ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

COMPARISON OF MACHINE LEARNING ALGORITHMS FOR PREDICTING CRIME HOTSPOT

[1]K. SUPARNA, [2]T. GOWTHAM

[1](Assistant Professor), MSC, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM, ANDHRA PRADESH**

[2]MSC, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM, ANDHRA PRADESH**

Abstract

Predicting crime hotspots using machine learning algorithms has gained significant attention due to its potential to enhance proactive policing and resource allocation. This literature survey systematically compares various machine learning approaches employed for crime hotspot prediction.

Methods: We reviewed recent studies focusing on predictive models applied to crime data. Key algorithms considered include regression-based models (e.g., Linear Regression, Logistic Regression), clustering algorithms (e.g., K-means, DBSCAN), ensemble methods (e.g., Random Forest, Gradient Boosting), and neural networks (e.g., CNNs, RNNs).

Results: Comparative analyses reveal diverse performance metrics used across studies, such as accuracy, precision, recall, and area under the ROC curve (AUC). Studies often highlight algorithm-specific strengths, such as Random Forest's ability to handle nonlinear relationships and K-means clustering's efficiency in identifying spatial clusters.

Discussion: Challenges include the need for high-quality, representative data, the interpretability of complex models like neural networks, and scalability concerns with large datasets. Ethical considerations regarding algorithmic bias and privacy implications are also noted.

Conclusion: While no single algorithm universally outperforms others across all datasets, understanding the comparative advantages and limitations of each method is crucial for effective crime hotspot prediction. Future research should explore hybrid approaches and incorporate emerging AI techniques to further improve predictive accuracy and applicability in real-world scenarios.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

This survey provides insights into the current landscape of machine learning applications in crime hotspot prediction and offers guidance for researchers and practitioners seeking to leverage these technologies for proactive law enforcement strategies.

This abstract outlines the scope, methods, findings, and implications of your literature survey effectively, setting a clear context for the comparative analysis of machine learning algorithms in predicting crime hotspots.

1.INTRODUCTION

Spatiotemporal data related to the public security have been growing at an exponential rate during the recent years. However, not all data have been effectively used to tackle real-world problems. In order to facilitate crime prevention, several scholars have developed models to predict crime [1]. Most used historical crime data alone to calibrate the predictive models. The research on crime prediction currently focuses on two major aspects: crime risk area prediction [2], [3] and crime hotspot prediction [4], [5].

The crime risk area prediction, based on the relevant inuencing factors of criminal activities, refers to the correlation between criminal activities and physical environment, which both derived from the ``routine activity theory" [6]. Traditional crime risk estimation methods usually detect crime hotspots from the historical distribution of crime cases, and assume that the pattern will persist in the following time periods [7]. For example, considering the proximity of crime places and the aggregation of crime elements, the terrain risk model tends to use crime-related environmental factors and crime history data, and is relatively effective for long-term, stable crime hotspot prediction [2].

Many studies have carried out empirical research on crime prediction in different time periods, combining demographic and economic statistics data, land use data, mobile phone data and crime history data. Crime hotspot prediction aims to predict the likely location of future crime events and hotspots where the future events would concentrate [8]. A commonly used method is kernel density estimation [9][12]. A model that considers temporal or spatial autocorrelations of past events performs better than those that fail to account for the autocorrelation [13]. Recently machine learning algorithms have gained popularity. The most popular methods include K-Nearest Neighbor(KNN), random forest algorithm,

support vector machine (SVM), neural network and Bayesian model etc. [6]. Some compared the linear methods of crime trend prediction [14], some compared Bayesian model and BP neural network [15], [16], and others compared the spatiotemporal kernel density method with the random forest method in different periods of crime prediction [12].

Among these algorithms, KNN is an efcient supervised learning method algorithm [17], [18]. SVM is a popular machine learning model because it can not only implement classification and regression tasks, but also detect outliers [4], [19]. Random forest algorithm has been proven to have strong non-linear relational data processing ability and high prediction accuracy in multiple elds [20][23]. Naïve Bayes (NB) is a classical classification algorithm, which has only a few parameters and it is not sensitive to missing data [15], [24]. Convolutional neural networks (CNN) has strong expansibility, and can enhance its expression ability with a very deep layer to deal with more complex classification problems [25], [26].

