CRIMINALITY PREDICTION NEURAL NETWORK USING DEEP LEARNING

Muskan Rao
(COMPUTER SCIENCE ENGINERING)
muskanrao200@gmail.com
Harshita Jain
(COMPUTER SCIENCE ENGINERING)
jainh6786@gmail.com
Chandni Lodhi
(COMPUTER SCIENCE ENGINERING)
thakurchandni034@gmail.com

Abstract: This paper dives into the realm of predicting crimes using smart technology like machine learning and deep learning. By carefully reviewing over 150 articles, it explores how different computer-based methods can anticipate where crimes might happen. The study not only shares where to find the information needed but also looks at what causes crimes to occur. It also points out where improvements can be made and suggests ways to make crime prediction more accurate. This paper is like a treasure map for researchers and police, showing them the best paths to follow. By understanding how to predict crimes better, we can make our communities safer

Keywords- Crime · Numerous safety problem · Data mining · KNN (K-Nearest Neighbor) · Safe route, Computer vision, Crime forecasting.

---

1. INTRODUCTION

In today's modern age, urbanization brings about significant benefits and challenges in managing cities, including traffic, environmental concerns, and public safety. Large cities, in particular, face high crime rates, making crime reduction a crucial social issue affecting security, youth development, and socio-economic status. Crime rate prediction involves using various algorithms to estimate future crime rates based on historical data. As we go about our daily activities, we often encounter security issues such as hijacking, kidnapping, and harassment. Despite relying on navigation tools like Google Maps for directions, we frequently opt for shortcuts without fully understanding their safety. This lack of awareness leads to unpleasant situations. Our research aims to address this by designing a strategy based on past crime data to analyze crime rates in different areas at various times. We gather primary data from individuals based on their past experiences with crime. Using different algorithms, we assess the accuracy of crime rate predictions, focusing on the last three years of crime data. Our methodology includes dataset collection, data processing, crime analysis, and prediction using various algorithms, notably the K-nearest neighbor (KNN) algorithm.

2. Machine Learning (ML) Techniques Employed in Crime Prediction

In one study, a comparative analysis was conducted using the Communities and Crime Unnormalized Dataset alongside actual crime statistics. The study utilized three algorithms—linear regression, additive regression, and decision stump—implemented through the Waikato Environment for Knowledge Analysis (WEKA) software. Among these, linear regression exhibited superior performance, demonstrating its efficacy in predicting violent crime patterns and supporting various applications such as identifying criminal hotspots, profiling offenders, and discerning criminal trends. WEKA, an open-source data mining tool, offers a Knowledge Flow interface, which enhances data mining processes by visually representing the flow of information using java beans. Additionally, the Experimenter interface facilitates the comparison of multiple learning schemes across various datasets. Another investigation focused on predictive crime analysis in urban settings, particularly home burglary, street robbery, and battery crimes. By aggregating crime data into grids and employing ensemble models synthesizing logistic regression and neural networks, the study achieved accurate fortnightly and monthly crime predictions for the year 2014. In a separate study, crime predictions were made based on machine learning algorithms applied to Vancouver's crime data spanning 15 years. Despite achieving moderate accuracy ranging between 39% and 44%, the study suggested that fine-tuning algorithms and crime data could enhance prediction accuracy for specific applications. Similarly, a machine learning approach was employed to predict crime-related statistics in Philadelphia, with algorithms such as logistic regression, K-nearest neighbor (KNN), ordinal regression, and tree methods utilized for training datasets. The approach yielded detailed quantitative crime predictions, achieving notable accuracies in predicting both the occurrence and number of crimes. Finally, an analysis of crime datasets from Chicago employed various ML and data science techniques to predict future crimes based on contextual conditions. The study tested multiple models, including KNN classification, logistic regression, decision trees, random forest, support vector machines (SVM), and Bayesian methods. KNN classification emerged as the most accurate model, emphasizing the potential of ML in aiding law enforcement agencies to predict, detect, and address crime effectively.

3. METHODOLOGY

1.Dataset- The crime dataset comprises approximately 500 entries spread across 10 rows. It was derived from primary data collection conducted through fieldwork. The dataset encompasses various key features including Name, Years, Months, Crime Type, Crime Areas, Victim Genders, Victim Ages, Victim Areas, and Months, which are designated as input features for the system. Additionally, the dataset includes characteristics such as Perpetrator Ages, Perpetrator Genders, and Victims Relations, which are identified as the system's target variables. These variables serve as focal points for analysis and prediction within the system (refer to Table 1 for a detailed breakdown).

