International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Crime Data Analysis and Prediction of Perpetrator Identity

A.Harshitha, P.Ramya, V.Yeshwanthi, V.Pranitha
Computer Science of Engineering, GRIET ,India
Department of CSE, GRIET, India
Dr K Butchi Raju, Professor, *Department of CSE, GRIET, Hyderabad, Telangana*

ABSTRACT

Crime is one of our society's most serious and distressing issues, and preventing it is a critical duty. Crime analysis is a method of finding and studying patterns and trends in crime that is done in a methodical manner. The purpose of this approach is to improve the efficiency of criminal justice systems. This model recognises criminal patterns based on inferences drawn from the crime scene and predicts the description of the offender who is most likely to have committed the crime. This project includes two primary components: crime analysis and perpetrator identity prediction.The Crime Analysis phase determines the number of unsolved crimes and investigates the impact of numerous characteristics such as year, month, and weapon on those crimes. The attackers' age, sex, and relationship with the victim are estimated during the prediction phase. The evidence gathered at the crime scene is used to make these predictions. Using methods like KNN Classifier and Neural Networks, the system guesses the perpetrator's description.It has undergone extensive training and testing using Kaggle Us open dataset and implemented using python.The implementation of this project results the speed up in the procedure time of the case solving based on the previous analyzed data which helps in the solving issues.So for the implementation is rather essential and that is the prime motive of the idea establishment. We use pycharm which supports constraints like matplotlib for future data visualization or the machine learning algorithms for analysis and prediction.To get the required result we consider sklearn, flask,html,sqlyog for requirements like webpage depiction and database for the execution.

Keywords

Multilinear Regression; K-Neighbors Classifier; Artificial Neural Networks; Kaggle Homicide Dataset;Python

.

INTRODUCTION

Criminality is a significant threat to mankind. Numerous crimes are committed on a regular basis. It's possible that it's increasing and spreading throughout a large area. From a tiny town to a huge urban region, crime can strike at any time. Robbery, murder, rape, assault, and other offences such as false imprisonment, kidnapping, and homicide can all be filed against you. Because the number of crimes is on the rise. It is necessary to settle the issues in a timely manner. a far more efficient technique Simultaneously, criminal activity has increased. It is the police department's job to proceed at a faster rate.. The primary challenges for the police force are criminal identification and criminal identification of criminals in order to keep track of and limit criminal activity. Because there is a substantial amount of crime, the police department is required. There is data that is currently available, but technology is required.With the appropriate technologies, the project can process the solution at a high rate. Machine learning techniques, which give regression and classification approaches to fit the objective, are being considered for the project's implementation. For analysis and prediction, we employ KNN and Neural Networks. We develop a web page that displays the user's identification as either an admin or a case follower, and then displays the pages depending on that identity.The admin uploads the criminality proof, and the user may follow up on the case by inputting particular data. Before using, the user must be approved, which means he must authenticate his identity. The government has to authorize the creation of an account on the website. For the examination of evidence and to identify the records for the projected culprit, data with various features is supplied. The criminal's identification and analysis aid in the finding of trends in instances involving a certain region or feature.

LITERATURE

Jyoti Agarwal, Renuka Nagpal, and colleagues (2013) employed k-means clustering to analyze the crime dataset. This model was made with the rapid miner tool. To investigate the clustered findings, the data are shown across time. The model indicates from the data that the number of homicides fell from 1990 to 2011.

Lawrence McClendon and Natarajan Meghanathan (2015) used the Communities and Crime Dataset to test multiple prediction methods, including Linear Regression, Additive Regression, and Decision Stump approaches, all with the same set of inputs (features).When compared to the other two methods, the linear regression method produced the best results. The linear regression method has the advantage of

being able to deal with some volatility in test data.

S. Sivaranjani, S. Sivakumari, et al., (2016) used different clustering approaches, including K-Means clustering, To cluster criminal activities in Tamil Nadu, researchers used a combination of agglomerative clustering and Density Based Spatial Clustering with Noise (DBSCAN) techniques.The performance of each clustering technique is evaluated using metrics like accuracy, recall, and F-measure, and the results are compared. Based on the given criteria, the DBSCAN approach generated the best results when compared to the other two methods.

Chirag Kansara, Rakhi Gupta, and colleagues (2016) developed a method for analyzing people's Twitter attitudes and predicting if they constitute a threat to a specific person or society. This model, which employs sentiment analysis to categorize people, is built using the Naive Bayes Classifier.

Ryan Heartfield, George Loukas, and colleagues (2016) [6] forecast the number of crimes committed as a result of Semantic Social Engineering Assaults and investigate the viability of predicting user sensitivity to deception-based attacks. The authors used logistic regression and a random forest prediction model to make their predictions, with accuracy rates of.68 and.71, respectively.

