International Journal of Digital Crime and Forensics
Volume 12 • Issue 4 • October-December 2020

Spatio-Temporal Crime Analysis Using KDE and ARIMA Models in the Indian Context

Prathap Rudra Boppuru, Christ University (Deemed), India
https://orcid.org/0000-0002-5161-4972

Ramesha K., Dr. Ambedkar Institute of Technology, India

ABSTRACT

In developing countries like India, crime plays a detrimental role in economic growth and prosperity. With the increase in delinquencies, law enforcement needs to deploy limited resources optimally to protect citizens. Data mining and predictive analytics provide the best options for the same. This paper examines the news feed data collected from various sources regarding crime in India and Bangalore city. The crimes are then classified on the geographic density and the crime patterns such as time of day to identify and visualize the distribution of national and regional crime such as theft, murder, alcoholism, assault, etc. In total, 68 types of crime-related dictionary keywords are classified into six classes based on the news feed data collected for one year. Kernel density estimation method is used to identify the hotspots of crime. With the help of the ARIMA model, time series prediction is performed on the data. The diversity of crime patterns is visualized in a customizable way with the help of a data mining platform.

KEYWORDS

ARIMA Model, Crime, Crime Analysis, Crime Density, Crime Prediction, Kernel Density Estimation, Law Enforcement, Newsfeed Crime Data, Predictive Analysis, Social Media

1. INTRODUCTION

Cities are considered complex systems and all the components are not working independently but interacting with each other. For the development of sustainable cities, urban planners and managers need to a more comprehensive and up-to-date understand of different aspects of cities. With the increase in migration to cities, crime rates in India have been steadily increasing. However, there has been very little research done on the crime patterns in India and cities such as Bangalore. While crime data is available from the National crime database, it is often outdated and does not have the current information at the rate in which the crime happens each day, as per the study (Chainey & Radcliffe, 2018). Research efforts are directed to use social media data which contains real-time information about crimes. Correlation and classification of data help in identifying the similarities and differences between data objects (Zhang & Wu, 2011).

DOI: 10.4018/IJDCF.2020100101

Historical crime data can be used to identify high crime areas and plan resources optimally. Predictive policing using data enables law enforcement authorities to take proactive decisions to improve response time to crime incidents (Angers, Biswas & Maiti, 2016). Knowledge acquired from the data mining techniques can be used to helping find criminals faster thereby reducing crime rate. Crime prediction, a subtask of crime analysis, considers all the past crime records, classifies the crime categories and predicts the future crime. Crime prediction using pattern and association rule mining determines the chances of performing crime by the same criminal.

This paper aims to solve this problem by using news feed data as the primary source. This source gives real-time information about the crime and the place in which it has happened, enabling criminologists to visualize data in real time. In the following sections, the research findings of crime rates in India and Bangalore are presented, and the implications of these findings for future research directions are discussed. The steps followed in this approach are data collection, classification, visualization and pattern prediction.

This paper outlines the development of a crime analytics system that analyses the news feed data and performs interactive data visualizations for generating insights. The remainder of this work consists of Section 2. Literature review, Section 3. Methodology used, Section 4. System architecture, Section 5. Results obtained. The last section 6 concludes our work and suggests future research scope.

2. LITERATURE REVIEW

Criminology has two research areas – one that seeks to understand the development of criminal offenders and the other that aims to understand the evolution of crime events (Fayyad, 2012). Rational choice theory of criminology explains that perpetrators select targets and identify the means to achieve their goals based on rational decisions. Routine activity theory explains that the offender must be present along with other crime favorable circumstances at the same time for the occurrence of the crime. Crime pattern theory combines rational choice and routine activity theory and emphasizes the importance of place in crime events (Green, 2002). Significantly dense population growth increases opportunities for crime. The complexities of large-scale, urban, residential development and the challenges of embedding crime prevention during this period of rapid and sustained population growth are studied in Australia (Clancey, Kent, Lyons & Westcott, 2017).

