ijamrsd.com JAMRSD ISSN: 2583-6404
Journal of Advanced Multidisciplinary Research Studies and Development
Peer Reviewed Referred Journal Bi-monthly Open Access

LSTM-BASED DEEP LEARNING FRAMEWORK FOR REGIONAL CRIME RATE PREDICTION

Bhausaheb Markande[1], Tanmay Garje[2], Jaydip Gadekar[3]
[1,2,3] Department of Computer Engineering, Sharadchandra Pawar College of Engineering, Pune, India.
Email: markandebhausaheb23@gmail.com[1], tanmaygarje7@gmail.com[2], gadekarjaydip7@gmail.com[3]

Abstract
Crime prediction has become an important application of data analytics and machine learning for improving public safety. With the rapid growth of urban populations, analyzing crime patterns and predicting future crime trends has become essential for effective law enforcement planning. Traditional crime analysis methods mainly rely on manual investigation and statistical techniques, which are often limited in identifying complex patterns in large datasets. This project presents a crime rate prediction system that uses a Long Short-Term Memory (LSTM) model to analyze historical crime data and forecast future crime occurrences across different regions. The system processes crime datasets through data preprocessing, feature extraction, and model training to identify temporal patterns in criminal activities. LSTM networks are particularly suitable for this task because they can learn long-term dependencies in time-series data. The proposed model aims to improve prediction accuracy compared to traditional machine learning techniques. By predicting potential crime rates in advance, the system can help law enforcement agencies allocate resources more efficiently and take preventive measures in high-risk areas. This approach contributes to data-driven decision-making and supports the development of safer communities.
Keywords: Crime Prediction, Machine Learning, LSTM, Time Series Analysis, Data Analytics.

► *Corresponding Author: Bhausaheb Markande*

I. Introduction

Crime is one of the major social challenges faced by modern societies. Rapid urbanization, population growth, and economic changes have led to an increase in various types of criminal activities. As cities continue to expand, maintaining public safety has become more complex for law enforcement agencies. Understanding crime patterns and predicting future crime occurrences are important steps toward preventing crime and improving community safety.

Traditionally, crime analysis has relied on historical records and manual statistical methods to identify trends and patterns. While these approaches provide useful insights, they often struggle to handle large volumes of data and complex relationships between different factors influencing crime. With the growth of digital data collection systems, large crime datasets are now available, which makes it possible to apply advanced data analytics and machine learning techniques for crime prediction.

Machine learning algorithms have shown significant potential in identifying hidden patterns in data and making accurate predictions. Techniques such as K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Random Forest, and Neural Networks have been widely used for predictive analysis. However, crime data often contains temporal dependencies, meaning that past crime events can influence future occurrences. Traditional machine learning models may not effectively capture these sequential patterns.

To address this challenge, deep learning models such as Long Short-Term Memory (LSTM) networks have been introduced. LSTM is a type of Recurrent Neural Network (RNN) designed to handle sequential and time-series data. It can learn long-term dependencies and patterns within historical crime records, making it suitable for predicting future crime trends across different regions.

In this project, a crime rate prediction system is developed using the LSTM model. The system analyzes historical crime data, performs preprocessing and feature extraction, and trains a deep learning model to forecast future crime rates. The predicted results can help authorities identify high-risk areas, allocate police resources effectively, and take preventive measures to reduce criminal activities.

The proposed approach aims to demonstrate how machine learning and data-driven techniques can support better decision-making in crime prevention and contribute to safer and smarter cities.

II. Related Work

A. Literature Survey

Paper Name: "LSTM-BASED PREDICTIVE MODELING FOR CRIME RATEFORECASTING ACROSS REGIONS

Markande Bhausaheb*1, Garje Tanmay*2, Gadekar Jaydip*3, Prof. Dumbre R.C.*4, Prof. Dr. Khatal S.S.*5

Abstract: - Urban and regional safety always matters and attracts so much interest up on crime and safety level predictions.Accu racy and presicion in the crime rate and safety level scores relate to data collected from previous occasions which remains very sensitive to time. The data dependency on time requires spacial care in handling and processing. Recurrent neural networks (RNN) have been used as alternative machine learning approach tostatistical models in handling data with time sensitivity properties. Recently, long short term memory (LSTM) model has been used successfully in prediction, especially with timesensitive data. In this study, crime rateprediction for city of Bristol is investigated and analysed using LSTM models in comparison to well-know machine learning and statistical approaches, where LSTM model is found significantly outperforming.

