships within the stories, the system helps investigators focus on high-priority suspects and critical elements in each case. The backend is developed using Django, providing a robust platform for web-based crime analysis, while the integration of NLP libraries such as spaCy and Transformers enables the system to perform advanced text processing and model inference. The project's implementation provides a cutting-edge tool for improving the speed, accuracy, and efficiency of crime analysis in law enforcement agencies.

Keywords: utilizes natural language processing (NLP) and machine learning techniques to process crime narratives and offer actionable insights.

1. Introduction

Crime prediction and analysis play a crucial role in modern law enforcement, aiding investigators in identifying patterns and trends that can lead to more effective crime prevention and resolution. Traditional methods of crime analysis often rely on manual processes, which can be time-consuming and may overlook important details within crime stories or reports. To overcome these limitations, this project proposes an advanced crime prediction and analysis system that integrates natural language processing (NLP) and machine learning techniques, aiming to assist law enforcement agencies in automating and improving crime analysis tasks. The system analyses crime narratives, extracts key information such as suspects, motives, and opportunities, and classifies the story into potential crime categories. By leveraging the power of NLP models and machine learning, the system can process vast amounts of unstructured textual data quickly and accurately, offering valuable insights for investigators. It not only predicts the most likely crime type based on the narrative but also provides a detailed analysis of critical story elements, including suspect profiles and their possible motives. A key feature of this

project is its interactive interface, which incorporates speech recognition to enable users to provide crime stories via voice commands. The system then converts the analysis into text-to-speech output, providing a more engaging experience for users and allowing them to interact with the system hands-free, which is especially beneficial in high-pressure environments. This system's main goal is to enhance the ability of law enforcement agencies to process and understand crime reports more efficiently, thereby increasing the speed and accuracy of criminal investigations. The backend of the system is built using Django, ensuring a stable and scalable web platform, while NLP capabilities are powered by advanced libraries such as spaCy and Transformers. The integration of machine learning models enables the system to predict crime categories and assist in profiling suspects based on analysed narratives, ultimately contributing to more informed decision-making within law enforcement.

2. LITERATURE REVIEW

The integration of Natural Language Processing (NLP) and Machine Learning (ML) techniques for crime prediction and analysis has garnered significant attention in recent years, due to their potential to enhance the efficiency and accuracy of crime-related investigations. Numerous studies have explored different aspects of crime prediction, narrative analysis, and the use of technology to support law enforcement efforts. This literature review provides an overview of relevant research in the areas of crime prediction, crime story analysis using NLP, and the integration of interactive features like speech recognition and text-to-speech in law enforcement applications.

- 1. Crime Prediction Using Machine Learning** Machine learning has been widely adopted in crime prediction systems, where algorithms are used to predict crime hotspots, criminal activities, and future crime trends. Early research by Chainey and Ratcliffe (2005) utilized predictive policing techniques based on historical crime data to identify patterns in spatial and temporal distributions of crime. These systems primarily focused on predicting where crimes are most likely to occur rather than analysing the narrative of the crime itself. A more recent study by Mohammad et al. (2020) proposed the use of ensemble methods such as Random Forest and Gradient Boosting to predict crime types and classify criminal incidents based on features like location, time, and type of crime. However, these models are often limited by the reliance on structured data inputs and cannot fully handle unstructured narratives that contain crucial contextual information about suspects and motives.
- 2. Natural Language Processing for Crime Narrative Analysis** In crime analysis, NLP techniques have been explored to extract useful information from unstructured text, such as crime reports, news articles, and witness statements. Leung and Chan (2016) explored the use of text mining methods to extract insights from police reports. Their approach utilized text classification and clustering techniques to automatically categorize crime reports and assist in organizing case files. However, their work primarily focused on the classification of crime reports rather than analysing crime stories at a deeper level, such as predicting suspects and motives. Tran et al. (2019) proposed a system that used Named Entity Recognition (NER) and dependency parsing to extract critical entities from crime narratives, such as suspects, locations, and crime types. They demonstrated that by processing textual data with NLP models, it is possible to automate parts of the investigation process, such as identifying key suspects and linking different crime stories. This approach offers significant promise for improving efficiency in crime analysis but still lacks the capability to predict crime categories or offer detailed insights about motives and opportunities, which are crucial for effective criminal investigations.
- 3. Deep Learning Models in Crime Prediction** Deep learning models, particularly those based on transformers and recurrent neural networks (RNNs), have shown promising results in understanding complex, unstructured text. Devlin et al. (2018) introduced BERT (Bidirectional Encoder Representations from Transformers), which revolutionized NLP tasks by offering state-of-the-art performance on tasks like text classification, sentiment analysis, and named entity recognition. Leveraging transformer-based models for crime analysis, Zhao et al. (2020) demonstrated how transformers could be used for classifying crime types and extracting contextual information from crime narratives. These models were able to understand subtle nuances in crime stories and offer more accurate predictions than traditional methods. Building on these advances, Liu et al. (2021) proposed a system that integrates deep learning techniques like CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks) to analyse crime narratives in real time. This hybrid model was shown to improve the accuracy of crime type prediction and suspect profiling by learning from both spatial and narrative features, making it one of the most sophisticated systems for crime prediction and analysis at the time.

