Improve Public Safety Through GIS: Analysis and Prediction of Criminal Activities

Khanjan J Shah
SCOPE
Vellore Institute Of Technology
Chennai,India
khanjanjayraj.shah2020@vitstudent.ac.in

Dr.Amrit Pal
SCOPE
Vellore Institute Of Technology
Chennai,India
amrit.pal@vit.ac.in

Abstract: **In today's world public safety is more important than anything else people needs to take care during outdoor activities. The increasing number of crimes makes public unsafe and suffer from mental disorders. Due to the illegal acts people muse over what to do to become safe during outdoor activities. As, past year reports says, people depends on police to take care of them, but now criminal activities are done in such a way that police didn't got a smell of any illegal act. So the paper here proposes some analysis and prediction methods that will be helpful to detect crime activities and criminals, and also help police to track criminals through GIS. This paper propose some new methods and techniques that are found accurate to make GIS helpful in improving public safety.**

Keywords – Crime analysis techniques, Crime prediction techniques, Criminal tracking, GIS, Public Safety through GIS.

I. INTRODUCTION

Public safety through GIS is a purpose to take care of public through latest GIS technologies, So that no one can feel unsafe during their outdoor hours.
To improve public safety, GIS techniques are mainly divided into three firms.
Analysis, comparison and prediction.

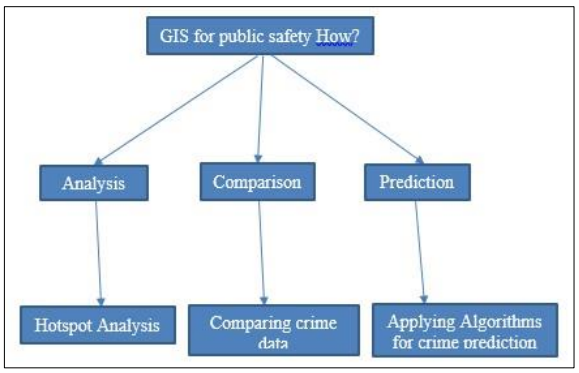

Fig:1 .Classification of GIS for public safety

Analysis: In analysis part, GIS focus on methods to reduce crime rates and focus on area where crimes are increasing day by day due to illetracy, low income and population growth [1]. Secondly, hotspot analysis are implemented into GIS, so that crime areas
are covered easily, for this crime pattern analysis must come into picture [2].

Comparison: In this part, GIS compare data received from police records with the feedbacks received from the public. The feedbacks are based on which type of crime and at what time crime is happened on that place.

Prediction: In this part some machine learning algorithms are used to predict the crime on certain places which are hotspots. Algorithms like: SVM, Gradient Boost Decision Tree, Regressions are used to predict and compare the crime data.

1.1 Motivation and Challenges

The idea of this paper came to the picture by seeing lots of problems and challenges faced by GIS. As we know public safety is very much important and due to increasing number and usage of technologies, day by day people will be more dependent on mobile phones. GIS came to the picture due to this tremendous use of technologies and for public safety it is found to be more important for all. In GIS there are lots of problems faced by people. Like:

1. Not getting accurate prediction of areas where crimes are happened.
2. No alarm system during any crime happens in a particular area.
3. No on-the-spot message sent to the police using GIS.
4. No previous records of crimes are publicly displayed on GIS.

To overcome such problems and challenges, a better and perfect survey is needed that resolve all these issues happening in GIS.
Also some tools, techniques and architectures are needed to complete this survey.

1.2 Tools, Techniques and Architectures.

Tools used for implementation of this survey:
R-studio – Used for analysis of crime and crime activities.

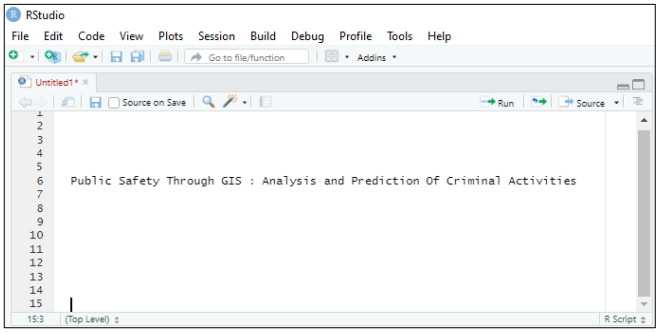

Fig: 2 R-studio Software

Google Collab – We are Using collab for machine learning and deep learning algorithm implementation in purpose of prediction of crime activities.
MS-word – Used for documentation and drawing architectures and figures.