III. Proposed System

The proposed system was implemented to predict future crime rates in different regions using a deep learning approach based on the Long Short-Term Memory (LSTM) model. The system analyzes historical crime data and identifies temporal patterns that help forecast possible future crime occurrences. By utilizing machine learning techniques, the implemented system provides a data-driven solution that can assist law enforcement agencies in understanding crime trends and improving preventive strategies.

In the first stage, historical crime datasets were collected containing information such as region, date, type of crime, and the number of crime incidents. The collected data was stored in a structured format such as CSV files, which served as the input for the system.

Next, data preprocessing was performed to prepare the raw dataset for analysis. This process included handling missing values, removing noise, normalizing data, and organizing the dataset into a structured format suitable for machine learning algorithms. Proper preprocessing ensured that the model received clean and consistent data for training.

After preprocessing, feature extraction was carried out to identify the most relevant attributes from the dataset. Features such as crime count, region identifiers, and time-related variables were selected to effectively represent crime patterns. These features helped the model learn the relationships between historical crime events and future trends.

The processed dataset was then used to train an

LSTM-based prediction model. LSTM networks are specifically designed for timeseries data and are capable of capturing longterm dependencies within sequential datasets.

The model learned patterns from past crime records and built a predictive model capable of estimating future crime rates.

After training, the model was evaluated using testing data to measure its prediction performance. Evaluation metrics such as R^2 score, Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) were used to assess the effectiveness of the model.

Finally, the predicted results were displayed through a user interface where users could visualize crime trends and forecasts. This information helps decision-makers identify high-risk areas and allocate resources more efficiently for crime prevention.

The implemented system demonstrates how deep learning techniques can be applied to crime data analysis to generate accurate predictions and support data-driven public safety planning.

IV. Objectives

The main objectives of this research are as follows:

1) To develop a deep learning–based model using Long Short-Term Memory (LSTM) networks for predicting crime rates across different regions.

2) To analyze historical crime datasets in order to identify temporal patterns and trends that influence future crime occurrences.

3) To preprocess and transform raw crime data using techniques such as data cleaning, normalization, and feature selection to improve model performance.

4) To design and train a time-series prediction model capable of learning sequential relationships in crime data.

5) To evaluate the effectiveness of the proposed prediction model using performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

6) To provide insights that can assist law enforcement agencies in identifying potential crime hotspots and planning preventive strategies.

7) To contribute to the development of intelligent crime analysis systems that support data-driven decision-making for public safety.

V. Methodology

The proposed crime prediction system was developed using a Convolutional Neural Network (CNN) model to analyze historical crime data and predict future crime rates across different regions. The overall implementation of the system was carried out through several stages including data collection, data preprocessing, feature preparation, model development, training, and prediction.

Initially, historical crime data was collected from available datasets containing information such as region name, date, total number of crimes, and different crime categories. This dataset served as the primary input for the prediction system. The collected data was organized in a structured format to make it suitable for machine learning processing.

In the preprocessing stage, the dataset was cleaned to remove missing values, redundant records, and inconsistencies. Data normalization techniques were applied to ensure that the input values were within a consistent range, which helps improve the learning capability of the neural network. The dataset was then divided into training and testing sets to evaluate the model performance.

After preprocessing, relevant features were selected from the dataset to represent crime patterns over time. These features included crime frequency, region identifiers, and timerelated variables. The prepared data was then converted into a format suitable for input to the CNN model.

A Convolutional Neural Network (CNN) was implemented to learn patterns and relationships within the crime dataset. CNN models are capable of automatically extracting important features from structured data and identifying complex patterns that may not be easily detected using traditional machine learning methods. The convolution layers helped in identifying meaningful patterns in the crime data, while pooling layers reduced dimensionality and improved computational efficiency.

The CNN model was trained using the prepared training dataset. During training, the model learned relationships between historical crime patterns and future crime occurrences. After the training process was completed, the model was evaluated using the testing dataset to measure prediction accuracy and reliability.

Finally, the trained CNN model was used to generate predictions for future crime rates in different regions. The predicted results were displayed as the final output of the system, helping identify possible crime trends and highrisk areas. This information can assist law enforcement authorities in taking preventive actions and improving resource allocation for crime control.