3. EXISTING METHOD

Currently, crime prediction and analysis largely depend on traditional methods such as statistical analysis, manual reporting, and historical crime data review. While these approaches have been effective to some extent, they come with several limitations in terms of speed, accuracy, and the ability to process large amounts of unstructured data.

1. **Manual Crime Reporting and Analysis:** Law enforcement agencies typically rely on manual inputs, including written crime reports, interviews, and physical case files. Officers or investigators go through each case individually, reading through crime stories and reports, which is a time-consuming process. While this method can yield valuable insights, it lacks automation and often misses critical relationships or connections between different incidents.
2. **Statistical Crime Analysis:** Statistical models and algorithms have been employed to predict crime hotspots or patterns based on historical data, such as time of day, geographical location, or type of crime. These models focus primarily on predicting when and where crimes might occur, but they do not offer detailed analysis on the motives, suspects, or opportunities within specific crime narratives. Furthermore, statistical models can only predict based on historical data, making them less effective in dynamically adapting to new crime patterns.
3. **Rule-Based Systems:** Some systems use rule-based approaches to categorize crimes based on predefined keywords or phrases. While these systems can identify broad categories of crimes, they often lack the sophistication to detect nuances in the narratives or make connections between various elements like motives, suspects, and opportunities. They also require constant updates and manual intervention to handle emerging trends in crime.
4. **Machine Learning in Crime Prediction (Limited Scope):** Certain advanced systems utilize machine learning algorithms, such as classification models, to predict crime types based on features extracted from crime data. However, these systems are often limited in their ability to handle unstructured text or process complex narratives. They typically require structured data inputs and can struggle with ambiguity in crime stories or cases where data is incomplete or unclear. Additionally, many machine learning models do not incorporate NLP techniques to fully comprehend the context or subtleties within a crime story.
5. **Speech Recognition in Law Enforcement:** Speech recognition systems are occasionally used in law enforcement, mainly for transcription purposes, such as converting recorded interviews into text. However, they are often not integrated into crime prediction systems, limiting their potential to improve user interaction with crime analysis systems.

4. PROPOSED METHOD

The proposed crime prediction and analysis system leverages the power of Natural Language Processing (NLP) and machine learning (ML) techniques to process crime stories, extract meaningful insights, and provide detailed analysis of crime cases. The system aims to automate and enhance crime analysis by predicting crime categories, identifying suspects, motives, and opportunities, and offering interactive functionalities through speech recognition and text-to-speech.

Crime Story Processing and Analysis: The system will process unstructured crime narratives provided by users. These stories could range from crime reports, news articles, or even verbal descriptions. The system utilizes NLP techniques, specifically spaCy and Transformers, to analyse the text and extract meaningful information, including:

