



## CRIME ANALYSIS AND PREDICTION

K SIREESHA<sup>1</sup>, B. RAMYA<sup>2</sup>, P. SRIJA<sup>3</sup>, A. VAISHNAVI<sup>4</sup>

ASSISTANT PROFESSOR<sup>1</sup>, UG SCHOLAR<sup>2,3&4</sup>

DEPARTMENT OF CSE, BHOJ REDDY ENGINEERING COLLEGE FOR WOMEN, VINAY NAGAR, HYDERABAD-59

### ABSTRACT:

Crime is one of the biggest and dominating problem in our society and its prevention is an important task. Daily there are huge numbers of crimes committed frequently. This requires keeping track of all the crimes and maintaining a database for same which may be used for future reference. The current problem faced are maintaining of proper dataset of crime and analyzing this data to help in predicting and solving crimes in future. The objective of Crime prediction is to formulating of policing strategies and the implementation of crime prevention and control. Therefore, a crime information system that is able to process large amount of data in a short period of time is needed for investigators to know crimes hotspots, crime patterns and to predict future ones. In this project, we will be using the technique of machine learning for crime prediction of Chicago crime data set. The crime data is extracted from the Kaggle. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training of the model data preprocessing will be done following this feature selection and scaling will be done so that accuracy obtain will be high. Random Forest algorithm will be tested for crime prediction and will be used for training. Visualization of dataset will be done in terms of graphical representation of many cases for example at which time the criminal rates are high or at which month the criminal activities are high. The soul purpose of this project is to give a jest idea of how machine learning can be used by the law enforcement agencies to detect, predict and solve crimes at a much faster rate and thus reduces the crime rate. It not restricted to Chicago, this can be used in other states or countries depending upon the availability of the dataset.

**KEY WORDS:** Random Forest, Autoregressive moving average, Crimes, Machine learning, Analysis, Prediction.

### INTRODUCTION:

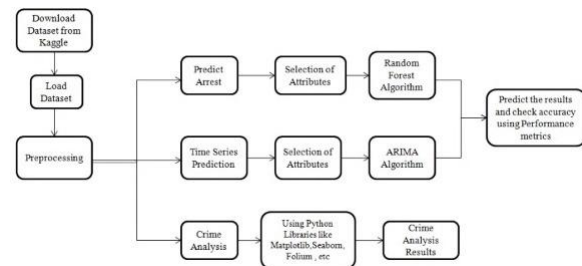
The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster. The rate of crime is rising on a daily basis as current technologies and high-tech ways assist criminals in carrying out their unlawful

activities. Crimes are neither systematic nor random otherwise crime cannot be analysis. When crimes like robbery, firebombing etc. have been decreased, crimes like murder, sex abuse, gang rape etc. have been increased. We cannot analyze the victims of crime but can analyze the place where crime occurred or happened. Data about crime will be gathered from a variety of blogs, news outlet, and websites. The massive data is used to create a crime report database as a record. The police department can benefit from the information gathered. The police

have discovered that identifying criminal “hotspots”, or regions with a high concentration of crime, is an effective method. Predictive modelling can be divided further into two areas: Regression and pattern classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables. The task of pattern classification is to assign discrete class labels to particular data value as output of a prediction. Machine learning algorithms can be used to extract useful information from crime report datasets. Crime analysis is the first phase in the investigation of a crime. Exploring, relating, and analyzing crime is what crime analysis is all about. Elements such as criminal intelligence, location security, and so on the effort followed the steps of data analysis, with data gathering data classification, pattern discovery, prediction, and visualization being the most essential parts. The proposed system employs a variety of visualization approaches to depict crime trends and several methods for predicting crime using a machine learning algorithm. The prediction is to know about arrested or not. Attributes selected for this are primary type, domestic, district, ward, community area. And arrest is chosen as output variable. In Time series forecasting, for each area - date wise crime count (several years) is plotted on graph and predict for future. For Time series prediction , algorithm like ARIMA is used. We perform multiple classification tasks– we first try to predict which of 6 classes of crimes are likely to have occurred, and later try to differentiate between violent and non-violent crimes.

## METHODOLOGY:

### PROJECT ARCHITECTURE:



## ALGORITHMS:

**Random Forest Algorithm:** It is a popular machine learning algorithm that belongs to the supervised learning technique. Random Forest algorithm is a combination of decision trees. Random Forest uses bagging.

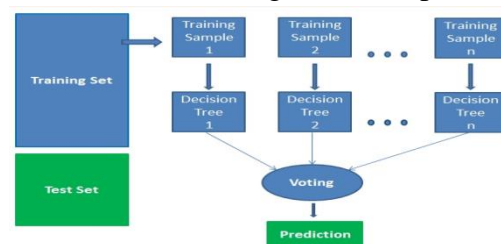
Steps involved in random forest algorithm:

Step 1: In Random forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.



