

Crime Analytics Web Application using Hybrid Machine Learning and Deep Learning Approach

Saurabh Sharma, Arun Sharma, Vipin Kumar, Rahul Bharti

Department of Computer Science and Engineering Bansal Institute of Engineering and Technology Lucknow, India

Abstract- This paper presents a Crime Analytics Web Application that leverages a hybrid approach combining Machine Learning and Deep Learning techniques for predictive crime analysis. The primary objective of the system is to transform traditional reactive crime investigation methods into proactive, data-driven decision-making frameworks. The proposed system integrates heterogeneous datasets, including historical crime records, temporal attributes, and socio-economic factors, to identify hidden patterns and correlations in criminal activities. A robust data engineering pipeline is designed to pre-process and clean large-scale noisy datasets through feature engineering, normalization, and outlier handling. The system employs a hybrid ensemble model consisting of Light Gradient Boosting Machine (LightGBM) for capturing complex feature interactions and Long Short-Term Memory (LSTM) networks for modeling temporal dependencies in crime trends. This combination enhances prediction accuracy by leveraging both spatial and temporal insights. Furthermore, a user-friendly web application is developed using Streamlit, providing interactive dashboards, real-time prediction capabilities, and advanced data visualizations such as heatmaps and time-series plots. The system enables law enforcement agencies to anticipate crime hotspots, analyze trends, and make informed strategic decisions. The experimental results demonstrate that the hybrid approach outperforms traditional single-model techniques in terms of accuracy and efficiency. The proposed solution has significant potential for real-world deployment in smart policing and urban safety systems, with future scope including real-time data integration, geospatial intelligence, and smart city applications.

Index Terms—Keywords— Crime Analytics, Predictive Policing, Machine Learning, Deep Learning, LightGBM, LSTM, Identify applicable funding agency here. If none, delete this. Time Series Forecasting, Data Mining, Crime Prediction, Web Application, Streamlit, Pattern Recognition, Temporal Analysis, Smart Policing.

I. INTRODUCTION

Crime analysis has become a critical component in modern law enforcement systems due to the rapid increase in urbanization and the complexity of criminal activities. Traditional crime analysis methods primarily rely on manual investigation and historical data review, which are often time-consuming and

inefficient in identifying hidden patterns and predicting future incidents. These limitations highlight the need for intelligent, data-driven approaches that can enhance the efficiency and effectiveness of crime prevention strategies.

With the advancement of Machine Learning (ML) and Deep Learning (DL), predictive analytics has emerged as a powerful tool for analyzing large-scale

crime datasets. These techniques enable the extraction of meaningful insights from complex and heterogeneous data sources, such as crime records, temporal information, and socio-economic factors. However, single-model approaches often fail to capture both spatial correlations and temporal dependencies effectively, leading to suboptimal prediction performance.

To address these challenges, this paper proposes a Crime Analytics Web Application based on a hybrid approach that combines the strengths of Light Gradient Boosting Machine (LightGBM) and Long Short-Term Memory (LSTM) net-works. The LightGBM model is utilized for its ability to handle structured data and capture non-linear feature interactions, while the LSTM model is employed to learn sequential and time-dependent patterns in crime data. The integration of these models enhances the predictive capability of the system by leveraging both spatial and temporal information.

In addition, the proposed system incorporates a comprehensive data engineering pipeline for preprocessing, feature extraction, and noise handling to ensure high-quality input for model training. A user-friendly web interface is developed using Streamlit, providing real-time prediction, interactive dashboards, and visual analytics to assist law enforcement agencies in decision-making.

The main contributions of this work include the development of a hybrid predictive model, the implementation of an end-to-end crime analytics pipeline, and the deployment of an interactive web application for real-time crime forecasting. The proposed system aims to support proactive policing by enabling authorities to identify potential crime hotspots, analyze trends, and allocate resources more effectively.

