International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predictive Analysis and Forecasting of Crime Patterns in Jharkhand Using Linear Regression and Random Forest Models

Anubhuti Srivastava

Research Scholar, RKDF University, Ranchi, Jharkhand, India.

Abstract

Crime prediction plays an important role in understanding crime trends and supporting decision-making in law enforcement. This paper presents a predictive analysis of crime patterns in Jharkhand using district-wise crime records collected during 2024 and 2025. The dataset contains information on several categories of cognizable offences, including murder, dacoity, robbery, burglary, theft, riot, kidnapping, rape, Arms Act cases, Naxal-related incidents, and miscellaneous crimes. After data pre-processing and feature selection, Linear Regression and Random Forest Regression models were developed to predict Total Cognizable Crime.

The performance of the models was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2. Experimental results showed that the Linear Regression model significantly outperformed the Random Forest Regression model. Linear Regression achieved an RMSE of 4.37 and a R^2 value of 0.999993, whereas Random Forest Regression produced an RMSE of 553.38 and a R^2 value of 0.894390. The results indicate a strong linear relationship between the selected crime variables and Total Cognizable Crime.

The results showed that Linear Regression provided more accurate predictions than Random Forest Regression for the available dataset. The findings demonstrate that historical crime records can be effectively used for crime forecasting and pattern analysis. The proposed approach offers a simple and reliable framework for predicting crime trends and highlights the potential of machine learning techniques for crime analytics in Jharkhand.

Keywords: Crime Prediction, Crime Forecasting, Machine Learning, Linear Regression, Random Forest Regression, Jharkhand.

1. Introduction:

Crime has become an important social issue that affects public safety, economic development, and the overall quality of life. Understanding crime patterns and identifying areas with higher crime incidence are essential for effective planning and crime prevention. With the increasing availability of digital crime records, data-driven techniques are being widely used to analyse crime trends and support decision-making.

Machine learning has emerged as a useful tool for analysing large datasets and discovering patterns that may not be easily identified through traditional statistical methods. In recent years, researchers have applied various machine learning techniques to crime prediction, hotspot detection, and crime forecasting. These approaches help in understanding historical crime trends and estimating future crime levels based on previously observed data.

Jharkhand is one of the major states of eastern India and exhibits considerable variation in crime levels across its districts. Crime records indicate that some districts consistently report higher numbers of offences than others. Such differences suggest that crime patterns vary across regions and may be studied using analytical and predictive techniques. Examining these patterns can provide useful insight into the distribution of crime and help identify districts that require greater attention.

The present work uses district-wise crime data collected from Jharkhand for the years 2024 and 2025. The dataset includes several categories of offences such as murder, dacoity, robbery, house burglary, theft, riot, kidnapping, rape, Arms Act cases, Naxal-related incidents, miscellaneous crimes, and Total Cognizable Crime. These variables are analysed to investigate their relationship with overall crime levels.

Two machine learning models, namely Linear Regression and Random Forest Regression, are employed for crime prediction. The performance of the models is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). A comparative analysis is carried out to determine the model that provides the most accurate predictions for the available dataset.

The main contribution of this work is the application of machine learning techniques to district-level crime data from Jharkhand and the evaluation of their effectiveness in crime forecasting. The findings provide insight into crime patterns within the state and demonstrate the usefulness of machine learning methods for predictive crime analysis.

2. Literature Review:

The application of machine learning techniques to crime analysis has attracted considerable attention in recent years. The increasing availability of digital crime records and advances in data analytics has enabled researchers to develop predictive models for understanding crime patterns and forecasting future criminal activities. Various statistical, machine learning and deep learning approaches have been proposed for crime prediction, hotspot identification, and decision support in law enforcement.

Several studies have investigated the use of traditional machine learning methods for crime prediction. Shah et al. (2021) developed a crime forecasting framework based on machine learning and computer vision techniques and demonstrated the potential of data-driven approaches in crime prevention. Similarly, Kim et al. (2018) applied machine learning algorithms to crime analysis and reported that predictive models can effectively identify relationships among crime variables. Agarwal et al. (2018) explored statistical approaches for crime prediction and showed that historical crime records contain useful information for forecasting future crime trends.

Research has also focused on comparing different machine learning algorithms for crime prediction. Kshatri et al. (2021) conducted an empirical analysis of several machine learning methods and proposed an ensemble-based framework using stacked generalization. Their findings indicated that combining multiple models can improve predictive performance. Bandekar and Vijayalakshmi (2020) examined the applicability of machine learning algorithms for crime reduction and highlighted the usefulness of predictive analytics in supporting law-enforcement activities. Kumar et al. (2020) investigated the K-nearest neighbour algorithm for crime prediction, while Sivanagaleela and Rajesh (2019) applied the fuzzy C-means algorithm to analyse crime patterns.

