iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49204

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

Suraksha Kavach - Crime Prediction and Analysis

Esha Gupta[1], Shashwat Samarpan[2], Sanjukta Sharma[3], Raj Pratap Singh[4], Prof Aruna S[5]
Department of Computer Science & Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India

Abstract: Suraksha Kavach is a word used to describe protection to ward off the evil eye in our lives. Our primary motivation behind taking up this work was our common interest in the fields of Machine Learning and contribution towards the field of police, the welfare of the public, and their safety, thus the name "Suraksha Kavach". In addition to this, observed that there was a lack of means for facilitating the safety of society. So we decided to take up this topic and try to solve this massive problem. Be using several machine learning algorithms to create an ensemble method to solve this problem. With the help of ensemble methods, the accuracy of the result that we will obtain will be very high. Our effort can bring about a revolution in technology regarding crime prediction and analysis and greatly help the crime branch and police personnel.

I. INTRODUCTION

A prevalent social issue that has a severe effect on a society's standard of living and economic development is a crime. It's seen as a vital factor in deciding if people have moved to a different city and which places shouldn't be preferred to travel to. Crime has an adverse impact on society by undermining a population's feeling of community, causing social ties to dissolve as a result of routinely avoiding certain places, making people reluctant to go out. People may stop visiting a town and also may leave if they believe that too many crimes are occuring in that particular town.

Crime has an adverse effect on the economy of a country by putting a finance-related burden on taxpayers and governments due to the increase in demand for enforcement of law, courts as well as other costs like stress and a lower quality of life for victims of crime. Reduction in crime results in much improved financial condition of a country.

Crime rate is increasing day by day and the older methods are now no longer useful in solving crime cases as criminals now have advanced technologies for committing crimes.

This calls for new methods of technologies that can counter-attack criminal technologies. Several machine learning algorithms are there which can aid in crime prediction and analysis. Several other technologies of Machine Learning such as Data Mining, Data Visualization, Trends forecasting, Deep Learning, and ensemble Methods are there which are very useful in crime prediction and analysis. Crime prediction and analysis greatly helps in reduction of crime and thus leads a country to prosperity.

II. LITERATURE SURVEY

Increasing rates of crimes are a big threat to this world. To understand the features of a crime and identify crime characteristics, scientists are actually investing time in studying crime and criminal activities. Analysing crime data is difficult since it develops in large amount quickly, which can lead to issues such as storage and processing.[1]

The dataset which is used for classification is divided into two different sets: the training set and the test set. The training set is where the machine learning algorithm is originally performed, and the test set is where the prediction model is afterward applied.

The categorization algorithm used here in predicting crime is-

A. Decision Tree

This is used to identify event outcomes by taking values that can be taken by target variables in the tree. It works well with monotonic transformation but it cannot handle linear relationships and unstable trees. It also becomes difficult to interpret the tree where the number of nodes is large. [1]

Since the labelled training data was given, the machine learning model being considered in the work was built using supervised learning (classification) approaches.

A new observation must be classified into one of a set of categories (subpopulations) based on a training set of data that contains observations (or instances) whose category membership is known. Our method includes the collection of data, data preprocessing, the development of a classification model using training data, and model evaluation using test data. After that, a model which is trained and validated was used to score the data that was incoming.

©IJRASET: All Rights are Reserved

B. Exploratory Data Analysis Applied For Smart Cities

The research uses SFPD crime data collected over a 15-year period (2003 to 2018) and analyses it to determine crime trends over time and forecast crimes that may occur in the future. The suggested data preparation approach enhances prediction for the highly skewed dataset as compared to other studies that used the same data [1, 4, 15, 23]. [2]

IncidntNum	Category	Descript	DayOfWeek
160919032 160920976	VANDALISM ASSAULT	MALICIOUS MISCHIEF, VANDAL THREATS AGAINST LIFE	Friday Saturday
Date	Time	PdDistrict	Address
11/11/16 11/12/16	7:00 2:58	MISSION CENTRAL	2600 Block of MASON ST FILLMORE ST / GEARY BL
X	Y	Resolution	Location
-122.4052518 -122.4140032	37.751525 37.8079695	NONE ARREST, BOOKED	(37.75152495730467, -122.4052517658 (37.80796947292687, -122.4140031783
PdId			
160919032281600 160920976190570			

San Francisco Police Complaints dataset (2003 - 2018)

A crime hotspot is a place with a higher-than-average crime rate relative to other places. The basics of spatial scan statistics, a crucial technique for hotspot identification, are also covered in the work.