Techniques used in this survey:
Statistics Central Tendency measures are used like : Median Which is used to locate latitudes and longitudes points on map. All points in dataset is not possible to locate on map, so we took median of all points and locate them on map.
The points shows the crime locations.
Other Techniques :

- Comparison of recorded crime with the public feedback about crime activities.
- Comparison is done using multi-variable regression technique.
- Applying Random Forest and Neural Networks algorithms for prediction of crimes in particular city and which type of crime happens in particular city.

Architecture:

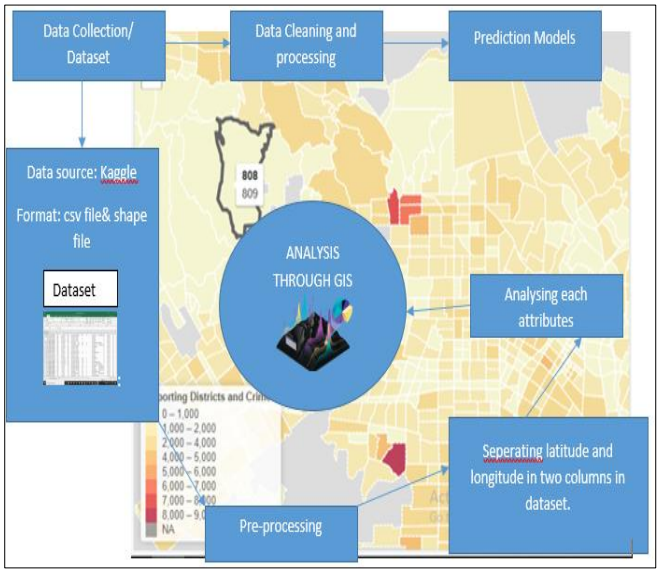

Fig: 3. Architecture of GIS Public safety

2. LITERATURE-SURVEY

Reducing Crime Rates [1]:
To reduce crime rates electre tri model is used. That help us to detect the area where crime is low or where crime rates are high. So where crime rates are more in that area government can take special survey and impose special operations to reduce crime rates and make area free of crime.

Survey on Hotspot Analysis [2]:
The paper here gives different methods to analyze hotspot areas like: Gird thematic mapping, Kernel Density, Local Moran's, Getis-Ords*. First of all raw data is collected and mapped on. After that hotspot analysis techniques should apply to analyze perfect hotspot area of crime. The Kernel density and grid thematic mapping technique is not much as accurate as Getis-Ords method to hotspot crime area.

Crowd Monitoring through GIS [9]: In this paper shows image and video processing to capture the images of people in the crowd and using this images buzzer setup is there to alarm people if there is any smell of crime comes under the camera.
The images captured here are sent to the police station to predict the crime, means that if anything wrong already happened at this place than with the help of crowd picture police can easily make safety of people by alarming them.

Track Nearest Police Station [8]: This paper propose to track nearest police station from crime scene. So that criminal can easily detected and captured. This is only be possible if we use GIS technology to send the shortest routes and notifications if any crime happens in surrounded area.

Resident's crime comparison [4]:
This paper survey on comparison of recorded crime from police files with the feedbacks taken from the public. Feedback about what type of crime and happens at what time is very necessary to predict and make the area as a crime hotspot area. The comparison is done with two factors High and Low.
Factors from both side are compared equally and regression graph is developed to put this comparison more accurate and predictable. We are giving colors to various comparison factors. The factor showing high density in crime shows red in color and others are shown in similar way below:

Some algorithms and techniques this paper shown to analyze and predict the crime [10]: Data Source From where the data we collect for this whole project. Pre-processing – The data will be more clear by pre- processing it or by filling missing values in data.
Regression – It is the most common method used in prediction of crime with the help of recorded data. Likewise we can also use, Gradient Boost Decision Tree, Random Forest, Reinforcement Learning where data are available and we have to make prediction from that data.