Traditional urban crime research focused on leveraging demographic data, which is insufficient to capture the complexity and dynamics of urban crimes. In the era of big data, we have witnessed advanced ways to collect and integrate fine-grained urban, mobile, and public service data that contains various crime-related sources as well as rich environmental and social information (Zhao & Tang, 2018). With the recent advancements in environmental criminology, spatial and temporal information has found its importance in criminology research. The ecological approach used by criminologists maintains the effects of place in the crime patterns in a neighborhood community (Jiang, Yang & Li, 2018). The crime event gives a lot more information about the criminal than the criminogenic causes such as social, developmental or biological characteristics of the criminal. There are three types of crime analysis - micro, meso, and macro. Microanalysis looks at crime in specific locations. Meso analysis analyses crime behavior at the neighborhood level (Eck & Weisburd, 2018). The macro analysis compares crime rates between different countries or between different regions.

According to Eck and Weisburd, the rise in information technology enables rapid recording and processing of crime incident information and the geographic location they occur. This data can help generate insights for a better understanding of criminal behavior. Spatial and temporal analysis methods use modeling and mapping to help law enforcement agencies to determine the distribution of crime and the likelihood of their occurrence. This methodology is increasingly becoming important for studying crime trends and activities. There are convincing data management and visualization toolboxes for analyzing crime within a theoretical framework. Crimes such as gang violence happen concentrated in time and space. Spatiotemporal hotspot is defined as a geographic location coupled with a time period where greater than normal amount of crimes occurs. It aims to incorporate temporal patterns on spatial hotspot for crime analysis.

As per the research from Yeran Sun and Yunyan Du, big data can help in disaster management and emergency planning in urban cities. New technologies enable faster response to natural hazards. Improving urban livability is a crucial issue in urban planning and development. As a fundamental part of development of a livable city, the settlement is facing new challenges in accommodating a fast-growing urban population. Using diverse spatial data (census data, environmental data, survey data, geosocial networking data), it is possible to assess the map settlement suitability in urban cities. Specifically, they take account of a variety of environmental and socio-economic factors and integrate and manage datasets with different forms and scales. Specifically, the environmental factors selected include distance from the built-up area, topographic (slope), distance from river, soil (bearing capacity), distance from road, and land use activities; while the socioeconomic factors selected include population density, land value, proximity to the security area and same ethnic area.

Crime analysis can provide relevant information regarding crime patterns and trends to help law enforcement personnel to plan the effective deployment of human resources to suppress criminal activities. This helps in informing law enforcement department about relevant insights about crime promptly. This also enables the end user to make use of data available in public domain to know about the areas more prone to crime. Crime analytics helps in identifying the future pattern of crime without necessarily understanding the underlying factors causing the crime in a particular region. A crime analytics system extracts insight from available crime data and predicts future occurrences based on spatial distribution. There are various frameworks explored for understanding crime trends including the new distance measure that profiles individuals and classifies them.

The existing systems used by Indian police include a query based interactive system and new e-governance initiatives for better analysis of crime to assist police in curbing crime incidents. The data collected in the National Crime Record Bureau (NCRB) are used to identify the crime hotspots, and they are classified, into groups, using various classification algorithms. This interface has been useful in identifying Indian crime records. Similar kinds of analytics systems are implemented in Malaysia such as Visual Interactive Malaysia Crime News Retrieval System (i-JEN). With the help of classification and clustering, crime data such as location and time are used to identify the trends. Effective and interactive ways of visualizing crime data help in combating crime at a higher level. The clustering technology can be used to determine the accounting frauds easily.

Use of clustering algorithm and missing value helps in predicting crime patterns and improve the process of solving the crime. City crime data can be used to identify the dangerous areas to avoid for citizens and intimate them accordingly. Crime analytics tool can collect and economically clean the data and analyze it to determine trends and patterns. Decision tree algorithms and k-means clustering are used to identify the crime patterns. This helps criminologists to discover trends and patterns, make forecasts, find relationships and map criminal networks that are normally hidden from plain sight. Data analytics helps in programming years of human experience and insight into computer models that assist in designing a simulation model.

3. METHODOLOGY

In this section, we are going to explain how we have collected the data, which algorithms we have used to classify the data as well as its analysis. Later we will be discussing the crime density, how we have found out the exact crime location, the longitude and the latitude of the precise spot where the crime density is highest. The exact spot is found based on the data collected for 1 year. Later we will be discussing how we have used the ARIMA model for crime forecasting and prediction.

