International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | April 2024 www.irjet.net p-ISSN: 2395-0072

Crime Analysis and Prediction Using Machine Learning

Shradha Rajput[1], Minal Thombare[2], Sawan Kumar[3], Aachal Gupta[4], Dr. Radhika Nanda[5]

[1,2,3,4]Student, Department of CSE(IOT, Cybersecurity and Blockchain), Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

[5]Professor, Department Of CSE (IOT, Cybersecurity and Blockchain), Smt. Indira Gandhi College of Engineering, Ghansoli, Navi Mumbai, Maharashtra, India

---

Abstract - *Crime represents a significant and pervasive challenge in our society, making the prevention of criminal activities is a paramount responsibility. To address this, it is essential to maintain comprehensive records of all offenses and establish a database for future reference. The central issue at hand revolves around maintaining a dependable crime database and leveraging data analysis to assist in forecasting and resolving potential future crimes. The primary aim of this project is to assess a dataset containing various criminal incidents and predict the potential type of crimes that might occur in the future based on different factors. In this project, we will harness the power of machine learning and data science techniques for forecasting crimes using Indian crime data. Crime analysis and prediction involve a systematic approach to identifying patterns of criminal activity. This algorithmic approach can anticipate and delineate areas susceptible to criminal incidents. By employing machine learning, we have the capability to uncover valuable insights from unstructured data, revealing previously unknown information. The extraction of new insights relies on the analysis of current datasets. Crime is a grave and widespread societal issue that impacts individuals globally. It influences people's well-being, economic prosperity, and a nation's reputation. To protect our communities from crime, we must employ cutting-edge technology and innovative crime analytics techniques. We introduce a system capable of analyzing, identifying, and predicting various crime probabilities within a given location. This project delves into various forms of criminal analysis and crime prediction using machine learning methodologies.*

Key Words: Crime Datasets, Crime Prediction, Machine Learning, Prophet.

1.INTRODUCTION

The frequency of criminal activities is on the rise due to the continuous advancement of technology, which provides criminals with more sophisticated tools to carry out their unlawful actions. Various types of crimes such as burglary, arson, and others, as reported by the Crime Record Bureau, have seen an increase. This includes more severe offenses like murder, rape, abuse, and gang rape, among others. Crime-related data is gathered from a wide range of sources, including blogs, news websites, and online platforms. This extensive data is utilized to construct a comprehensive crime report database [8]. The insights derived through data mining techniques play a pivotal role in reducing crime by facilitating the identification of culprits and areas most affected by criminal activities [7].

Crime analysis serves as the initial phase in the examination of criminal activities. It involves the exploration, analysis, and identification of connections among various crimes and crime-related variables. The machine learning algorithm trains the data to make predictions based on the provided dataset [2]. We can train the data and create models to effectively analyze and predict the crime in a certain area. This analytical process aids in generating real-time statistics, queries, and maps. It also helps in determining whether a crime has taken place in a particular, well-defined location. The prediction of crime helps in the security of the area and thus lowers the crime rate increasing safety of the citizens.

1.1 OBJECTIVES

- The primary objective is to provide valuable insights into crime trends and patterns using historical crime data.
- It will also predict future crime locations, etc. on the basis of historical crime data
- The web application will display the crime rate in various forms such as graph, etc.
- It will also display the predictions in the same manner.

1.2 SCOPE

The primary goal of this crime evaluation is to successfully track crime rate in various areas. It will display crime rate of a particular area. This will help the police forces and other related defense forces to effectively track and stop crime in a particular area and eventually in the society. IT will also predict the crime rate of different types of crime using various types of machine learning algorithms. The prediction will be based on various historical crime data. This will also help the defense personnel to keep checking the area with high crime rate to prevent crimes in future.

2. LITERATURE REVIEW

Crime analysis and future prediction using machine learning have garnered significant attention from both the academic and law enforcement communities in recent years. These applications have the potential to revolutionize the way we approach crime prevention and public safety. A significant body of literature exists in this field, highlighting various methodologies, algorithms, and the integration of predictive analytics into web applications. Machine learning algorithms, such as support vector machines, decision trees, and neural networks, have been widely employed in the development of these web applications [1]. For instance, the methodology involved using a Logistic regression model for crime classification, followed by k-means clustering to group districts based on their crime rates, demonstrating the feasibility of this technology [2]. Additionally, deep learning techniques, particularly convolutional neural networks, have been employed for image-based crime prediction, further expanding the scope of ML applications in this domain [3]. The integration of spatial and temporal data has been a major focus in the literature. Researchers have investigated the correlation between environmental factors, urban development, and crime patterns, allowing the development of predictive models that consider not only historical data but also contextual information. Recent studies have also explored the fusion of real-time data, such as social media updates and weather conditions, to enhance prediction accuracy.