A number of studies have explored the use of Random Forest and related ensemble techniques for predictive crime analytics. Raza and Victor (2021) employed Random Forest for crime region prediction and reported encouraging forecasting performance. Yao et al. (2020) utilized Random Forest for hotspot prediction based on spatial factors and demonstrated its effectiveness in identifying crime-prone areas. Das et al. (2021) proposed an incremental classification framework for crime prediction and showed that optimization-based ensemble methods can improve prediction accuracy. These studies indicate that Random Forest and ensemble learning approaches are widely used for handling complex crime datasets.

Spatial and temporal characteristics of crime have received significant attention in the literature. Catlett et al. (2018) proposed a data-driven framework for spatio-temporal crime prediction in smart cities, while Catlett et al. (2019) extended this work through experimental evaluation using real-world datasets. Dash et al. (2018) applied network analytic methods to spatio-temporal crime prediction and demonstrated the importance of considering both location and time in predictive modeling. Hossain et al. (2020) further emphasized the value of spatio-temporal information by developing crime prediction models based on spatial and temporal data attributes.

The growing interest in hotspot detection has led to the development of several spatial forecasting frameworks. Araujo et al. (2018) proposed a crime hotspot detection framework for patrol planning and resource allocation. Kounadi et al. (2020) conducted a systematic review of spatial crime forecasting studies and highlighted the importance of geographical information in crime prediction. Saraiva et al. (2022) combined machine learning, spatial analytics, and text analysis for crime monitoring and prediction in Porto, Portugal. Their work demonstrated that integrating multiple data sources can improve understanding of crime distribution.

Recent studies have increasingly adopted deep learning approaches for crime forecasting. Chun et al. (2019) developed a crime prediction model using deep neural networks and reported promising predictive performance. Wang et al. (2019) introduced a deep learning framework for real-time crime forecasting and demonstrated its applicability to large-scale crime datasets. Han et al. (2020) proposed an integrated LSTM and graph convolutional network model for theft crime prediction in urban communities. Similarly, Liang et al. (2022) developed a neural attention framework that incorporated spatial, temporal, and categorical information for crime prediction.

Advances in artificial intelligence have also encouraged the development of more sophisticated predictive models. Li et al. (2022) proposed a spatial-temporal hypergraph learning framework for crime prediction, while Tasnim et al. (2022) introduced a multi-module architecture based on attention mechanisms and sequential fusion models. Zhou et al. (2022) investigated domain adaptation techniques for crime risk prediction across different cities. These studies demonstrate the growing use of advanced machine learning methods for handling complex crime datasets.

Several researchers have examined the role of predictive analytics in smart policing and public safety. Elluri et al. (2019) developed machine learning models to support smart policing applications. Meijer and Wessels (2019) reviewed the benefits and challenges of predictive policing and discussed its practical implications for law-enforcement agencies. ToppiReddy et al. (2018) proposed a crime monitoring framework based on spatial analysis and emphasized the usefulness of predictive systems in crime management.

Although substantial progress has been made in crime prediction research, several limitations remain. Many existing studies focus on large metropolitan areas or utilize complex deep learning architectures that require extensive datasets and computational resources. Comparatively fewer studies have examined district-level crime prediction using relatively simple and interpretable machine learning models. In addition, limited attention has been given to crime forecasting using recent crime data from Indian states such as Jharkhand.

The present work seeks to address this gap by applying Linear Regression and Random Forest Regression models to district-wise crime records from Jharkhand. Unlike many previous studies that emphasize highly complex prediction frameworks, the present analysis focuses on evaluating the effectiveness of two widely used machine learning approaches for forecasting Total Cognizable Crime. The study also provides insight into district-level crime patterns and contributes to the growing body of research on machine learning-based crime prediction.

3. Dataset Description:

The analysis presented in this paper is based on district-wise crime data collected from Jharkhand for the years 2024 and 2025. The dataset contains monthly crime records reported from different districts of the state and includes information on several categories of cognizable offences.

The crime categories considered in this work are Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, Miscellaneous crimes, and Total Cognizable Crime. These variables provide a broad picture of crime occurrence across the state and make it possible to examine the relationship between different types of offences and overall crime levels.

The records from 2024 and 2025 were combined to create a single dataset for analysis. Before applying the machine learning models, the data were checked for consistency and prepared for further processing. The crime categories were used as predictor variables, while Total Cognizable Crime was selected as the target variable for forecasting.

The dataset shows noticeable differences in crime levels across districts. Some districts reported a larger number of crime cases, while others recorded comparatively fewer incidents. Such variation makes the dataset suitable for studying crime patterns and evaluating predictive models.

The use of district-wise monthly records allows the analysis to capture both regional differences and changes over time. This provides a useful basis for applying machine learning techniques and examining their effectiveness in crime prediction and forecasting.

4. Proposed Prediction Framework:

4.1 Data Pre-processing:

Before applying the machine learning models, the crime data were prepared for analysis through a pre-processing stage. The records collected for 2024 and 2025 were combined into a single dataset to provide a unified source for model development.

