IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205016 |

Crime Prediction Using Machine Learning

Rakhi Waykole[1], Aditi Naik[2], Sarika Patil[3], Dr Akash D. Waghmare[4]

UG Student, Department of Computer Engineering, SSBT's College of Engineering & Technology Jalgaon, Maharashtra, India[1,2,3]

Associate Professor, Department of Computer Engineering, SSBT's College of Engineering & Technology Jalgaon, Maharashtra, India[4]

ABSTRACT: Crime analysis and prediction is a systematic approach for identifying the crime. The project focuses on predict region which have high probability for crime occurrences and visualize crime prone area. Using the concept of machine learning we can extract previously unknown, useful information from an unstructured data. By collecting and preprocessing a comprehensive dataset, and employing algorithm like random forest, the project aims to develop an accurate predictive model. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problem faced worldwide. Crimes affect the quality of life, economic growth and reputation of nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. The project focuses on propose a system which can analysis, detect, and predict various crime probability in given region. The project explains various types of criminal analysis and crime prediction using several machine learning techniques.

KEYWORDS: Crime analysis and prediction, Machine Learning, Random Forest, Comprehensive Dataset.

I. INTRODUCTION

Day by day crime data rate is increasing because the modern technologies and hi-tech methods helps the criminals to achieving the illegal activities. According to the Crime Record Bureau, there has been an increase in crimes such as burglary and arson, as well as crimes like murder, sexual abuse, gang rape, and others. Crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the Machine Learning techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime. Machine learning helps in solving the crimes faster and this technique gives good results when applied on crime dataset, the information obtained from the data mining techniques can help the police department. A particular approach has been found to be useful by the police, which is the identification of crime 'hot spots' which indicates areas with a high concentration of crime. The occurrences of crime depended on several factors such as intelligence of criminals, security of a location, etc. The work has followed the steps that used in data analysis, in which the important phases are Data collection, data classification, pattern identification, prediction and visualization. The proposed framework uses different visualization techniques to show the trends of crimes and various ways that can predicts the crime using machine learning algorithm.

II. RELATED WORK

Amar Shukl et al., in [1], the issue of crime is addressed through the application of various machine learning techniques. The primary objective is to minimize crime rates by identifying crime patterns and predicting the probability of criminal activities. Focused on crime datasets from North Carolina, the study employs mathematical and statistical models for analysis. Methodology encompasses data preprocessing, exploratory data analysis, feature selection, and model testing. Noteworthy findings include the identification of factors such as population density and urbanity highly correlated with crime rates. The paper concludes that the developed models can aid in crime prevention and contribute to creating safer communities. However, it underscores the necessity for standardized datasets and further research in understanding crime behaviors.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205016 |

Alkesh et al., in [2], the project aims to predict and analyze crime using machine learning techniques, particularly focusing on Chicago crime data. The project involves preprocessing the data, selecting relevant features, training various machine learning models, and visualizing the dataset through graphs. After testing different algorithms, the K-Nearest Neighbor (KNN) algorithm yielded the highest accuracy for crime prediction. The results showed insights into crime patterns, arrest ratios, and crime occurrence across different locations and times. Overall, machine learning technology facilitates crime prediction, analysis, and potential reduction in crime rates.

Lawrence et al., in [3], the research paper explores the application of machine learning algorithms to analyse crime data, focusing on violent crime patterns. The study compares algorithms like Linear Regression, Additive Regression, and Decision Stump using datasets from the University of California-Irvine repository and actual crime statistics for Mississippi. Linear regression proves to be the most accurate algorithm for predicting crime patterns. The paper highlights the importance of data applications for law enforcement.mining in crime detection and prevention, offe ring insights into crime trends and potential

Shrinath et al., [4], the research explores the application of machine learning algorithms for crime prediction and prevention. It emphasizes the importance of leveraging artificial intelligence to analyze crime data and enhance law enforcement efforts. By employing various techniques such as data mining, statistical modeling, and neural networks, the study aims to predict criminal behavior and identify strategies for crime deterrence. The research highlights the significance of technology in addressing contemporary security challenges and improving public safety.

