IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022

# AN ANALYTICAL STUDY ON RECOGNITION AND ELIMINATION OF CRIME PATTERNS USING MACHINE LEARNING ALGORITHMS

**[1]P. Kalyani, [2]G. Vidya Sagar, [3]M. Kanchana, [4]P.Sravan Kumar Reddy**

[1,2,3]Dept of Computer Science and Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.

[4]Dept of Electronics and Communication Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.

## ABSTRACT

The development of policing strategies and the implementation of measures for crime control and reduction depend heavily on crime prediction. Today, machine learning is the most popular prediction technique. However, there haven't been many studies that thoroughly contrasted various machine learning approaches for criminal behaviour prediction. substantial seaside city in the southeast In this study, The ability of several machine learning algorithms to forecast crime is assessed using public property crime data from China from 2015 to 2018. The LSTM model appears to outperform KNN, Random Forest, Support Vector Machine, Naive Bayes, and Convolution Neural Networks based only on results from historical crime data. As a result, it is advised that characteristics linked to criminological theories and knowledge of previous crimes be used to predict future crime. Not all machine learning methods for predicting crimes are equally effective.

**Keywords:** naïve bayes, convolution neural networks, KNN and RNN algorithms, crime prediction

## I. INTRODUCTION

Public security-related spatial and temporal data have been expanding rapidly in recent years. Even said, not all data have been successfully used to address issues in the real world. A number of researchers have developed crime prediction models to help with crime reduction. The majority of them essentially made minor adjustments to their forecast algorithms utilising crime data from the past. The two primary areas of research in crime prediction at the moment are identifying crime hotspots and estimating crime risk. According to the "routine activity theory," which is based on the key factors that influence criminal behaviour, "crime risk area prediction" refers to the relationship between criminal activity and the local region. By looking at the historical distribution of crime cases and assuming that the tendency will continue in the subsequent time periods, traditional crime risk estimating methods commonly pinpoint crime hotspots. For instance, the terrain risk model considers the proximity of crime places and the buildup of crime components. Typically, it also incorporates data on prior crimes as well as environmental characteristics that are associated to crime. Predicting hotspots for crime numerous studies have combined data from land use, mobile phones, demographic and economic statistics, and crime history in their empirical research on crime prediction over a range of time periods. The goal of crime hotspot prediction is to identify potential hotspots for future crime events and their likely locations. Kernel density estimation is a frequently used technique. The performance of models differs depending on whether they account for the spatial or temporal autocorrelations of past events. Machine learning algorithms are now being used more often. A few of the most widely used techniques are K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), Neural Network, and Bayesian Model.

One of these efficient algorithms for supervised learning is KNN. Being able to perform classification, regression, and outlier detection tasks makes SVM a popular machine learning model. It has been demonstrated

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022

that the Random Forest method offers excellent non-linear relational data processing skills as well as good prediction accuracy across a wide range of fields. Only a few parameters and a method known as Naive Bayes(NB) are used in traditional classification techniques because they are not sensitive to missing data. Strongly extensible, convolution neural networks (CNNs) can improve their expression capability with a very deep layer to handle more challenging classification issues. Processing of data with strong time series trends is significantly impacted by the LSTM neural network, which separates time-series features from features. The primary focus of this study will be a comparison of the six machine learning algorithms stated above, which will identify the algorithm that performs the best in terms of displaying both predictive power and covariate-free predictive power.

## II. LITERATURE SURVEY

To predict future geographic hotspots of criminal activity is the aim of crime hotspot prediction. Theoretical criminology offers the essential theoretical underpinning. Furthermore, they provide a fundamental framework for how the police can use knowledge of crime hot spots for crime prevention or control. Numerous associated criminological theories, in particular, help us comprehend the crucial part that location factors play in the emergence and accumulation of criminal events. The main elements are the theories of routine activity, rational choice, and criminal patterns. Situational crime prevention is primarily based on these three theories, theoretically. Routine activity theory was first put forth by Cohen and Felson in 1979, and it has since been improved by being combined with other concepts.This theory contends that the convergence of the three factors—motivated perpetrators, appropriate targets, and incapacity to defend in time and space—is necessary for the majority of crimes, particularly predatory crimes, to occur. Cornish and Clarke proposed the theory of rational choice. According to the hypothesis, the offender's decisions about location, objectives, and tactics can be attributed to a logical trade-off between effort, danger, and reward. The rational choice theory and routine activity theory are combined in crime pattern theory, which more thoroughly describes the spatial distribution of criminal events. People create their "cognitive map" and "activity space" through their everyday routines.Additionally, potential offenders must make decisions about where to commit crimes in a location that is generally familiar by using their cognitive maps. When committing a crime, a person usually avoids unfamiliar areas and instead opts for locations where, in their opinion, there is a "criminal opportunity that overlaps with cognitive space." These locations "produce" or "attract" crime, which is why they end up being crime hotspots. For the purpose of predicting crime hotspots, local environmental elements must also be taken into account in addition to historical crime data.