Outcomes

The performance of the proposed crime prediction system is evaluated by analyzing how accurately the trained model predicts crime trends for different regions. After training the CNN model on historical crime datasets, the system generates predicted crime values for future time periods. These predicted results are then compared with the actual values to evaluate the reliability and effectiveness of the model.

To measure the model’s performance, evaluation metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and prediction accuracy are used. These metrics help determine how closely the predicted crime rates match the actual crime data. A lower error value indicates better prediction performance and higher model reliability.

The expected outcome of the system is to generate visual analysis charts that represent crime predictions for different regions. The system produces graphical outputs such as line charts and bar charts that show crime trends over time. These charts help visualize how crime rates change in a particular region based on historical data and predicted results.

The expected outcome of the system is to generate visual analysis charts that represent crime prediction trends. The system produces graphical outputs such as line charts and bar charts that show crime trends over time. These charts help visualize how crime rates change over different years based on historical data and predicted results.

1) R^2 (R-Squared)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

2) MSE (Mean Squared Error)

$$MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2$$

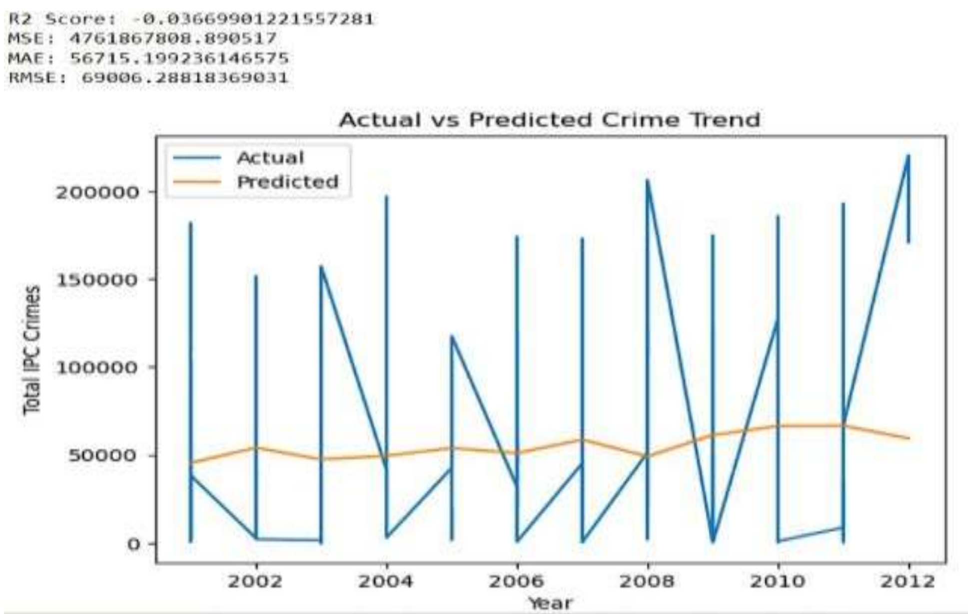
Evaluation and Expected

3) MAE (Mean Absolute Error)

$$MAE = \frac{1}{n}\sum|y_i - \hat{y}_i|$$

4) RMSE (Root Mean Squared Error)

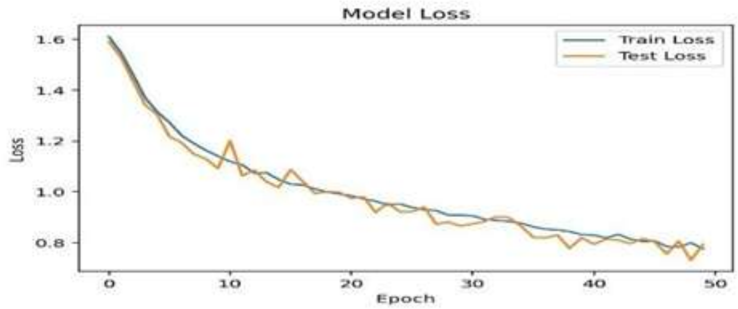
$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$$