Manish et al., [5], India's high population density, coupled with factors like unemployment, poverty, and illiteracy, contribute to varying crime rates across states. The National Crime Record Bureau data shows a steady increase in crime rates, particularly in urban areas. A study suggests that time series models effectively forecast crime rates, with recent data falling within the 95% confidence interval. Future research aims to analyze crimes against women and children to inform strategies for reducing crime rates

III. PROPOSED SYSTEM ARCHITECTURE

In the initial work on holistic model feature extraction for data selection and a KNN algorithm has been used to predict crimes. The Indian Crime dataset is used for identifying types of crimes rates and hotspot areas at a particular region. Proposed system architecture is depicted in Figure 1 with a detailed description of each step.

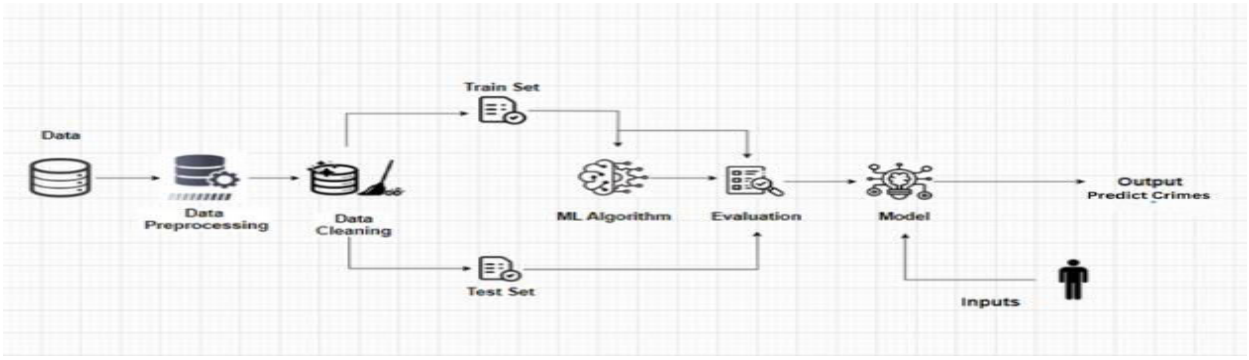

Figure 1: System architecture for Crime Prediction using Machine Learning

Data Collection: Gather various types of data relevant to crime prediction, including historical crime data (e.g., types of crimes, locations, times), demographic data (e.g., population density, income levels), environmental data (e.g., weather conditions, time of day), and possibly social media data (e.g., posts related to crime activities).

Data Preprocessing: Cleanse and preprocess the collected data. This may involve handling missing values, outlier detection, normalization, and feature engineering to extract meaningful features from raw data.

Feature Selection: Select the most relevant features for crime prediction. This step helps reduce the dimensionality of the data and improves the efficiency of the machine learning algorithms.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205016 |

Machine Learning Model Selection: Choose appropriate machine learning algorithms for crime prediction. This could involve supervised learning algorithms such as logistic regression, decision trees, random forests, support vector machines (SVM), or more advanced techniques like deep learning models (e.g., convolutional neural networks, recurrent neural networks) if dealing with complex data.

Model Training: Train the selected machine learning models using the preprocessed data. This involves splitting the data into training and validation sets, tuning hyperparameters, and evaluating the models using appropriate metrics (e.g., accuracy, precision, recall).

Model Deployment: Deploy the trained models into a production environment where they can make real-time predictions. This could involve integrating the models into a web application, mobile app, or API.

Scalability and Efficiency: Ensure that the system architecture is scalable and efficient to handle large volumes of data and real-time prediction requests. This may involve using distributed computing frameworks (e.g., Apache Spark) and optimizing the performance of machine learning algorithms.

Monitoring and Evaluation : Regularly monitor the system's performance and evaluate its effectiveness in predicting crime rates. This involves analyzing prediction accuracy, assessing the impact of predictive policing on crime prevention, and making adjustments to the system as needed.

IV. ALGORITHM DESIGN

In our project on "Crime Prediction Using Machine Learning," we have implemented a KNN (K-Nearest Neighbour) model for recognising crime hotspot areas and crime rates at various regions. In the context of sequence prediction or sequential data analysis, the utilization of KNN (K-Nearest Neighbour) algorithms holds significant promise. KNN is one of the simplest Supervised Learning technique. KNN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.KNN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to available categories .

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Among these k neighbors, count the number of the data points in each category. • Step-5: Assign the new data points to that category for which the number of the neighbor is maximum.
- Step-6: Final model is ready.