Use of Data mining helps to find hidden patterns in large crime datasets quickly and efficiently [4]. An essential aspect of these web applications is the interpretability of ML models. Researchers have explored methods to make these models more transparent and interpretable to law enforcement personnel and the public [5]. This includes research on explainable AI and feature importance analysis to understand the factors contributing to predictions, ensuring accountability and trust in the technology. Moreover, ethical considerations and potential biases in crime prediction algorithms have gained prominence in recent literature [6]. Scholars have emphasized the importance of fairness, accountability, and transparency in the development of these applications, addressing issues related to bias in historical crime data and the potential for reinforcing existing inequalities in law enforcement practices. Community engagement and collaboration have been another key area of research. Developing web applications that empower communities to participate in crime reporting and safety concerns fosters a more comprehensive and inclusive approach to crime prevention. Researchers have explored ways to facilitate information sharing and feedback mechanisms between law enforcement agencies and the public.

3. SYSTEM DESIGN

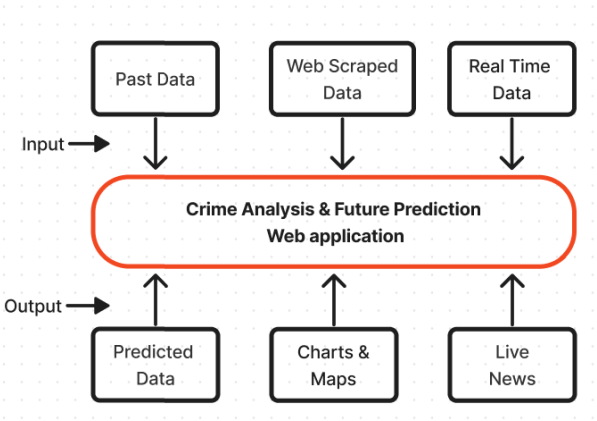

Fig.3.1. *System Design*

Machine learning systems architecture involves creating a blueprint for the software, infrastructure, algorithms, and data required to fulfill specific requirements. This blueprint guides the development of software for web applications by detailing the intricacies of how the program should be constructed. The system uses various technologies and methods for gathering, processing displaying data and making predictions based on the given data.

4. METHODOLOGY

The working of the system is based on various data analyzing, data visualization and machine learning algorithms for accurate analyses of data and also to make predictions based on the given data. We will also look at tools and algorithms used in this system.

4.1. DATA COLLECTION

The data is collected from various government, non-government websites, available datasets, and other websites. Also some of the datasets are collected from other resources such as online news using web scraping, etc.

4.2. DATA PREPROCESSING

The data is cleaned and pre-processed to remove redundancy and fill the gaps in the data for achieving a smooth and complete data set. This dataset results in a smooth and accurate prediction. The data is arranged as required.

4.3. DATA ANALYSIS

The data is analyzed for required information which will become an input to the predicting algorithm later. Data analyses helps to know the data and take required measures for the machine learning model to perform accurately.

4.4. DATA PREDICTION

The data is then feed to the prophet tool which predicts the crime rate of certain crimes in a specific area. This tool works on the date time column i.e., the time series, to produce its output.

4.5. DATA VISUALIZATION

This website provides various forms in which the data can be visualized such as heat map, pie chart, bar graph, etc. It helps in understanding large datasets.

4.6. TOOLS

4.6.1. PROPHET

Prophet is a method for predicting time series data using an additive model that accommodates non-linear trends alongside yearly, weekly, and daily seasonality, as well as holiday impacts. It performs most effectively with time series characterized by robust seasonal patterns and ample historical data spanning multiple seasons. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well. Prophet is open source software released by Facebook's Core Data Science team [9].

Prophet's input always consists of a dataframe containing two columns: 'ds' and 'y'. The 'ds' (datestamp) column should adhere to the format anticipated by Pandas, preferably YYYY-MM-DD for dates or YYYY-MM-DD HH:MM:SS for timestamps.The 'y' column must be numeric, and represents the measurement we wish to forecast [10].