3.1. Newsfeed Data Aggregation

RSS news feed data are rich in both location and context for prediction of crime incidents (Behrens & Robert-Nicoud, 2014). This method has two components. The first component has a spatiotemporal model that uses feature-based extractions to predict the future crime rates in a particular location. The second component involves the extraction of textual information through semantic role labeling. The essential features in the news feed are extracted using linguistic analysis and mathematical topic analysis. Addition of news feed data to traditional crime data sources increases the crime prediction accuracy. This method can be extended to form a decision support system. A sampling approach is used to handle the missing data over time. Some crime types seem to have a close relationship with the internet and social media data.

3.2. Classification of Text Based on Features

Machine Learning for Language Toolkit (MALLET) is used for extracting features from text and classify them into various clusters. It is a sophisticated natural language processing tool for topic analysis, document classification, information extraction, and machine learning. It consists of multiple algorithms which remain useful in doing the classification.

3.3. Spatio-Temporal Pattern Analysis

The crime patterns in urban areas are not randomly or evenly distributed. The typical pattern is that the crime occurs rather dense in some regions of a city and sparse in other areas (Bowers & Newton, 2018). With the help of spatial pattern analysis, it is possible to identify the hotspots, i.e., the area in which there is a high aggregation of crime. Also, the environmental context plays a vital role in the occurrence of crime. The definitions of spatial pattern analysis are as follows: crime hotspots are defined as the geographic locations in which criminal activities repeatedly occur (Marzan & C. Baculo, 2018). Individuals have a higher risk of victimization in these areas compared to other places. Cold spots are locations in which there is decidedly less criminal activity. Spatial clustering is used to identify the hotspot patterns in crime. The spatial analysis correlates with the environmental contexts.

3.4. Crime Density Detection

Crime hotspot detection is the spatial mapping technique that identifies the concentration of various crimes in the urban area (Berestycki, Wei & Winter, 2014). The most widely used crime hotspot detection method is the Kernel Density Estimation method. KDE is a non-parametric method of estimation in which the probability density of crimes is calculated. KDE uses grid cell size, interpolation methods, and bandwidth to identify the precision of kernel density. The interpolation process has various user-defined settings thus increasing the quality of KDE hotspots (Wang & Luo, 2018). This analytical technique is used for multiple types of crime such as burglary, robbery, and assault, etc. The KDE hotspot with low resolution is converted into one with contour lines. The hotspot is generated with smooth boundaries; hence the generation speed is increased. Parameters such as bandwidth and grid cell size are specified for generating hotspots.

3.4.1. Crime Density Identification Using KDE

Kernel Density Estimation (KDE) is the research method used for estimating the data. There are various terminologies used in KDE as follows.

3.4.1.1. Probability Density Function

Probability theory deals with quantities that have a random distribution. The probability density function (PDF) is defined as the probability of a random value fitting into a range of values in the function. In this methodology, the integral of the value's density is identified. The resultant value gives the probability of the new random value. The probability of the random number is given by the area covered by the density function:

$$P(a < X < b)\int_a^b \left(d\right)dx \ \forall a < b \tag{1}$$

Equation (1) shows the relationship between crime type X, crime density D and the bandwidth of the crime from a to b:

- **Bandwidth:** Bandwidth is a smoothing parameter that denotes the width of the sample in KDE. It determines the search radius of the function. The function can be over or under smoothed. Bandwidth can be estimated on the basis of thumb rules. Perfect computational solutions cannot be applied.
- **Grid Size:** The grid cell size defines the resolution of the KDE algorithm. Large grid sizes lead to the low-quality hotspot and low visualization. The right grid size is identified by the standard deviation of latitude and longitude. Grid size is denoted by a and b.

3.4.1.2. Kernel Density Estimation

Kernel density estimation is used to identify the crime hotspot of a city easily. There is a spatial mapping of the various parts of the city such as K.R Market in Bengaluru. The strength of the hotspot is measured by counting the events in a particular area. The area of the circle divides the static slide circle and the number of events in the area. The event density d(S) is given by:

$$d(s) = \# S \in C\left(s, r\right) / \pi r^2 \tag{2}$$

Equation (2), C(s,r) denotes the center of the circle, #S is the event count of crimes and r is the circle radius. Kernel Density Estimation is defined as:

$$1 / nh \sum_{i=1}^{n} \frac{k\left(x - Xi\right)}{h} = f\left(x\right) \tag{3}$$

In Equation (3), x is the variable, h is the bandwidth, x-Xi is the distance between the events Xi and the estimated point x. k() depicts the kernel function, X1….Xn gives the newsfeed data that is randomly selected. Based on the spatial study, KDE is defined as:

$$\sum_{i=1}^{n} \frac{\frac{1}{\tau 2} k\left(s - si\right)}{\tau} = \lambda \tau \left(s\right) \tag{4}$$

In Equation (4), n is the sample size, s-Si is the distance between locations, Si is the mean value, τλ(s) is the Crime data, K is the coefficient. S1……..Sn is the newsfeed data that is randomly selected. k() is the kernel function, S-Si is the distance between the estimated points and the event Xi.