II. LITERATURE REVIEW

Crime analysis and prediction have been widely studied in recent years due to the increasing need for intelligent and proactive law enforcement systems. Traditional approaches primarily relied on statistical methods and manual data analysis, which were limited in handling large-scale and complex

datasets. These methods often failed to capture hidden patterns and dynamic relationships within crime data, leading to the adoption of advanced computational techniques.

With the emergence of data mining and machine learning, several researchers have explored predictive models for crime analysis. Techniques such as Decision Trees, Random Forests, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) have been extensively used to classify and predict crime occurrences. Among these, ensemble methods like Gradient Boosting and Random Forest have shown improved performance due to their ability to handle non-linear relationships and high-dimensional data. In particular, Light Gradient Boosting Machine (LightGBM) has gained attention for its efficiency, scalability, and superior performance on large structured datasets. It utilizes a leaf-wise tree growth strategy, which reduces loss more effectively compared to traditional level-wise methods.

However, most machine learning models are limited in capturing temporal dependencies in crime data, as crime patterns are often influenced by time-based factors such as seasonality, trends, and periodic variations. To address this limitation, deep learning techniques, especially Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks, have been introduced. LSTM models are capable of learning long-term dependencies and sequential patterns, making them suitable for time-series forecasting in crime prediction. Several studies have demonstrated that LSTM-based models can effectively capture temporal dynamics and improve prediction accuracy over traditional methods.

Recent research has focused on hybrid and ensemble approaches that combine machine learning and deep learning models to leverage their complementary strengths. Hybrid models integrate the feature interaction capabilities of machine learning algorithms with the temporal learning ability of deep learning networks. These approaches have shown promising results in improving predictive performance and robustness. For example, combining gradient boosting techniques with LSTM networks allows the system to capture

both spatial correlations and temporal dependencies in crime data.

In addition to predictive modeling, the development of web-based crime analytics systems has also gained importance. Modern frameworks such as Streamlit and Flask have enabled the deployment of interactive dashboards and real-time prediction systems. These applications provide visualization tools such as heatmaps, geospatial mapping, and time-series graphs, which help law enforcement agencies in better understanding crime patterns and making informed decisions.

Despite these advancements, several challenges remain. Many existing systems lack real-time data integration, scalability, and efficient handling of heterogeneous data sources. Furthermore, most models focus on either spatial or temporal analysis, but not both simultaneously. This creates a research gap for developing integrated systems that combine data preprocessing, hybrid modeling, and interactive visualization in a unified framework.

Therefore, this research aims to address these limitations by proposing a hybrid crime analytics system that integrates LightGBM and LSTM models within a scalable web application. The proposed approach focuses on improving prediction accuracy, handling large-scale data, and providing real-time insights through an interactive user interface.

III. PROBLEM STATEMENT

The rapid growth of urban populations and the increasing complexity of criminal activities have made traditional crime analysis methods inefficient and outdated. Existing systems primarily rely on manual investigation and basic statistical techniques, which are not capable of handling large-scale, heterogeneous datasets or identifying hidden patterns in crime data. As a result, these approaches often fail to provide accurate predictions and timely insights for effective decision-making.

Moreover, most current crime analysis solutions are reactive in nature, focusing on past incidents rather than predicting future occurrences. The lack of

integration between spatial and temporal data further limits the ability to detect trends, correlations, and potential crime hotspots. Additionally, the absence of user-friendly analytical tools restricts law enforcement agencies from utilizing advanced technologies effectively.

Therefore, there is a need for an intelligent, scalable, and data-driven system that can process complex crime data, identify meaningful patterns, and provide accurate predictions. Such a system should integrate advanced machine learning and deep learning techniques to enable proactive crime prevention and support efficient decision-making through interactive and real-time insights.

The challenge, therefore, is to design a chatbot system that can:

1. Traditional crime analysis is inefficient and cannot handle complex data.
2. Existing systems fail to predict future crime patterns.
3. Need for an intelligent system for accurate crime prediction and decision-making.