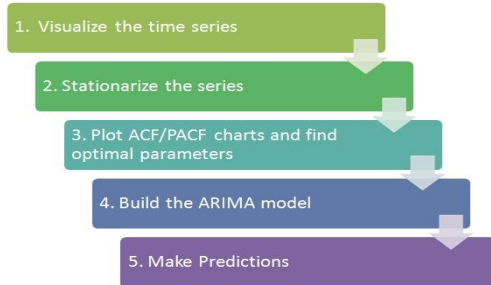
This prediction is to know about arrested or not. Attributes selected for this are primary type, domestic, district, ward, community area. And arrest is chosen as output variable.

## ARIMA:

An **Auto Regressive Integrated Moving Average**, or **ARIMA**, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A statistical model is

autoregressive if it predicts future values based on past values.

- **AR:** an **Auto Regressive** model which represents a type of random process. The output of the model is linearly dependent on its own previous value i.e. some number of lagged data points or the number of past observations.
- **MA:** a **Moving Average** model which output is dependent linearly on the current and various past observations of a stochastic term .
- **I:** integrated here means the **differencing step** to generate stationary time series data, i.e. removing the seasonal and trend components.



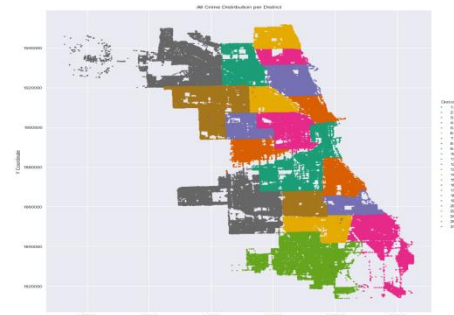
**Input parameters :** Time series data - crime count for each date in the past years

**Output parameter :** predicted crime count for each date in the future years .

Output of this model is line graph representing future and past crime count. The graph is plotted between crime count vs dates for each year.

### CRIME ANALYSIS:

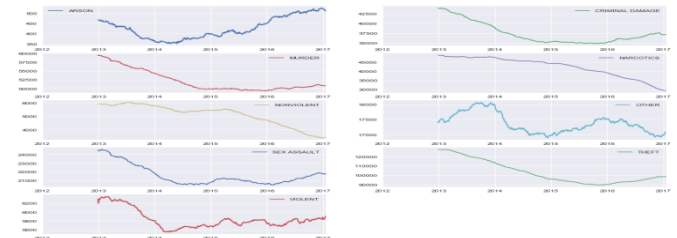
- 1) Scatter plot representing number of cases in each places and each district is represented with unique color.



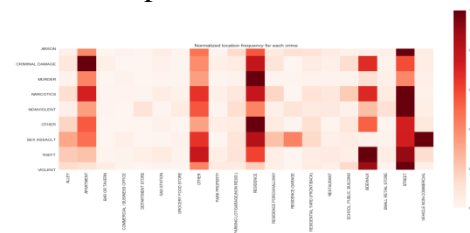
2) Bar graph representing crime type vs count of cases



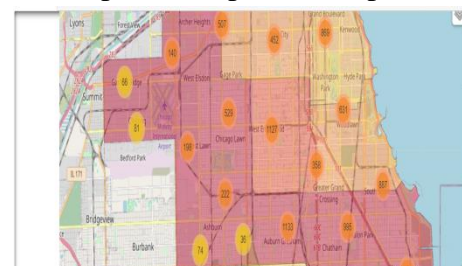
3) Line graph representing crime rate by its types



4) Heat plot between area vs type relationship

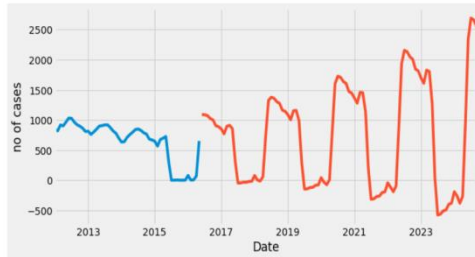


5) Geospatial Graph for hotspots



### CRIME PREDICTION:

### Time Series Forecasting:



### Arrest Prediction:

Primary Type	Domestic	District	Ward	Community Area	Arrest
Battery	False	4	7	Chicago Lawn Chicago	False
Kidnapping	False	7	0	Uptown Chicago	True
Homicide	True	3	3	Brighton Park Chicago	False
Assault	True	4	3	Pullman Chicago	False
Narcotics	False	9	16	New City Chicago	True

### CONCLUSION

1. The data analysis will give us a scientific view about the security status and crime rate of the Chicago city. From these analysis result and visualization , we can view the most occurring crimes and locations where crime happened. We believe if our data analytics can give us all these information about the security status of the Chicago city, a bigger data analytics project will provide much more valuable information which can be used as a powerful source for taking wise actions that increases the security status of our cities.
2. With the help of machine learning technology, it has become easy to find out relation and patterns among various data's. The work in this project mainly revolves around predicting the arrest of crime which may happen if we know the primary type, district, ward, community area, domestic. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data transformation. The model predicts the type of crime with accuracy of 0.86. Data

visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helped in understanding Chicago crimes datasets that can help in capturing the factors and also help in keeping society safe.