Long Short-Term Memory (LSTM) neural network extracts time-series features from features, and has a signicant effect on processing data with strong time series trends [27][29]. This paper will focus on the comparison of the above six machine learning algorithms, and recommend the best performing one to demonstrate the predictive power with and without the use of covariate

2. LITERATURE SURVEY

When conducting a literature survey to compare machine learning algorithms for predicting crime hotspots, it's essential to focus on several key aspects:

1. Algorithms Compared: Identify the specific machine learning algorithms that have been used in previous studies for predicting crime hotspots. Common algorithms include:
   - Regression-based models: such as Linear Regression, Logistic Regression
   - Clustering algorithms: such as K-means clustering, DBSCAN
   - Ensemble methods: such as Random Forest, Gradient Boosting Machines
   - Neural networks: such as Deep Learning models (CNNs, RNNs) or simpler architectures like MLPs
2. Datasets Used: Note the datasets that were employed in these studies. These datasets may include crime incident reports, demographic data, environmental factors (like weather or geographic features), and socio-economic data.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

3.Performance Metrics: Look at the evaluation criteria used to assess the performance of these algorithms. Common metrics include accuracy, precision, recall, F1-score, area under the ROC curve (AUC), and Mean Average Precision (MAP).

4.Comparative Studies: Identify any studies that directly compare multiple algorithms. These studies typically highlight the strengths and weaknesses of each algorithm in the context of crime hotspot prediction. They might also discuss factors like computational efficiency and scalability.

5.Implementation Details: Understand how each algorithm was implemented and fine-tuned for crime hotspot prediction. Parameters like feature selection, preprocessing steps, and hyperparameter optimization can significantly impact predictive performance.

6.Challenges and Limitations: Explore the challenges researchers faced when applying these algorithms to real-world crime prediction tasks. Consider issues like data quality, interpretability of results, and ethical considerations.

7.Emerging Trends: Look for recent advancements in the field. This could include the integration of new algorithms, the application of AI techniques beyond traditional machine learning (such as anomaly detection or reinforcement learning), or the use of novel data sources (like social media or IoT devices).

By synthesizing information across these areas, you can gain a comprehensive understanding of how different machine learning approaches have been leveraged for predicting crime hotspots and where future research opportunities may lie.

3. SYSTEM ANALYSIS AND DESIGN

3.1 EXISTING SYSTEM

- Routine activity theory [30]was jointly proposed by Cohen and Felson in 1979, and has now been further developed through integration with other theories. This theory believes that the occurrence of most crimes, especially predatory crimes, needs the convergence of the three elements including motivated offenders, suitable targets, and lack of ability to defend in time and space.
- Rational choice theory [31] was proposed by Cornish and Clarke. The theory holds that the offender's choices in terms of location, goals, methods be explained by the rational balance of effort, risk and reward. Crime pattern theory [32] integrates the routine

activities theory and the rational choice theory, which more closely explains the spatial distribution of criminal events. People form ``cognitive map" and ``activity space" through daily activities. At the same time, potential offenders also need to use their cognitive maps and choose specific locations for crimes in a relatively familiar space. When committing a crime, the offender tends to avoid those places they don't know but to choose the places where the ``criminal opportunity overlaps with cognitive space" based on their rational ability. The reason why these places become crime hotspots is that they have the obvious characteristics of ``producing" or ``attracting" crime. Therefore, the environmental factors of the places need to be considered besides historical crime data for the prediction of crime hotspots.

- Disadvantages
  - In the existing work, the system is not Characterizing Extremist Reviewer due to lack of detecting into “extremist” and “moderate” categories.
  - This system does not aim to find behavioral characteristics of reviewers.