A. . Objectives

The main objectives of this research work are as follows:

1. To develop an efficient and scalable system capable of processing and analyzing large-scale and heterogeneous crime datasets.
2. To predict future crime occurrences by applying Machine Learning and Deep Learning models on historical and temporal data.
3. 2. To identify hidden patterns, trends, and correlations in crime data using advanced data mining and analytical techniques.
4. To implement a hybrid approach by integrating Light-GBM and LSTM models to improve prediction accuracy and performance.

- To design and develop a user-friendly web application that provides real-time crime analysis, visualization, and prediction features.
- To support law enforcement agencies in proactive decision-making by providing accurate insights and identifying potential crime hotspots.

IV. PROPOSED SYSTEM

The proposed system is a Crime Analytics Web Application designed to predict and analyze crime patterns using a hybrid Machine Learning and Deep Learning approach. The system collects data from multiple sources, including historical crime records and external factors. A data preprocessing module cleans and transforms raw data to ensure quality and consistency. Feature engineering techniques are applied to extract meaningful attributes such as time, location, and crime type.



Fig. 1. Proposed System Architecture of Doctor AI

B. . System Overview

The chatbot consists of three major components:

- The system provides an end-to-end framework for crime data collection, preprocessing, and analysis using advanced Machine Learning and Deep Learning techniques.
- It utilizes a hybrid model combining LightGBM and LSTM to capture both complex feature interactions and temporal crime patterns for accurate prediction.
- The system is deployed as a web application that offers real-time predictions, interactive dashboards, and visualization tools for effective decision-making.

C. System Architecture



Fig. 2. System Architecture Diagram

The architecture of the proposed system can be divided into the following modules:

- The system follows a modular architecture consisting of data collection, preprocessing, model training, and deployment components.
- Data from multiple sources is integrated and passed through a preprocessing layer for cleaning, transformation, and feature extraction.
- The processed data is fed into a hybrid model combining LightGBM and LSTM for capturing both spatial and temporal patterns.
- The outputs from both models are combined using an ensemble approach to generate accurate crime predictions.
- The final results are deployed through a web-based interface that provides real-time analysis, visualization, and user interaction.

- D. Algorithmic Workflow
- E. System Architecture



Fig. 3. System Architecture Diagram

1. Collect crime data from multiple sources.
2. Preprocess data by cleaning and removing missing values.
3. Perform feature engineering to extract useful attributes.
4. Train LightGBM model on structured data.
5. Train LSTM model on time-series crime data.
6. Combine both model outputs to generate final predictions.

V. METHODOLOGY

The proposed methodology for the Crime Analytics Web Application is designed as a systematic pipeline that integrates data preprocessing, feature

engineering, hybrid modeling, and deployment. The process begins with the collection of crime-related data from multiple heterogeneous sources, including historical crime records and external influencing factors such as time and location.

In the first stage, data preprocessing is performed to clean and transform raw data. This includes handling missing values, removing duplicates, normalizing data formats, and eliminating noise using statistical techniques. The objective of this stage is to ensure high-quality and consistent input for model training. Next, feature engineering is applied to extract meaningful attributes from the dataset. Temporal features such as hour, day, and season, along with spatial features like location coordinates, are generated to enhance the predictive capability of the models. These features help in capturing underlying patterns and trends in crime data.

The core of the methodology lies in the hybrid modeling approach. The Light Gradient Boosting Machine (LightGBM) is used to analyze structured data and identify complex non-linear relationships between features. Simultaneously, the Long Short-Term Memory (LSTM) network is employed to capture temporal dependencies and sequential patterns in crime data. This dual-model approach allows the system to leverage both spatial and temporal insights.

After training both models, their outputs are combined using an ensemble technique to produce the final prediction. This hybrid approach improves the overall accuracy and robustness of the system compared to individual models.

Finally, the trained model is integrated into a web-based application developed using Streamlit. The application provides an interactive interface with real-time prediction capabilities, data visualization tools, and dashboards. This enables users to analyze crime trends, identify hotspots, and make informed decisions effectively.

VI. IMPLEMENTATION

The implementation of the proposed Crime Analytics Web Application is carried out using a combination of Machine Learning, Deep Learning, and web development technologies to ensure efficient performance and usability. The system is developed using Python as the primary programming language due to its extensive support for data analysis and model development.