Number	Details		
	Name	Type of columns	Descriptions
1	Person_Id	Value type	Person ID in dataset
2	Name	String	Victim person name
3	Year	Numeric	Crime occurs year
4	Ages	Numeric	The ages of the victim
5	Gender	String	Victims neuter
6	Time	String	Time when the crime has occurred
7	Victim area	String	Area where the crime has occurred.
8	Region	String	Region of the victim
9	Home town	String	Home town of the victim
10	Month	String	The year in which the crime has occurred

2 . Data Preparation-To streamline our analysis, we opted to exclude unknown values, which serve as indicators of incomplete data, from our dataset. Dates and times were recorded in the MM/DD/YY HH/MM format, posing challenges for direct matching due to variations in classification systems. To address this, we categorized dates into three groups—weekends, weekdays, and unspecified—based on the features of the date-time windows.

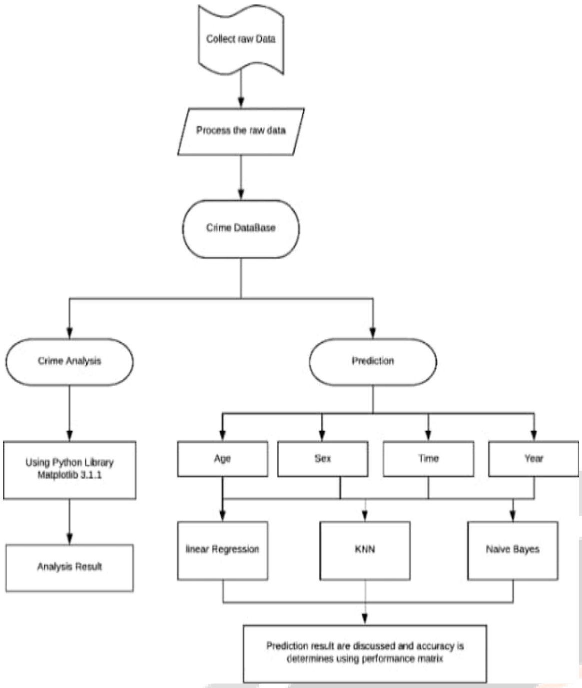
System Workflow:

Our workflow commences with data extraction from a repository housing datasets with various roles. Subsequently, primary data undergoes preprocessing to transform it into criminal data.

Four Target Variables:

1. Aged-Based Crime Accuracy: Utilizing Linear Regression to ascertain the accuracy of age-based crime predictions.
2. Sex Estimation: Employing K-Nearest Neighbors classification to estimate the sex of individuals involved.
3. Perpetrator Gender Estimation: Utilizing K-Nearest Neighbors classification to predict the gender of perpetrators.
4. iv. Final Prediction Rate: Utilizing the K-Nearest Neighbors algorithm to predict the rate for upcoming years based on age, sex, time, and year.

The workflow diagram is depicted in Figure 1

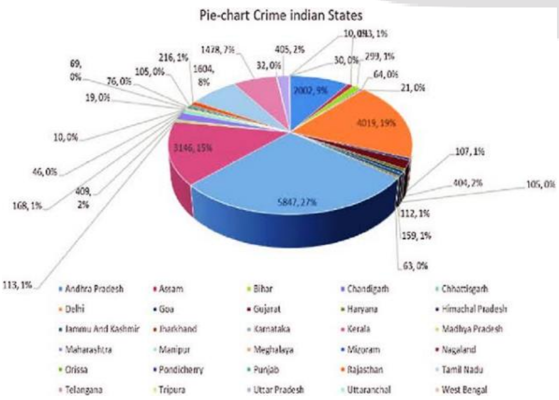


3. K-Nearest Neighbors – K-nearest neighbors (KNN) is employed when the target variable necessitates classification into more than two classes. In this dataset, the target variable pertains to the perpetrator's sex, with three distinct classes: male, female, and unknown. Similarly, age is categorized into three groups: young, old, and kid. To classify these target variables, the K-nearest neighbors Classifier is utilized.