### BUILT ENVIRONMENT DATA:

Numerous studies that are currently available show that opportunities to deter and prevent crime have an impact on the built environment of cities, which has a significant impact on urban criminal behaviour. According to the 2007 Global Habitat Report, the built environment has a major impact on the frequency of criminal activity. The crime prediction model takes into account road network density data and point of interests (POIs) data as covariates..

1) POIDATA:
2) The location data and attribute data of numerous urban facilities are included in the urban infrastructure data POI. While entertainment venues draw criminals, catering facilities, shopping malls, and stores are typically situated in areas with easy access to transportation and high pedestrian traffic. As a result, these locations attract a variety of different target populations. These POIs are chosen to serve as the prediction model's covariates.
3) ROADNETWORKDENSITY:

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022

Road network density is often defined as the total number of roads divided by the area of an areal unit. A more densely populated region draws more visitors, including criminals and prospective victims. Previous research has demonstrated that the density of the road network affects crime rates, particularly in public areas.

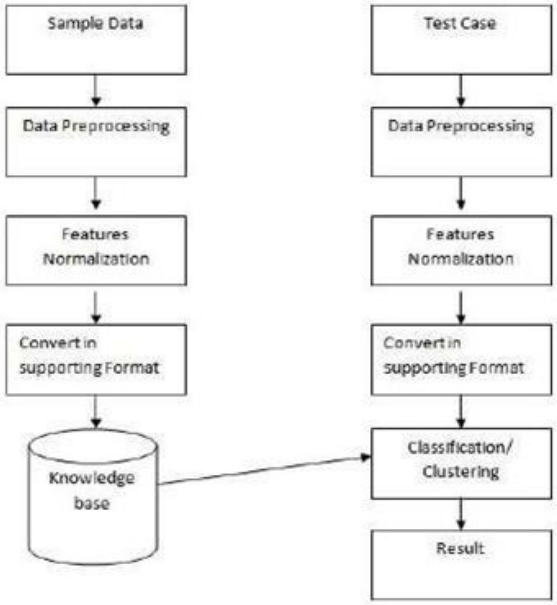
## III.EXISTING SYSTEM

To predict future geographic hotspots of criminal activity is the aim of crime hotspot prediction. Theoretical criminology offers the essential theoretical underpinning. Additionally, they offer a fundamental framework for how the police can use their understanding of crime hotspots to reduce or prevent crime. Numerous associated criminological theories, in particular, help us comprehend the crucial part that location factors play in the emergence and accumulation of criminal events. The ideas of routine behaviour, deliberate decision, and criminal patterns make up the key components. Datasets can be found on Kaggle.com.

## IV.PROPOSED SYSTEM

In this study, crime prediction is done using the random forest, KNN, and SVM algorithms. First, the models are calibrated using only historical crime data as the input. Comparison would show which model is most successful. Second, to investigate if prediction accuracy can be further increased, built environment variables like road network density and poi are introduced as covariates to the predictive model.

## V.SYSTEM ARCHITECTURE:




### 5.1 IMPLEMENTATION: PREDICTION MODEL:

Some of the techniques employed in this study to predict crime include the random forest approach, KNN algorithm, SVM algorithm, and LSTM algorithm. First, the models are calibrated using only historical crime data as the input. Comparison would show which model is most successful. Second, to investigate if prediction accuracy can be further increased, built environment parameters like road network density and poi are introduced as covariates to the predictive model.

KNN

The class with the nearest neighbours is the one whose data must be validated if k = 1. KNN uses weighted

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022

voting based on distance or a majority vote as a classification method for data. The input instance is mostly classified by the k neighbouring training instances based on its closeness.

**RANDOM FOREST**

Several tree classifiers make up the random forest. {h(x, βk),k = 1. . . }, In this case, the output of the forest is determined by voting; the input vector x; the independent random vector k; and the meta classifier h(x, k) are all uncut regression trees built using the CART algorithm. The split attribute set and the training sample set are both picked randomly using the bagging technique, mirroring the randomness of the random forest in both cases. Assuming that there are M attributes total in the training sample, In order to identify which of the f attributes has the best split mode.

**SVM**

Based on statistical learning theory, SVM is a data mining technique that successfully addresses a wide range of issues, including pattern recognition and regression (classification problem, discriminant analysis). The SVM mechanism identifies a superior classification hyperplane that complies with the classification standards in order to guarantee classification accuracy and maximise the blank space on both sides of the hyperplane. The best classification of linear systems may theoretically be accomplished via SVM.

**LSTM**

A deep neural network built on RNN is called LSTM. The fundamental idea behind LSTM is to include a special unit (memory module) that can be used to learn the most recent data as well as to extract any correlations and rules that exist between the data in order to convey the information. The LSTM is more suitable for deep neural network calculations because of its memory module, which minimises information loss.

They are applied to selectively memorise the feedback error function's correction parameters when the gradient decreases. Figure 1 displays the precise structure..

