IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

A Crime Prediction Model/Tool for Hotspot Mapping

Sabin T T[1], Archana R[2], Lithisha K J[3], Vikhyath P[4]
Assistant Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4]
SJC Institute of Technology, Chikabalapura, India

Abstract: *The hotspot map is a popular analytics tool used to help law enforcement and crime reduction resources determine where to focus. Essentially, the hotspot map is used as a data-driven crime map to identify high-crime areas and places where police and other criminal activities should be used. The project uses crime data from a period before (already past) a certain date to create heat maps and test their accuracy in predicting what will happen next. Compare hotspot mapping accuracy with maps used to measure crime issues (thematic mapping of the census output area, spatial ellipses, grid thematic mapping, and core density estimation) and by crime type – the four crime types (crime, crime, car theft, and car theft)..*

Keywords: Crime Prediction

I. INTRODUCTION

An important part of solving a crime is investigating the crime scene. This is based on the recognition that crime is spatial (Chainey and Ratcliffe, 2005) and that crime occurs in specific locations.Crime does not occur in the same way. It tends to focus on specific areas that can be explained by the interaction between victims and offenders and the opportunities for crime that exist (Cohen and Felson, 1979; Brantingham and Brantingham, 1984; Cornish and Clarke, 1986). These crime areas are often called hotspots - "high crime areas by overall distribution of crime" (Chainey and Ratcliffe, 2005, p. 147). Many different techniques can be used to identify and investigate crime patterns. These concepts can be as simple as representing each crime sceneas a point and tracking the geographic distribution of that content; applying features in geographic information systems (GIS) to a shadow office (such as a census or police department), or representing the distribution of crime as a continuous area up to the volume of distribution of crime.

II. PROBLEM IDENTIFICATION

Violence and Crime are a threat to justice and must be contained. Accurate crime forecasting and future predictions can help improve the security of major cities through computing. Limiting people's ability to process complex information from big data affects early and accurate predictions and crime predictions. Accurate crime prediction is difficult, but necessary to prevent crime. Accurately predicting crime rates, types, and hotspots based on previous models presents many computational and time-consuming challenges.

III. RELATED WORK

[1]. Title: Semantic Reasoning-Based Effectiveness for Criminal Information Extraction from Metin
Authors: Vládia Pinheiro; Vasco Furtado; Tarcicio Pequeno; Douglas Nogueira
Summary: This article describes the Architecture for Network Information Extraction Systems, specifically the operation (N. to search for information on crimes).The main feature of this architecture is the NLP module based on a semantic cause model. We demonstrate the effectiveness of the proposed model by using it to provide input to a crime-sharing website called WikiCrimes.
Warning: Predicting the future of crime is not supported.

[2]. Title: Deep Learning and triangulation challenge for real-time crime Prediction Author: Bao Wang, Penghang Yin, Andrea L.Bertozzi, P. Jeffrey Brantingham, Stanley J. Osher, Jack Xin
Summary: Estimating true crime is crucial. However, it is difficult to accurately predict when and where the next crime will occur. Unknown physical models provide reasonable predictions for complex processes.Criminal history data is

both spatially and temporally sparse with an unsatisfactory signal. In this study, we will first talk about the proper representation of criminal records. Next, we modified the spatiotemporal residual network of well-represented data to estimate the crime rate in the Los Angeles metropolitan area. These tests and comparisons with various available estimation methods show that the best of the proposed models are correct. Finally, we provide an internalization process to resolve resource usage for real-world deployment.
Note: It takes more time to guess

[3]. Title: Proposed Hybrid Project for Big Data Mobile Applications for Reporting and Crime Prevention Authors: Abdi Fidow, Ahmed Hassan, Mahamed Iman, X. Cheng, M. Petridis & Clifford Sule
Summary: Traditional Crime Prediction Techniques rely on crime data for a specific location. However, relying solely on criminal records is dubious because this information is limited and often does not reflect the complexity of the action.This chapter presents a new approach to crime reporting and prevention using data collected from multiple sources using a mobile application with big data analytics called Hybrid Intelligent Crime Reporting Application (HIVICRA). It is an infographic-intelligent crime reporting analysis app that combines crime data from local police, social media, and the public, including sentiment analysis and criminal history police records via Twitter feed. Evaluation of the method shows that crime prediction can be improved by combining emotional intelligence with intelligent crime reports.
Warning: Predicting the future of crime is not supported.