Pseudo Code:

1. Initialize KNN Classifier (Data Entry).
2. Specify the Number of Clusters in of K
3. Select a set of K instances to serve as cluster centers
4. Associate data points with each output
5. Calculate the Euclidean distance for each data point
6. Assign the data point to the nearest cluster
7. Continuously compute centroids and reassign cluster variables
8. Repeat until an optimal clustering result is achieved
9. Return the clusters and their respective values.

This process facilitates the classification of target variables based on their proximity to neighboring instances within the dataset, ensuring accurate categorization across multiple classes.



4. Naïve Bayes

Naïve Bayes classifiers encompass a family of classification algorithms based on Bayes' theorem. Rather than being a single algorithm, it represents a group of algorithms that share a common principle: each combination of features being classified is treated as independent from one another. Bayes' theorem calculates the probability of one event occurring given the occurrence of another event. The theorem is expressed by the following equation:- $p(A|B)=P(B|A).P(A)/P(B)$ Naïve Bayes algorithms find extensive application in tasks such as sentiment analysis, spam detection, and recommendation systems. They are known for their speed and simplicity in implementation. However, a major limitation lies in their assumption of feature independence, which may not hold true in many real-world scenarios where predictors are interdependent. Although we did not utilize this algorithm to derive the final result in our study, it has demonstrated high effectiveness in numerous previous cases and has yielded the best accuracy rates for crime prediction.

4. Conclusion

In summary, our paper endeavors to bridge the divide between theoretical crime prediction concepts and their practical implementation using cutting-edge technology. We propose an integrated framework that combines machine learning, deep learning, and computer vision to develop a holistic system for predicting and detecting crimes. Through the adoption of these technologies, law enforcement agencies stand to transform their crime prevention and management strategies. We acknowledge the intricacies involved in realizing such a system and are mindful of the challenges inherent in its development, deployment, and utilization. Nevertheless, we remain optimistic that with concerted effort, these obstacles can be surmounted, paving the way for a security apparatus capable of continuous monitoring of crime-prone areas and delivering more reliable crime forecasts. Our study showcases the efficacy of machine learning algorithms, particularly the K-nearest neighbor approach, in accurately predicting crime rates. Leveraging these predictive tools, our aim is to enhance the precision of future crime predictions and pinpoint areas with heightened criminal activity. Moreover, our research contributes valuable insights to the existing body of literature by shedding light on the practical applications of machine learning and deep learning techniques in crime prediction. We underscore the potential of these advancements to bolster the accuracy and effectiveness of crime prediction models, thereby empowering law enforcement agencies to address the ever-evolving landscape of crime detection and prevention. We express our gratitude to Daffodil International University for its unwavering support of our research endeavors, and extend our thanks to Mr. Abdus Sattar for his invaluable guidance and input throughout the project.

5. References

1. Shah D, Dixit R, Shah A, Shah P, Shah M (2020) A comprehensive analysis regarding several breakthroughs based on computer intelligence targeting various syndromes. Augment Hum Res 5(1):14. https://doi.org/10.1007/ s41133-020-00033-z .

2.patel H, Prajapati D, Mahida D, Shah M (2020) Transforming petroleum downstream sector through big data: a holistic review. J Pet Explor Prod Technol 10(6):2601–2611. https://doi.org/10.1007/s13202-020-00889-2 3.

3.Szeliski R (2010) Computer vision: algorithms and applications. SpringerVerlag, Berlin, pp 1–979 .

4. Vedaldi A, Fulkerson B (2010) Vlfeat: an open and portable library of computer vision algorithms. Paper presented at the 18th ACM international conference on multimedia. ACM, Firenze. https://doi.org/10.1145/1873951.1 874249 .

5. Le TL, Nguyen MQ, Nguyen TTM (2013) Human posture recognition using human skeleton provided by Kinect. In: Paper presented at the 2013 international conference on computing, management and telecommunications. IEEE, Ho Chi Minh City. https://doi.org/10.1109/ComMa nTel.2013.6482417.

6. Ahir K, Govani K, Gajera R, Shah M (2020) Application on virtual reality for enhanced education learning, military training and sports. Augment Hum Res 5(1):7. (https://doi.org/10.1007/s41133-019-0025-2