Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

CRIME FORECASTING WITH MACHINE LEARNING AND DEEP LEARNING: A SYSTEMATIC ANALYSIS

G. Pavan, Assistant Professor, Seshadri Rao Gudlavalleru Engineering college
(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada), Seshadri Rao Knowledge Village, Gudlavalleru-521356, Andhra Pradesh, India
S.Uday Kiran, S.Harshitha, S.Leela Sai Kumar, VVDS Phanindra Naidu, IV- B. Tech CSE
Department of Computer Science and Engineering, Seshadri Rao Gudlavalleru Engineering college
(An Autonomous Institute with Permanent Affiliation to JNTUK, Kakinada), Seshadri Rao Knowledge Village, Gudlavalleru-521356, Andhra Pradesh, India

***Abstract*— This study analyzes and predicts crime patterns using a dataset of crimes from Chicago starting from 2001. Through exploratory data analysis (EDA), we uncovered unique and non-unique case numbers and crime distributions. The data was preprocessed for machine learning by handling duplications, extracting date components, and encoding categorical data. Visualization techniques highlighted key trends and high-crime areas. We developed a predictive model using an artificial neural network (ANN), equipped with dense and dropout layers to prevent overfitting, and optimized using the Adam optimizer for regression tasks. The model's effectiveness was assessed with mean squared error (MSE) and mean absolute error (MAE), confirming its ability to accurately predict crime counts. Additionally, we integrated the model into a Gradio web interface, allowing easy user interaction for crime count predictions. This research enhances understanding of crime dynamics and offers a practical tool for resource allocation in public safety, demonstrating the impact of machine learning and deep learning in predictive analytics within the public safety sector.**

***Index Terms*— Crime Data Analysis, Predictive Modeling, Machine Learning, Deep Learning, Exploratory Data Analysis, Data Visualization, Neural Networks, Dropout Regularization, Adam Optimizer, Gradio Interface, Public Safety, Crime Trends.**

I. INTRODUCTION

Artificial Intelligence & Crime Prediction: An Efficient Writing Survey This paper embraces an systematic literature review (SLR) to look at the viability of computerized reasoning in wrongdoing expectation. It talks about the different AI procedures used over late years, zeroing in on the assorted sorts of violations [1], expectation strategies, and the qualities and shortcomings of various models. The review expects to give extensive experiences into the job of computer-based intelligence in anticipating crimes, offering important direction for future exploration around here.

Crime Analysis and Forecasting Using Machine Learning This research focuses on applying various machine learning algorithms to predict crime locations
to assist in crime prevention and reduction. By analyzing historical crime data from New York State, the study evaluates the performance of multiple machine learning techniques [2], including decision trees and neural networks, highlighting their effectiveness in forecasting crime with high accuracy. The paper emphasizes the importance of predictive analytics in law enforcement strategies.

Crime Analysis and Prediction Using Machine Learning This study develops a machine learning-based application for analyzing crime data across different districts in India. Utilizing logistic regression and k-means clustering [3], the paper categorizes districts by crime frequency, offering preventive

https://doi.org/10.36893/JNAO.2025.V16.I01.025

measures and precautions. The application aims to provide users with actionable insights into crime rates, enhancing public awareness and safety preparedness.
Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions This review paper synthesizes over 150 articles to explore the application of machine learning and deep learning in crime prediction [4]. It assesses various predictive models and their effectiveness in identifying crime patterns, providing a critical analysis of current trends and methodologies. The paper also discusses the challenges of data quality and model interpretability, proposing future directions to enhance the accuracy and applicability of predictive policing.
Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review Addressing the complex problem of crime prediction, this paper reviews the application of machine learning and deep learning technologies for predictive policing [5]. By analyzing and predicting crime data, including environmental and demographic factors, the study explores the potential of these technologies to improve public safety. It also highlights the significant challenges such as data availability and ethical concerns, providing a comprehensive overview of the state-of-the-art and suggesting paths forward for research and application in the field.
This research explores the use of machine learning
techniques in predicting crime, specifically analyzing two data-processing methods applied to Vancouver
crime data from the past 15 years. The study employs
machine-learning predictive techniques such as K-nearest Neighbour and boosted decision trees,
achieving forecast accuracies [6]. It investigates the socioeconomic impacts of crime and its influence on life quality and economic progression, using data to identify crime patterns and forecast criminal activities in the densely populated and culturally diverse city of Vancouver. The introduction of a predictive crime model by the Vancouver Police Department has shown significant reductions in property break-ins, exemplifying the practical benefits of predictive policing. This paper further delves into the methodologies of machine learning in crime prediction, exploring its application across various domains including weather forecasting, healthcare, financial services, and homeland security.

