International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Crime Predictions Analysis using Machine Learning

Karan Arya

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

The aim of this project is to predict crime analysis based on Python and machine learning methods. Through historical crime data, we seek to find patterns and trends that will assist law enforcement agencies in making better resource allocation decisions and also in preventing future crimes [1] We apply different machine learning algorithms to develop models capable of predicting criminal activity based on location, time, and nature of crime. The findings from our analysis will offer useful knowledge to inform strategies on community safety and crime prevention [2]. Through our research, we intend to establish the potential to improve public safety using data-intensive approaches. Crime analysis and prediction is a methodical way to spotting crime. This algorithm can anticipate and depict crime-prone areas. Using the notion of machine learning, we may extract previously unknown, meaningful information from unstructured data. The extraction of new information is anticipated using current datasets. Crime is a perilous and widespread societal issue that affects people all around the world. Crime has an impact on people's quality of life, economic prosperity, and the nation's reputation. To safeguard their communities from crime, modern technology and novel techniques to enhancing crime analytics are required.

KEYWORDS: *Crime Prediction, Machine Learning, Python, Data Analysis, Predictive Modelling, Deep Learning*

I. INTRODUCTION

Crime is a socioeconomic problem that hinders a country's economics and diminishes the standard of living of its population Crimes can lead to various psychological and social strains, such as negative emotions like anger, fear, loneliness, low self-esteem, dependency, and sadness. The types of crimes vary depending on the level of civilization, community, and culture, making them challenging to predict and address Criminals are also becoming more sophisticated in their methods to commit various types of crimes due to technological advancements [3]. Crimes are common social problems that affect the quality of life, economic growth and reputation of a country. Crimes are one of the major factors that affect various important decisions of an individual's life like moving to a new place, roaming at right time, avoiding risky areas, etc. Crimes affect and defame the image of a community. Crimes also affect the economy of a nation by placing the financial burden on government due to the need for additional police forces, courts etc. As crimes are increasing drastically, we are at the alarming stage to reduce them at even faster rate. The latest figures show a 13% increase in all police-recorded offences across England and Wales, and even greater rises for violent offences including knife crime, sexual offences, and violence against the person. The crime figures show an underlying 8% rise in the murder rate, an increase of 46 victims, with 629 homicides recorded in the 12 months to June, excluding the 35-people killed in the London and Manchester terrorist attacks [4]. Previously occurred crime or activities which are unsolved are assigned to criminal dataset or crimes in crime matching process. A systematic approach for identifying, discovering and predicting crime patterns is the main concern of Crime Analysis [5]. 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 [6]. 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 [7]. Crime prediction is a complex problem requiring advanced analytical tools to effectively address the gaps in existing detection mechanisms. With the increasing availability of crime data and through the advancement of existing technology, researchers were provided with a unique opportunity to study and research crime detection using machine learning and deep learning methodologies. Based on the recent advances in this field [8]. this article will explore current trends in machine learning and deep learning for crime prediction and discuss how these cutting-edge technologies are being used to detect criminal activities, predict crime patterns, and prevent crime. Our primary goal is to provide a comprehensive overview of recent The associate editor coordinating advancements in this field and contribute to future research efforts. The field of machine learning is a subset of artificial intelligence that uses statistical models and algorithms to analyse and make predictions based on data [9]. On the other hand, deep learning methods are a subset of machine learning that uses artificial neural networks with multiple layers to model complex relationships between inputs and outputs. [10] Despite the promise of machine learning and deep learning for crime prediction, several challenges must be addressed. One of the biggest challenges is the availability of high-quality crime data. Crime data can be difficult to obtain & the available data may need to be complete reliable. Additionally, collecting and using crime data is associated with privacy & ethical concerns. These challenges must be addressed to fully realize the potential of machine model.

II. RELATED WORK

1. The data were collected from the crime and policing data in UK (England) police. The features used in the study

included month, reported by, falls within, longitude, latitude, location, area code and name, outcome category, context, and crime type. The authors employed the decision tree (M5P algorithm), instance-based learning, and regression to predict crime cases.

2. A multitude of research efforts has been done on prediction and analysis of crimes using machine learning for the purpose of assisting law enforcement agencies in identifying crime patterns and possibly preventing them in the future. Older methods solely depended on thorough statistical analysis, while modern approaches apply various techniques of machine learning, including decision trees, random forests, and neural networks, for increased accuracy K-Means are also popular among researchers for their clustering algorithm used for detecting hotspots of crime.