1) Naive Bayes

It is an algorithm that follows the mathematical concept of Bayes Theorem. This algorithm requires less computational time with respect to other algorithms and can also handle categorical input variables well. [2]

2) Random Forest

It uses the decision tree algorithm on data, collects predictions from each of them, and tries to find the best solution. It consists of low bias and components of this model are minimised by averaging the prediction in the tree. [2]

3) K-Nearest Neighbour (KNN)

This algorithm takes k instances of the dataset and evaluates the variable y of interception and concludes an output. This algorithm also requires less computational time because it does not need training initially as it learns from the data set while making a prediction.[2]

C. Clustering Based Hotspot Identification

Crime is an illegal conduct that is subject to governmental or other punishment. This recurring recurrence of specific criminal activities in a particular location throughout time can therefore be referred to as a temporal event. The word "Hotspot" refers to a specific geographic area within the research area where crime is more concentrated than it is elsewhere.

Predicting future concentrations of crimes in a certain area is the main goal of crime hotspot analysis. Data on points of interest (POIs) and the density of the road network are taken into account as variables in the model for predicting crime.[18]

This increasing crime rate can be reduced if we can properly analyse and predict the crime based on the previous history of crimes in the area[6].There are several factors that lead to the occurrence of a crime, we should try to devise mechanisms to take care of all such factors.

D. Spatial Analysis Method

Web mapping and visualisation of crime prediction using R language and its various libraries RgoogleMaps, and googleVis is a very useful method. This framework shows the trends of crimes and various ways of predicting crimes using various algorithms[6]. The various important phases in it are Data Collection, Data processing, Visualisation, and Model Building. Each phase has its own importance and together they create a very effective model for the prediction analysis of crimes.





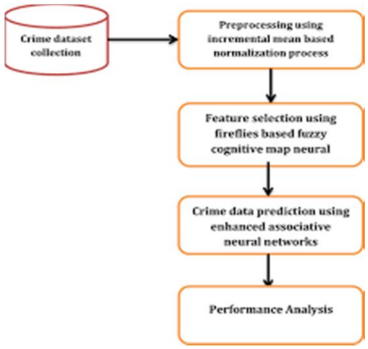
3D Visualisation

Spatial Analysis can be used to solve this problem. This framework is based on GUI-based tools and R programming language. The Various phases in this methodology are

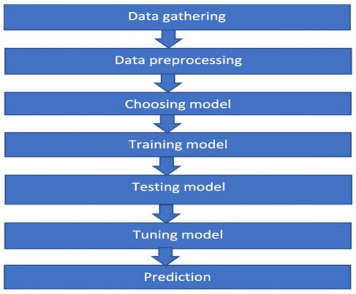
- Data Collection and Processing
- data Visualization
- crime Prediction

E. Data Mining Method

With increase in computerised system for reducing crimes, the criminals are also coming with more advanced system for committing crimes. So the increase in new and more improved techniques for catching crimes should go on. This method will greatly help police to catch criminals. An extremely useful method in this regard is data mining. Data mining is a strong tool to extract useful information from large datasets[5].



Crime Prediction and Crime clustering based on the input dataset



Steps followed in prediction

The crime Analysis methods are pointed out as follows:

- Text and Content Based Methods
- NLP Based Methods
- Crime Patterns Analysis And Evidence-Based Methods
- Geo-Location Based Methods
- Spatial Analysis
- Prisoner Based Methods
- Communication-Based Methods[5]

Profiling of criminals Additionally, a fresh idea known as criminal profiling aids crime scene detectives in documenting the traits of criminals. It is a behavioural technique that is meant to assist investigators in accurately predicting and profiling the features of unidentified criminal subjects or offenders. It is a very accurate instrument for profiling the characteristics or details of criminals.[10]

Conventional crime prediction techniques have severe drawbacks such as lack of factors. In addition to other factors, this paper takes into consideration the social media content. The data generated by social media content can serve as an important measure in crime analysis and prediction.[4]With the help of social media content, the sentiment analysis of the people can be done and hence this will be an important factor in reducing and predicting crimes.