Algorithm for KNN Classification is as follows:
Input:

1. A finite set D of points to be classified,
2. A finite set T of points,
3. A function c: T –>{1,...,m},
4. A natural number k. Output: A function r: D{1,...,m}

Method:

1. Begin
2. For each x in D do
3. Let U <–{ }
4. For each t in T add the pair (d(x,t) , c(t)) to U;
5. Sort the pairs in U using the first components;
6. Count the class labels from the first k elements from U;
7. Let r(x) be the class with the highest number of occurrence;
8. End For each

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205016 |

9. Return r
10. End
11. Thus proposed system will predict the crime and their patterns of occurrences that security can be provided for those areas. It can be applicable for particular region.

V. RESULTS AND DISCUSSIONS

The project aims to provide a a flexible system that can be easily used by user. The results of a Crime Prediction system typically depend on the accuracy of the model used and the data it's trained on. A successful system would provide reliable predictions, aiding in detection. Below are the results after completion of the proposed system:

Figure 2 displays a graphical representation where the y-axis denotes the names of states, and the x-axis represents accuracy. The plotted data illustrates the crime rate across various states. Such graphs serve to provide users with a clear understanding of the crime rates in different regions or states.

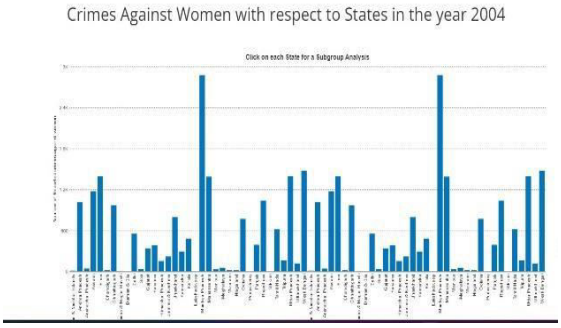

Figure 2Crime against women with respect to states

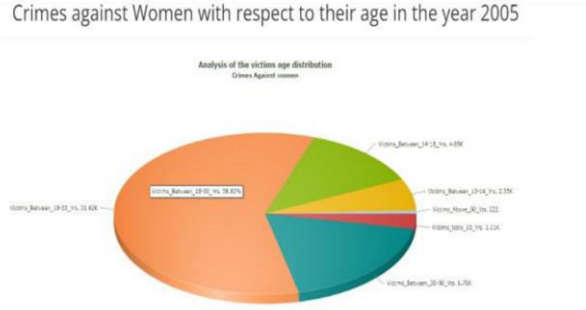

Figure 3: Crime against women with respect to age

Figure 3 displays a pie chart which represents the age group of criminals involved in crime during the specific year. This helps the crime department to judge which age group commits certain crime frequently and this will also help the government to take necessary steps.

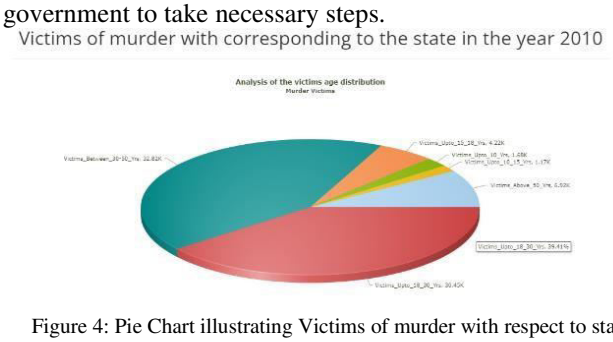

Figure 4: Pie Chart illustrating Victims of murder with respect to states

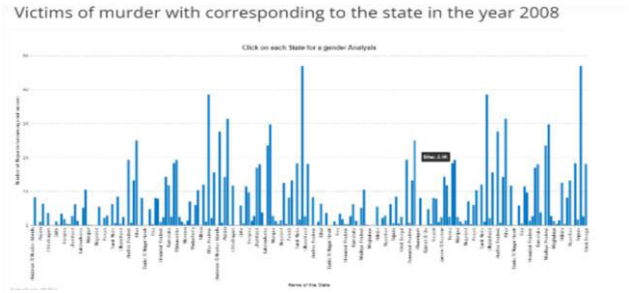

Figure 5: Victims of murder according to states

Fig 4 depicts the pie chart which illustrates murder victims categorized by age group, providing a valuable tool for controlling the crime rate more effectively.