The formula shows that KDE is influenced by bandwidth. When τ increases, the point density change is smooth. When τ decreases, the change is rough. Point process smooth intensity represents the kernel density estimation. In this paper, 64 types of crime are grouped into 6 classes and the crime density is identified for each one of them. The effectiveness of the model is then evaluated using various parameters.

3.5. ARIMA Forecast Modelling

Time series forecasting is a method in which data is collected regarding a particular event, and a model is generated to represent the underlying relationship (Chen & Yuan, 2008). The model is then used to forecast the future values of the event through time series extrapolation. This method is useful in estimating future behavior when there is no existing correlation identified. Autoregressive integrated moving average (ARIMA) model is the most widely used time series models. In ARIMA model, the future values are the linear projections of the past value. The application of nonlinear forecasting is minimal.

The underlying process that generates the time series in the ARIMA model is as follows:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \varepsilon_t - \varphi_1 \varepsilon_{t-1} - \varphi_2 \varepsilon_{t-2} -\cdots-\varphi_q \varepsilon_{t-q} \quad (5)$$

In Equation (5) φ_i (i=1; 2;::: ; p) and φ_j (j=0; 1; 2;::: ; q) are the model parameters. Y_t and ε_t are the actual value and random error at a particular time t. The orders of the model are expressed as p and q. ARIMA model determines the appropriate model order of (p, q). The ARIMA model is given by:

$$\left(1-\sum_{i=1}^{p'} \propto iLi\right) Xt = (1+\sum_{i=1}^{q}\theta iLi)\varepsilon t \quad (6)$$

In Equation (6), L shows the lag operator, θi is the parameter for the moving average, $\propto i$ is the parameter for the autoregressive model, εt is the error term. These error terms are the independent, identically distributed variables. It is sampled from a normal distribution with zero mean.

The work by Box and Jenkins has developed a practical method to implement ARIMA models. This method works in three iterative steps. It includes model identification as the first step, parameter estimation as the second step and diagnostic checking as the third step. Model identification ensures that the time series generated will have auto correlational properties. The data is transformed into the model identification step to make the stationary time series. Once the approximate model is developed, parameter estimation is done to reduce the overall amount of errors. The model adequacy is then checked with the help of diagnostic checking. This ensures that the model's future predictions fit with the historical data. This three-step iterative process is performed multiple times to identify the right model fit. The final selected model can then be used for prediction purposes.

3.6. Types of Crime Mapped

In this research paper, 68 types of different crimes are grouped into 6 classes which are given below. The data analytics system can view forecasting data for each type of crime on both National level and city level namely Bangalore (Table 1).

4. SYSTEM ARCHITECTURE

The methodology used is to collect the RSS news feed data in a database and clean them to remove duplicity. The data is then classified based on the type of crime string and location. Data that is inconsistent, incomplete and lacking in trends are turned into actionable information. Time series forecasting is done with the help of the ARIMA model.