4.7. FRAMEWORK

4.7.1. FLASK

Flask is a web framework that allows developers to build lightweight web applications quickly and easily with Flask Libraries. It was developed by Armin Ronacher, leader of the International Group of Python Enthusiasts(POCCO). It is based on the WSGI toolkit and Jinja2 templating engine [11].

4.8. LIBRARIES

4.8.1. PANDAS

Pandas is a Python library designed for data manipulation and analysis, providing specialized data structures and functions tailored for working with numerical tables and time series data. The library is built upon another library, NumPy [12].

4.8.2. NUMPY

In 2005, Travis Oliphant amalgamated features from Numarray into Numeric, implementing extensive modifications to create NumPy, a Python library. NumPy enhances Python with support for large, multi-dimensional arrays and matrices, complemented by a vast array of high-level mathematical functions tailored for manipulating these arrays. NumPy is open-source software and has many contributors [13].

4.8.3. MATPLOTLIB

Matplotlib, a plotting library compatible with both Python and its numerical mathematics extension NumPy, stands as a potent tool for individuals engaged in Python and NumPy-based endeavors, offering extensive capabilities for creating visualizations. And for making statistical interference, it becomes very necessary to visualize our data and Matplotlib is the tool that can be very helpful for this purpose. It provides MATLAB like interface only difference is that it uses Python and is open source [14].

4.8.4. SEABORN

Seaborn, built upon matplotlib, serves as a Python data visualization library offering a sophisticated interface for crafting visually appealing and informative statistical graphics. Seaborn is a library for making statistical graphics in Python. Expanding upon matplotlib and tightly integrating with pandas data structures, Seaborn's plotting functions are tailored to operate seamlessly on dataframes and arrays encompassing entire datasets. Internally, they execute essential semantic mapping and statistical aggregation, culminating in the creation of insightful plots. Its dataset-oriented, declarative API lets us focus on what the different elements of our plots mean, rather than on the details of how to draw them [16].

4.9. ALGORITHM

4.9.1. SUPERVISED MACHINE LEARNING

This system uses supervised learning algorithm for prediction. Supervised learning falls within the realm of machine learning, where labeled datasets are employed to train algorithms, enabling them to predict outcomes and identify patterns. Labelled data refers to input data that has been pre-assigned with corresponding correct output values. In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly [15].

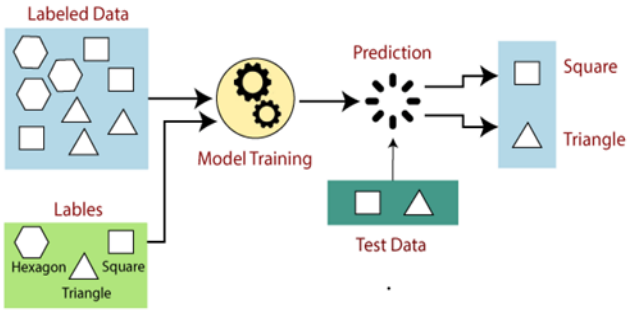

Fig.4.9.1. *Supervised Machine Learning*

4.9.1.1. PIECEWISE LINEAR REGRSSION

Prophet utilizes a form of piecewise linear regression to model the trend component of the time-series data. But Prophet's approach to linear regression is not the same as traditional linear regression models. In Prophet, the trend component is modeled as a piecewise linear function that allows for changes in the trend direction at specific change points. These change points are automatically selected based on historical data and represent times when the trend undergoes significant shifts.

The piecewise linear regression model in Prophet captures the overall trend in the data while allowing for flexibility and adaptability to changes over time. This approach differs from traditional linear regression models, which assume a single linear relationship between the predictor variables and the target variable.

By incorporating piecewise linear regression, Prophet can capture complex trends and patterns in the time-series data, making it particularly suitable for forecasting tasks where the trend may exhibit non-linear behavior or undergo changes over time.

