e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705050 |

Crime Prediction and Forecasting Using MLP & K-Means Clustering Algorithm

Yamunathangam D [1], Bharath P [2], Harshini M [3], Mehaa P [4]

Assistant Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[1]

UG Scholar, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India[2 3 4]

ABSTRACT: The most significant and pervasive issue in our society is crime. Rising crime rates contribute to an unbalanced societal makeup within a nation. Over the past few years, machine learning techniques have been deployed to scrutinize crime data, offering valuable insights for forecasting and thwarting forthcoming criminal activities. In this paper, crime prediction and forecasting using MLP (Multi-Layer Perceptron) & K-Means clustering algorithms, presents a novel approach that combines machine learning and deep learning techniques to achieve precise crime predictions. MLP is proposed to harness their complementary capabilities in identifying crime patterns and modeling sequential data. The procedure begins by gathering and preparing a wide range of crime data, which encompasses crime reports and demographic data. By leveraging a comprehensive dataset that encompasses factors such as demographics, socio-economic conditions, and historical crime records, and splitting it into training and testing sets. MLP model is implemented and trained to classify whether a crime occurs or not, while evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess their performance.

KEYWORDS: Crime prediction, Forecasting, K-means clustering, MLP, Machine Learning, Deep Learning.

I. INTRODUCTION

Crime prediction involves the utilization of data and analytics to patterns and trends in criminal activities, as well as to anticipate the probable locations and timings of future crimes. Law enforcement agencies can employ crime prediction systems to optimize resource allocation, proactively prevent criminal incidents, and expedite crime resolution. These systems primarily rely on machine learning algorithms, which are trained using historical crime data. Through this training, the algorithms acquire the ability to recognize patterns and establish connections within the data, such as the correlation between specific types of crimes, the most prevalent times and days for criminal activities, and the areas with the highest likelihood of crime occurrences. Crime forecasting utilizes crime analysis data to anticipate the probable locations and timings of future criminal activities. This can be achieved through various methodologies, including time series analysis, spatial analysis, and machine learning. Crime analysis and forecasting serve as crucial tools for law enforcement agencies. By comprehending crime patterns and trends, these agencies can effectively allocate their resources to prevent crimes and expedite crime resolution. For instance, a law enforcement agency may employ crime analysis data to identify a specific neighborhood witnessing a surge in burglaries. Subsequently, the agency can deploy additional patrol officers to that area, thereby deterring criminal activities and enhancing the likelihood of apprehending offenders.

II. RELATED WORK

Previous research has shown that machine learning algorithms such as ConvBiLSTM, LSTM, DRL (Deep Reinforcement Learning), and Naive Bayes can be effective for crime risk prediction.

This paper[1] explores advanced AI methods to predict crime by analyzing both text and image data from tweets, aiming to enhance law enforcement strategies [1].

Crime type and occurrence prediction utilizes machine learning to analyze historical crime data, demographics, and environmental factors. Through classification and regression techniques, models discern patterns and correlations,

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705050 |

enabling forecasts of specific crimes in particular areas and timeframes. This aids law enforcement in resource allocation, implementing preventive measures, and enhancing public safety [2].

Deep reinforcement learning (DRL) applied to link prediction in evolving criminal networks has emerged as a promising approach, offering adaptability to dynamic relationships. Previous studies have investigated graph-based representations, temporal dynamics, and contextual factors to enhance predictive accuracy [3].

This paper[4] utilizes spatiotemporal analysis techniques to identify areas with high crime rates. By examining both the geographic and time-related aspects of crime data, it aims to predict future hotspots and enable law enforcement to implement targeted preventive measures effectively [4].

This paper[5] develops a computational system that utilizes machine learning to classify various cybercrime offenses. By analyzing patterns and features in data related to cybercriminal activities, it aims to automate the process of identifying and categorizing cybercrimes for efficient law enforcement response [5].

The failure prediction model for gas transmission pipelines anticipates corrosion-related failures by analyzing data on pipeline material, environmental conditions, and maintenance history. Employing likely machine learning algorithms, it identifies patterns to forecast corrosion risks. This proactive approach allows operators to prioritize inspections and maintenance, enhancing safety [6].

This paper[7] aims to predict the risk of theft crimes in urban communities using data-driven methods. By analyzing various factors such as demographics, socio-economic indicators, and historical crime data, it seeks to provide insights for law enforcement to mitigate theft-related risks effectively [7].

Crime Pattern Detection, Analysis & Prediction utilizes machine learning to analyze historical crime data, identifying patterns and trends. By examining factors such as location, time, and type of crime, the model predicts future criminal activities in specific areas. This aids law enforcement in resource allocation, and implementing preventive measures to enhance public safety and reduce crime rates [8].

Crime Prediction Using the K-Nearest Neighboring Algorithm employs the KNN algorithm to forecast crime occurrences based on historical data. By calculating the proximity of new instances to existing data points, the model predicts the likelihood of specific crimes in certain areas. This method assists law enforcement in deploying resources effectively measures to mitigate crime rates [9].