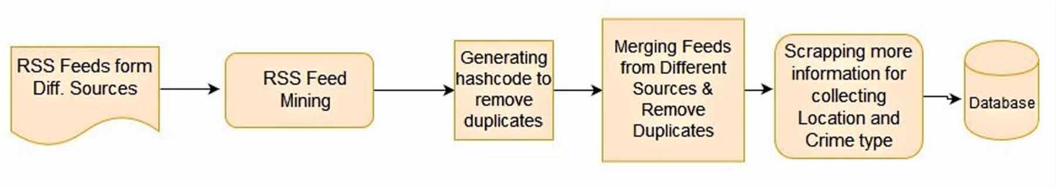
Table 1. Types of crimes

Drug Related Crimes	Drug Trafficking, Drug dealing, Alcohol Drinking, Alcohol dealing, Liquor law violation arson, Alcohol, Drug, Narcotics, liquor law violation arson
Violent Crimes	Assault, Rape, Murder, Robbery with Firearms, Terrorism, Kidnapping, Sexual Harassment, Employee Abuse, President Abuse, Intentional Killing Peoples, sexual assault, sex offense, homicide, gambling, gunshot, shootout, gang rape, harassment, murder, attempt to murder, kidnapping & abduction, attack, dishonor, lash out, outrage, sexual abuse, snatch, putting to death, belt down, obliterate, hit and run
Commercial Crimes	Official Document Forgery, Currency Forgery, Official Seal Forgery, Official Stamp Forgery, Bribery, counterfeiting, cheating
Property Crimes	Arson, Robbery without Firearms, Motor vehicle theft, Theft, battery, burglary, robbery, Deceptive practice, riots, criminal breach of trust, larceny, stealing, assault and battery, barrage, barrage fire, bombardment, electric battery, shelling, stamp battery, looting, criminal, embezzlement, trespass, incendiarism, shoplifting, vandalism
Traffic Offences	Speeding, Signal Jump, Running a Red Light, drunk driving
Other Offences	Employing Illegal Worker, Prostitution, Illegal Gambling, Begging, Adultery, Homosexuality, weapons violation, offense involving children, public peace violation, stalking, cheating, hurt, counterfeiting, dowry deaths, outrage her modesty, causing death by negligence, suicide, criminal damage, weapons violation, harlotry, whoredom, homicide, espionage, pickpocketing, pilfering, poaching.

4.1. Data Mining, Cleaning, and Exploratory Data Analytics

First layer Figure 1 of data analytics, the RSS feeds are mined from different sources, and duplicate feeds are removed with the help of hash code. The feeds are then stored in a database. An algorithm is used to remove the duplicates, and new information is merged with the old data. Using a scraping algorithm textual information related to crime type and location are extracted and stored in a database. An Xml parser is used to retrieve the crime related text from the news feeds that are in xml format. The preprocessing of the news feeds are done using tools such as POS, Lemmatizing, Stop words and stemming. The noisy data is removed and cleaned using tools like plyr and diplyr. The output data is then stored in xml format in a local database.

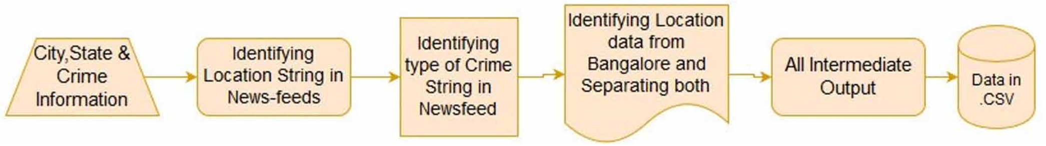
Figure 1. Data preprocessing architecture


RSS Feeds form Diff. Sources
RSS Feed Mining
Generating hashcode to remove duplicates
Merging Feeds from Different Sources & Remove Duplicates
Scrapping more information for collecting Location and Crime type
Database

4.2. Preprocessing and Classification

Second layer Figure 2 shows from the news feed data, specific feature-based information such as city, and state information and the type of crime are extracted. The data is preprocessed by identifying the location string in the news feeds. More specifically, crime related text strings are identified in the news feeds. Then crime data associated with Bangalore is identified and separated from the overall data. All the intermediate output is converted to excel form and stored in a database. Digest package has been