METHODOLOGY

1. Dataset

Criminal records were collected by Kaggle, the source of records across multiple domains. The dataset you choose to use is the US Homicide open database for running your project's test cases. It is publicly accessible. State, year, month, type of crime, state of crime, gender of victim, age of victim, weapons are all important aspects of the dataset used as input characteristics to the algorithm.

The attributes Perpetrator Age, Perpetrator Sex, and Perpetrator-Victim Relationship are chosen as the goal variables for the algorithm to forecast.

NO	Name	Type of Column	Description
1	State	String	The state in which the crime has occurred.
2	Year	Numeric	The year in which the crime has occured.
3	Month	String	The month in which the crime has occurred.

4	Crime Type	String	The type of the crime
5	Crime Solved	String	Status of the investigation
6	Victim Sex	String	The gender of the victim
7	Victim Age	Numeric	Age of the victim
8	Weapon	String	Weapon used for committing the crime
9	Victim Count	Numeric	No of victims in the case
10	Perpetrator Age	Numeric	Age of the perpetrator
11	Perpetrator Sex	String	Gender of the perpetrator
12	Relationship	String	Relationship of the perpetrator with the victim

Crime analysis helps you analyze open crimes in your db. This is done using a Python library called Matplotlib. This library visualizes data about different aspects given. Forecasts are made for 3 target variables. X perpetrator age is predicted or identified using multiple linear regression. x perpetrator gender is predicted using the KNeighborsClassifier and Neural Networks. The x relationship is predicted using the KNeighborsClassifier and Neural Networks. Explain the results of the prediction and test the accuracy.

WorkFlow Diagram

2. WorkFlow Diagram

The following figure illustrates the workflow of the system. The workflow begins with extracting murder data from Kaggle, a repository of records across different streams. The raw data is then further preprocessed and converted to a criminal databaseSystem. The db is then provided and given for the crime analysis and forecasting phases.

3. Preprocessing

In the above dataset, Month, type of crime, resolved crime, victim's gender, victim's age, victim's race, weapons, Perpetrator's age, Perpetrator's gender, relationship, and so on are all included in the aforementioned dataset. Some of them are of the same high quality. To apply the

mathematical model to the predictions, the qualitative data must be categorised or separated into 0s and 1s. The addition of a dummy column [9] is used to preprocess the data. For N unique values, this function inserts N1 dummy columns to the supplied column.

TABLE II PREPROCESSING

States	**Dummy_A**	**Dummy_B**
Alaska	1	0
Alabama	0	1
Anchorage	0	0

The preprocessing of the data is described in the table above. Three states are chosen from the state of the columns in the dataset: Alaska, Alabama, and Anchorage. The number of unique values in the column state is N = 3 in this case. You'll need to add a N1 dummy column to categorise the data. As a result, there are two dummy columns, dummy A and dummy B. Assign 1 to Dummy A and 0 to Dummy B if state x is Alaska. Assign 0 to Dummy A and 1 to Dummy B if the x state is Alabama. Assign 0 to both Dummy A and Dummy B if the x state is anchorage. LabelEncoder and OneHotEncoder are used to implement it in Python. N dummy columns will be created as a result of this. You can delete one of them at random.

IMPLEMENTATION

1. Analysis

Data analysis is the process of collecting and organizing information to draw valuable conclusions. The process of data analysis combines analytic and logical thinking to extract information from the data.

We used a dataset containing criminal records from 1980 to 2014. The analysis and identification phase examines and guesses that the count value of open crimes, about the weapons used in open crimes, the month with the highest counts of open crimes, and the investigative agency with the highest number of open crimes to do.

Fig.. Years vs. No of Unsolved Crimes

2.Algorithms

a) MultiLinear Regression

Multiple linear regression is a statistical method for determining the association between a dependent variable (criminal age) and a series of independent variables (input evidence collected from the crime scene). This approach estimates the criminal's age based on the input information displayed in Figure 1's metadata column. The Multilinear Regression Line's equation or formula is as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x \quad (1)$$

Where, Y is the dependent variable

x is the independent variable

β_i are coefficients for the regression equations

In crime prediction system scenarios, data from crime scenes is used to use linear regression to calculate the most likely criminal age.

b) KNeighbors Classification: When there are three or more classes to classify into a target variable, the KNeighbors classifier is utilised. For the target variable, there are three kinds of perpetrator genders in this dataset: male, female, and unknown. Similarly, target variable relationships are divided into 27 distinct categories, including boyfriend, husband, and wife. As a result, the KNeighbor classifier is used to identify or decide these target variables (gender-criminal relationship).

c)Neural Networks: These are a subset of machine learning and deep learning techniques and are also known as artificial neural networks (ANNs) or simulated neural networks (SNNs). Neural networks are built using neural networks.Data is stored at different levels. Users can view data based on the type of login. Each node is considered as the separate linear regression then the derived formula i.e., equation is

$\sum wixi + bias = w1x1 + w2x2 + w3x3 + bias$

and the output is

output = f(x) = 1 if $\sum w1x1 + b >= 0$; 0 if $\sum w1x1 + b < 0$

RESULTS AND OUTCOMES

The idea behind this project is that crime is relatively predictable. I just need the ability to sort

Examine the vast amount of data to find patterns that help law enforcement. This type of data analysis was technically impossible decades ago, but it is expected that recent developments in machine learning will be a challenge.