### FUTURE SCOPE:

Using a combination of ML and computer vision can substantially impact the overall functionality of law enforcement agencies. security equipment such as surveillance cameras and spotting scopes, a machine can learn the pattern of previous crimes, understand what crime actually is, and predict future crimes accurately without human intervention. A possible automation would be to create a system that can predict and anticipate the zones of crime hotspots in a city. Law enforcement agencies can be warned and prevent crime.

### REFERENCES:

1. Pratibha Kumari, Akanksha Gahalot, Suraina, Lokesh Chouhan (2020, february). An overview on 'crime prediction and analysis'. In Conference: 2020 2nd International Conference on Data, Engineering and Applications (IDEA)
2. Shamsuddin, N. H. M., Ali, N. A., & Alwee, R. (2017, May). An overview on crime prediction methods. In Student Project Conference (ICT-ISPC), 2017 6th ICT International (pp. 1-5). IEEE.
3. Yadav, S., Timbadia, M., Yadav, A., Vishwakarma, R., & Yadav, N. (2017, April). Crime pattern detection, analysis & prediction. In Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of (Vol. 1, pp. 225-230). IEEE.



4. Sivaranjani, S., Sivakumari, S., & Aasha, M. (2016, October). Crime prediction and forecasting in Tamilnadu using clustering approaches. In *Emerging Technological Trends (ICETT), International Conference on* (pp. 1-6). IEEE.
5. Sathyadevan, S., & Gangadharan, S. (2014, August). Crime analysis and prediction using data mining. In *Networks & Soft Computing (ICNSC), 2014 First International Conference on* (pp. 406-412). IEEE.
6. Nath, S. V. (2006, December). Crime pattern detection using data mining. In *Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 IEEE/WIC/ACM International Conference on* (pp. 41-44). IEEE.
7. [https://www.academia.edu/74533300/Crime\\_Prediction\\_and\\_Analysis\\_Using\\_Machine\\_Learning](https://www.academia.edu/74533300/Crime_Prediction_and_Analysis_Using_Machine_Learning)
8. <https://mifratech.com/public/blog-page/Crime+Prediction+and+Analysis+Using+Machine+Learning>
9. Zhao, X., & Tang, J. (2017, November). Exploring Transfer Learning for Crime Prediction. In *Data Mining Workshops (ICDMW), 2017 IEEE International Conference on* (pp. 1158-1159). IEEE.
10. Al Boni, M., & Gerber, M. S. (2016, December). Area Specific Crime Prediction Models. In *Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on* (pp. 671-676). IEEE.
11. Tayebi, M. A., Gla, U., & Brantingham, P. L. (2015, May). Learning where to inspect: location learning for crime prediction. In *Intelligence and Security Informatics (ISI), 2015 IEEE International Conference on* (pp. 25-30). IEEE.
12. K. Zakir Hussain, M. Durairaj and G. R. J. Farzana, "Criminal behavior analysis by using data mining techniques", *IEEE-International Conference On Advances In Engineering Science And Management (ICAESM -2012)*, pp. 656-658, 2012.
13. Mohammad Keyvanpour, Mostafa Javideh and Mohammadreza. Ebrahimi, "Detecting and investigating crime by means of data mining: A general crime matching framework", *Procedia CS. 3*, pp. 872-880, 2011.
14. Sunil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav, "Crime pattern detection, analysis and prediction", *International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2017.
15. Amanpreet Singh, Narina Thakur, Aakanksha Sharma, "A review of supervised machine learning algorithms", *3rd International Conference on Computing for Sustainable Global Development*, 2016.
16. Varshitha D N Vidyashree K P, Aishwarya P Janya T S, K R Dhananjay Gupta Sahana R, "Paper on Different Approaches for Crime Prediction system", *International Journal of Engineering Research Technology (IJERT)*, ISSN: 2278-0181, 2017
17. R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. Shariat Panahy, and N. Khanahmadliravi, "An experimental study of classification algorithms for crime prediction," *Indian J. of Sci. and Technol.*, vol. 6, no. 3, pp. 4219- 4225, Mar. 2013.
18. K.B.S. Al-Janabi, "A Proposed Framework for Analyzing Crime Data Set using Decision Tree and Simple K-Means Mining Algorithm," in *Journal of Kufa for Mathematics and Computer*, Vol. 1, No. 3, 2011, pp. 8-24.



# International Journal For Advanced Research In Science & Technology

A peer reviewed international journal

[www.ijarst.in](http://www.ijarst.in)

**IJARST**

ISSN: 2457-0362

19.A. Malathi, S.S. Baboo, “An Enhanced Algorithm to Predict a Future Crime using Data Mining,” in International Journal of Computer Applications, Vol. 21, 2011, pp. 1-6

20.Ned Horning . Introduction to Decision trees and Random Forests, American Museum of Natural History's Center for Biodiversity and Conservation