The dataset was examined to identify missing values, inconsistencies, and formatting issues. Column names were standardized, and the records were organized in a form suitable for machine learning implementation. Since the data were collected from multiple files, pre-processing helped ensure consistency across all observations.

After cleaning the data, the crime categories were selected as predictor variables. These variables included Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, and Miscellaneous crimes. Total Cognizable Crime was chosen as the target variable for prediction.

The processed dataset was then used for training and testing the machine learning models. Proper data pre-processing improved the quality of the dataset and helped ensure reliable model performance during the prediction process.

4.2 Feature Selection:

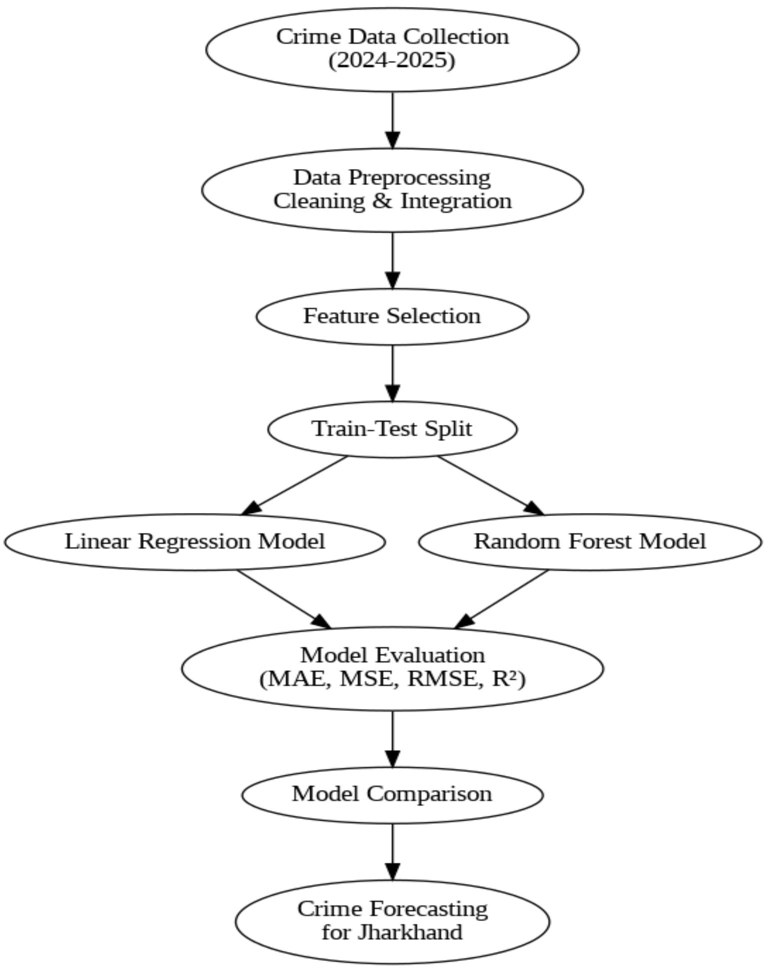
Feature selection is an important step in machine learning because the quality of the input variables directly affects the performance of the predictive models. In the present work, the variables were selected from the available crime records based on their relevance to overall crime occurrence.

The selected features included Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, and Miscellaneous crimes. These variables represent major categories of reported offences and provide a broad picture of crime patterns across the districts of Jharkhand.

Total Cognizable Crime was chosen as the target variable because it reflects the overall level of crime recorded in a district. The selected crime categories were used as predictor variables to examine how they contribute to variations in total crime.

The feature selection process was based on the structure of the available dataset and the objective of the study. By using multiple crime categories as input variables, the machine learning models were able to learn the relationship between individual offences and overall crime levels. These selected features formed the basis for the development of the prediction models described in the following sections.

Figure 1: illustrates the overall workflow of the proposed crime prediction framework.



4.3 Train-Test Split:

After selecting the predictor and target variables, the dataset was divided into training and testing subsets. This step is important because it allows the predictive models to be evaluated on data that were not used during the learning process.

In the present study, 80% of the available records were used for model training, while the remaining 20% were reserved for testing. The training dataset was used to develop the machine learning models and identify relationships among the selected crime variables. The testing dataset was then used to examine how accurately the trained models could predict Total Cognizable Crime for previously unseen observations.

The train-test split helps provide an objective assessment of model performance and reduces the possibility of evaluating the models only on the data used during training. This approach was applied to both the Linear Regression and Random Forest Regression models used in the study.

4.4 Linear Regression Model:

Linear Regression was used as one of the prediction models in this study. The model examines the relationship between the selected crime variables and Total Cognizable Crime. By using historical crime records, it estimates how variations in different crime categories influence the overall level of crime.

The predictor variables included Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, and Miscellaneous crimes. Total Cognizable Crime was considered as the target variable.