F. Innovation In Crime Prediction Using Twitter Ana Social Media Sentiments

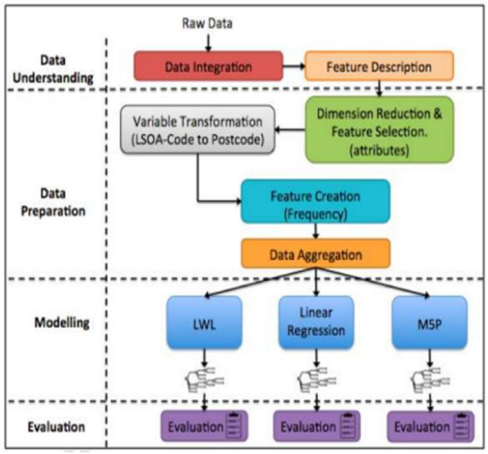
The data considered for this paper was twitter data, taking into consideration tweets and GPS coordinates. The data is taken from chicago data portal website. Another important and unique factor this paper considers is weather. The weather data was collected from Weather Underground, a website. The modelling of the data is done in a very unique and helpful way. To construct the crime prediction and analysis model, the training set of theft density, twitter data and weather data are considered. The data also consists of latitude and longitude information. The crime points are considered to be small sectors.[4]

Some of the important explanatory variables used are:

- Density of Crime
- Polarity score of tweets
- 3 –day Trend of polarity score
- Temperature (Fahrenheit)
- Dew points (Fahrenheit)
- Mean humidity (%)
- Mean sea level pressure (Inch)

G. Data Mining On Open Data

The police have been consistently compiling crime data for many years,but over the past ten years, Open Crime Data1 and web-based applications that display crime statistics on maps from official sources, such as the police, have proliferated. The most widely used method is hot spot analysis. [9]

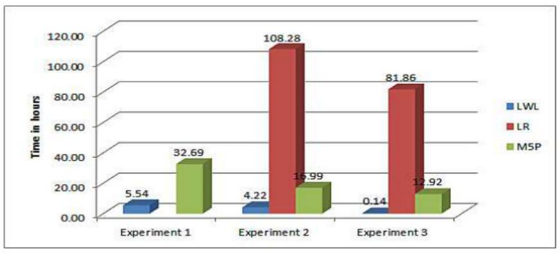


Six phases make up the CRISP-DM approach, and they are briefly detailed below:

- Knowing a business's goals is essential to doing business with it.
- Familiarising yourself with the data is a necessary step in data interpretation.
- The process of formatting the data required for the modelling process is known as data preparation.
- Applying a variety of modelling approaches or algorithms is part of modelling.
- Evaluation is the process of determining how well the models created in the preceding phase performed.

Algorithms:

- Instance-based education
- Linear Regression
- Decision trees[9]



In this study, we used official data from the Police Department of the UK to create forecast models for crime counts by crime category, month etc. Investigating the performance level that may be obtained from 40 months of criminal records was the major objective of the analysis.

H. Machine Learning Techniques

A socio-economic issue, crime has an impact on both economic growth and life quality. The specifics of how crime is perpetrated vary depending on the type of civilisation and culture. Earlier research on crime prediction showed that factors including education, poverty etc have an effect on crime rates. The authentic datasets had been retrieved from Vancouver's open records repository. For this study, datasets—crime and neighbourhood—are employed. The VPD has been recording crime records since 2003, and updates it on every Sunday morning. It gives info at the type of crime that changed into dedicated in addition to the event and placing of the offence.

Criminology is one of the most important areas that can produce significant results with the help of data mining. Datasets have different various attributes such as age, education, gender, employment status, civil status, location of occurrence, etc.[11]

The methods used were ANN machine learning prediction models and a boosted decision tree to achieve decent accuracy. Prediction accuracy can be enhanced by using both the algorithm and the data for specific use. In spite of not being very accurate, it provides a good framework for more analysis.[8]

I. Use Of Stacked Generalisation

Stacking and bagging ensemble approaches allow the integration of a variety of algorithms to create an ensemble model. The output of an ensemble model is produced by several classifiers with different output capabilities.