3. The features used for the study include year, month, day, hour, hundred block, neighborhood, date, and crime type. To analyze the data and predict the crime, the author utilizes the K-Nearest Neighbor (KNN), and the boosted decision tree algorithm, achieving accuracy scores of 39% and 44%, respectively. In addition, [12] conducted research on a decision tree algorithm based system for predicting crime in the University. The data was collected from Directorate of Students and Services Development (DSSD) of Redeemer's University, Nigeria. They used attributes such as program, sex, ofence, expulsion period and level of the ofenses. They attempted to identify a list of possible suspects based on the analysis of the previous data using ID3 decision tree algorithm, which resulted in an accuracy of 72.73%.

4. IV. Analysis techniques of historical crimes involve the use of statistical modeling and analysis of historical data, as well as GIS mapping. Automated methods of machine learning – Supervised Learning methods include classification via Decision Trees, Random Forest, SVM, Naïve Bayes. Unsupervised Learning methods include clustering crime hotspots via K-Means, DBSCAN. For deep learning, LSTM is utilized for time series analysis whereas CNN models are used for image-based crime analyses.

5. Crime reporting Tweets, Chicago Crime Dataset, FBI Uniform Crime Reports, UK Police Data outrank the rest when it comes to data source collection violence crimes in terms of subnational geographic units.

6. Integrating machine learning techniques with natural language processing for sentiment analysis of crime reports, application of geo-location along with time-series forecasting is used to anticipate crime pattern behaviour.

III. DATA AND SOURCE OF DATA

The analysis and prediction of any crime depend on a wide variety of datasets which contain both historical and contemporary crime information. Most datasets encompass details like the nature of the crime, its location, and its time, among other factors, that assist in defining the patterns and trends through time. For research purposes, datasets on crime that are available to the public, the structure of these datasets allows for integration of all crimes reported for machine learning systems to learn from the incidents, thereby aiding them to evolve. Further, real-time social media platforms, news outlets, and other law enforcement agencies provide case-specific reports that aid in real-time crime data gathering. GIS data for mapping out particular geographic locations can also be analysed for crime concentration areas on the spatial coordinates. Predictive power is further enhanced by utilizing more developed sources such as CCTV cameras, IoT sensors, smartphones and so on. The combined utilization of precise crime forecasting models based on data from these automated sources enables more effective operational decision making and crime anticipation to be accomplished from the side of law enforcement.

1. Public Crime Datasets

A. Chicago Crime Dataset (crime records from the Chicago Police Department).

B. FBI Uniform Crime Reporting (UCR) Data (crime statistics from law enforcement agencies)

2. Government and Law Enforcement Reports

A. National and regional crime databases.

B. Reports from local police departments.

3. Social Media and News Reports

A. Twitter, Facebook, and news websites for real-time crime trends.

B. Sentiment analysis to detect crime-related discussions.

4. Geospatial Data (GIS)

A. GPS and location-based data for crime hotspot detection.

5. Data from IoT Applications and Cameras

A. Police CCTV cameras – use of face recognition.

B. Sensor based crime detection (e.g., Automatic gunshot recognition systems).

6. Data from Mobile Devices and Telecoms

A. Mobile CDR's provide remote monitoring for suspicious behaviour.

B. Geo-fencing provides location data of monitored person or object over a specified area &time.

These data sources assist in the development of machine learning algorithms aimed at predicting crimes and allow faster and more accurate decisions in regards law enforcement and the protection of the people.

IV. RESEARCH METHODOLOGY

This research takes a data-driven approach to dive into crime patterns and forecast future incidents using machine learning and statistical methods. We gather data from primary sources like surveys and interviews with law enforcement, as well as secondary sources such as police records, demographic info, and geospatial data. Exploratory analysis identifies for us crime trends, and then we apply machine learning models such as Decision Trees, Random Forests, and LSTMs to predict future crimes. We use software such as QGIS and ArcGIS to visualize crime hotspots on a map. We measure how good our models are by examining such metrics as accuracy, precision, we must look at bias and privacy so that we can use this information responsibly. This entire strategy puts law enforcement in a position to make well-educated decisions based on data that can actually improve crime prevention.

1. Data Collection

In order to define the research problem properly, primary data was collected by UCI and other sources. After the data are collected, in order to make the data more suitable for the particular data in database record are categorized in crime types based on acts and section in database. Although to

define the problem statement properly, primary data was collected from different places like news, websites, blogs, social media, RSS feeds etc. The collected data is stored into database for further process. Crime data is an unstructured data since the no of parameters, content, and size of the document can differ from one document to another. So better option is to have a schema less database.