The utilization of AI to distinguish wrongdoing by means of sound or cameras right now exists for work demonstration. It has been shown to work and is expected to continue to grow. However, it is still largely unknown to use AI / ML to predict fraudulent activity or the likelihood that a person will commit fraudulent activity. The biggest test is probably to "demonstrate" to government officials that it works. It's hard to show a negative when the framework is supposed to prevent something from happening. Organizations that are legally involved in arming governments with AI tools to screen territories and predict fraud could benefit from a circle of positive criticism. .

Enhancements in wrongdoing avoidance innovation will probably spike expanded complete spending on this innovation. Potential roads through which to broaden this work incorporate time-arrangement demonstrating the information to comprehend fleeting relationships in it, which would then be able to be utilized to anticipate floods in various classifications of wrongdoing. It would likewise be fascinating to investigate connections between floods in various classifications of wrongdoings – for instance, the facts could confirm that at least two classes of violations flood and sink together, which would be an intriguing relationship to reveal. Different zones to take a shot at incorporate actualizing a progressively exact multi-class classifier, and investigating better approaches to picture our outcomes

CONCLUSION AND FUTURE SCOPE

Conclusion:

In our study, we proposed a method that predicts the number of undetected cases (crimes) and analyses the impact of numerous elements such as year, month, and weapon. Additionally, the guess, or prediction phase, assesses the perpetrator's age, sex, and relationship with the victim. To forecast the perpetrator's description, the suggested method employs Multilinear Regression, K-Nearest Classifier, and Neural Networks. These assumptions, or predictions, were made based on evidence collected at the crime site. The proposed system can be put in place during the course of a criminal inquiry.

We came up with a solution in which implementing this model, By using Neural network it produces an accuracy of 0.87 and By using K-Nearest Neighbor it produces an accuracy of 0.85

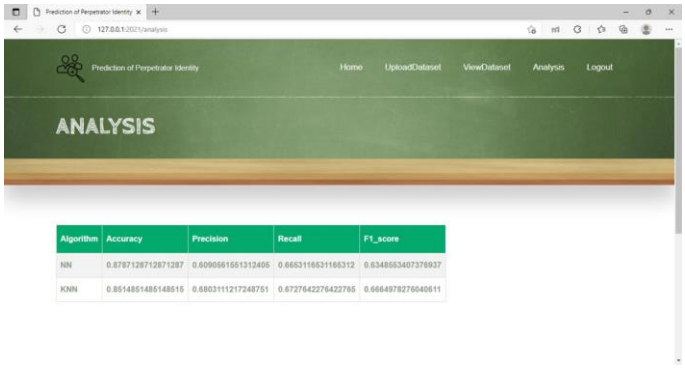

Fig: Analysis for Prediction of Perpetrator identity

Future Scope

The purpose of society is to prevent crimes from occurring in the first place, not to arrest offenders. Use these predictive systems to find more accuracy in the future, and then use that precision to detect and locate criminal hot zones. Use the CNN technique to evaluate the image data and a Google API to display the hot zone to do this job. I'd like to provide a geographical place as well .

Predicting Future Crime Spots:

By utilizing chronicled information and seeing where late violations occurred so we can anticipate where future wrongdoings will probably occur. For instance a rash of robberies in a single region could correspond with more thefts in encompassing zones sooner rather than later.The framework contains potential hotspots that police need to monitor more closely. The system detects suspicious changes in behavior or unusual behavior. For example, if a person appears to be keeping pace back and forth in an area, it suggests that the person may be pickpocketed or may be covering the area for future crimes. It also tracks individual overtime hours.

Pretrial Release and Parole:

After being accused of wrongdoing, most people are discharged until they really stand preliminary. In the past, deciding who should be released before trial or what bail should be for a person is now essentially the best judgment exercise.They found, "That dark litigants who didn't recidivate over a two-year time frame were almost twice as prone to be misclassified as higher hazard contrasted with their white partners (45 percent versus 23npercent)." The report brings up the issue of whether better AI/ML can in the end produce progressively precise expectations or on the off chance that it would fortify existing issues.Each frame is based on true information, but if this current reality information is created by a one-sided police officer, the AI / ML can be one-sided.

REFERENCES

[1] Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In Networks & Soft Computing (ICNSC), 2014 First International Conference on (pp. 406-412). IEEE.

[2] Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International(pp. 1-5). IEEE.

[3]https://www.bing.com/search?q=crime+prediction+using+ml&qs=n&form=QBR%20%20E&msbsrank=1_10&sp=-1&pq=crime+prediction+using+m&sc=1-%20%2024&sk=&cvid=09ECB8B410754A58923136C693DAF026