Figure 2. Data classification architecture

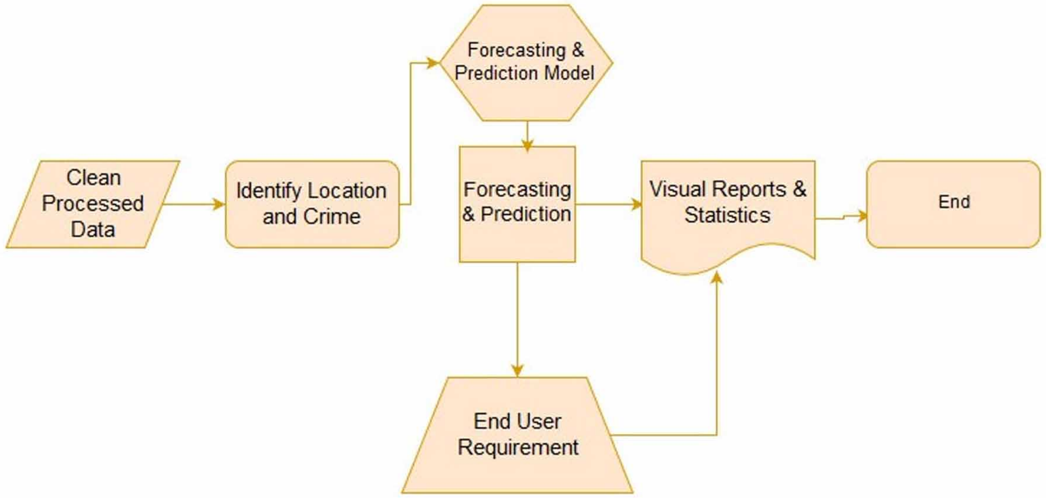

City,State & Crime Information
Identifying Location String in News-feeds
Identifying type of Crime String in Newsfeed
Identifying Location data from Bangalore and Separating both
All Intermediate Output
Data in .CSV

used for duplicate detection with the help of hash code. The collected data is periodically checked for adequacy. A minimum of 3000 trial and error procedures are done to review the data. Ggmap is used for locating hotspots in the data. Then the classification model is prepared for preprocessing the data. The extracted data is then stored in both csv and xml format and sent to the database.

4.3. Geospatial Analysis and Visualization

Figure 3 shows the data post processing and visualization methodology. Tools like plyr and diplyr are used for cleaning the processed data. This is done to ensure that the collected data does not have any outliers. The data is analyzed multiple times and the latitude and longitude are processed. A model is built to analyze the crime. In some cases, the data are missing. These missing values are inserted in case of location not available. The final results are shown in a hotspot visualization on the map of India and Bangalore. The data analytics system also has the options to add filters to visualize the crime hotspots in different ways.

Figure 3. Data post-processing and visualization architecture


Forecasting & Prediction Model
Clean Processed Data
Identify Location and Crime
Forecasting & Prediction
Visual Reports & Statistics
End
End User Requirement

4.4. Time Series Forecasting Using ARIMA Model

The data is populated on linear time scale and ARIMA forecasting model is used to extrapolate the data. ARIMA model identifies the correlation between crime data over last one year and predicts the future value of the crime data.