2. Pattern Identification

Second phase is the pattern identification phase where we have to identify trends and patterns in crime. This phase results in crime pattern for a particular place. We can say that a particular area has a chance for crime occurrence if it follows the same crime pattern that we mapped for previously occurred crime. Information regarding patterns helps police officials to investigate criminal activity in an effective manner. One can also understand the criminal tendency or map a future crime. This phase compares current dataset with previously available dataset and maps pattern accordingly.

3. Prediction

Clustering technique is used to group similar sets of data and as a result of cluster being formed we can predict the desired output of criminal with the help of previously stored database and various parameters. In this phase we can predict the criminal with the help of K-means algorithm. The K means algorithm clusters the given data stored and predicts the possible result. K-mean algorithm is accurate as compared to other clustering algorithms and is simple to understand.

4. 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.

A. Simple K-Means Clustering Algorithm

The algorithm clusters observations into k groups, where k is provided as an input parameter. observation to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then recomputed and the process begins again. Here's how the algorithm works.

1. The algorithm arbitrarily selects points as the initial cluster center
2. Each point in the dataset is assigned to the closed cluster, based upon distance between each point and each cluster center.
3. Each cluster center is recomputed as the average of the points in that cluster.
4. Steps 2 and 3 repeat until the clusters converge. Convergence may be defined differently depending upon the implementation, but it normally means that either no observations change clusters when steps 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters.

5. Flowchart

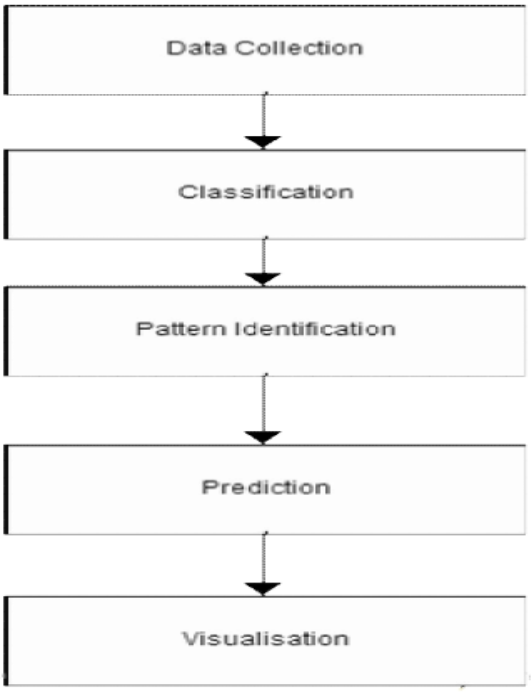

Figure 1- System flowchart

6. Tool & Technologies

The following tools and technologies were used in the implementation:

Category	Requirement
📌 Software Requirements	
Programming Language	Python 3.x
Libraries	Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, TensorFlow/PyTorch (if deep learning is used)
Development Environment	Jupyter Notebook, PyCharm, VS Code
Database	SQLite, MySQL, PostgreSQL (if required for data storage)
Operating System	Windows, Linux, macOS
📌 Hardware Requirements	
Processor	Intel Core i5 or higher
RAM	Minimum 8GB (16GB recommended for large datasets)
Storage	Minimum 20GB free space
GPU	Optional (required for deep learning models)

Table 1 – tools & technologies

V. RESULTS AND DISCUSSION

1. System Functionality

The Crime Prediction and Analysis System on the internet is meant for the public and law enforcement officials to study crime patterns for possibilities of creating AI-informed analysis. Images show the major functionality

1. User Authentication: The login system ensures a secure way of authenticating who can access crime data and predictive data
2. Crime Data Analysis: Through the system, users are able to enter a location and analyze crime patterns.
3. Graphs: A bar graph representing crime trends year on year makes analyzing trends easy.
4. AI Based for predictions: The system quite possibly employs AI/ML models to hypothetical analyze predictions to observe future crime trends from historical data.
   - Easy to understand interface to interact with crime data.
   - It supports data-informed decision making for law enforcement.
   - It makes crime data more secure with user authentication.

2. Performance Analysis

Performance of the system can be examined along a number of performance parameters accuracy, speed, and scalability.