3. MY CONTRIBUTION

The survey of this paper begins from the collection of datasets that I got from the internet. Dataset consists of many such useful information that are needed in implementation of this proposed model. Like: Types of crime, Date (When crime happens), Day (when crime happens), time of crime, images of area where crime happens, images of crime, information of victim, information of suspect etc.

Step 1:
After getting dataset, we convert our raw data into required data, some pre-processing steps must required in cleaning the data.

Algorithm For Data Pre-Processing

Input:

1. GIS DATA
2. Removing Null Values
3. Convert Text Data to numeric (after visualization)
4. Normalize and One-Hot encoding for neural networks i/p.

```
Binary_encode = oneHotEncoder (sparse = False, categories = 'auto')
Encoded_x = Binary_encod.fit_transform(x)
```

5. Separate latitude and longitude in two columns.

```
LACrime_latlong = gsub('\\)','',LACrime_latlong)

LACrime_latlong = str_split(LACrime_latlong,",")
LACrime_latlong = do.call(rbind.data.frame, LACrime_latlong)
colnames(LACrime_latlong) = c("latitude","longitude")
```

Step 2:

Now dataset is imported and plotted on map. On applying central tendency measures and ggplot library, we will easily analysis crimes, types of crimes, at which location crime happens, from which instrument crime happened, what age people did crime. All these analysis are done using R-studio tool.

Note: For analysis we are not using tableau because the dataset is too big and have lots of varied attributes. So for purpose of accurate result, we are doing analysis with R-studio tool only, some part of our code we analysed in python also. So the results and formulas for analysis we are describing step by step.

Type of crime and reporting districts:

```
# Type of Crime and Reporting District

LACrime %>%
  group_by(CrimeCodeDescription,ReportingDistrict) %>%
  tally() %>%
  arrange(desc(n)) %>%
  head(20) %>%
  mutate(ReportingDistrict = reorder(ReportingDistrict,n))  %>%

  ggplot(aes(x = ReportingDistrict,y = n,fill = CrimeCodeDescription)) +
  geom_bar(stat='identity',colour="white") +
  labs(x = 'ReportingDistrict', y = 'Count of Incidents',
       title = 'ReportingDistrict and CrimeCodeDescription') +
  coord_flip() +
  theme_bw() +  theme(legend.position="top")
```

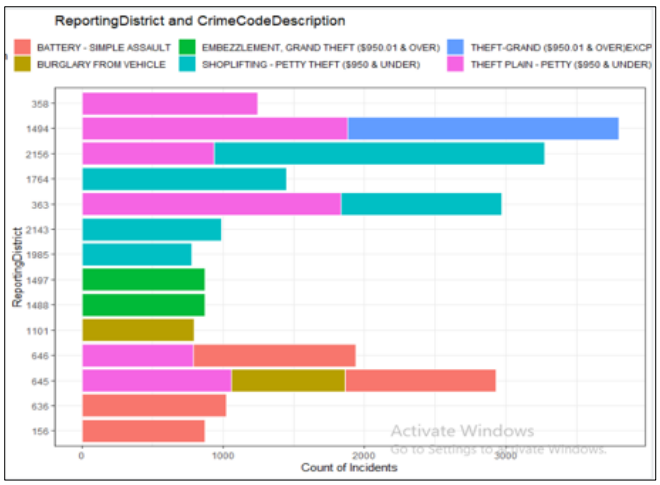

Fig:4 Result of Type Of crime and their registered districts.

Crime in particular District with total counts:

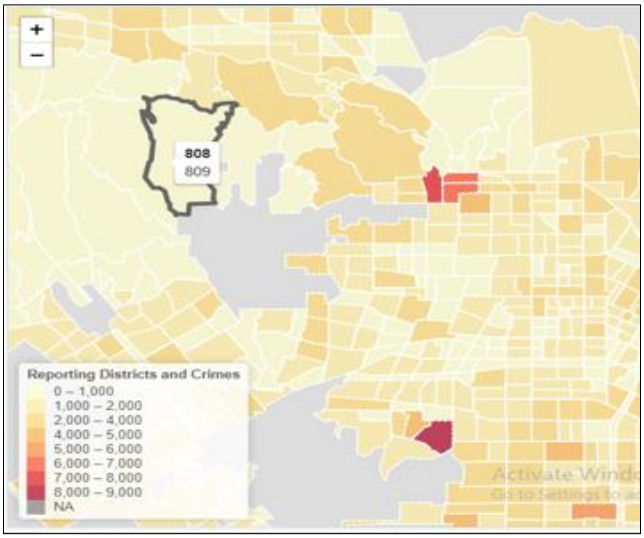

Fig : 5 Crime in Districts with total counts

Proportion of crime according to Premise type:

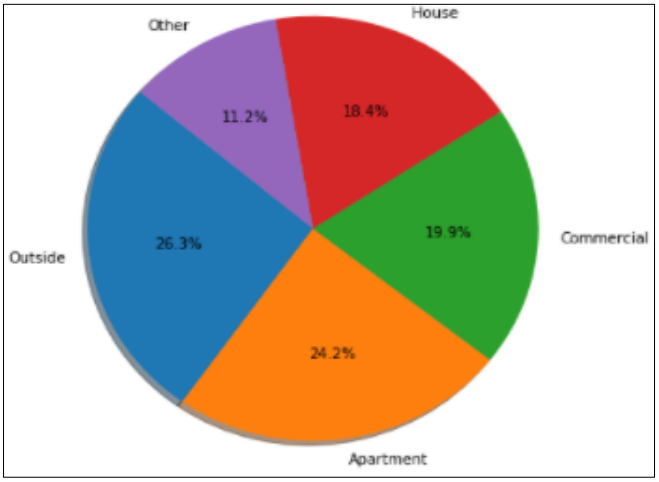

Fig : 6 crime according to Premise type

Analysis of crime location in particular city:

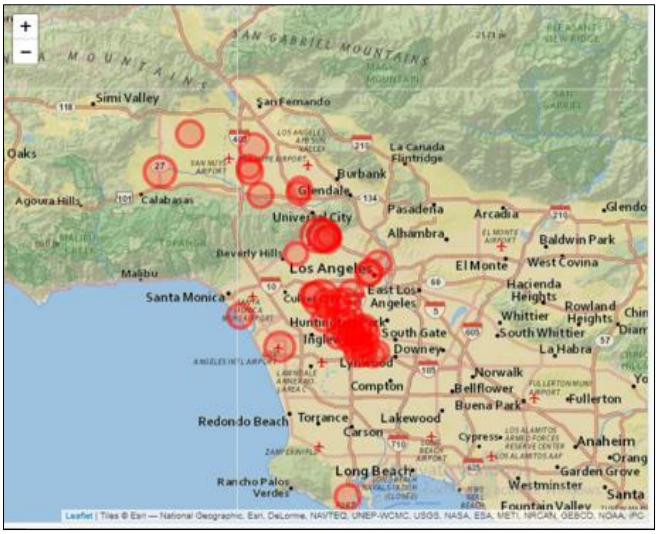

Fig: 7 Analysis of crime location in particular city

Step 3:
Now, after all analysis done on dataset, we move to the prediction of crimes in particular area and also which type of crimes predicted in that area.

Methods for calculating Accuracy, Precision and Recall:

$$\text{Base Rate} = \#incedents \left(\frac{\#incedents}{Relevant\ Population\ Size}\right) \times Base$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Random Forest Implementation and Results:

```
#One Hot Encoded Model
Classifier = RandomForestClassifier( n_estimators = 100,
criterion = 'entropy', random_state = 42)

Classifier.fit( x_train_OH, y_train_OH )
y_pred_OH = Classifier.predict( x_test_OH )

Print ( accuracy_score( y_test, y_pred )
Print ( confusion_matrix( y_test_OH, y_pred_OH )
```

Results :

```
Accuracy of Random Forest with OneHotEncoder :  0.6352691356100602
[[24536  1948   499    28   730]
 [ 5223  5157    33    54   233]
 [ 3107   253  1673     6   321]
 [ 1168   319    25    24   143]
 [ 3012   621   176    17  2025]]
                precision    recall  f1-score   support

       Assault       0.66      0.88      0.76     27741
Break and Enter       0.62      0.48      0.54     10700
       Robbery       0.70      0.31      0.43      5360
    Theft Over       0.19      0.01      0.03      1679
    Auto Theft       0.59      0.35      0.44      5851

      accuracy                           0.65     51331
     macro avg       0.55      0.41      0.44     51331
  weighted avg       0.63      0.65      0.62     51331
```

Fig: 8 Random Forest results:

The Results above shown the accuracy of RandomForest algorithm for crime prediction. It describes that the prediction accuracy of crime in districts is 63% and also which type of crime happens are also predicted. So the crimes in the results suggest that the possibility of Assaut, Robbery, Theft Over, Auto Theft and Break and Enter types of crimes happens in districts, so we from this public came to know that they have to take precautions against certain crime happening.