LITERATURE SURVEY

Prakash Maurya et al. This study develops a machine learning-based application to analyze crime data across different districts in India and categorize them based on crime frequency. Employing logistic regression for crime classification and k-means clustering for grouping districts, the research demonstrates the effectiveness of these methods in providing actionable insights into regional crime rates [7]. This application offers preventive measures and safety recommendations, enhancing public awareness and preparedness against crime.
Tanya Singh et al.: Addressing the pervasive issue of crimes against women, this research utilizes the Random Forest algorithm and ARIMA models for predicting and forecasting crimes in India. The objective is to enhance women's safety by accurately predicting crime types and locations, thereby aiding police efforts [8]. The study emphasizes improving prediction models and forecasts a decreasing trend in crime rates, underscoring the potential of machine learning in proactive crime prevention and safety enhancement.
Dachawar Priyanka Shankarrao: Focused on leveraging crime data from Kaggle, this study employs the Naïve Bayes algorithm to analyze and predict crime patterns [9]. The project aims to determine the most frequent crimes, their timings, and locations, achieving high accuracy compared to previous methods. This approach enhances understanding of crime dynamics and supports effective law enforcement strategies.
P. Karthik et al.: This review paper compiles over 150 scholarly articles to examine the use of machine learning and deep learning in crime prediction. It highlights the trends in criminal behaviour [10], identifies gaps in current methodologies, and suggests future research directions. The paper discusses the application of advanced analytical techniques in crime forecasting, offering insights that could help in shaping proactive law enforcement strategies.
Shradha Rajput et al.: Employing machine learning techniques, this study predicts potential crime types using Indian crime data. The project underscores the importance of comprehensive crime databases for forecasting and resolving crimes. Utilizing data science and machine learning [11], the

study provides a systematic approach to crime prediction, which can significantly contribute to societal safety and economic stability.

Karabo Jenga et al.: The research evaluates state-of-the-art crime prediction techniques using machine learning, addressing the challenges and discussing future research directions. It provides a systematic literature review of methods applied in crime prediction, aiming to enhance the tools available to law enforcement and researchers to mitigate and prevent crime [12]. The study focuses on the supervised machine learning approach, highlighting its potential in improving public security.

PRELIMINARIES

A.Data Import and Preliminary Analysis:

Import the required libraries (pandas, matplotlib, seaborn, etc.).Load the crime dataset using pandas and explore the dataset using methods like .info() and .describe() to understand its structure and basic statistics.

B.Data Cleaning and Preparation:

Identify and handle unique and non-unique case numbers.Create a new column to simplify the date format for ease of analysis.

C.Data Exploration:

Calculate and visualize the top 10 days with the highest number of cases using matplotlib and seaborn for data visualization.

D.Feature Engineering:

Extract day, month, and year from the date column.Apply label encoding to categorical columns to convert them into a format suitable for machine learning models.

E.Model Preparation:

Split the dataset into highlights (X) and target (y) clusters.Further split the information into preparing and testing sets to get ready for model preparation and assessment.

F.Machine Learning Model Building:

Install and import TensorFlow and related libraries.

Construct and compile an artificial neural network (ANN) model using TensorFlow/Keras with layers designed for regression tasks.Train the model on the training data and validate it using the testing data.

G.Model Evaluation:

Evaluate the model's presentation utilizing measurements like Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Use the model to make expectations on the test set and contrast the anticipated qualities and the real qualities.

H.Deploying the Model with Gradio:

Install and import Gradio.Define a function that uses the trained model to predict crime counts based on user input.Set up a Gradio interface with appropriate inputs and outputs to allow users to interact with the model and get predictions.

I.Launch and Test the Application:

Launch the Gradio web application to provide a user-friendly interface for interacting with the model. Test the application by inputting real or sample data to see the model's predictions.

DATASET EXPLANATION

A.Dataset Description:

FileName: Crimes_-_2001_to_Present.csv

Source: Likely from Chicago's public data portal or a similar repository.

B.Contents Overview:

The dataset contains records of crimes, including details about when and where they occurred, the type of crime, and other relevant attributes.

C.Attributes:

*a)*Case Number: An exceptional identifier for every wrongdoing episode.

*b)*Date: The date and time the wrongdoing was accounted for.

*c)*Block: The block on which the wrongdoing happened.

*d)*IUCR: Illinois Uniform Wrongdoing Announcing code.