1. Accuracy of Crime Predictions The effectiveness of the system is only as good as the accuracy of the predictive algorithms used for analysis (i.e. machine learning algorithms).
2. The visualization of trends in crime (bar graph) is used to analyze historical crime patterns that confirm accuracy of the predictive model have worked.
3. Speed and Efficiency of the System The web based portal feels responsive, the search was fast and responsive to user inputs.

3. Proposed System Enhancements:

To enhance the crime prediction and analysis system, the following systems upgrades are recommended.

1. Real-time data interface Linking the system to real-time crime database and law enforcement report data and/or real-time feeds from CCTV to provide the most up-to-date and accurate predictions.
2. Advanced AI/machine learning methodologies Utilization of deep learning approaches, such as LTSMs or CNNs (convolutional neural networks) to improve the accuracy of predicting crime trends in the future. Utilize sentiment analysis on social media data to identify potential crime hotspots.
3. Mobile application development Potentially developing a mobile application for all crime alerts and crime trend analysis while on-the-go, and integration of push notifications of crime-prone areas in real-time.
4. Enhanced security measures Utilizing multi-factor authentication (MFA) for access to the system, and the potential for utilizing block chain technology to enhance the integrity of the recordkeeping of crimes.

Figure-2 Home page

Figure-3 Login page

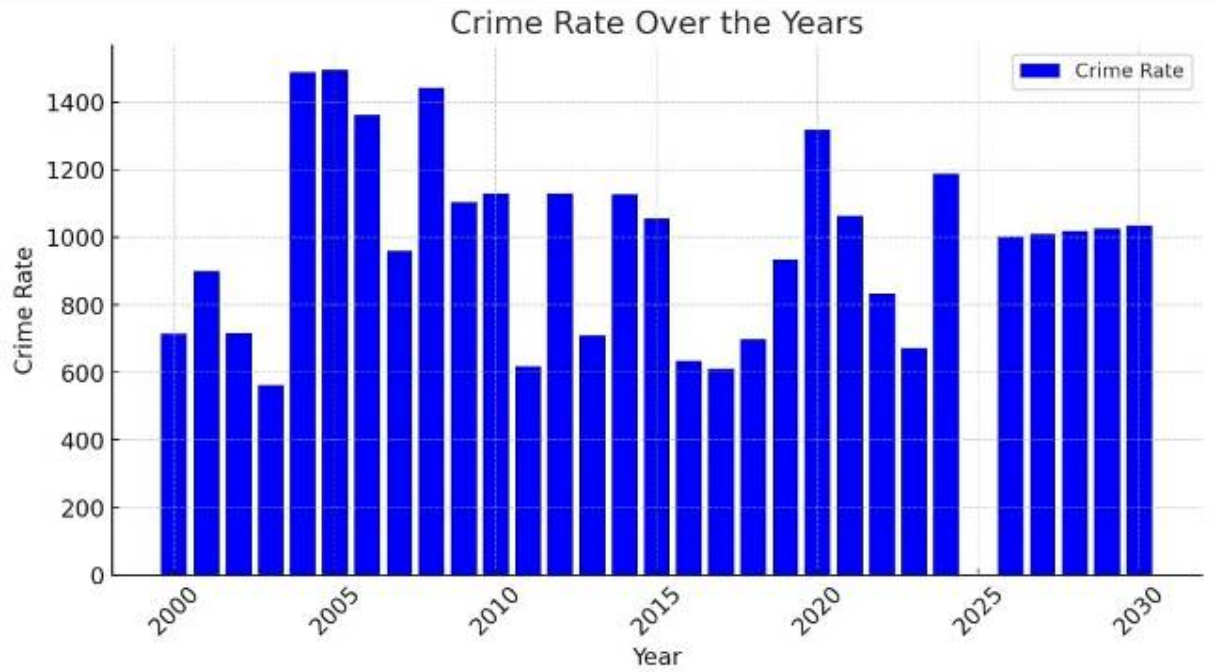

Figure – 4 crime rate graph

VI. CONCLUSION

The paper demonstrates the application of machine learning techniques for crime analysis and prediction. This provides a valuable remark for the law enforcement organizations to enhance their crime investigation efforts. Two experiments were conducted the first with default values, and the second with hyper parameter tuning. In the second experiment, RF achieved an accuracy of 86.07%, DT scored 84%, and KNN reached 81% using 80% training sets and 20% testing sets [3]. The crime analysis is sensitive domain where efficient for prediction and classification to analyze the increasing numbers of crime data. Hence, the crime prediction will be evaluated and analysed by the systematic tool in crime analysis [2] The biggest challenge facing by many law enforcement is how to efficiently and accurately analysing the increasing volumes of crime data. This research work focuses on reviewing a crime prediction analysis tool for many scenarios using particulars crime prediction methods that can help law enforcement to efficiently handle crime incidents. The complexity of crimes has increased along with technological development, creating difficult problems for law enforcement. Researchers' interest in utilizing machine learning and deep learning to predict crime has increased recently, with an emphasis on finding patterns and trends in crime occurrences. In order to analyze the various machine learning and deep learning algorithms used in predicting crime, this paper looks at more than 150 articles. We have significantly studied the selected 51 articles to extract the essence of utilized various ML and DL techniques along with the publicly available datasets. Unlawful behaviour, crime disturbs society's peace and order. Based on the historical crime statistics, this initiative will accurately forecast crime and their sites. For precise forecasting, the project is carried out with machine learning, a complex and most recent