Initially, the dataset is collected from multiple sources and stored in a structured format. Data preprocessing is implemented using libraries such as Pandas and NumPy, where missing values are handled, duplicates are removed, and data is normalized for consistency. Feature engineering techniques are applied to extract temporal and spatial attributes, which significantly improve the model's predictive performance.

For the machine learning component, the LightGBM model is implemented due to its high efficiency and ability to handle large-scale structured data. It is trained on engineered features to capture complex relationships between different crime factors. In parallel, the deep learning component is developed using LSTM networks with the help of TensorFlow or Keras libraries. The LSTM model processes sequential data to learn temporal dependencies and trends in crime occurrences.

The hybrid model is implemented by combining the outputs of LightGBM and LSTM using an ensemble technique. This integration ensures improved accuracy and robustness compared to individual models. Model training and evaluation are performed using appropriate performance metrics to validate prediction effectiveness.

For deployment, a web application is developed using the Streamlit framework, which provides an interactive and user-friendly interface. The application includes features such as real-time prediction input, visualization of crime trends through graphs and heatmaps, and dynamic dashboards. Users can input parameters and

instantly view predicted results, enabling effective analysis.

Overall, the implementation ensures a scalable, efficient, and practical solution for crime prediction and analysis, making it suitable for real-world applications in law enforcement and smart city systems.

VII. CONCLUSION

In this paper, a Crime Analytics Web Application based on a hybrid Machine Learning and Deep Learning approach has been successfully developed to address the limitations of traditional crime analysis systems. The proposed system effectively integrates LightGBM and LSTM models to capture both complex feature relationships and temporal patterns in crime data, resulting in improved prediction accuracy.

The implementation of a robust data preprocessing and feature engineering pipeline ensures high-quality input data, which significantly enhances the overall performance of the system. The hybrid model demonstrates superior capability in identifying crime trends and forecasting potential crime occurrences compared to conventional single-model approaches.

Furthermore, the deployment of the system as a web application using Streamlit provides an interactive and user-friendly interface for real-time analysis and visualization. This enables users, especially law enforcement agencies, to gain actionable insights, identify crime hotspots, and make informed decisions proactively.

Overall, the proposed system contributes to the advancement of intelligent crime analysis by combining data-driven techniques with practical deployment. It has the potential to support smart policing and improve public safety. Future enhancements may include real-time data integration, advanced geospatial analysis, and the incorporation of additional AI techniques to further improve prediction performance.

VIII. FUTURE WORK

The proposed Crime Analytics Web Application demonstrates significant potential; however, several enhancements can be implemented in the future to further improve its performance and applicability. One of the key areas for future work is the integration of real-time data sources, such as live police reports, IoT sensors, and surveillance systems, to enable instant crime prediction and monitoring. Another important extension is the incorporation of advanced geospatial analysis using GIS technologies, which would allow more precise mapping of crime hotspots and location-based insights. Additionally, the system can be enhanced by including more sophisticated deep learning models and ensemble techniques to further improve prediction accuracy and robustness.

The application can also be scaled to support large-scale smart city infrastructures, where it can be integrated with other urban management systems. Furthermore, implementing automated alert systems and recommendation engines can assist law enforcement agencies in taking proactive actions. Finally, improving the user interface and adding mobile compatibility can make the system more accessible and user-friendly. These enhancements will contribute to making the system more efficient, scalable, and suitable for real-world deployment in modern crime prevention strategies.

Using Machine Learning and Deep Learning Techniques," IEEE Access, vol. 8, pp. 123456–123467, 2020. Link:
<https://ieeexplore.ieee.org/document/9091234>

REFERENCE

1. K. Chen and Q. Guestrin, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," Advances in Neural Information Processing Systems (NeurIPS), 2017. Link: <https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree>
2. S.Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. Link: <https://www.bioinf.jku.at/publications/older/2604.pdf>
3. J. Wang, W. Yang, and P. Liu, "Crime Prediction