Two phases make up classification. the creation of the model, and the model's use. Model for testing is used in the beginning procedure. The top model is chosen in the subsequent phase. Based on the model's precision and timeliness, these models will be selected.

Through combining numerous classifiers, often by polling, utilising related data, and enhancing their presentation, ensemble learning approaches seek to produce a meta classifier.

There are three steps in this model:

- Sub-models and aggregators.
- Combining predictions.
- Organised crime data set.

1) Move 1: The above-mentioned models are configured not dependent on one another using data set training during the cycle, which is then followed by a cross-validation technique that creates authentication value.
2) Move 2: Entire training dataset is used in the testing cycle to produce the forecasts
3) Move 3: In the new training package, the whole validations for layer 2 are successively aggregated.
4) Move 4: Dataset in this stage contains four verified model parameters and four observations.
5) Move 5: Similarly, the dataset for the study's second cycle included three models and observations.
6) Move 6: The resultant model is then used to calculate the combinational multimodal simulation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue II Feb 2023- Available at www.ijraset.com

J. Predictive Model For Smart Cities

Over the past few years, there has been an increase in crime. The accuracy of crime projections contributes to a drop in crime.[13] Using 2018 crime datasets, this study uses statistical analysis techniques and machine learning approaches to predict various sorts of violence in New York City. It also combines weather information with temporal factors like cloud cover, lighting, and time of day to determine how relevant weather information is to crime data.

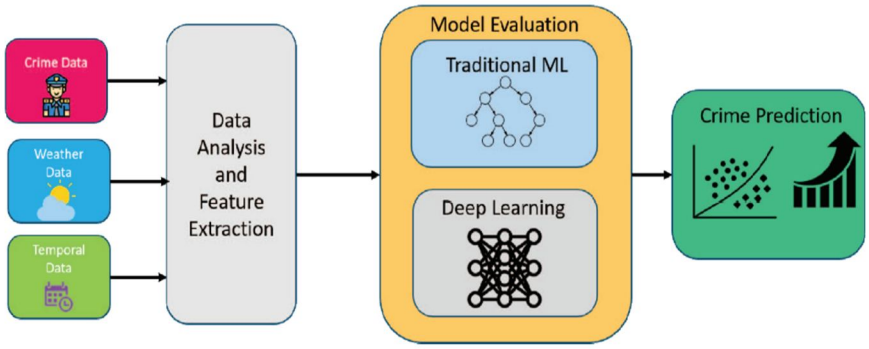


Figure 1: Architecture Flow

The study discovered data suggesting a connection between crime and the kind of visitors that came there for various purposes.

- *Dataset Extraction:* DataSet includes information on 220K events of all crime kinds.
- *Data Cleaning:* Before using any analytical, machine learning, or deep learning techniques, all the empty or null values were changed to "NA".
- *Feature Selection Approaches:* To find the best subset from a dataset, we employed a variety of feature selection strategies, including best subset selection, advance stepwise collection, and backward stepwise selection.
- *Machine Learning Models:* We used a number of different approaches for classifying data in order to assess the confusion matrix and track various performance indicators.[13]

Traditional machine Learning Models performance for predicting crime used are

- AutoMLP
- Decision Tree
- SVM
- Neutral net
- Logistic regression
- Random Forest[13]

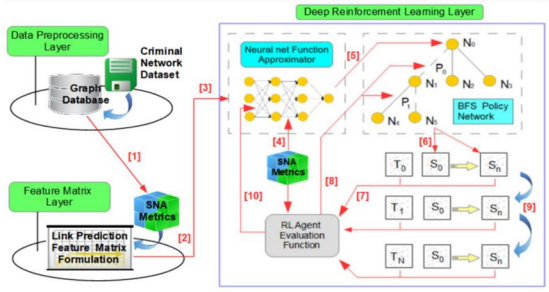
Deep Machine Learning models used are:

- CNN
- Simple RNN
- LSTM
- GRU[13]

K. Deep Reinforcement Learning Techniques

The experimental findings show that the DRL model's predicted accuracy is much higher compared to other Ensemble methods that are developed solely on the dataset at a certain point in time.[14]
The link prediction model employed in this study takes into account the impact of wavelet transform on the development of criminal networks in order to achieve the following goals:
First, a link prediction method is formed and tested using data on the social network dataset's temporal properties.