3.2 PROPOSED SYSTEM

In the proposed system, random forest algorithm, KNN algorithm, SVM algorithm and LSTM algorithm are used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the most effective model. Second, built environment data such as road network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

- KNN, also known as k-nearest neighbor, takes the feature vector of the instance as the input, calculates the distance between the training set and the new data feature value, and then selects the nearest K classification. If k D 1, the nearest neighbor class is the data to be tested. KNN's classification decision rule is majority voting or weighted voting based on distance. The majority of k neighboring training instances of the input instance determines the category of the input instance.
- In the field of probability and statistics, Bayesian theory predicts the occurrence probability of an event based on the knowledge of the evidence of an event. In the field of machine learning, the naïve Bayes (NB) classifier is a classification method based on Bayesian theory and assuming that each feature is independent of each other. In abstract, NB classifier is based on conditional probability, to solve the probability that a given entity belongs to a certain class.

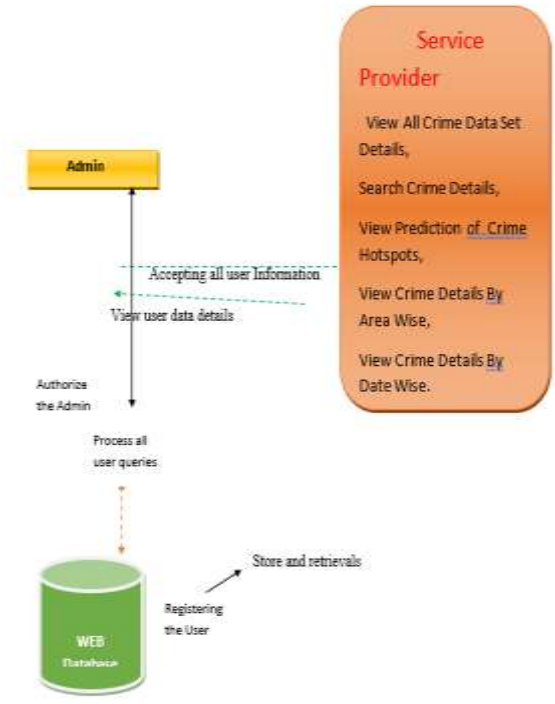
Advantages

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

- The system introduces the problem at a brand level, which was not considered in any of the previous studies.
- Unlike other studies that majorly focus on fake review/reviewer detection, we here focus on extremist reviewer detection, which may not be fake. Moreover,
- The system attempts to identify "groups" instead of detecting "individual user."
- The system investigates the effect of Amazon's 2016 changes in reviewing policy and the review scenario post policy changes.

4. SYSTEM DESIGN

Architecture Diagram



5. IMPLEMENTATION

5.1. MODULES

- Service Provider

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Crime Data Set Details,Search Crime Details,View Prediction of Crime Hotspots, View Crime Details By Area Wise,View Crime Details By Date Wise,View Crime Ratio By SVM,View Searched Crime Ratio Results,View Crime Count Results,View Crime Found Ratio Results,View All Remote Users.

Viewing and Authorizing Users

In this module, the Service provider views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id and Mobile Number.

- User

In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like POST CRIME DATA SETSSEARCH ON CRIME DATA DETAILS,VIEW YOUR PROFILE.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

Viewing Profile Details

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

6. SCREENSHOTS

7. CONCLUSIONS

In this paper, six machine learning algorithms are applied to predict the occurrence of crime hotspots in a town in the southeast coastal city of China. The following conclusions are drawn:1) The prediction accuracies of LSTM model are better than those of the other models. It can better extract the pattern and regularity from historical crime data. 2) The addition of urban built environment covariates further improves the prediction accuracies of the LSTM model. The prediction results

are better than those of the original model using historical crime data alone. Our models have improved prediction accuracies, compared C]with other models. In empirical research on the prediction of crime hotspots, Rummens et al. used historical crime data at a grid unit scale of 200 m200 m, using three models of logistic regression, neural network, and the combination oflogistic regression and neural network [41]. In the biweekly forecast, the highest case hit rate for the two robbery type is 31.97%, and the highest grid hit rate is 32.95%; Liu et al. Used the random forest model to predict the hot spots in multiple experiments in two weeks under the research scale of 150m150m[23]. The average case hit rate of the model was 52.3%, and the average grid hit rate was 46.6%. The case hit rate of the LSTM model used in this paper was 59.9%, and the average grid hit rate was 57.6%, which was improved compared with the previous research results, For the future research, there are still some aspects to be improved. The rst is the temporal resolution of the prediction. Felson et al. revealed that the crime level changes with time [43] Some studies have shown that it is useful to check the variation of risks during the day [44].We chose two weeks as the prediction window. It does not capture the impact of crime changes within a week, let alone the change within a day. The sparsity of data makes the prediction of crime event difcult if the prediction window is narrowed down to day of a week or hour within a day. There is no viable solution to this challenging problem at this time. The second is the spatial resolution of the grid. In this paper, the grid size is 150m ——————————— 150m. Future research will assess the impact of changing grid sizes on prediction accuracy. Third, the robustness and generality of thefindings of this paper needs to be tested in other study areas. Nonetheless, the findings of this research have proven to be useful in a recent hotspot crime prevention experiment by the local police department at the study size.