The Linear Regression model can be expressed as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon$$

Where Y represents the predicted value of Total Cognizable Crime, $(X_1, X_2, ..., X_n)$ denotes the predictor variables, β_0 is the intercept term, $(\beta_1, \beta_2, ..., \beta_n)$ is the regression coefficients, and ϵ represents the error term.

The model was trained using the training dataset and subsequently tested on the remaining observations. The predicted values obtained from the model were compared with the actual crime values to evaluate forecasting performance.

4.5 Random Forest Regression Model:

In addition to Linear Regression, Random Forest Regression was used to examine whether crime patterns could be predicted more effectively through an ensemble learning approach. Random Forest is a machine learning technique that combines the predictions of multiple decision trees to produce a final prediction.

Unlike Linear Regression, which assumes a linear relationship among variables, Random Forest can capture more complex relationships within the data. This makes it useful for datasets where different variables may interact with one another in a non-linear manner.

In the present study, the same set of crime variables used in the Linear Regression model was employed as input features. These variables included Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, and Miscellaneous crimes. Total Cognizable Crime was considered as the target variable.

The prediction generated by the Random Forest model is obtained by combining the outputs of several decision trees and can be represented as

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} T_i\,(X)$$

Where $\hat{Y}$ denotes the predicted crime value, N represents the number of decision trees, and $T_i(X)$denotes the prediction produced by the i-th decision tree.

The model was trained using the training dataset and subsequently evaluated using the testing dataset. The predicted values were then compared with the actual crime values to assess the forecasting performance of the model. The results obtained from Random Forest Regression were further compared with those of Linear Regression to identify the model that provided more accurate predictions for the Jharkhand crime dataset.

4.6 Model Evaluation Metrics:

After training the machine learning models, their performance was evaluated by comparing the predicted crime values with the actual values recorded in the dataset. The purpose of this evaluation was to determine how accurately the models could forecast Total Cognizable Crime and to identify the model that provided the most reliable predictions.

To assess model performance, four commonly used evaluation measures were employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2. These measures provide quantitative information about the accuracy of the predictions and allow a fair comparison between different machine learning models.

Mean Absolute Error (MAE):

Mean Absolute Error measures the average difference between the actual and predicted values. It provides an indication of how far the predictions deviate from the observed values on average.

$$MAE = \frac{1}{N}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Lower MAE values indicate better prediction accuracy.

Mean Squared Error (MSE):

Mean Squared Error measures the average of the squared differences between actual and predicted values. Since larger errors are squared, this measure gives greater importance to substantial prediction errors.

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

A smaller MSE value indicates better model performance.

Root Mean Squared Error (RMSE):

Root Mean Squared Error is obtained by taking the square root of MSE. It is expressed in the same unit as the original data and is therefore easier to interpret.

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

Lower RMSE values indicate that the predicted values are closer to the actual observations.

Coefficient of Determination R^2:

The coefficient of determination measures how well the model explains the variation in the target variable. It provides an indication of the overall goodness of fit of the model.

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

A R^2 value closer to 1 indicates a better fit and stronger predictive performance.

The values of these evaluation measures were calculated for both the Linear Regression and Random Forest Regression models. The results obtained from these measures were then used to compare the performance of the models and identify the most suitable approach for crime prediction and forecasting.

5. Experimental Setup:

The experiments were conducted using the Python programming language in the Google Colab environment. Python was selected because of its extensive support for data analysis, machine learning, and visualization. The crime datasets for 2024 and 2025 were processed and analysed using widely used Python libraries.

Data manipulation and pre-processing were performed using the Pandas and NumPy libraries. Machine learning models were implemented using the Scikit-learn library, while graphical representations of the results were generated using Matplotlib. These tools provided an efficient platform for developing and evaluating the prediction models.

The combined dataset consisted of district-wise monthly crime records collected from different districts of Jharkhand during 2024 and 2025. The final dataset contained 54 observations representing crime information from 24 districts of the state. A total of 11 crime categories, namely Total Murder, Total Dacoity, Total Robbery, House Burglary, Theft, Riot, Total Kidnapping, Rape, Arms Act cases, Naxal-related cases, and Miscellaneous crimes, were used as predictor variables. Total Cognizable Crime was considered as the target variable for prediction.

Before model development, the dataset was pre-processed and organized into a suitable format for machine learning analysis. The selected features provided information on different categories of crime and were used to examine their relationship with the overall level of crime recorded in each district.

To evaluate the predictive capability of the models, the dataset was divided into training and testing subsets using a $80 : 20$ ratio. The training data were used to develop the Linear Regression and Random Forest Regression models, while the testing data were used to assess their prediction accuracy.

The performance of the models was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2. These measures provided an objective basis for comparing the models and identifying the approach that produced the most accurate crime forecasts.

6. Results and Discussion:

This section presents the results obtained from the machine learning models and discusses their effectiveness in predicting Total Cognizable Crime in Jharkhand. The performance of Linear Regression and Random Forest Regression was evaluated using MAE, MSE, RMSE, and the coefficient of determination R^2.