©IJRASET: All Rights are Reserved



Uses the breadth-first search (BFS) algorithm, which is a a very general-purpose algorithm. To calculate the score that is made by each instance of the virtual system during the process of link prediction, the SNA function approximation metrics are built into something similar to the feature matrix as feed to the RL policy network. The SNA metrics are used to provide the powerful training defined as a vector of likelihood distributed over prediction of link in a time-evolving criminal networking system with DRL sets consisting of node pairs and edges by the neural network amount of additional (value network). The highest-ranking node pair will serve as the starting point for the DRL model's search.

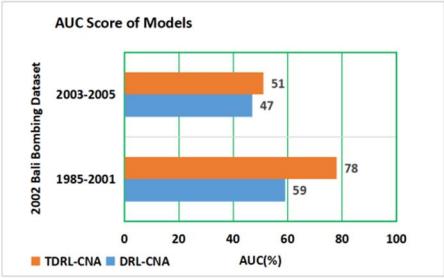


FIGURE 8. AUC scores of DRL-CNA and TDRL-CNA link prediction model.

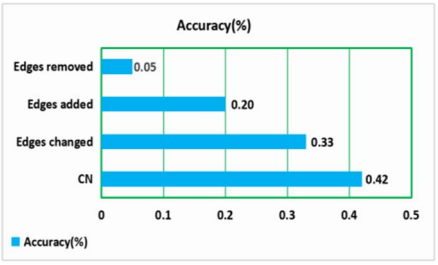
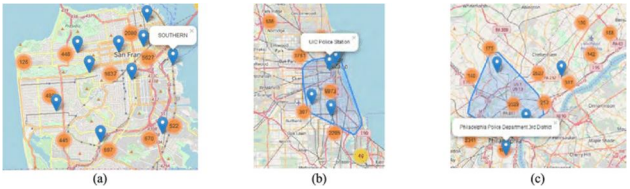


FIGURE 9. Analysis of features contributing to the TDRL-CNA model predictive accuracy.

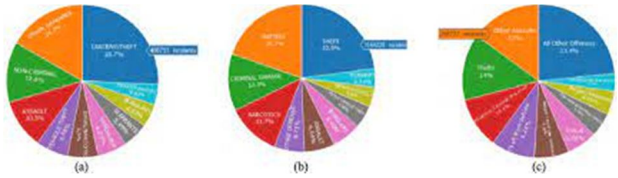
To pinpoint the linkages that may develop and also vanish from the network over a period of time, the suggested TDRL-CNA function approximation model was created. There doesn't seem to be much proof in the field of CNA that the proposed moment link prediction model has now been built using DRL.[14]

L. Use Of Big Data And Mining

Big Data Analytics is used to analyse different relations among a huge amount of data. Here we use BDA to plot criminal data-based plots. The Prophet model and Keras-based techniques have been shown to be more effective than neural network models through a variety of findings.

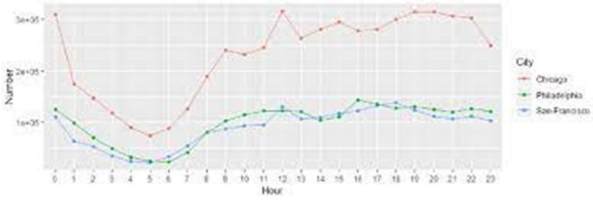


Visualisation of criminal activity for the cities of (a) San Francisco, (b) Chicago, and (c) Philadelphia on open street maps



op-10 crime cases for cities of (a) San-Francisco, (b) Chicago, and (c) Philadelphia.

Another important method for crime data analysis is crime data mining, which explains the extraction of useful and meaningful data from huge amounts of data .Visualisation plots help in studying the data with the help of various plots, which gives an amazing statistical analysis.[16]Trends forecasting deals with the future prediction by taking into consideration past and present data.



The hourly trend of crime in each city

The prediction models used are

- Prophet Model
- Neural Network Model
- LSTM Model[16]

The results of the experiment are given with the help of the Decomposition of Time Series.A time series may be divided into smaller pieces, each of which reflects a different kind of pattern.