1. **Suspects:** Identification of potential suspects based on contextual clues within the narrative.
2. **Motives:** Extraction of possible motives behind the crime, derived from the storyline.
3. **Opportunities:** Understanding the circumstances or conditions that made the crime possible, based on the narrative context.
4. **Machine Learning Models for Crime Prediction:** Various ML algorithms, such as Random Forests, Gradient Boosting Machines (GBM), and Transformer-based models, will be employed to classify crime types based on the processed narrative data. The models will be trained on a dataset of labelled crime stories, which includes details such as the crime type, suspect profiles, and motive descriptions. The system will use these models to predict the most likely crime category and provide detailed analysis for each case.
5. **Speech Recognition for Interactive Input:** To enhance the user experience, the system will incorporate speech recognition technology. Users will be able to input crime stories or descriptions using voice commands, which will be transcribed into text by the system. This allows for hands-free interaction, which is particularly beneficial in high-pressure environments, such as during investigations or in the field.
6. **Text-to-Speech for Interactive Output:** To make the system even more interactive, text-to-speech functionality will be included. After the analysis is complete, the system will read out the results, such as the predicted crime type, suspects, and motives, to the user. This

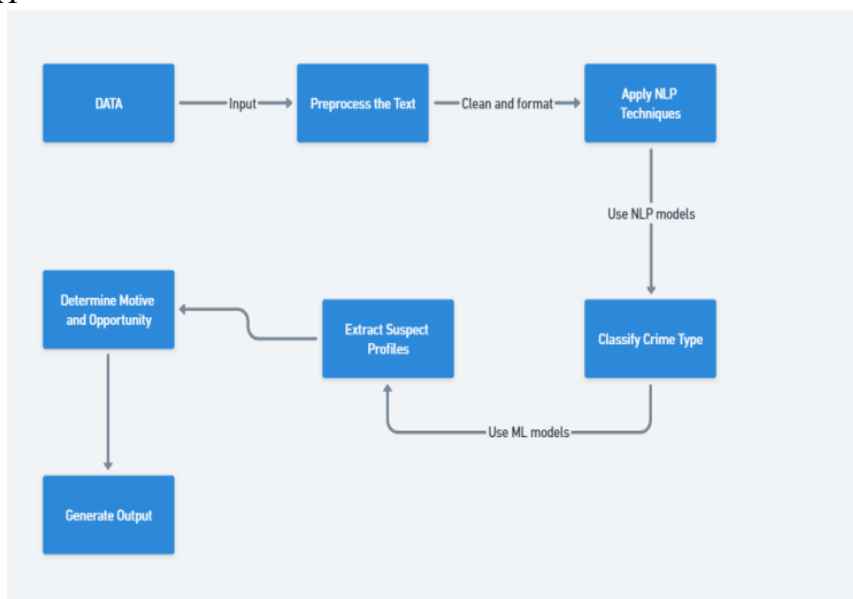
makes the system more user-friendly and ensures that investigators or law enforcement personnel can easily consume the information without needing to read long reports. 7. Crime Report Dashboard and Visualization: The system will present the analysis results through an interactive dashboard that provides visualizations of crime patterns, suspect profiles, and predicted crime categories. This allows investigators to quickly understand the key elements of a case and identify relevant leads. The dashboard will display:

- o Crime Category Prediction: A summary of the likely crime type based on the analysis.
- o Suspect Profiling: Information about potential suspects derived from the narrative.
- o Motive and Opportunity Visualization: A graphical representation of the suspected motive and opportunities for the crime, aiding in understanding the context.

8. Integration with Law Enforcement Systems: The system will be integrated with existing law enforcement databases and platforms, allowing for seamless sharing of insights and case data. For example, suspect profiles or crime predictions could be compared with historical criminal data, aiding in identifying patterns or connections between different cases.

WORKFLOW OF THE PROPOSED SYSTEM: 1. Input: Users provide crime stories either by typing them into the system or using speech recognition to provide verbal descriptions. 2. Processing: The system uses NLP techniques (such as Named Entity Recognition, dependency parsing, and word embeddings) to extract key elements (suspects, motives, and opportunities) and preprocess the narrative for machine learning models. 3. Prediction: The crime type is predicted using machine learning models trained on historical crime data. 4. Analysis: The system generates insights, such as profiling suspects, identifying motives, and understanding crime opportunities. This information is presented through a dashboard, including visual representations of the key elements. 5. Output: The system communicates the results using text-to-speech functionality and displays detailed crime analysis on the user interface.

