# ENHANCE ALGORITHAM TO PREDICT A CRIME USING DATA MINING

**[1]HARDI. M. PATEL, [2]RIPAL PATEL**
[1]P.G.Scholar, [2]Asst. Professor,
[1]Computer Engineering Department,
[1]Silver oak college of engineering and technology, Ahmedabad, India

*Abstract— Data mining is used extensively in terms of analysis and investigation of patterns for occurrence of different crime. Data mining can be used to model crime detection problems. Crimes are a social nuisance and cost our society dearly in several ways. Our aim is to predict the crime and location in which specific types of crime will occur. Our approach is based on sentimental analysis by applying lexicon-based method. We have to use here twitter data set. We are predicting future crime in particular area (illinoise, Chicago).*
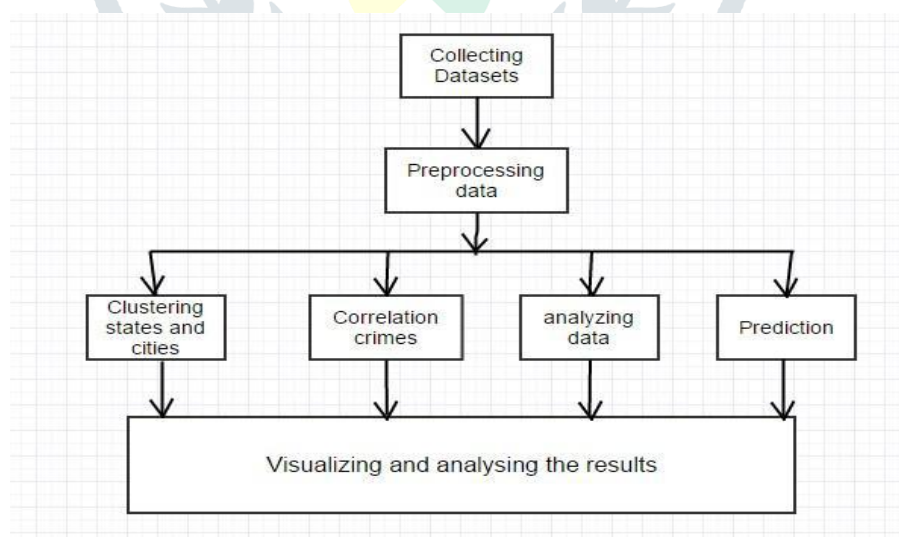
*Index Terms— Data mining, crime prediction, twitter sentiment analysis, clustering*

## I. INTRODUCTION

Crime has been a part of society ever since laws were first approved. It is defined as an act committed or omitted in violation of a law forbidding or commanding it and for which punishment is imposed upon conviction. In the world today crime analysis is gaining significance and one of the most popular disciplines is crime prediction. Crime is classically unpredictable. It is not necessarily random; neither does it take place persistently in space or time. A Good theoretical understanding is needed to provide practical crime prevention solutions that equivalent to specific places and times. Crime analysis takes past crime data to predict future crime locations and time. Crime prediction for future crime is process that finds out crime rate change from one year to the next and projects those changes into the future. Crime predictions can be made through both qualitative and quantitative methods. Qualitative approaches to forecasting crime, as environmental scanning, scenario writing, are useful in identifying the future nature of criminal activity. There are multiple factors that could affect future crime incident other than crime hot spot. As per twitter data we considered environmental factors affect the criminal incidents. We collected twitter data from official twitter streaming API. Along with the twitter data we also obtain weather data from weather underground website to build a crime prediction model.

An ideal crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. In the present scenario the major challenges are encountered. Following figure shows the crime pattern analysis.

- Increase in the size of crime information that has to be stored and analysed.
- Problem of identify techniques that can accurately analyse this grooving value of crime data different methods and structures used for recording crime data.
- The data available is inconsistent and are complete thus making the task of formal analysis a far more difficult.
- Investigation of crime takes longer duration to complexity of issues.



1. **Advantages of Crime Prediction :**
   The advantages of crime prediction techniques are as follow:
   - Visualize criminal network.
   - Risk is reduced.
   - Increase crime analysts work productivity.
2. **Disadvantages of Crime Prediction :**
   The disadvantages of crime prediction techniques are as follow.
   - Unexpected occurrence.
   - Takes lot of computational time.

3. **Challenges of Crime Prediction :**
   The challenges of crime prediction techniques are as follow:
   - Provide an accurate prediction for the location of the criminal.
   - Collecting and managing large volumes of accurate data.
   - Provide good performance by combining prior knowledge.
   - Fast evolution of crime site data.
   - Maintain effective crime analysis resource.

## II. REVIEW WORKS

Xinyu chen, youngwoon, and suk young jang in[1] have analyzed textual content in twitter data by using sentiment analysis to score the positivity / negativity of tweets and their trends in different neighborhoods. They have used lexicon based method and understanding of categorized weather data combined with kernel density estimation.

The studies are more focused on temperature, humidity and other weather factors contribute to crime incidents. Here sentiment analysis also part of lexicon based method.

S. Yamuna, N.sudha bhauvaneswari in [2] used k-means and DB-Scan algorithm are considered for further analysis. Here also classification algorithm applied to predict the future crime pattern. In this paper researcher used clustering technique for accuracy, efficiency in indentifying crime trend, identifying crime zones, crime density of a state and efficiency of a state in controlling crime rate.