M. Situation Aware Deep Reinforcement Learning

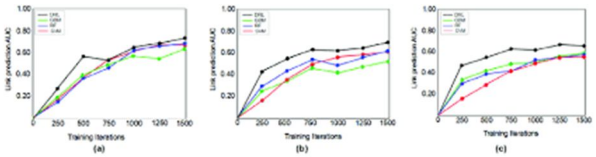
With the increase in technology for catching crimes, the increase in technology for committing crimes is also increasing. This paper takes the help of ML models such as support vector machines(SVM). It has also been proved that FDRL-CNA model's accuracy is higher than the TDRL-CNA model.[15]



Proposed FDRL-CNA link prediction model with SNA and metadata metrics

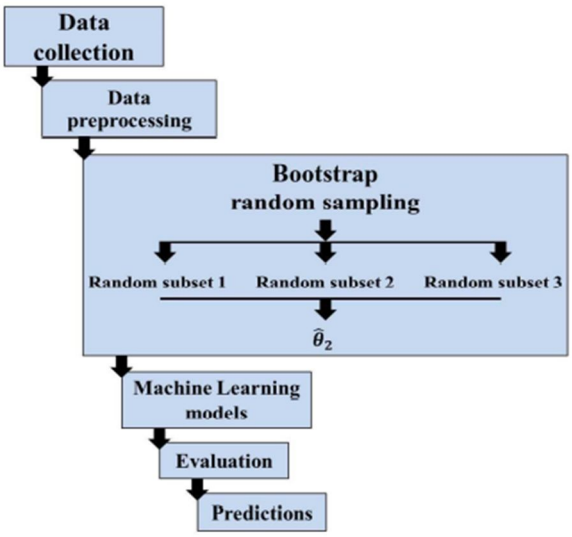
The list of most popular deep learning algorithms are:[15]

- Convolutional Neural Networks
- Long Short Term Memory Networks
- Recurrent Neural Networks
- Generative Adversarial Networks



FDRL, GBM, RF, SVM link prediction models(AUC metrics)

Globally, both developed and poor countries are harmed by criminal activities. Criminal conduct may have a detrimental effect on the economy as well.
Time series analysis is necessary to make visual patterns in addition to a deep LSTM is an investigative algorithm that delivers the improved criminal categorisation over time based on sufficient measures.
They used two techniques—convolutional neural network (CNN),recurrent neural network (RNN). An analytical tool called RTM (risk terrain modelling) was employed.
The following two components make up this study:
Crime forecasting and prediction are both categories.
As a result, the method of bootstrap random sampling approach is preferred over feature selection because it produces estimates of population parameters that are less biassed, especially when the dataset is large.



Bootstrap random sampling method.

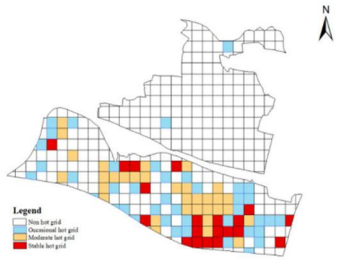
The RNN design has a beautiful variant called LSTM that is a technique that may be used to represent data that is sequential. LSTM offers greater learning for time series by precisely capturing the structure of data that is sequential.

N. Predicting Crime Hotspots

Many academics have created crime prediction models to aid in crime prevention. The majority of them just calibrated their prediction models using past crime data. A current focus of crime prediction research is on two key elements : the forecast of crime hotspots and crime risk areas. Estimating kernel density is a frequently used technique.

1) *POI:* Information on the location and characteristics of various urban facilities is included. Criminals collect many diverse groups of individuals in locations with easy access to transportation and high traffic, whereas criminals are drawn to entertainment venues.[18]
2) *Road Network Density:* Means to divide the length of all roads by the size of a geographic unit. Greater flows of people, including criminals and potential victims, are drawn to areas with denser road networks.

When compared to the standard KDE approach, the temporal autocorrelation-based KDE method frequently performs better for stable hotspot identification.[18]



Hotspot grid pattern of the study area XT.



Distribution of POIs of the study area XT.

The ideal number of clusters using the K-means clustering approach is found to be 4, therefore the grid is split into four groups: stable high-risk hot grids, high-risk hot grids, occasional hot grids, and non-hot grid for the study of area XT in this paper.

O. An Integrated Model of LSTM and ST-GCN

Urbanisation has increased a lot in recent years.Urbanisation creates issues for urban management and related security. Here we suggest a model of everyday crime prediction that combines LSTM and ST-GCN to effectively identify the more risky areas in a city so that crime issues can be resolved.