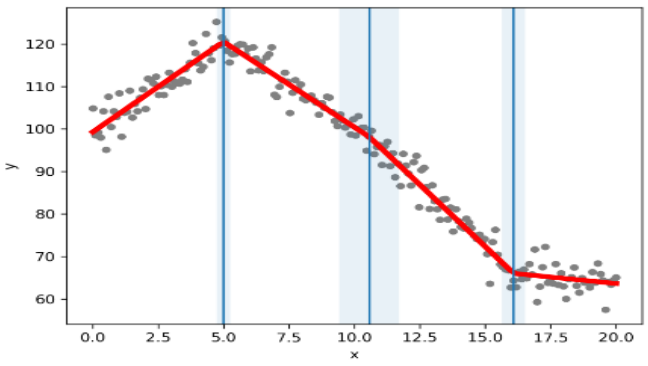

Fig.4.9.1.1. *Piecewise Linear Regression*

5. RESULTS

The required data is collected and preprocessed as required. Then we have created a heat map for depictions of areas with high crime rate.

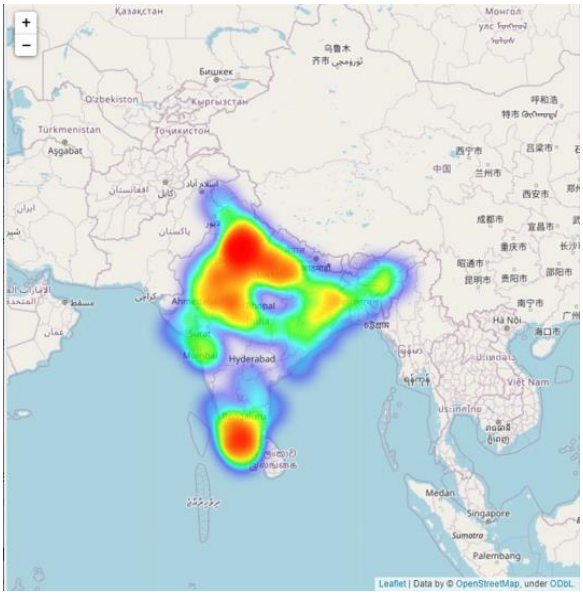

Fig.5.1. *Heat Map*

We have also created crime intensity metrices graph to depict intensity of crime in a specific area.

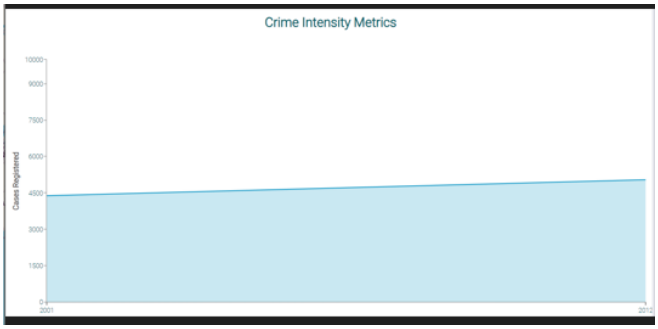

Fig.5.2. *Crime Intensity Metrices*

Then we have also created graphs for prediction of crime rate in the future using prophet.

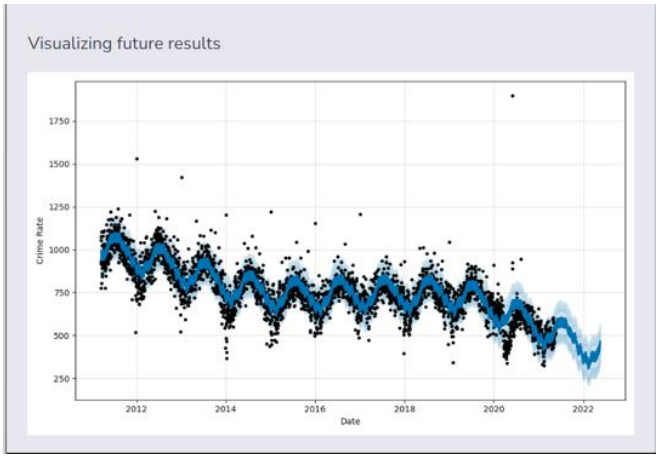

Fig.5.3. *Prediction Graph*

Thus we have created a system which predicts the crime rate and visualizes crime intensity of different areas in various ways.