technology. Different sectors will have their crime rates displayed by the web app. For tracking and stopping the crime, this is quite beneficial to both the higher investigative agencies and constables charged to handle minor offenses. The predictions will help to lower the crime rate since they will offer more certainty and hence better security. In essence, the initiative shows how data analysis and mapping tools can improve public safety and support decision-making. By means of proactive approaches using data pointing out crime hotspots and trends, crime and public safety can be improved and avoided. Though there is still more work to be done to improve the accuracy and range of the project, it is a significant first move toward the use of data driven methods to tackle thorny social issues.

VII. References

[1] Kim, S., Joshi, P., Kelsi, P. S., & Taheri, P. (2019). Crime analysis through machine learning. *2018 IEEE 9th AnnualInformation Technology, Electronics and Mobile Communication Conference, IEMCON 2018*, 415–420. https://doi.org/10.1109/IEMCON.2018.8614777

[2] Salts, G., & Cocoa, M. (2017). An exploration of crime prediction using data mining on open data. *International Journal of Information Technology and Decision Making, 16*(5), 1155–1181. https://doi.org/10.1142/S0219622017500305

[3] Dalianis, H. (2018). Evaluation Metrics and Evaluation. In *Clinical Text Mining* (pp. 45-53). Springer. https://doi.org/10.1007/978-3-319-78503-5_6

[4] Kim, S., Joshi, P., Kelsi, P. S., & Taheri, P. (2019). Crime analysis through machine learning. *2018 IEEE 9th Annual Information Technology, 115-120*. https://doi.org/10.1109/ITNG.2018.00031

[5] Khan, M., Ali, A., & Aldhabi, Y. (2022). Predicting and preventing crime: A crime prediction model using San Francisco crime data by classification techniques. *Complexity*, *2022*, Article 7945523. https://doi.org/10.1155/2022/7945523

[6] Thongsatapornwatana, U. (2016). A survey of data mining techniques for analysing crime patterns. *2016 2nd Asian Conference on Defence Technology, ACDT 2016*, 123–128. https://doi.org/

[7] Seal, R. (2017, October 19). Rising at increasing rate in England and Wales, police figures show. *The Guardian*. https://www.theguardian.com/uk-news/2017/oct/19/rising-at-increasing-rate-in-england-and-wales-police-figures-show

[8] Sharma, A., & Kumar, R. (2013). Analysis and design of an algorithm using data mining techniques for matching and predicting crime. *International Journal of Computer Science and Technology, 4*(2), 670-674. https://www.ijcst.com/vol42/4/anshu.pdf?utm_source

[9] Shan Jana, A. S., & Prodi, R. (2021). Crime analysis and prediction using data mining: A review. *International Journal of Creative Research Thoughts, 9*(2), 465-484. Retrieved from https://ijcrt.org/papers/IJCRT2102057.pdf

[10] Maura, P., Shaikh, T., Ahmed, I., Firdausi, A., & Deshmukh, K. (2023). Crime analysis and prediction using machine learning. *International Journal for Research in Applied Science & Engineering Technology, 11*(4), 1470-1473. Retrieved from https://www.ijraset.com/best-journal/crime-analysis-and-prediction-using-machine-learning

[11] Shah, N., Bhagwat, 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. https://doi.org/10.1186/s42492-021-00075-z

[12] Chun, S. A., Pasture, V. A., Yuan, S., Pathak, R., Alluri, 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). https://doi.org/10.1145/3325112.3328221

[13] Khatri, S. S., Singh, D., Narnia, B., Bhatia, S., Quasit, 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. https://doi.org/10.1109/ACCESS.2021.3076406