- *Spatial-Temporal Extraction Module*

Community is built like a graph node with the following characteristics in this module:
To deal with the graph data, we used ST-GCN. ST-ResNet and Graph Convolutional Networks (GCN) were employed to deal with this graph,employed to provide the spatial-temporal properties from each graph collection[19]

- *Temporal Feature Extraction Module*

When theLSTM [19] is computing the number of crimes, crime cases are accumulated as the timing data. This model makes use of an enhanced multilayer perceptron network based on recurrent neural networks (RNN). By assessing what happened to the new input , LSTM can learn the related information between the serial data.[19]

P. Proposed Model Framework

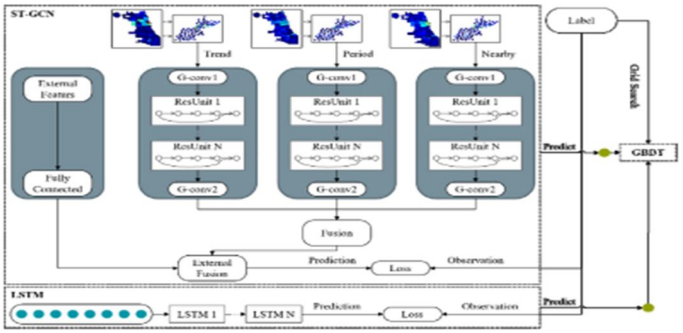


FIGURE 3. The structure of the integrated model, where blue points and green points refer to inputted temporal data and the predicted data, respectively.

III. CONCLUSION

We can see just how easy it is to get started with a revolution to safeguard our society from the evil eye, through Suraksha Kavach that can change the whole crime scene in the world. It is remarkable to see the success of machine learning in such varied real-world problems. In this paper, we have shown different models of both machine learning and deep learning which can help in the process of crime prediction. The aim of society must not be just to catch criminals but to prevent crimes from occurring in the first place. It cannot be said with precision which method is better than the other or which is the best method.It wholly depends on the number of datasets taken into consideration and the type of statistical analysis that has to be done.