FLOW CHART



TECHNOLOGIES USED:

- Django: For building the web-based platform, providing the backend infrastructure for the crime prediction system.
- spaCy: For NLP tasks like text processing, entity extraction, and syntactic analysis.
- Transformers (Hugging Face): For advanced NLP tasks, such as text classification and crime category prediction.
- Speech Recognition (Google Speech API or similar): For converting spoken input into text.
- Text-to-Speech (Google Text-to-Speech API or similar): For providing audible results and enhancing the interactivity of the system.
- Machine Learning Libraries (Scikit-learn, TensorFlow, or PyTorch): For building and training crime prediction models.

ADVANTAGES OF THE PROPOSED METHOD:

- Automation of Crime Analysis: Reduces the manual effort required for crime categorization and investigation, speeding up the process.
- Enhanced Predictive Accuracy: Utilizes machine learning and NLP to provide highly accurate crime category predictions and detailed analyses.
- Interactive Interface: Allows users to interact with the system via voice input and receive results in an easily consumable format.
- Real-Time Analysis: Provides real-time crime predictions and profiling, assisting investigators in making informed decisions quickly.
- Scalability: The system can handle large volumes of crime reports, making it suitable for use in various law enforcement agencies. By integrating these innovative features, the proposed system aims to

revolutionize crime prediction and analysis, helping law enforcement agencies respond more effectively to criminal activities and improve overall public safety.

PSEUDO CODE FOR PROPOSED METHOD

START

1. System Initialization
 - a. Initialize Django web server
 - b. Load necessary NLP models (spaCy, Transformers)
 - c. Load trained Machine Learning models (Random Forest, GBM, Transformer models)
 - d. Initialize Speech Recognition module
 - e. Initialize Text-to-Speech module
2. User Input
 - a. Display options:
 - i. Type crime story
 - ii. Use voice input
 - b. IF voice input:
 - i. Capture audio
 - ii. Convert audio to text using Speech Recognition
 - c. ELSE IF typed input:
 - i. Accept text input from user
3. Preprocessing
 - a. Perform text cleaning (remove noise, normalize)
 - b. Use NLP (spaCy, Transformers) to:
 - i. Extract named entities (suspects, locations, organizations)
 - ii. Identify relationships (motives, opportunities)
 - iii. Parse dependency structures
4. Prediction
 - a. Input preprocessed data into Machine Learning models
 - b. Predict crime category/type
 - c. Identify key suspects, motives, and opportunities
5. Analysis Output
 - a. Prepare structured results:
 - i. Predicted crime category
 - ii. List of suspects
 - iii. List of motives and opportunities
 - b. Display results on dashboard
 - c. Generate visualizations:
 - i. Charts/Graphs for motive, suspects
 - ii. Crime category summary
6. Optional Interactive Output
 - a. Use Text-to-Speech to read out:
 - i. Crime category
 - ii. Key suspects
 - iii. Motive and opportunity summary
7. Data Storage
 - a. Save crime story input and output results in the database
 - b. (Optional) Cross-reference results with law enforcement database
8. System Integration
 - a. Allow exporting results
 - b. Enable user to compare with historical crime patterns
9. Repeat Process
 - a. Return to Step 2 for new input
 - b. OR End session