Step 4:
After accuracy of Random Forest we got, we are applying neural networks algorithm on our dataset that gives us most accurate result.

Neural Networks Methodology:

Libraries Used – numpy, pandas, seaborn, matplotlib, sklearn.

HeatMap Visualization of major crime indicators by month.

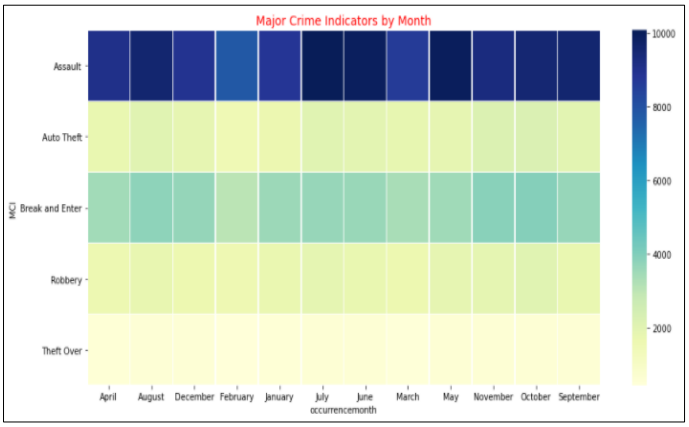

Fig: 9 Heat map Visualization (month – crimes):

Data Preprocessing for Neural networks:
Convert categorical data to numeric data.

```
Crime_var = pd.factorize(df2['MCI'])
Df2['MCI'] = crime_var[0]
Definition_list_MCI = crime_var[1]
```

Train_test_split:
Test_size = 0.30, random_state=15

Neural Network Model:

```
#Neural Network … Create Model config.
nn_model = MLPClassifier(solver = "adam", alpha = 1e – 5,
hidden_layer_sizes = (500,), random_state = 1, max_iter
=1000)

#Model training
nn_model.fit(X=x1,y = x2)

#prediction
result = nn_model.predict(y[Features])
```

```
# Model Evaluation
ac_sc = accuracy_score(y2, result)
rc_sc = recall_score(y2, result, average = "weighted")
pr_sc = precision_score(y2, result, average = "weighted")
f1_sc = f1_score( y2, result, average = "micro")
confusion_m = confusion_matrix(y2, result)
```

```
========== Neural Network Results ==========
Accuracy     :  0.5571479014720744
Recall       :  0.5571479014720744
Precision    :  0.461578424796213
F1 Score     :  0.5571479014720744
Confusion Matrix:
[[41819  1493     0     0   832]
 [14335  2851     0     0    79]
 [ 8133    66     0     0   533]
 [ 2413   180     0     0    70]
 [ 7420   817     0     0  1088]]
```

Fig: 10 Neural Networks Results

Classification_Report:

```
================ Classification Report ================

                precision    recall  f1-score   support

       Assault       0.56      0.95      0.71     44144
Break and Enter      0.53      0.17      0.25     17265
       Robbery       0.00      0.00      0.00      8732
    Theft Over       0.00      0.00      0.00      2663
    Auto Theft       0.42      0.12      0.18      9325

      accuracy                           0.56     82129
     macro avg       0.30      0.25      0.23     82129
  weighted avg       0.46      0.56      0.45     82129
```