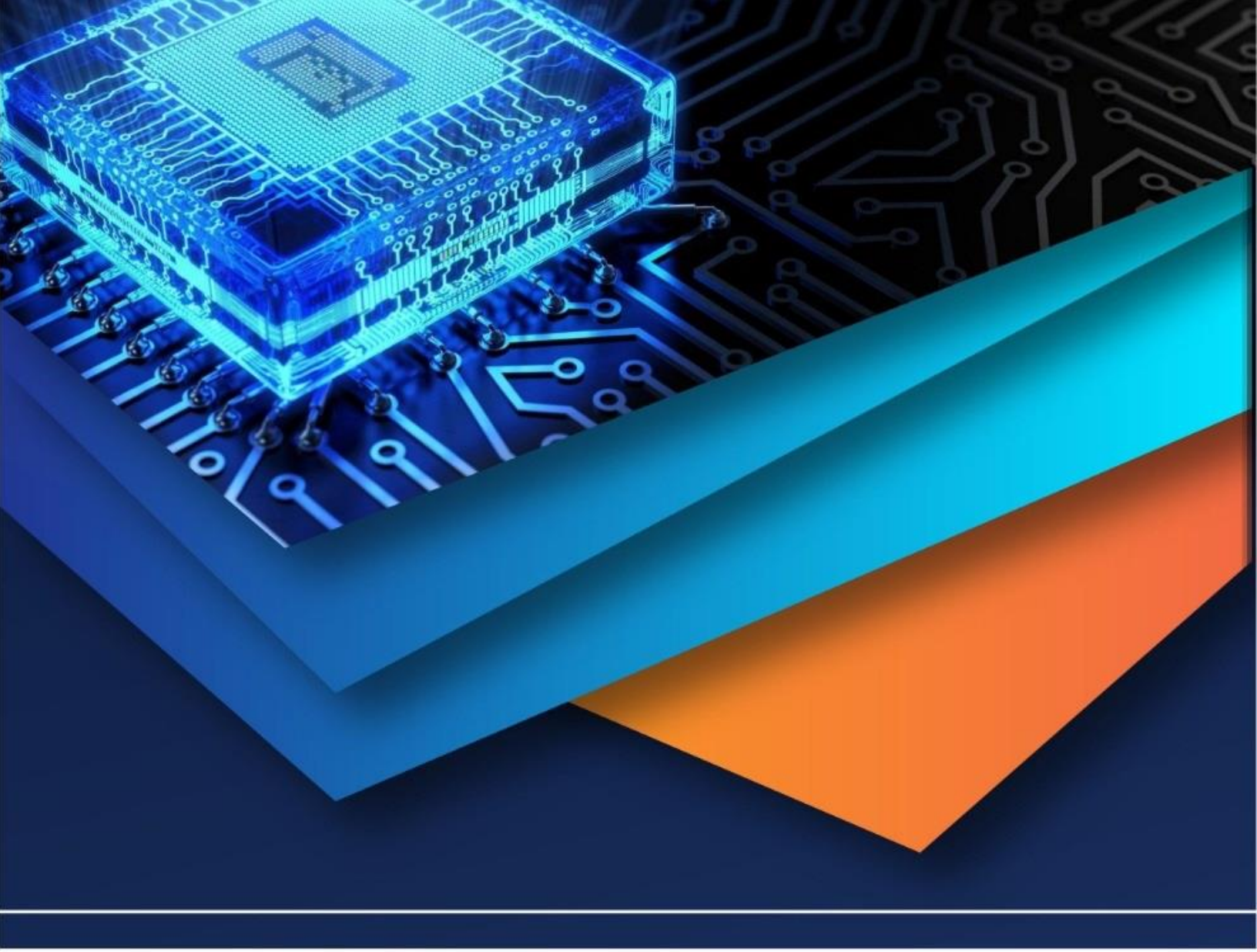
IV. FUTURE WORK

While coming up with the idea of making of this project, we came up with many ideas about how to prevent and analyse crime, but we stuck to the making mobile app and sending of alerts to emergency contacts , as it was the most convenient for people to use .Another modification that can be done is that if the user alerts that he is unsafe then his details, location can be sent to the nearest police station so that police can take immediate response and prevent the crime from happening.

REFERENCES

[1] “Crime Prediction Using Decision Tree (J48) Classification Algorithm” - International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 06 – Issue 03, May 2017
[2] “Exploratory data analysis and crime prediction for smart cities” - Conference Paper, June 2019· The 23rd International Database Applications & Engineering Symposium
[3] “A Clustering Based Hotspot Identification Approach For Crime Prediction” - International Conference on Computational Intelligence and Data Science (ICCIDS 2019)
[4] “Crime Prediction Using Twitter Sentiment and Weather” - 2015 IEEE Systems and Information Engineering Design Symposium
[5] “Crime Analysis and prediction using Data Mining Techniques” - ICTACT JOURNAL on soft computing, April 2017
[6] “Crime Prediction & Monitoring Framework Based on Spatial Analysis” - International Conference on Computational Intelligence and Data Science (ICCIDS 2018)
[7] “Crime Analysis and Prediction Using Data Mining” - First International Conference on Networks & Soft Computing (ICNSC), 2014
[8] “Crime Analysis Through Machine Learning” - 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)
[9] “An Exploration of Crime Prediction using Data Mining on OPEN DATA” - 7 May 2015
[10] “Crime analysis and prediction using Data Mining” - International Journal of Creative Research Thoughts(IJCRT) February 2021
[11] “Crime analysis and prediction using Machine Learning” - 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)
[12] “An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalisation: An Ensemble Approach” - IEEE April 22, 2021
[13] “Developing Machine Learning-based Predictive Models for Smart Policing” - 2019 IEEE International Conference on Smart Computing (SMARTCOMP)
[14] “Link Prediction in Time-Evolving Criminal Network With Deep Reinforcement Learning Technique” - IEEE December 10, 2019
[15] “Situation-Aware Deep Reinforcement Learning Link Prediction Model for Evolving Criminal Networks” - IEEE December 6, 2019
[16] “Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data” - IEEE July 22, 2019
[17] “Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques”- IEEE May 6, 2021
[18] “Comparison For Machine Learning Algorithms For Predicting Crime Hotspots”- IEEE October 2, 2020
[19] “Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN”

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)