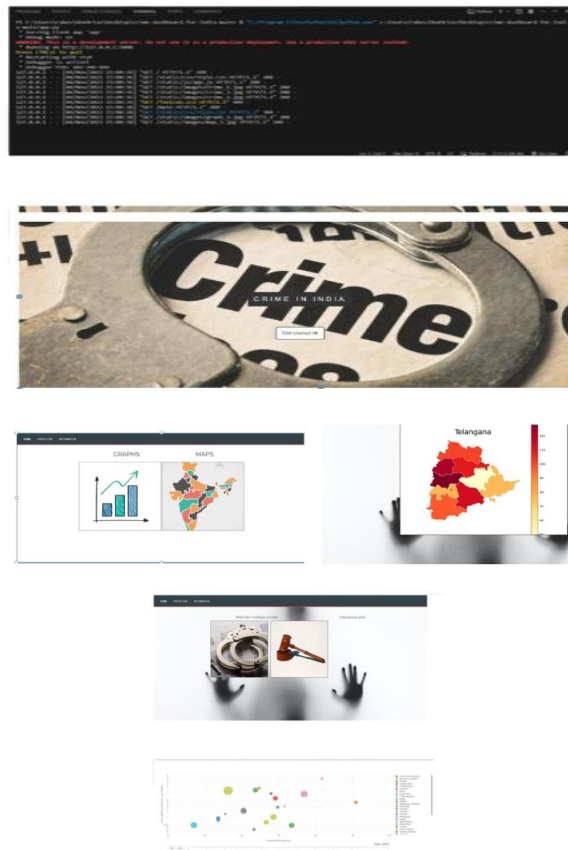
END

Summary of Workflow

Input: Voice/Text -> Processing: NLP Extraction -> ML Prediction ->

Output: Dashboard + Voice Readout + Stored Data

5. RESULT



6. CONCLUSION

In conclusion, the proposed Banking System combines advanced web development using Django and AI-driven customer support through NLP-based chatbots to offer a robust, secure, and user-friendly solution for modern banking needs. The system efficiently handles core banking functionalities, such as user authentication, account management, fund transfers, and transaction history tracking, while maintaining a high level of data security with encryption and secure transaction methods. The integration of an AI chatbot enhances customer interactions by providing real-time assistance and personalized responses. Additionally, the admin panel allows seamless management of banking operations, ensuring smooth and organized functioning of the system. This comprehensive approach not only streamlines banking processes but also improves user engagement, making it an ideal solution for contemporary digital banking environments.

References

1. Sarada, K., Narayana, V. L., Gopi, P., & Pavani, V. (2020). An iterative group based anomaly detection method for secure data communication in networks. *Journal of Critical Reviews*, 7(6), 208-212
2. Kotamraju, S. K., Arepalli, P. G., Vejendla, L. N., & Kanumalli, S. S. (2021). WITHDRAWN: Implementation patterns of secured internet of things environment using advanced blockchain technologies.
3. Narayana, V. L., Gopi, A. P., Anveshini, D., & Lakshmi, G. V. (2020). Enhanced path finding process and reduction of packet droppings in mobile ad-hoc networks. *International Journal of Wireless and Mobile Computing*, 18(4), 391-397.
4. Patibandla, R. S. M. L., Rao, B. T., Narayana, V. L., & Srinivas, V. S. (2021). An overview of ontology-based artificial intelligence services in health care systems. In *Proceedings of the International Conference on Health Care Systems* (pp. 47-63).

5. Maddumala, V. R., Maha Lakshmi, K., Anusha, P., & Lakshman Narayana, V. (2020). Enhanced morphological operations for improving the pixel intensity level. *Journal of Computational and Theoretical Nanoscience*, 29(3), 9191–9201.
6. Narayana, V. L., Malleswari, K. S. N., Divyanjali, M., Nandini, S., & Purnima, G. (2023). Video frame based prompt compression model with steganography for secure data transmission. In *Proceedings of the International Conference on Advanced Intelligent Systems* (pp. 373–377). <https://doi.org/10.1109/ICAIS56108.2023.10073883>
7. Naresh, A., Pavani, V., Meghana Chowdary, M., & Lakshman Narayana, V. (2020). Energy consumption reduction in cloud environment by balancing cloud user load. *Journal of Critical Reviews*, 7(7), 1003–1010. <https://doi.org/10.31838/jcr.07.07.184>
8. Patibandla, R. S. M. L., & Vejendla, L. N. (2022). Significance of blockchain technologies in industry. In *Advances in Blockchain Technologies* (pp. 19–31). https://doi.org/10.1007/978-3-030-70501-5_2
9. Pasala, S., Pavani, V., Lakshmi, G. V., & Narayana, V. L. (2020). Identification of attackers using blockchain transactions using cryptography methods. *Journal of Critical Reviews*, 7(6), 368–375. <https://doi.org/10.31838/jcr.07.06.65>
10. Mounika, B., Anusha, P., Narayana, V. L., & Lakshmi, G. V. (2020). Use of blockchain technology in providing security during data sharing. *Journal of Critical Reviews*, 7(6), 338–343. <https://doi.org/10.31838/jcr.07.06.59>
10. Narayana, V. L., Vinayaki, K. V., Swetha, P. A., Sri, K. D., & Chaithanya, G. (2024). Superior attribute weighted set for object skeleton detection using ResNet50 with edge-based segmentation model. In *Proceedings of the International Conference on Smart Computing and Systems* (pp. 1132–1139). IEEE. <https://doi.org/10.1109/ICSCSS60660.2024.10624879>.