[4].Title: Predicting the Occurrence of Crime from Multimodal Data Using Deep Learning Authors: Hyeon-Woo Kang and Hang-Bong Kang
Summary: In recent years, many studies have been conducted on accurate crime prediction. This predictive capability is designed to help prevent crime by increasing the effectiveness of police surveillance. Previous research has used data from a variety of sources, including demographics, economics, and education. Predictive models treat data from different sources equally. These methods have problems in predicting crime occurrences, such as difficulty in finding nonlinear relationships, redundancies, and dependencies between multiple documents.To develop the crime prediction model, we consider environmental background information such as broken glass theory and crime prevention by the built environment. In this paper, we propose a deep neural network (DNN) as a feature-level data fusion method with the environmental context. Our data includes information gathered from various online crime statistics databases, demographic and weather data, and pictures of Chicago, Illinois. Before creating the training data, we select the data related to violence by performing data analysis. Finally, we train our DNN, which consists of four layers: layer-by-layer, physical layer, layer-by-layer, and layer-by-layer representation.Combining valuable data from multiple sources, our unified DNN is the product of a complex decision-making process that analyzes repetitive data. Experimental results show that our DNN model is more accurate in crime prediction than other prediction models.
Warning: Training takes a lot of time and does not support future crime prediction.

[5]. Title: Investigating Crime in Smart Cities Using Interactive Learning Authors: Sharmila Chakravarthy; Steven Schmidt; Li Yang
Summary: Fast and accurate detection of crime is important to ensure the safety of any place.With the rapid development of smart cities, this is the integration of crime detection to improve security. In the past, people relied on video surveillance techniques to achieve this. This usually creates a backlog of video files that must be watched by supervisors. For large urban areas, this increases the workload of the inspectors, causing more errors. Workarounds are used to help reduce the workload.Currently, autoregressive models are used to better predict violent behavior, but these also have shortcomings. We propose a solution to analyze video stream data using a combination of neural networks and hybrid deep learning algorithms. Our system will be able to quickly detect and evaluate violations and reduce the workload of administrative staff. When used in the smart infrastructure of the city, it will increase the efficiency and flexibility of crime detection.
Warning: Predicting the future of crime is not supported.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

IV. SYSTEM ARCHITECTURE

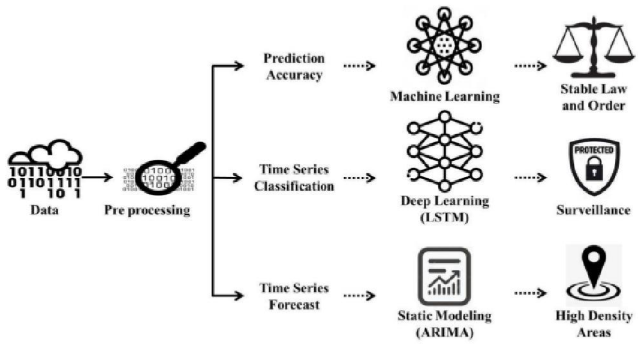


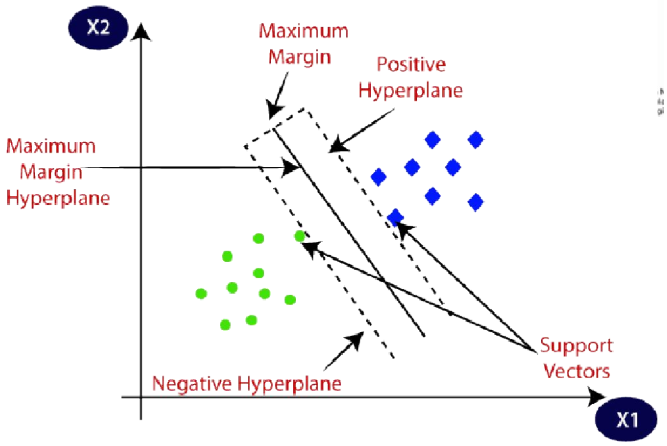
Figure 4.1 System Architecture

In the system, a user of the architecture uses a breach dataset as input. The system will prioritize and use machine learning algorithms to detect crime and ARIMA models to predict crime in the area. This project uses different learning techniques such as logistic regression, support vector machine (SVM), Naive Bayes, k-Nearest Neighbors (KNN), decision trees, and Long Short Term Memory (LSTM) for time series analysis and automatic Regression. Integrated Moving Average (ARIMA) models to better-fit crime data.