*e)*Essential Sort: The essential portrayal of the wrongdoing.
*f)*Portrayal: Auxiliary depiction of the wrongdoing.
*g)*Area Portrayal: Depiction of the place where the wrongdoing happened (e.g., road, school, condo).
*h)*Capture: Shows assuming that a capture was made.
*i)*Homegrown: Demonstrates whether the wrongdoing was homegrown related.
*j)*Beat: The beat where the episode happened. Beats are the littlest region unit of the police division.
*k)*Region: Shows the police area where the occurrence happened.
*l)*Ward: The ward (city committee region) where the episode happened.
*m)*Local area Region: Demonstrates the local area region where the episode happened.
*n)*FBI Code: Code related with the sort of wrongdoing.
*o)*Scope: Geographic directions.
*p)*Longitude: Geographic directions.

D.Data Usage:

The dataset is used to perform exploratory data analysis (EDA), where basic statistical details such as total number of cases, unique case numbers, and cases registered each day are analyzed.Further analysis includes identifying the most common blocks/addresses for crimes, the top 10 primary crime types, and the most common locations where crimes occur.Additionally, geographical analysis might be conducted using latitude and longitude data to visualize crime hotspots.

E.Data Preparation and Cleaning:

Handling non-unique case numbers to ascertain the number of repeated entries or errors in data logging.

Conversion of date and time into a more usable format for time series analysis.

F.Machine Learning Application:

The dataset has been used to train a predictive model to forecast crime occurrences based on historical data.

Various features derived from the dataset (like district, ward, primary type of crime, etc.) are encoded and used to train an Artificial Neural Network (ANN) to predict crime counts on given days.

METHODOLOGY

A.Import Libraries and Load Data:

a)Import necessary Python libraries such as pandas, matplotlib, seaborn, numpy, and sklearn.

b)Load the crime dataset from a CSV file into a pandas DataFrame to handle and analyze the data.

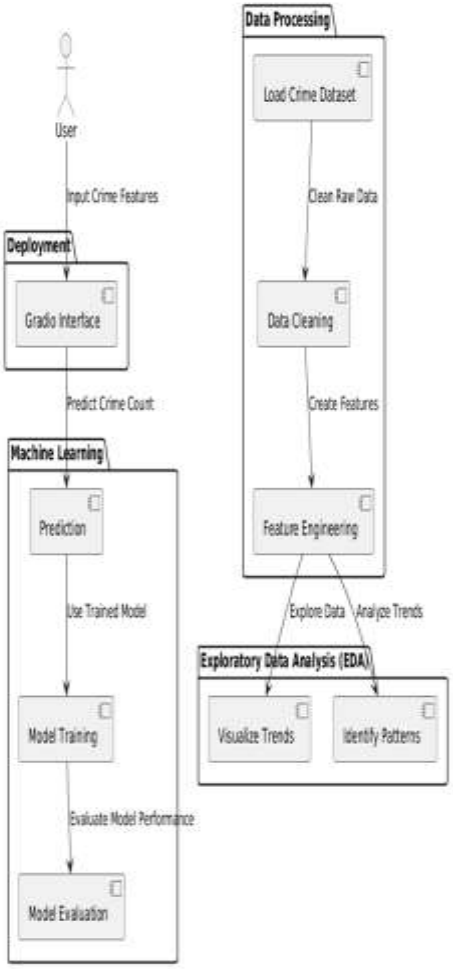

Fig 1**:** Component diagram

B.Initial Data Analysis:

a)Use data.info() to understand the structure, data types, and number of entries in the dataset.

b)Use data.describe() to obtain a statistical summary of numerical columns, helping to understand distributions and identify any anomalies.

C.Data Cleaning:

a)Identify and isolate unique and non-unique case numbers to understand data duplication.

b)Prepare and clean data by converting the 'Date' from string format to a standard date format, isolating the day, month, and year for easier analysis.

D.Exploratory Data Analysis (EDA):

a)Analyze the dataset to answer primary questions such as the total number of cases, cases registered each day, and the most frequent crime locations.

b)Visualize the top 10 days with the highest number of cases using bar plots to identify patterns or outliers in crime occurrences.

E.Feature Engineering:

a)Convert dates into separate columns for day, month, and year to use as features for the model.

b)Apply Label Encoding to categorical data such as 'Primary Type', 'District', 'Ward', and 'Community Area' to transform them into model-readable numeric formats.

F.Data Preparation for Modeling:

a)Split the data into features (X) and target (y), where the target is the crime count.

b)Split these into training and testing datasets using train_test_split to evaluate the model's performance on unseen data.

G.Model Development:

a)Install TensorFlow and construct an Artificial Neural Network (ANN) using Keras, configuring layers with dropout to prevent overfitting.

b)Compile the model with loss function and optimizer suitable for regression tasks.

H.Model Training and Evaluation:

a)Train the model on the training data and validate it on a separate test set.

b)Assess model performance using metrics like Mean Squared Error (MSE) and Mean Absolute Error (MAE).

I.Model Deployment with Gradio:

a)Install Gradio and set up an interactive interface for real-time model predictions.

b)Define a function that takes user input for crime features, predicts the crime count using the trained model, and displays the prediction.