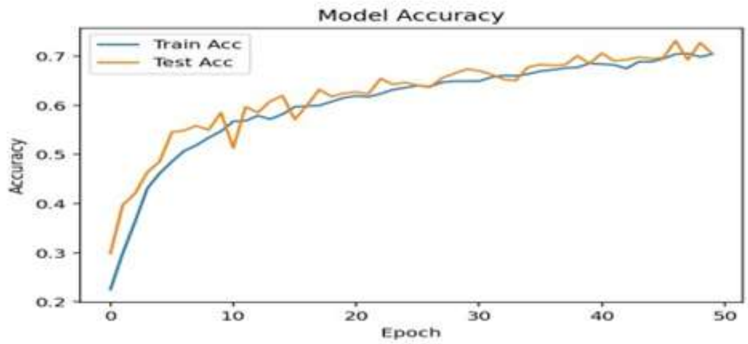


In addition to individual region analysis, the system also generates comparative charts between multiple regions. These comparison charts allow users to analyze crime levels across different areas and identify regions with higher or lower predicted crime rates. Such visualization helps in understanding regional crime patterns and identifying potential crime hotspots.

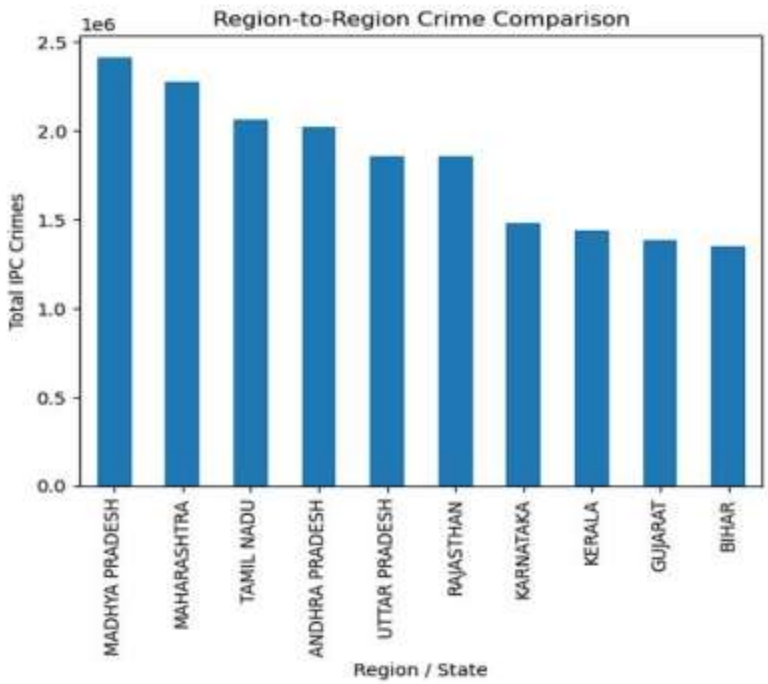
The final output of the system includes:

- Crime trend prediction charts for each region showing historical and predicted crime data.

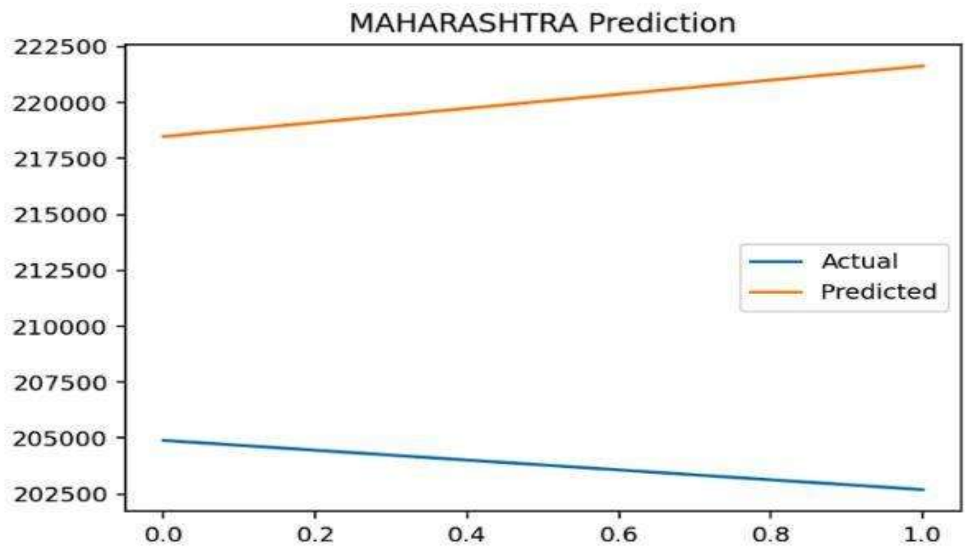




- Region-to-region comparison charts that highlight differences in crime levels across multiple regions.