V. METHODOLOGY

5.1 Support Vector Machine

Support Vector Machines, or SVMs, are one of the most popular supervised learning algorithms for classification and regression problems. However, it is often used in classification problems in machine learning. The purpose of the DVM algorithm is to create a good line or decision boundary that can divide the n-dimensional space into classes so that we can easily add new data to the class. This well-defined boundary is called the hyperplane. The SVM selects high points/vectors that help create general planes. These conditions are called support vectors and hence the algorithm is called a vector machine. Consider the following figure, where two different classes are divided into a decision boundary or a general plane:



5.2 Naive Bayes

The feature is independent of the occurrence of other features. For example, if fruits are defined by their color, shape, and taste, red, spherical, and sweet fruits are defined as apples. Thus, each feature alone helps to define it as an apple without being dependent on each other.

Bayes: It is called Bayesian because it is based on the principle of Bayes' theorem.

Bayes' theorem: Also known as Bayes' rule or Bayes' rule, Bayes' theorem is used to determine the probability of a previously informed hypothesis.Depends on whatit is.

The formula of Bayes' theorem is:

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

Where,
P(AB) is the posterior probability: is the probability of hypothesis A over hypothesis B.
P(BA) is the probability: it is the probability that a hypothesis will be proven true.

5.3 Decision Tree

In a decision tree, the algorithm starts at the root of the tree to predict the class of given data. This algorithm compares the value of the root element with the value of the data (real data) attribute and follows the branch to jump to the next line according to the match rate. For the next node, the algorithm compares the attribute value again with the other child nodes and continues on its way. He continued this process until he reached a single leaf on the tree.
The completion process can be better understood using the following procedure:
Step 1: Start the tree from the root node S, which contains all the information.
Step 2: Find the best feature in the data using the feature selection index (ASM). Step 3: Divide S into smaller chunks with the best values of the feature.
Step-4: Build decision trees with the best features.
Step 5: Iteratively builds a new decision tree using a subset of the dataset created in Step -3.

5.4 ARIMA

Autoregressive Integrated Moving Average - ARIMA model is a generalization of the Simple Autoregressive Moving Average - ARMA model. Both of these models are used to predict or predict future content in time series. ARIMA is a form of regression analysis that shows the strength of one variable over other variables. The ultimate goal of the model is to predict the future time series movement by analyzing the deviation in the mean value of the series rather than the true value.ARIMA models are appropriate when the data show evidence of non-stationarity. In time series analysis, non-stationary data is always converted to stationary data. Based on the name of the Model, we can divide it into the following subcomponents:
AR: Represents the autoregressive model of the stochastic process. The output of the model is linearly dependent on its previous value, i.e.Some lagging data or previous observations.
MA: A moving average model whose output is linearly dependent on various current and past observations of the stochastic period.
I: Integration here refers to the different steps that make up a continuous time series, eg removing seasonal items and patterns.
The ARIMA model is usually expressed as ARIMA(p,d,q) and the p,d,q parameters are defined as follows:
p: Delay order or delay time (p) of the autoregressive AR model
d: Degree of difference or data from previous values subtracted a number
q: Order of the moving average model MA(q)

VI. IMPLEMENTATION

DATA COLLECTION: In this model, we collect data from kaggle.com

6.1 PRELIMINARY DATA OPERATION

In this model, cleaning, removing stop words, and rooting operations are performed.
Input: - Dataset
Output: - Cleaned dataset.
Pseudocode/Algorithm:
1: Read data.
2: Split the text into one word.
3: Remove stop words like "it", "what", "if", "you", "will", "of", "then", "me", "and".

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

4: Remove wildcards such as \ and <>. /#$?@! ( ) - % ^& [ ] {};: * _ ~ '.
5: Remove values
6: Save a clean-up file.

6.2 CRIME FEATURES

In this model, we use various machine learning algorithms to predict crime.
Input: - Information and user ideas
Output: - Crime estimation.
Pseudocode/Algorithm:
1: Read data.
2: Read user input.
3: Divide the data into training and testing.
4: Design ML algorithms Support Vector Machines (SVM), Naive Bayes, Decision trees, Random Forests, and Extreme Gradient Boosting (Boost).
5: Train the model using a dataset
6: Crime prediction based on user input
7:View crime

6.3 CRIME ACTIVITY

In this model, we use ARIMA and CNN-LSTM and estimate the crime rate for the next 5 days.
Input: - Data set and user input
Output: - Predicted error.
Pseudocode/Algorithm:
1: Read data.
2: Read user reviews.
3: Separate data into training and testing.
4: Build ARIMA and CNN-LSTM models.
5: Demonstrate the model using a dataset
6: Crime prediction based on user input
7: Report crime.