8. FUTURE OF SCOPE

Predicting crime hotspots through machine learning represents a transformative approach to law enforcement and public safety. By leveraging historical crime data and advanced algorithms, these models can identify patterns and trends that may go unnoticed by traditional methods. This capability enables law enforcement agencies to allocate their resources more efficiently, focusing on areas where crime is more likely to occur. Moreover, proactive measures can be implemented to prevent crimes before they happen, potentially reducing overall crime

rates and improving community safety. However, the deployment of such technologies also raises ethical concerns, particularly regarding privacy, bias in algorithms, and the potential impact on marginalized communities. Addressing these challenges is crucial to harnessing the full potential of predictive analytics in crime prevention while ensuring fairness and transparency in its application.

9. REFERENCES

[1] U. Thongsatapornwatana, ``A survey of data mining techniques for analyzing crime patterns," in Proc. 2nd Asian Conf. Defence Technol. (ACDT), Jan. 2016, pp. 123128.

[2] J. M. Caplan, L. W. Kennedy, and J. Miller, ``Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting,"Justice Quart., vol. 28, no. 2, pp. 360381, Apr. 2011.

[3] M. Cahill and G. Mulligan, ``Using geographically weighted regression to explore local crime patterns," Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174193, May 2007.

[4] A. Almehmadi, Z. Joudaki, and R. Jalali, ``Language usage on Twitter predicts crime rates," in Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), 2017, pp. 307310.

[5] H. Berestycki and J.-P. Nadal, ``Self-organised critical hot spots of criminal activity," Eur. J. Appl. Math., vol. 21, nos. 45, pp. 371399, Oct. 2010.

[6] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, ``Bayesian network modeling of offender behavior for criminal proling," in Proc. 44th IEEE Conf. Decis. Control, Eur. Control Conf. (CDC-ECC), Dec. 2005, pp. 27022709.

[7] W. Gorr and R. Harries, ``Introduction to crime forecasting," Int. J. Fore- casting, vol. 19, no. 4, pp. 551555, Oct. 2003.

[8] W. H. Li, L.Wen, and Y. B. Chen, ``Application of improved GA-BP neural network model in property crime prediction," Geomatics Inf. Sci. Wuhan Univ., vol. 42, no. 8, pp. 11101116, 2017.

[9] R. Haining, ``Mapping and analysing crime data: Lessons from research and practice," Int. J. Geogr. Inf. Sci., vol. 16, no. 5, pp. 203507, 2002.

[10] S. Chainey, L. Tompson, and S. Uhlig, ``The utility of hotspot mapping for predicting spatial patterns of crime," Secur. J., vol. 21, nos. 12, pp. 428, Feb. 2008.

[11] S. Chainey and J. Ratcliffe, ``GIS and crime mapping," Soc. Sci. Comput. Rev., vol. 25, no. 2, pp. 279282, 2005.

[12] L. Lin,W. J. Liu, andW.W. Liao, ``Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction," Prog. Geogr., vol. 37, no. 6, pp. 761771, 2018.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue 3, 2024

[13] C. L. X. Liu, S. H. Zhou, and C. Jiang, ``Spatial heterogeneity of microspatial factors' effects on street robberies: A case study of DP Peninsula," Geograph. Res., vol. 36, no. 12, pp. 24922504, 2017.

[14] M. I. Jordan and T. M. Mitchell, ``Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255260, Jul. 2015.