• Visual analysis graphs that make it easier to interpret crime patterns and trends.



VII. Conclusion and Future Enhancements

In this project, a crime rate prediction system was developed to analyze historical crime data and forecast future crime trends across different regions. The system uses a Convolutional Neural Network (CNN) model to learn patterns from past crime records and generate predictions for upcoming time periods. By processing and analyzing the dataset, the model is able to identify important relationships between crime occurrences and regional patterns.

The implementation of the system demonstrates that machine learning techniques can be effectively used to analyze large crime datasets and provide meaningful predictions. The generated prediction charts and region-toregion comparison graphs help visualize crime trends and identify areas with higher potential crime rates. This analysis can assist law enforcement agencies and decision-makers in understanding crime patterns and planning preventive strategies.

Overall, the proposed system shows how datadriven approaches and deep learning models can support crime analysis and improve public safety by providing insights into possible future crime trends.

Future Enhancements:

Although the developed system successfully predicts crime trends and provides comparative analysis between regions, several improvements can be made to enhance the system in the future.

First, the system can be extended by incorporating larger and more diverse datasets, including socioeconomic factors, population density, and environmental data. These additional features may improve prediction accuracy and provide deeper insights into the causes of crime patterns.

Second, the model can be improved by experimenting with advanced deep learning architectures such as hybrid models combining CNN with LSTM or other timeseries models. These models may capture both spatial and temporal relationships in crime data more effectively.

Third, the system can be integrated with a realtime dashboard or web application that allows authorities to monitor crime predictions dynamically. This would enable easier access to visual analysis charts and comparison reports for different regions.

Finally, the system could be expanded to support multiple cities or national-level datasets, allowing broader crime analysis and helping authorities make better decisions for large-scale crime prevention strategies.

These enhancements can further improve the efficiency, scalability, and practical use of the crime prediction system in real-world applications.

References

1. B. Markande, T. Garje, J. Gadekar, R. C. Dumbre, and S. S. Khatal, "LSTM-Based Predictive Modeling for Crime Rate Forecasting Across Regions," Department of Computer Science Engineering, Sharadchandra Pawar College of Engineering, Dumbarwadi, India, 2026.
2. K. Hanaoka, New insights on relationships between street crimes and ambient population: Use of hourly population data estimated from mo bile phone users' locations, Environment and planning. B, Urban ana lytics and city science 45 (2) (2018) 295–311.
3. M. Zeng, Y. Mao, C.Wang, Therelationship between street environment and street crime: A case study of pudong new area, shanghai, china, Cities 112 (2021) 103143.
4. W. Safat, S. Asghar, S. A. Gillani, Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques, IEEE Access 9 (2021) 70080– 70094. doi:10.1109/access.2021.3078117.
5. A. Sherstinsky, Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network, Physica D: Nonlinear Phenomena 404 (2020) 132306.
6. G. Van Houdt, C. Mosquera, G. N´apoles, A review on the long short term memory model, Artificial Intelligence Review 53 (8) (2020) 5929 5955.
7. O. R. Shah, L. Willoughby, N. Bowersox, Tackling homelessness through ai powered social innovations: A novel and groundbreaking assess ment of criminal victimization of homeless populations in los angeles employing predictive analytics and machine learning models such as arima and lstm, Issues in Information Systems 22 (3) (2021) 264–277. doi:10.48009/3 iis 2021 283-297.
8. Khalid, S., Khan, S. A., & Ifzal, S. Q. (2021). A Fuzzy Logic-Based Framework for Mapping Crime Data on Established Sociological Hypothesis for Societal Disorder Identification and Prevention. IEEE Access, 9, 80197 80207. doi:10.1109/access.2021.3083542
9. D. M. Raza and D. B. Victor, "Data mining and Region Prediction Based on Crime Using Random Forest," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021, pp. 980-987, doi: 10.1109/ICAIS50930.2021.9395989.