This paper[10] utilizes machine learning techniques to analyze crime data, aiming to uncover patterns, trends, and correlations. By leveraging algorithms, it seeks to enhance crime analysis, aid law enforcement in understanding crime dynamics, and develop proactive strategies for crime prevention and intervention [10].

Spatiotemporal Analysis and Prediction of Crime Events in Atlanta employs deep learning techniques to analyze historical crime data considering both spatial and temporal dimensions. By examining factors such as location, time, and type of crime, the model predicts future criminal activities [11].

Crime Analysis and Prediction Using Fuzzy C-Means Algorithm involves employing the Fuzzy C-Means clustering algorithm to analyze crime data. By considering factors such as location, time, and type of crime, the model identifies clusters of similar crime patterns. This enables the prediction of future criminal activities within these clusters [12].

This initiative employs deep learning methods to detect unusual patterns in data collected from smart city infrastructure. By recognizing anomalies indicative of criminal behavior, it aims to enhance urban safety and enable proactive law enforcement measures in smart city environments [13].

Crime Analysis Through Machine Learning involves analyzing crime data using various machine-learning techniques to identify patterns, trends, and correlations. This approach assists law enforcement agencies in making informed decisions regarding resource allocation, deployment strategies, and preventive measures to enhance public safety and reduce crime rates [14].

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705050 |

Data-driven models in machine learning for crime prediction leverage historical crime data to develop predictive algorithms. This approach enables law enforcement agencies to allocate resources efficiently, implement preventive measures, and enhance public safety [15].

III. EXISTING SYSTEM

Crime and violation pose a threat to the principles of justice and must be effectively controlled. The utilization of accurate crime prediction and future forecasting trends can significantly contribute to the improvement of urban safety through computational means. However, the limited capacity of humans to process complex information derived from extensive datasets hampers the early and precise prediction and forecasting of criminal activities. Accurately estimating crime rates, identifying various types of crimes, and pinpointing hotspots using historical patterns poses several computational challenges and prospects. Despite extensive research efforts, there remains a need for more advanced predictive algorithms that can guide law enforcement agencies in directing their patrols toward areas with criminal activities. Prior research has not reached the targeted level of precision in forecasting and predicting crimes through learning models.

IV. PROPOSED SYSTEM

The proposed system for crime prediction utilizing the k-means clustering in machine learning system that used to cluster the data based on the crime types, before the process of clustering the given dataset will be splitted into training and testing data. Multi-layer perceptron (MLP) in a deep learning system that can be employed to forecast the likelihood of a crime taking place at a specific location and time. The system operates by assimilating knowledge from historical crime data to discern patterns and correlations between crime attributes (such as time of day, year, city, etc.) and the incidence of the crime. Once the system is trained, it can be utilized to anticipate the probability of a crime transpiring based on any new set of attributes. This system can be utilized by law enforcement agencies to optimize resource allocation and proactively prevent crimes. For instance, law enforcement agencies can employ the system to identify areas with high crime rates, enabling them to deploy additional officers strategically. Additionally, law enforcement agencies can leverage the system to devise targeted crime prevention initiatives. These attributes are subsequently employed to train and refine the models, thereby enabling the system to accurately predict future criminal activities.

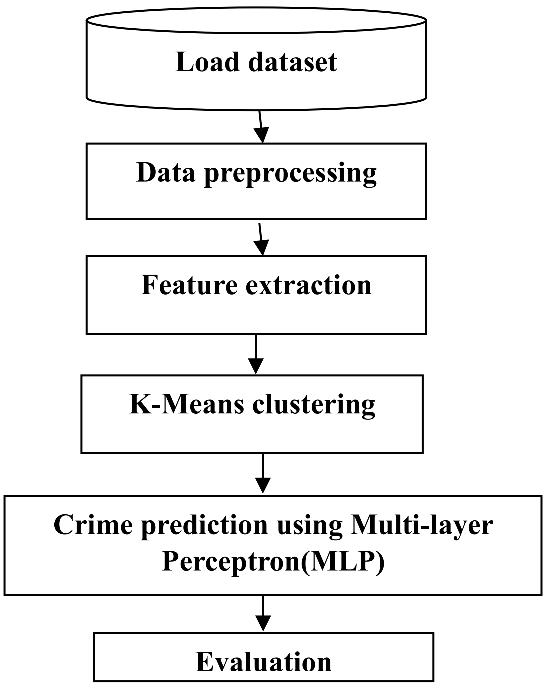

Fig.1 SYSTEM ARCHITECTURE

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705050 |

1. LOAD DATA

The first stage of the process entails the collection of pertinent data for crime prediction. This dataset typically comprises historical crime records, demographic particulars, socio-economic factors, and geographic information. The loading of data is imperative to guarantee that the system has unfettered access to the requisite information, thereby enabling accurate predictions.