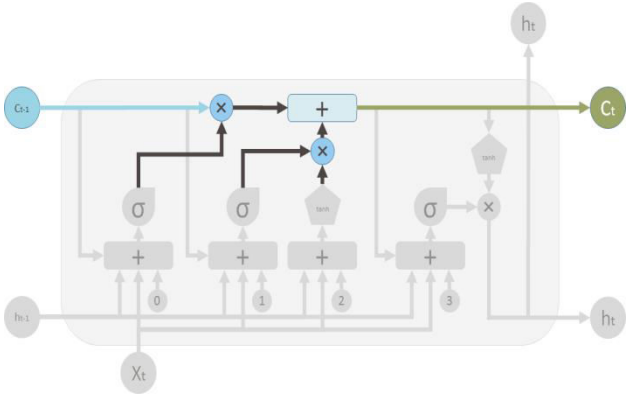



FIGURE1.The LSTM algorithm's structural diagram

Unlike the RNN, which simply communicates the cell state C over time, the LSTM comprises two state chains that are transmitted over time: h (hidden layer state) and C (cell state). Ct-1, Ct, and ht are the state values from the previous time, the current time, and the current time as communicated from the hidden layer from the previous time, respectively, in the LSTM memory cell. Xt is the input value as of this moment. As ht-1 and Xt pass through the forgetting gate, the information that should be forgotten is decided.The output value for the cell state can be between 0 and 1, with 1 representing full reserve of all information and 0 complete forgetting. The hint supplied by the "forgetting gate ft"

$$ft = \sigma(wf \cdot [ht-1, xt] + bf)$$

Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022

where the activation function is a sigmoid, and w and b are the forgetting gate's weight matrix and bias vector, respectively. New data can be updated into a cell using one of two methods. The information that has to be updated is first calculated using the Sigmoid function's input gate, and then the cell state is updated with a new value, kt, produced by the Tanh layer:

$it = \sigma(wi \cdot [ht-1, xt] + bi)$

$kt = \tanh(wk \cdot [ht-1, xt] + bk)$

The cell status affects the output's final result. The hidden layer transmits the state value ht to the next time after the sigmoid function has first classified the output results, picked the data to be produced, and used the tanh function to process the cell state. After sigmoid processing, Ht may currently recover the pre output value y, as indicated in equations (5) through (7):

$Ot = \sigma(wO \cdot [ht-1, xt] + bO) ht = Ot * \tanh(Ct)$

$y = \sigma(w0ht)$

## VI. Results


Research paper  UGC CARE Listed ( Group -I) Journal Volume 11, Iss 12, Dec 2022



## CONCLUSION

Six this study uses machine learning techniques to forecast the emergence of crime hotspots in a city on China's southeast coast. The findings are as follows: 1) The LSTM model outperforms the other models in terms of prediction accuracy. It is better at finding patterns and regularities in historical crime data. 2) Including factors linked to the urban built environment improves the LSTM model's ability to predict outcomes. Predictions that are just based on previous crime data perform better than those that are based on the original model. Our models have improved forecast accuracy over other models. In their study, There are still several areas that could be improved for the research in the future. The prediction's temporal resolution is the first. According to Felson et al., crime levels fluctuate throughout time. According to several studies, it is helpful to keep track of how hazards change during the day. The two-week forecast timeframe was chosen. The model's typical grid and case hit rates are 52.3% and 46.6%, respectively. Both the average grid hit rate and the case hit rate of the LSTM model utilised in this investigation—59.9% and 57.6%, respectively—were higher than the findings of previous investigations. The research could still be improved in a variety of areas in the future. The prediction's temporal resolution is the first. Crime rates change over time, according to Felson et al. Numerous studies have shown the value of monitoring changes in hazards throughout the day.

## REFERENCES

1. U.Thongsatapornwatana," A survey of data mining techniques for analyzing crime patterns", Proc. 2nd Asian Conf. Defence Technol. (ACDT),pp. 123-128, Jan.2016.
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. 360-381, Apr. 2011.
3. M.CahillandG.Mulligan,"Using geographically weighted regression to explorelocal crime patterns",

Social Sci. Comput. Rev.,vol. 25, no. 2, pp. 174-193, May2007.

4. A.Almehmadi, Z.JoudakiandR.Jalali, Language usage on Twitter predicts crimerates", Proc. 10th Int. Conf. Secur. Inf. Netw.(SIN),pp. 307-310, 2017.
5. H.BerestyckiandJ.-P.Nadal,"Self-organised critical hot spots of criminal activity",Eur. J. Appl. Math., vol. 21, no. 4, pp. 371-399,Oct.2010.
6. K.C.Baumgartner, S.FerrariandC.G.Salfati, "Bayesian network modeling of offenderbehaviorforcriminalprofiling",Proc.44thIEEE Conf. Decis. Control Eur. Control Conf.(CDC-ECC),pp. 2702-2709, Dec.2005.
7. W.Gorrand R. Harries, "Introduction tocrime forecasting", Int. J. Forecasting, vol. 19,no.4, pp. 551-555, Oct. 2003