Rasoul kiani, siamak mahdavi, amin keshavarzi in [3] used clustering by K-means algo and GA algo is used for optimizing of outlier detection operator parameters using rapidminer tool. In this paper researcher used datamining to detect much more complex pattern since in real life there are many attributes factors for crime and often future crime pattern can be detected clustering techniques work better for future crime. In this paper researcher used clustering technique only and also classification. Researcher main purpose of crime analysis are extraction of crime patterns by crime analysis and based on available criminal information, focus on existing data and prediction of crime frequency using data mining technique. Here researcher also used classification it means the process of dividing the data to some groups that can act either dependently or independently.

A Malathi & Dr. S. Santhosh Baboo in [4] defined clustering methods and used data cleaning algorithm. Crime detection and preventation techniques are applied to different applications ranging from cross border security, internet security to household crimes. They analyzed the result to produce profiles, which can be used to perceive the behavior of criminal. In this paper the existing algorithm predicted the crime as 83%. The new algorithm predicted the crime as 89%. In this paper different type of crime discussed like murder, robbery, burglary, theft etc. in this paper researcher discussed process of crime analysis for identifying and reducing crime in India. The aim of this research work is consist of developing analytical data mining method that can systematically address the complex problem related to various form of crime.

Yu-yuch huang, cheng-teli, shyh-kang jeng in [5] focus on location based social networks for criminal activity prediction. Also focus in environmental criminology. In this paper researcher discuss crime in all urban areas. Environmental criminology defined root causes of crime and why people become criminals. In this work researcher used large scale database. Location based social network data for a variety of applications, such as urban planning and predicting human mobility. This paper aims at exploiting the location based social network data to investigate the criminal activities in urban areas. Researcher focus on using only location based social network data to investigate the regional criminal behaviors and none of the demographic information like population, race, income and education are used. Here technically researcher used heterogeneous data. This paper aim is to employ the combination of features in order to predict the crime frequency and find which type of criminal activity is easier or more difficult to be accurately predicted.

A.malathi, A. Ganthi mathi in [6] focus on crime zone. In this paper researcher defined high crime zone, medium crime zone and low crime zone. In this paper two algorithms k-means & DB-scan algorithm are enhanced to identify the crime zones. This main focus is the identification of crime. The results of process are used to identify trend city wise for each type of crime. In this paper two algorithms used for best results and then compare two algorithms. As a output results DB-scan algorithm chosen for the clustering crime data in the proposed crime analysis tool. Classification also data mining techniques but all clustering techniques are very useful for this paper research.

Table 1 COMPARISION TABLE

| Sr. no. | Description | Approach | Prons. | Cons. |
|---|---|---|---|---|
| 1 | Crime prediction using twitter sentiment & weather | Sentiment analysis by applying lexicon based method | Correlation between crime & predictors of weather & sentiment | Collected data from selected group of people to predict their future response in certain circumstances. |
| 2 | Data mining technique to analyze & predict crimes | K-means & DB-scan predefine. | Better crime pattern | Data analysis can be difficult |
| 3 | Analysis & prediction of crimes by clustering & classification | Classify Clustered crimes based on occurrence frequency during different years. | Increase accuracy & classification | Classification is the process of dividing date to some groups that can act either dependently or independently. |
| 4 | Evolving data mining algo on the prevailing crime trend an intelligent | clustering method | Predict crime 83 % (nearest result) | Used common crime pattern |

| | | | | |
|---|---|---|---|---|
| | crime prediction model | | | |
| 5 | Minning location based social networks for criminal activity prediction | Location based social network | Correlated with crime prediction | The constraints of used datasets. |
| 6 | Design & development of enhanced algo to identify crime zones using data mining technique | Enhanced DB-scan | Produced near to unity values indicating that clustering quality has significantly improved | The accuracy was lower with small sized datasets. |

## III. CONCLUSION

This paper has presented frequency of crime, accuracy of crime and also preliminary investigation of criminal incident prediction. Also defined different type of clustering algorithm and also this algorithm enhanced for predict future crime. This all papers gives its best ideas for the predicting future crime but there are plenty of way to extend this work. The most direct way to improve the predicting accuracy is to obtain weather forecast data in every 6 hour's time period, and also correlated with positive and negative opinions.

## IV. FUTURE SCOPE

We will improve hypothesis will correlate twitter sentiment data to the specific find of crime. And to find better alternative to replace logistic regression model as its coefficient of sentiment polarity is insignificant. In this way the observation we taken of one month or one weak we will put this data in Chi-square equation to calculate crime density with clustering technique and finally link analysis will be added to check the relativity of data.

## V. REFERENCES

[1] Xinyu chen, Youngwoon Cho, and Suk young Jang, "Crime prediction using twitter sentiment and weather", IEEE systems and information engineering design symposium, 2015.

[2] S.Yamuna, N.Sudha Bhuvneswari, "Datamining techniques to analyze and predict crimes", The International Journal of Engineering And Science (IJES) – volume 1 Issue 2.

[3] Rasoul Kianin, Siamak Mahdavi, Amin Keshavarzi, "Analysis and prediction of crimes by clustering and classification", International Journal of Advanced Research in Artificial Intelligence (IJARAI) – volume 4 No. 8 2015.

[4] A.Malathi, Dr. S. Santhosh Baboo, "Evolving data mining algorithms on the prevailing crime trend – An intelligent crime prediction model", International Journal of Scientific & Engineering Research (IJSER) – volume 2 Issue 6 - June-2011.

[5] Yu-Yueh Huang, Cheng-Te Li, Shyh-kang Jeng, "Mining location-based social networks for criminal Activity prediction", Wireless and optical communication conference(WOCC) – 2015.

[6] Malathi, A. Ganthi mathi, "Design and Development of enhanced algorithms to identify crime zones using data mining techniques", International Conference on Information Technology & society – June 2015.