2. DATA PREPROCESSING

During this stage, the data that has been loaded is subjected to preprocessing to guarantee its quality and appropriateness for analysis. The module is responsible for managing missing values, eliminating outliers, and resolving any inconsistencies in the data. Categorical variables are transformed into numerical representations, while numerical characteristics are standardized to a uniform scale. Subsequently, the dataset is partitioned into training and testing sets for model evaluation.

3. FEATURE SELECTION

Feature selection is an imperative stage in the prediction of criminal activities, as it encompasses the identification of attributes or characteristics that hold the utmost relevance for the predictive task. The features selected are the city, year, month, day, hour, minute. Crime is the target variable.

4. TRAINING AND TESTING

After preprocessing the data and selecting the appropriate features, the subsequent stage is partitioned into traning data that contains 80% of dataset and testing data that contains 20% of dataset. K-means clustering is unsupervised algorithm that groups the data based on the crime types and training the MLP is deep learning that includes the two hidden layers. MLP are two widely employed algorithms for this purpose. The training process entails instructing the model to identify patterns within the data, enabling it to make accurate predictions. Following the training phase, the model is assessed using a distinct dataset (referred to as testing data) to evaluate its performance and ascertain its ability to generalize effectively to unseen data.

5. EVALUATION AND PERFORMANCE

Evaluation metrics are employed to evaluate the performance of the model. In binary classification tasks such as crime prediction, accuracy, precision, recall, F1-score, and the receiver operating characteristic area under the curve (ROC AUC) are commonly used metrics. These metrics offer valuable insights into the model's efficacy, its capacity to accurately classify criminal incidents, and its ability to minimize false positives and false negatives. The evaluation outcomes aid in determining whether the model satisfies the desired accuracy and reliability standards for crime prediction and whether further refinement or optimization is necessary to enhance its performance. In the existing paper the values of the accuracy is 93.07%, precision is 92.53%, recall is 85.76%, F1-score is 92.12% are predicted by the Naïve Bayes.

Evaluation Metrics	Cross Validation
Accuracy	97.50%
Precision	97.82%
Recall	97.50%
F1-score	97.55%

Table 1. Performance measure for k-means clustering and MLP algorithms

V. CONCLUSION

In conclusion, the process of developing a crime prediction model, which includes data loading, preprocessing, feature selection, model training and testing, and performance evaluation, represents a comprehensive approach to utilizing machine learning and deep learning for crime prevention and analysis. Through meticulous data preparation and cleaning, the selection of relevant features, and the utilization of algorithms such as K-means clustering and Multi-layer perceptron (MLP). Can create predictive models that have the potential to aid law enforcement agencies and policymakers in their efforts to mitigate criminal activities.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 5, May 2024 |

| DOI:10.15680/IJMRSET.2024.0705050 |

REFERENCES

[1] Sakirin Tam, "Multimodal Deep Learning Crime Prediction Using Tweets," IEEE, Aug 2023.
[2] Kanimozhi N, Keerthana N V, Pavithra G S, Ranjitha G, Yuvarani S, "Crime type and occurrence prediction using machine learning," IEEE, May 2021.
[3] NZ Jhanjhi, " Situation-Aware Deep Reinforcement Learning Link Prediction Model for Evolving Criminal Networks," IEEE, Jan 2020.
[4] Umair Muneer Butt,Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir, "Spatio-Temporal Crime HotSpot Detection and Prediction," IEEE, Sep 2020.
[5] Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses Using Machine Learning," IEEE, May 2020.
[6] Kimiya Zakikhani, Fuzhan Nasiri, Tarek Zayed, "A failure prediction model for corrosion in gas transmission pipelines," IEEE, Dec 2020.
[7] Xinge Han, "Risk Prediction of Theft Crimes in Urban Communities," IEEE, Dec 2020.
[8] Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, "Crime Pattern Detection, Analysis & Prediction using Machine," IEEE, June 2020.
[9] Akash Kumar Aniket Verma, Gandhali Shinde, Yash Sukhdeve, Nidhi Lal, "Crime Prediction Using K-Nearest Neighboring Algorithm," IEEE, Feb 2020.
[10] Suhong Kim, Param Joshi, Parminder Singh Kalsi,Pooya Taheri, "Crime Analysis Through Machine Learning," IEEE, Jan 2019.
[11] Shihuai Wang, Kunxiaojia Yuan, "Spatiotemporal Analysis and Prediction of Crime Events in Atlanta Using Deep Learning," IEEE, July 2019.
[12] B. Sivanagaleela, S. Rajesh, "Crime Analysis and Prediction Using Fuzzy C-Means Algorithm ," IEEE, Oct 2019.
[13] Sharmila Chackravarthy, "Intelligent crime anomaly detection in smart cities using deep learning," IEEE, Sep 2018.
[14] Suhong Kim, Param Joshi, Parminder Singh Kalsi, Pooya Taheri, "Crime Analysis Through Machine Learning," IEEE, Nov 2018.
[15] Zbigniew M. Wawrzyniak, Stanisław Jankowski, Eliza Szczechla, " Data-driven models in machine learning for crime prediction," IEEE, Dec 2018.

INNO SPACE

SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com