Fig: 11 Classification Report Table

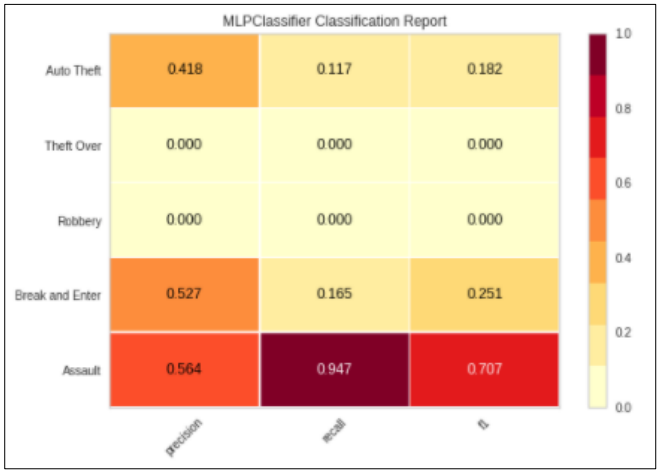

Fig: 12 Classification Report Visualization

Also our next goal for this research is to built alarm system and get crime data publicly on GIS. Up till now we are done with analysis of types of crimes and happening of crimes at particular locations. And prediction of crime rate and type of crimes in particular district.

4 CONCLUSION

From this Whole research, finally I come to my conclusion that GIS model I developed here is used for public safety, there are many pros and cons of model. Like information of certain people using GIS we get is a cons as there is no security threat yet to be developed to kept information highly secured is a pros of this model. So, I have made analysis of crimes using R-studio tool on crime dataset and also did prediction using Random Forest on dataset that give me accuracy that how much your data will be accurate to predict crime in particular area. Now I decide to increase the accuracy of my model using neural network algorithms.

My **Future Work** includes some more research on this topic and building AI system for automatic alarm for public safety.

5. REFERENCES

[1]. André Gurgel / Caroline Mota & Íris Pimenta "Public Safety Planning in Natal city: an application based on ELECTRE TRI model " 2014 IEEE International Conference on Systems, Man, and Cybernetics October 5-8, 2014, San Diego, CA, USA.

[2]. MA Wei, CHEN Jianguo, CHEN Peng "Illegal Activities Hotspot Analysis Based on GIS Methods " date of publication May, 2018, date of current version September, 2019.

[3] JIANMING ZHOU, ZHENG LI, JACK J.MA , AND FEIFENG JIANG "Exploration of the Hidden Influential Factors on Crime Activities" Received May 13, 2020, accepted June 17, 2020, date of publication July 17, 2020, date of current version August 12, 2020.

[4]. Jiřˇí Pánek, Igor Ivan and Lucie Macková, "Comparing Residents' Fear of Crime with Recorded Crime Data—Case Study of Ostrava, Czech Republic" ISPRS Int. J. Geo-Inf. 2019, 8, 401; doi:10.3390/ijgi8090401.

[5]. Guiyun Zhou1, Jiayuan Lin, Wenfeng Zheng., - "A Web-based Geographical Information System for Crime Mapping and Decision Support." IEEE Transactions on Consumer Electronics, Vol. 57, No. 4, November 2011.

[6]. Srikanth Vadlamani, Mahdi Hashemi. "Studying the impact of streetlights on street crime rate using geo-statistics.2020".IEEE 21st

[7]. M. Saravanan, Rakhi Thayyil and Shwetha Narayanan. Enabling Real Time Crime Intelligence Using Mobile GIS and Prediction Methods. 2013 European Intelligence and Security Informatics Conference 978-07695-5062-6/13 $26.00 © 2013 IEEE DOI 10.1109/EISIC.2013.27.

[8] Eman Mohamed Abd, El-Aziz, Saleh Meshah and Khaled Mahar " GIS-Based Description Support System for Criminal Tracking" ICCTA 2012, 13-15 october 2012, Alexandria, Egypt. 978-1-4673-28241/12/$31.00. 2012 IEEE.

[9] Hongquan Song, Xuejun Liu, Xingguo Zhang, Jiapei Hu. " Real-time Monitoring for Crowd Counting using Video Surveillance and GIS." 978-1-4673-0875-5/12/$31.00, 2012 IEEE.

[10] Meiliana, Dedi Trisnawarman, Muhammad Choirul Imam. " prediction Analysis Of Crime Data Using Machine Learning." IOP Conf. Series: Materials Science and Engineering (2020) 012164 doi: 10.1088/1757-899X/852/1/012164.