Fig 5 displays a graphical representation where the y-axis denotes the names of states, and the x-axis represents accuracy. The plotted data illustrates the crime rate across various states. Such graphs serve to provide users with a clear understanding of the crime rates in different regions or states.

The implementation has done on open-source python environment. The device runs with an INTEL Pentium 4 or higher processor and 4 GB RAM with a distributed manner on the Python 3.8, Jupyter analytic platform. Indian Crime dataset has used for crime prediction analysis. For the validation of results. The below Figure 6 demonstrates the classification accuracy of the proposed system and comparative analysis with various state-of-art systems. In this proposed system, K-Nearest Neighbour algorithm is used because it has more accuracy than other algorithms.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205016 |

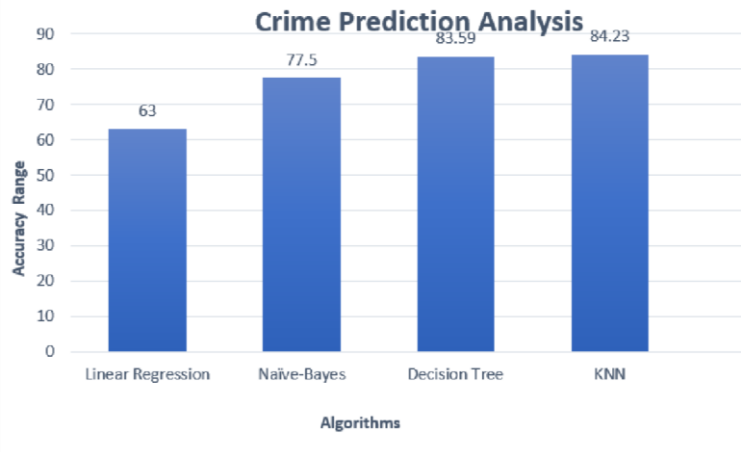

Figure 6: Comparative analysis of accuracy for crime prediction using proposed algorithm with other machine learning algorithms

To improve classification accuracy, a variety of feature extraction approaches have been used to build a solid training module. In the proposed work, we extract the numerous features such as age, gender, location, types of crimes and feed them to classify for validation. The proposed classifier has been used for crime prediction and gives highest accuracy, up to 84.23%.

VI. CONCLUSION AND FUTURE WORK

It is clear that basic details of criminal activities in a neighborhood contain indicators that will be employed by machine learning agents to classify a criminal activity given a location and date. The training agent suffers from imbalanced categories of the data set; it had been ready to overcome the problem by oversampling and under-sampling the dataset. This paper presents a crime data prediction by taking the types of crimes as input and giving are in which these crimes are committed as output using Colab notebook having python as a core language and python provide inbuilt libraries such as Mathplotlib, CSV and Plotly through which the work will be completed faster and Scikit provides all the processes of how to use different libraries providing by the python. Results of prediction are different for different algorithms and the accuracy of K-Nearest Neighbour found to be good with the accuracy of 84.2%.As future work is to keep improving the system and make experiments with complete crime datasets. The accuracy can be improved by using a complex neural network such as the recurrent neural network. Also, the deep learning approach can be used to enhance the accuracy of the model.

REFERENCES

1. Amar Shukl, Avita Katal, Saurav Raghuvanshi and Shivam Sharma. "Criminal Combat: Crime Analysis and Prediction Using Machine Learning."2021 International Conference on Intelligent Technologies (CONIT).IEEE 2021.
2. Alkesh Bharati, Dr Sarvanaguru RA., ”Crime Prediction and Analysis Using Machine Learning” in International Research Journal of Engineering and Technology (IRJET). September 2018.
3. Lawrence McClendon, Natarajan Meghanathan. “Using Machine Learning Algorithms to Analyze Crime data”, Machine Learning and International Journal (MLAIJ) Vol.2, No.1, March 2015..
4. Shrinath Dalave, Prajakta Lawande, Prachi Panasare. "Crime Prediction Using Machine Learning Algorithms" 2022 International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) May 2022.
5. Manish Kumar, Athulya S and Mary Minu MB, ”Forecasting of Annual Crime Rate in India: A case Study”-2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details