6.1 Performance Comparison of Machine Learning Models:

Table 1 presents the performance measures obtained for the two prediction models.

Model	MAE	MSE	RMSE	R^2
Linear Regression	2.3508	19.12	4.3727	0.999993
Random Forest Regression	277.9509	306229.20	553.3798	0.894390

The results show a clear difference in the performance of the two models. Linear Regression achieved very low error values and produced a R^2 value close to 1. This indicates that the model was able to predict Total Cognizable Crime with a high degree of accuracy.

Random Forest Regression also produced reasonable results, but its prediction errors were considerably larger than those of Linear Regression. Although the model explained a substantial portion of the variation in the data, its forecasts were less accurate for the available dataset.

The superior performance of Linear Regression suggests that the relationship between the selected crime variables and Total Cognizable Crime is largely linear. Since Total Cognizable Crime is closely related to the combined contribution of the individual crime categories, the Linear Regression model was able to capture this relationship effectively.

6.2 Comparison of Actual and Predicted Crime Values:

To further examine the performance of the best-performing model, the predicted values obtained from Linear Regression were compared with the actual crime values. The comparison showed that the predicted values were very close to the corresponding observed values.

Table 2: Sample Prediction Results Obtained Using Linear Regression

Actual Crime	Predicted Crime	Error
303	301.71	1.29
917	915.80	1.20
947	946.88	0.12
647	647.65	0.65
199	198.85	0.15
4250	4252.86	2.86
3427	3431.11	4.11
1055	1054.60	0.40
5426	5412.59	13.41
460	460.52	0.52

The prediction errors are small for most observations, indicating that the model was able to estimate crime values accurately. The close agreement between actual and predicted values supports the high R^2 value and low error measures reported earlier.

Figure 2: Actual versus Predicted Crime Values Obtained Using Linear Regression

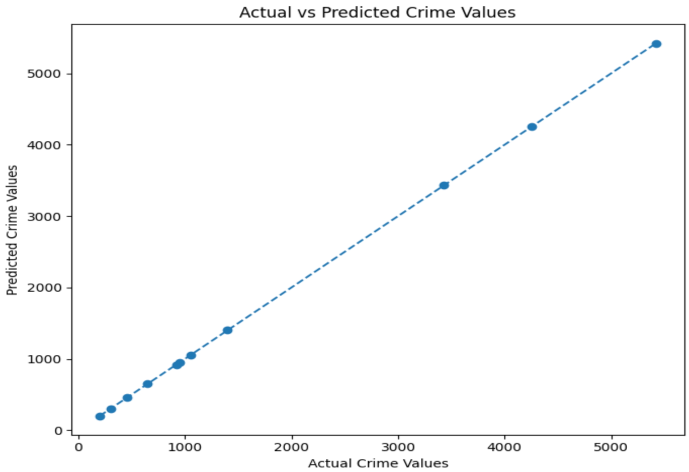
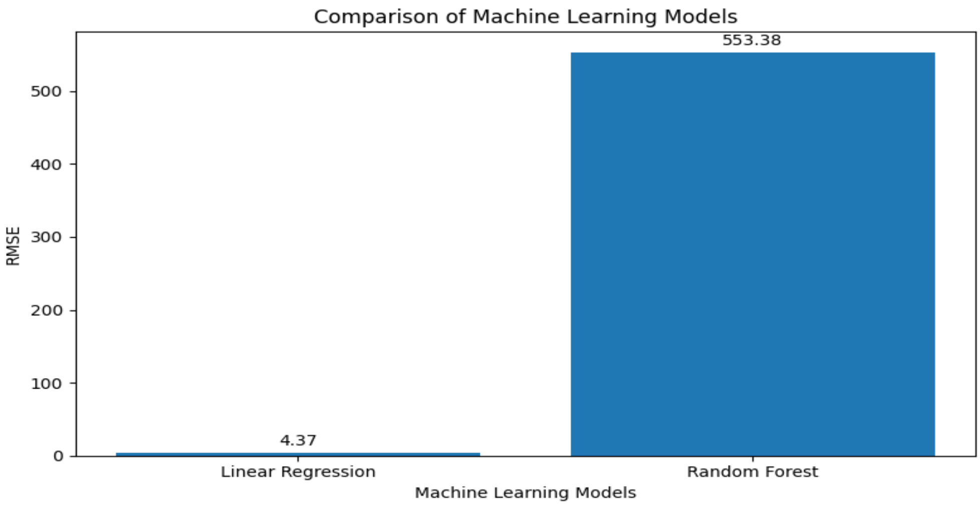


Figure 2 shows the relationship between the actual crime values and the values predicted by the Linear Regression model. The predicted values closely follow the actual observations, indicating a strong agreement between the two. Most of the points lie near the diagonal trend, which suggests that the model was able to estimate crime levels with a high degree of accuracy.