6. CONCLUSION

Crime is an unlawful act which disturbs the peace and harmony of the society. This projects aims to successfully predict crime and their locations based on the historical crime data. The project uses machine learning which is an advanced and latest technology for accurate prediction. The web application will display crime rate in various areas. It is extremely useful for both the higher investigating authorities and officers designated to handle low level crime for tracking and stopping the crime. The predictions will help to ensure increased security and thus could help in lowering the crime rate. Overall, the project demonstrates the potential of data analysis and mapping technologies to improve public safety and inform decision-making. Proactive measures can be taken to prevent crime and improve public safety by using data to identify crime hotspots and trends. Although there's more work needed to enhance the precision and breadth of the project, it marks a significant stride towards employing data-driven strategies to tackle intricate social challenges.

7. ACKNOWLEDGEMENT

Every project requires the guidance and support of the experts for its completion. Therefore, we would like to take a moment to express our gratitude to all the individuals who played a crucial role in shaping and realizing this project.

Our heartfelt appreciation goes out to our project mentor, Dr. Radhika Nanda, from the Department of Computer Science and Engineering (with a specialization in IOT and Cybersecurity including Blockchain Technology) at Smt. Indira Gandhi College of Engineering, University of Mumbai. Her guidance, unwavering support, and dedication throughout the project have been invaluable.

We'd also like to extend our thanks to our project coordinator, Prof. Sarita Bhopalkar, for her assistance in selecting this project and providing us with essential insights on how to present it effectively.

Our sincere recognition goes to all our professors at Smt. Indira Gandhi College of Engineering, who provided us with valuable guidance and tips during the project's design phase. Their contributions have been so significant that it's challenging to individually acknowledge each one.

We are grateful for the support, both direct and indirect, from our Head of Department, Dr. Madhu Nashipudimath. Her assistance was instrumental in various aspects of our project, and we genuinely appreciate her contributions through various channels.

I express my heartfelt gratitude for the warm and caring environment provided by Smt. Indira Gandhi College of Engineering. This nurturing atmosphere and the excellent working conditions have been truly encouraging and appreciated.

8. REFERENCES

[1] Raza, D. M. & Victor, D. B. ''Data mining and region prediction based on crime using random forest'', 980–987 (IEEE, 2021).

[2] Prakash Maurya, Tahir Shaikh, Imran Ahmed, Amaan Firdosi, Prof. Kiran Deshmukh, "Crime Analysis and Prediction Using Machine Learning", IJRASET, Volume 11, Issue 4.

[3] Varun Mandalapu , Lavanya Elluri, Piyush Vyas, and Nirmala Roy, "Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions" , IEEE, Volume 11.

[4] Suhong Kim , Param Joshi, Parminder Singh Kalsi, and Pooya Taheri, "Crime Analysis Through Machine Learning", IEEE, Conference: November 2018.

[5] Yujunrong Ma; Kiminori Nakamura; Eung-Joo Lee; Shuvra S. Bhattacharyya, "EADTC: An Approach to Interpretable and Accurate Crime Prediction", IEEE 2022.

[6] Tzu-Wei Hung1 ,Chun-Ping Yen, "Predictive policing and algorithmic fairness", Synthese (2023)

[7] Ruaa Mohammed Saeed, Husam Ali Abdulmohsin, "A study on predicting crime rates through machine learning and data mining using text", Journal of Intelligent Systems, Volume 32, Issue 1.

[8] Shanjana A.S, Dr.R.Porkodi, "CRIME ANALYSIS AND PREDICTION USING DATAMINING: A REVIEW", 2021 IJCRT; Volume 9, Issue 2; February 2021.

[9] Prashant Banerjee," Tutorial: Time Series Forecasting with Prophet", Kaggle.

URL: https://www.kaggle.com/code/prashant111/tutorial-time-series-forecasting-with-prophet

[10] Prophet Documentation, facebook.github.

URL:https://facebook.github.io/prophet/docs/quick_start.html

[11] "Flask Tutorial", geeksforgeeks.org.

URL: https://www.geeksforgeeks.org/flask-tutorial/

[12] "pandas(software)", en.wikipedia.org

URL: https://en.wikipedia.org/wiki/Pandas_(software)

[13] "NumPy", en.wikipedia.org

URL: https://en.wikipedia.org/wiki/NumPy

[14] "Matplotlib",en.wikipedia.org

URL: https://en.wikipedia.org/wiki/Matplotlib

[15] "Supervised Machine Learning", javatpoint.com

URL: https://www.javatpoint.com/supervised-machine-learning

[16] "An introduction to seaborn", seaborn.pydata.org

URL: https://seaborn.pydata.org/tutorial/introduction.html