VII. TESTING

Table 7.1: User Input Format

Test Case#	UTC01
Test Name	User input format
Input	Crime dataset
Expected Output	Read and Display Path
Actual Output	Read Selected Dataset and Display Path
Test Result	Success

TABLE 7.2: User input format

Test Case#	UTC02
Test Name	User input format
Input	Null Value Information
Expected Output	Report and Select Data
Actual Output	Report to Select Dataset
Test Result	Success

DOI: 10.48175/IJARSCT-9735

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, May 2023

TABLE 7.3: Pre-process

Test Case#	UTC03
Test Name	Pre-process
Input	Enter Crime DataWants to Release
Expected Output	Should remove unwritten data and remove important features
Actual Output	Release Cleans data
Test Result	Success

TABLE 7.4: Crime prediction

Test Case#	UTC04
Test Name	Crime prediction
Input	Latitude, longitude, date, and time
Expected Output	Forecast Report It is necessary to use the ML algorithm to predict the crime that will occur on the hotspot
Actual Output	predicts the crime and shows the results of the Tested
Test Result	Success

VIII. RESULTS

- Select a method to save data.
- The user must enter the location and length of the hotspot from which the crime should be estimated.
- Include the date and time at which the offense should be estimated.
- After providing input to the user, machine learning algorithms such as decision trees, random forests, and deep learning algorithms for ARIMA and LSTM are used.
- The application then predicts violations and generates reports of potential violations.

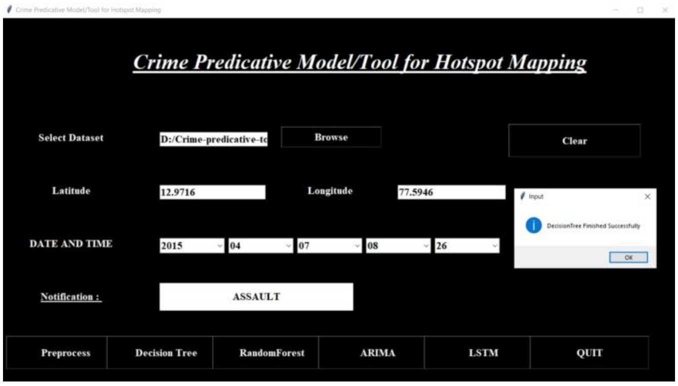


Figure 8.1: Main page

IX. CONCLUSION

Crime is a serious threat to human life, security, and sustainable development and therefore needs to be contained. Estimates and forecasts are often needed by the police to improve crime investigations improve the safety and security of cities and help prevent violence. Here we increase the accuracy of crime by using different machine learning on Chicago and Los Angeles crime data. Among the different algorithms, Random Forest has the highest accuracy on the Chicago dataset while Random Forest has the highest accuracy on the Los Angeles dataset. After data preprocessing, divide the dataset into training and test sets and check the performance parameters.

REFERENCES

[1]. G. Mohler, "Mapping the Token Point System for Murder and Armed Crime in Chicago," Int. J. Estimate, vo l. 30

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 3, Issue 2, May 2023

[2]. A. Iriberri ve G. A. McCarthy.Leroy, "Natural Language Processing and eGovernment: Extracting Reusable Crime Reporting Information", Proc. IEEE International Meeting. information.

[3]. V. Pinheiro, V.Furtado, T. Pequeno, thiab D. Nogueira, "Natural Language Processing Based on Semantic R easoning to Extract Criminal Information from Text",

[4]. S. Chakravarty, S. Schmitt, and L.Yang, "Intelligent Criminal Anomaly Detection Using Deep Learning in S mart Cities", Proc. IEEE Fourth International.

[5]. H.W. Kang ve H.B. Kang, "Siv Deep Learning for Predicting Criminality from Multimodal Data", PLoS ON E, vol. 12, no. Lub Plaub Hlis 4, 2017, Art.

[6]. Imran, X. Cheng, C. Petridis, and C. Sule, "Recommendation for integrated application of mobile phone with big data analytics for crime prevention and prevention" in Polising Social Media Strategies(Security Informa tics and Law Enforcement)

[7]. P. J. Brantingham, M. Valasik and G. O.Mohler, "Does Voting Contribute to DecisionMaking? Results from a Randomly Controlled Experiment,"

[8]. A. Nasridinov ve Y.H.Park, "Study on Performance Evaluation of Machine Learning Algorithms on Crimina l Datasets",

[9]. A. Stec and D.Klabjan, "Crime Crimes with Interactive Learning", 2018, arXiv: 1806.01486. [online].