The close correspondence between actual and predicted values is consistent with the low error measures and high R^2 value obtained during model evaluation. These results indicate that the Linear Regression model successfully captured the relationship between the selected crime variables and Total Cognizable Crime. The figure provides visual evidence of the effectiveness of the model for crime forecasting using the available dataset.

Figure 3: presents a graphical comparison of the machine learning models based on RMSE.



The figure shows that Linear Regression achieved a substantially lower RMSE than Random Forest Regression. A lower RMSE indicates that the predicted values are closer to the actual crime values. This result further confirms the effectiveness of Linear Regression for the present dataset.

6.3 Crime Forecasting Results:

After evaluating the prediction models, Linear Regression was selected for crime forecasting because it produced the lowest prediction error and the highest R^2 value. The model was used to generate predicted values of Total Cognizable Crime, and the results were compared with the corresponding actual observations. Figure 4 presents the forecasting results obtained from the model.

Figure 4: Crime Forecasting Results Obtained Using Linear Regression

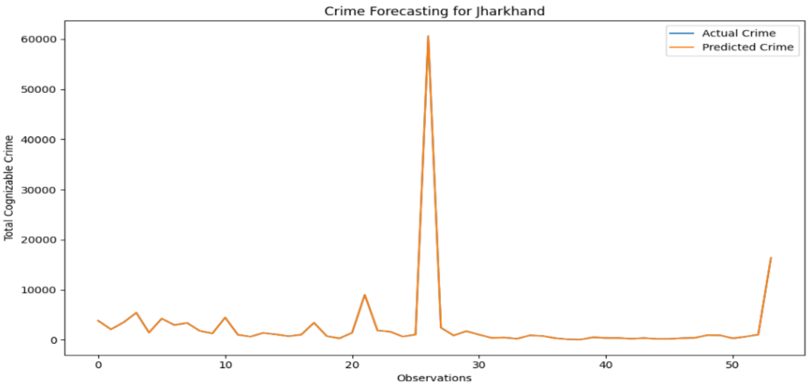


Figure 4 shows the comparison between the actual crime values and the values predicted by the Linear Regression model. The two curves follow a very similar pattern throughout the observations, indicating that the model was able to capture the underlying crime trends present in the dataset.

The close agreement between the actual and predicted values further supports the performance measures reported earlier. Even for observations with relatively high crime levels, the model produced predictions that remained close to the observed values. These results suggest that historical crime records contain useful information for forecasting Total Cognizable Crime and demonstrate the effectiveness of the proposed approach for crime prediction in Jharkhand.

6.4 Discussion of Findings:

The results indicate that machine learning techniques can be successfully applied to crime data for predictive analysis. Among the two models considered, Linear Regression provided the most reliable forecasts for the Jharkhand dataset.

One possible reason for this performance is that Total Cognizable Crime is strongly related to the individual crime categories included in the dataset. As a result, a linear model was sufficient to describe the relationship between the predictor variables and the target variable.

The findings also highlight the usefulness of district-level crime records for understanding crime patterns and supporting forecasting activities. The ability of the models to generate accurate predictions suggests that historical crime data can provide valuable information for future crime analysis and planning.

6.5 Limitations of the Study:

The present paper is based on officially reported crime records for 2024 and 2025. The dataset covers a relatively short time period and does not include socio-economic, demographic, or geographical variables that may influence crime occurrence. In addition, Total Cognizable Crime was used as the target variable, which is closely related to the selected crime categories. Future studies may consider longer time periods and additional explanatory variables to develop more comprehensive forecasting models.

7. Conclusion:

This paper presented a machine learning-based approach for predicting crime patterns in Jharkhand using district-wise crime data from 2024 and 2025. Two prediction models, namely Linear Regression and Random Forest Regression, were developed and evaluated using standard performance measures including MAE, MSE, RMSE, and R^2.

The results showed that Linear Regression outperformed Random Forest Regression for the available dataset. The Linear Regression model achieved very low prediction errors and a R^2value close to 1, indicating a strong relationship between the selected crime variables and Total Cognizable Crime. The comparison of actual and predicted values further confirmed the accuracy of the model.

The findings demonstrate that historical crime records can be effectively used for crime forecasting and pattern analysis. The results also indicate that machine learning techniques can provide useful insight into crime behaviour and support data-driven analysis of crime data.

Although the present work considered only two machine learning models, the results highlight the potential of predictive analytics in crime forecasting. Future work may explore larger datasets, longer time periods, additional socio-economic variables, and more advanced machine learning techniques to further improve prediction performance.

Overall, the proposed approach provides a simple and effective framework for crime prediction in Jharkhand and demonstrates the usefulness of machine learning methods for analysing crime-related data.