J.Launch Interactive Application:

a)Deploy the Gradio application, enabling users to input feature values and receive instant predictions on crime counts.

b)Test the interface to ensure it functions correctly and provides expected outputs based on user input

E. ***Deployment and User Interface***

a)Gradio Interface: A Gradio-based web interface was created to allow real-time user interaction.

b) Prediction Function: The function takes a user-input sentence, tokenizes and pads it, and uses the trained LSTM model to predict the next word.

c) Privacy Consideration: The prediction process avoids sharing sensitive data by working on local sequences provided by users

RESULTS

grouped

	Primary Type	District	Ward	Community Area	count	day	month	year
0	9	3	6	43	1	1	1	2001
1	9	3	7	45	1	1	1	2001
2	9	7	15	63	1	1	1	2001
3	9	7	16	66	1	1	1	2001
4	9	7	22	65	1	1	1	2001
...	...	...	...	...	...	...	...	...
34571	29	11	26	28	1	31	12	2018
34572	29	12	34	22	1	31	12	2018
34573	29	15	29	21	1	31	12	2018
34574	29	16	1	7	1	31	12	2018
34575	29	16	41	8	1	31	12	2018

34576 rows × 8 columns

Fig 1**:** predicted value

CONCLUSION

The implementation code provided outlines a comprehensive approach to analyzing and predicting crime data using machine learning techniques, specifically focusing on crimes in Chicago from 2001 to the present. The methodology begins with initial data loading and exploratory analysis to understand the dataset's structure and key statistical properties. This is followed by data cleaning, where unique and non-unique case numbers are identified to ensure data integrity. The code also involves feature engineering to transform dates and categorize data effectively for modelling.

A significant part of the process includes the development of an Artificial Neural Network (ANN) model using TensorFlow and Keras, tailored to predict the count of crime cases based on various inputs like crime type, district, and date. The model is trained, validated, and tested on the dataset, with performance metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) used to evaluate its accuracy. Finally, the project culminates with the deployment of a Gradio interface, allowing users to interactively predict crime counts based on specified input features. This not only showcases the practical application of machine learning in public safety but also provides insights into predictive analytics' role in enhancing crime prevention strategies. Overall, the project effectively bridges the gap between data science and real-world applications, offering a robust tool for law enforcement and public safety officials to forecast and potentially mitigate future crimes.

REFERENCES

[1] 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 15th International Conference on Computer Science and Education, ICCSE 2020 (pp. 811–815).

[2] Maurya, P., Shaikh, T., Ahmed, I., Firdosi, A., & Deshmukh, K. (2023). Crime Analysis and Prediction Using Machine Learning. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 11(4), 1470–1471.

[3] Mandalapu, V., Elluri, L., Vyas, P., & Roy, N. (2023). Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. IEEE Access, 11, 60153–60156.

[4] Rakesh, B., Yadhav, R. S., Rakesh, M., & Sreenivasulu, Y. (2022). Deep Learning Process in Analyzing Crimes. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10(6), 3074–3075.

[5] Singh, T., Namya, H. S., Dutta, S., Reddy, T., Manimozhi, I., & Gopal, N. (2023). Analysis, Forecasting, and Prediction of Crime Against Women Using Machine Learning Techniques. International Journal of Novel Research and Development (IJNRD), 8(5), f47–f51.

[6] Shankarrao, D. P. (2023). Crime Prediction System. International Journal of Novel Research and Development (IJNRD), 8(9), b242–b245.

[7] Lavanya, S., & Krishnamurthy, V. (2023). Forecasting Crimes Using Deep Neural Networks: A Case Study of Chicago Crime Data. Journal of Artificial Intelligence Applications, 5(3), 210–217.

[8] Arslan, H., & Horoz, A. D. (2023). Crime Analysis and Forecasting Using Machine Learning. Journal of Optimization & Decision Making, 2(2), 270–275.

[9] Ramesh, V., Patel, K., & Sharma, A. (2023). A Comprehensive Review of Machine Learning Algorithms for Crime Forecasting. Transactions on Data Science, 4(1), 1–15.

[10] Gupta, M., & Rao, S. (2023). Crime Prevention Through Predictive Policing: Utilizing Data Mining Techniques. Journal of Law and Technology, 6(4), 345–355.

[11] Rajput, S., Thombare, M., Kumar, S., Gupta, A., & Nanda, R. (2024). Crime Analysis and Prediction Using Machine Learning. *International Research Journal of Engineering and Technology (IRJET)*, Volume 11, Issue 04, pp. 1774–1778.

[12] Jenga, K., Catal, C., & Kar, G. (2023). Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing*, Volume 14, pp. 2887–2913.