References:

1. Agarwal, S., Yadav, L., & Thakur, M. K. (2018), Crime prediction based on statistical models, In Proceedings of the 11th International Conference on Contemporary Computing (IC) (pp. 1–3).
2. Almuhanna, A. A., Alrehili, M. M., Alsubhi, S. H., & Syed, L. (2021), Prediction of crime in neighbourhoods of New York City using spatial data analysis, In Proceedings of the 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) (pp. 23–30).
3. Alves, L. G. A., Ribeiro, H. V., & Rodrigues, F. A. (2018), Crime prediction through urban metrics and statistical learning, Physica A: Statistical Mechanics and Its Applications, 505, 435–443.
4. Araujo, A., Cacho, N., Bezerra, L., Vieira, C., & Borges, J. (2018), Towards a crime hotspot detection framework for patrol planning, In Proceedings of the IEEE 20th International Conference on High Performance Computing and Communications (pp. 1256–1263).
5. Asaro, P. M. (2019). AI ethics in predictive policing: From models of threat to an ethics of care, IEEE Technology and Society Magazine, 38(2), 40–53.
6. Bandekar, S. R., & Vijayalakshmi, C. (2020), Design and analysis of machine learning algorithms for the reduction of crime rates in India, Procedia Computer Science, 172, 122–127.
7. Baqir, A., Rehman, S. U., Malik, S., Mustafa, F. U., & Ahmad, U. (2020), Evaluating the performance of hierarchical clustering algorithms to detect spatio-temporal crime hot-spots, In

Proceedings of the 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1–5).
8. Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A., Koh, T. W., & Sherazi, H. H. R. (2021), Spatio-temporal crime predictions by leveraging artificial intelligence for citizens security in smart cities, IEEE Access, 9, 47516–47529.
9. Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2018), A data-driven approach for spatio-temporal crime predictions in smart cities, In Proceedings of the IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 17–24).
10. Catlett, C., Cesario, E., Talia, D., & Vinci, A. (2019), Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments, Pervasive and Mobile Computing, 53, 62–74.
11. Chun, S. A., Paturu, V. A., Yuan, S., Pathak, R., Atluri, V., & Adam, N. R. (2019), Crime prediction model using deep neural networks, In Proceedings of the 20th Annual International Conference on Digital Government Research (pp. 512–514).
12. Dash, S. K., Safro, I., & Srinivasamurthy, R. S. (2018), Spatio-temporal prediction of crimes using network analytic approach, In Proceedings of the IEEE International Conference on Big Data (pp. 1912–1917).
13. Das, A. K., & Das, P. (2022), Graph based ensemble classification for crime report prediction, Applied Soft Computing, 125, Article 109215.
14. Das, P., Das, A. K., Nayak, J., Pelusi, D., & Ding, W. (2021), Incremental classifier in crime prediction using bi-objective particle swarm optimization Information Sciences, 562, 279–303.
15. Elluri, L., Mandalapu, V., & Roy, N. (2019), Developing machine learning based predictive models for smart policing In Proceedings of the IEEE International Conference on Smart Computing (SMARTCOMP) (pp. 198–204).
16. Gahalot, A., Dhiman, S., & Chouhan, L. (2020), Crime prediction and analysis In Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA) (pp. 1–6).
17. Han, X., Hu, X., Wu, H., Shen, B., & Wu, J. (2020), Risk prediction of theft crimes in urban communities: An integrated model of LSTM and ST-GCN IEEE Access, 8, 217222–217230.
18. He, J., & Zheng, H. (2021), Prediction of crime rate in urban neighborhoods based on machine learning Engineering Applications of Artificial Intelligence, 106, Article 104460.
19. Hossain, S., Abtahee, A., Kashem, I., Hoque, M. M., & Sarker, I. H. (2020), Crime prediction using spatio-temporal data In Computing Science, Communication and Security (pp. 277–289),Springer.
20. Janiesch, C., Zschech, P., & Heinrich, K. (2021), Machine learning and deep learning. Electronic Markets, 31(3), 685–695.
21. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014), Large-scale video classification with convolutional neural networks, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1725–1732).
22. Kim, S., Joshi, P., Kalsi, P. S., & Taheri, P. (2018), Crime analysis through machine learning, In Proceedings of the IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 415–420).
23. Kounadi, O., Ristea, A., Araujo, A., & Leitner, M. (2020), A systematic review on spatial crime forecasting, Crime Science, 9, 1–22.

24. Kshatri, S. S., Singh, D., Narain, B., Bhatia, S., Quasim, M. T., & Sinha, G. R. (2021), An empirical analysis of machine learning algorithms for crime prediction using stacked generalization: An ensemble approach, IEEE Access, 9, 67488–67500.
25. Kumar, A., Verma, A., Shinde, G., Sukhdeve, Y., & Lal, N. (2020), Crime prediction using K-nearest neighboring algorithm, In Proceedings of the International Conference on Emerging Trends in Information Technology and Engineering (pp. 1–4).
26. Li, Z., Huang, C., Xia, L., Xu, Y., & Pei, J. (2022), Spatial-temporal hypergraph self-supervised learning for crime prediction, In Proceedings of the IEEE 38th International Conference on Data Engineering (ICDE) (pp. 2984–2996).
27. Liang, W., Wang, Y., Tao, H., & Cao, J. (2022), Towards hour-level crime prediction: A neural attentive framework with spatial–temporal–categorical fusion, Neurocomputing, 486, 286–297.
28. Ma, Y., Nakamura, K., Lee, E., & Bhattacharyya, S. S. (2022), EADTC: An approach to interpretable and accurate crime prediction, In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 170–177).
29. Raza, D. M., & Victor, D. B. (2021), Data mining and region prediction based on crime using random forest, In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 980–987).
30. Ristea, A., Al Boni, M., Resch, B., Gerber, M. S., & Leitner, M. (2020), Spatial crime distribution and prediction for sporting events using social media, International Journal of Geographical Information Science, 34(9), 1708–1739.
31. Safat, W., Asghar, S., & Gillani, S. A. (2021), Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques, IEEE Access, 9, 70080–70094.
32. Saraiva, M., Matijosaitiene, I., Mishra, S., & Amante, A. (2022), Crime prediction and monitoring in Porto, Portugal, using machine learning, spatial and text analytics, ISPRS International Journal of Geo-Information, 11(7), 400.
33. Sarker, I. H. (2021), Machine learning: Algorithms, real-world applications and research directions, S N Computer Science, 2(3), 160.
34. Shah, N., Bhagat, N., & Shah, M. (2021), Crime forecasting: A machine learning and computer vision approach to crime prediction and prevention, Visual Computing for Industry, Biomedicine, and Art, 4(1), 1–14.
35. Shenoy, R., Yadav, D., Lakhotiya, H., & Sisodia, J. (2022), An intelligent framework for crime prediction using behavioural tracking and motion analysis, In Proceedings of the International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1–6).
36. Shermila, A. M., Bellarmine, A. B., & Santiago, N. (2018), Crime data analysis and prediction of perpetrator identity using machine learning approach, In Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 107–114).
37. Sivanagaleela, B., & Rajesh, S. (2019), Crime analysis and prediction using fuzzy C-means algorithm, In Proceedings of the 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 595–599).
38. ToppiReddy, H. K. R., Saini, B., & Mahajan, G. (2018), Crime prediction and monitoring framework based on spatial analysis, Procedia Computer Science, 132, 696–705.

39. Wang, B., Yin, P., Bertozzi, A. L., Brantingham, P. J., Osher, S. J., & Xin, J. (2019), Deep learning for real-time crime forecasting and its ternarization, Chinese Annals of Mathematics, Series B, 40(6), 949–966.
40. Yao, S., Wei, M., Yan, L., Wang, C., Dong, X., Liu, F., & Xiong, Y. (2020), Prediction of crime hotspots based on spatial factors of random forest, In Proceedings of the 15th International Conference on Computer Science and Education (ICCSE) (pp. 811–815).
41. Abraham, J., Ng, R., Morelato, M., Tahtouh, M., & Roux, C. (2021), Automatically classifying crime scene images using machine learning methodologies, Forensic Science International: Digital Investigation, 39, Article 301273.
42. Alikhademi, K., Drobina, E., Prioleau, D., Richardson, B., Purves, D., & Gilbert, J. E. (2022), A review of predictive policing from the perspective of fairness, Artificial Intelligence and Law, 30(1), 1–17.
43. Chackravarthy, S., Schmitt, S., & Yang, L. (2018), Intelligent crime anomaly detection in smart cities using deep learning, In Proceedings of the IEEE 4th International Conference on Collaboration and Internet Computing (CIC) (pp. 399–404).
44. Das, A. K., & Das, P. (2022), Graph based ensemble classification for crime report prediction, Applied Soft Computing, 125, Article 109215.
45. Muthamizharasan, M., & Ponnusamy, R. (2022), Forecasting crime event rate with a CNN-LSTM model, In Innovative Data Communication Technologies and Application (pp. 461–470). Springer.
46. Tasnim, N., Imam, I. T., & Hashem, M. M. A. (2022), A novel multi-module approach to predict crime based on multivariate spatio-temporal data using attention and sequential fusion model, IEEE Access, 10, 48009–48030.
47. Wang, C., Han, B., Patel, B., & Rudin, C. (2022), In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction, Journal of Quantitative Criminology, 39, 519–581.
48. Yang, D., Heaney, T., Tonon, A., Wang, L., & Cudré-Mauroux, P. (2018), CrimeTelescope: Crime hotspot prediction based on urban and social media data fusion, World Wide Web, 21(5), 1323–1347.
49. Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022), Interpretable machine learning models for crime prediction, Computers, Environment and Urban Systems, 94, Article 101789.
50. Zhou, B., Chen, L., Zhao, S., Li, S., Zheng, Z., & Pan, G. (2022), Unsupervised domain adaptation for crime risk prediction across cities, IEEE Transactions on Computational Social Systems.