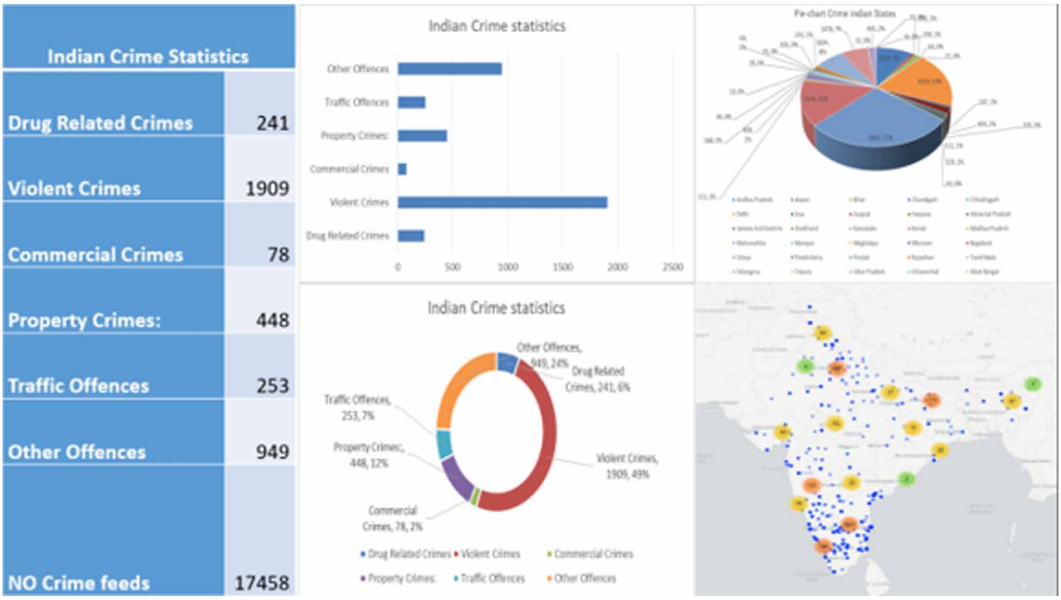
5. RESULT AND DISCUSSION

The results are divided into KDE analysis of existing data and time series forecasting using the ARIMA model. The spatial analysis of the crime data in India and Bangalore have shown great variance in the type of crime and their occurrence rates between regions. The spatial patterns can be explained with the help of social disorganization theory having factors of population density, family stability, ethnic heterogeneity, and residential instability. It is found that areas of high population density are most prone to high crime rates. A city with large populations such as Delhi, Bangalore, and Hyderabad are expected to have high crime rates. Regions with a high prevalence of drug and alcohol activity are strong indicators for crime rates in areas. Areas with theft and assault are highly correlated with each other.

5.1. Geospatial Analysis of Crime – India

Figure 4 gives the overall picture of the crime rates in India. This analysis shows the various types of crime that occur in the Bangalore city. The data can be visualized in various forms such as hotspots or pie charts for in-depth analysis.

Figure 4. Geo-Spatial analysis of crime – India



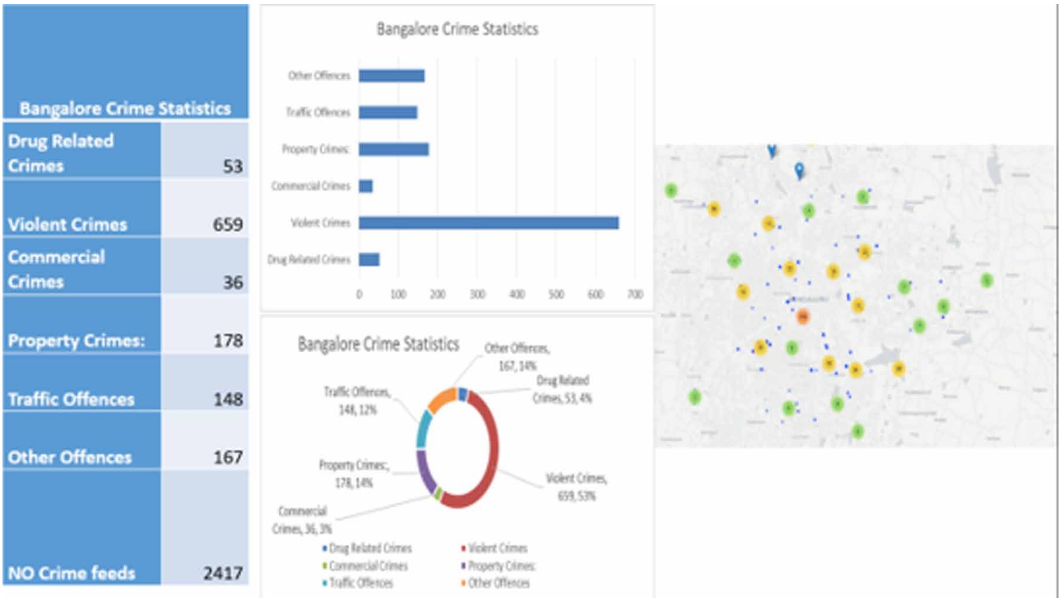
Indian Crime Statistics	
Drug Related Crimes	241
Violent Crimes	1909
Commercial Crimes	78
Property Crimes:	448
Traffic Offences	253
Other Offences	949
NO Crime feeds	17458

Figure 5 gives the geospatial analysis of crime rates in Bangalore. It is found that assault and trespass are the most reported crime in the city.

KDE algorithm shows the density of all crimes happening in India in Figure 6. It can be found that the crime hotspots are concentrated around the urban areas.

Figure 7 to 10 identified the density of classified crimes using KDE algorithm. It can be found that the crime hotspots are concentrated around the urban areas. KDE algorithm shows the density of individual crimes such as Drug related, Property related, Violent Crimes happening in India. Based on our experiment violent crimes is more compared other type of crimes in Indian context.

Figure 5. Geo-Spatial analysis of crime - Bangalore



Bangalore Crime Statistics	
Drug Related Crimes	53
Violent Crimes	659
Commercial Crimes	36
Property Crimes:	178
Traffic Offences	148
Other Offences	167
NO Crime feeds	2417

Figure 6. Density of ALL crimes using KDE algorithm-India

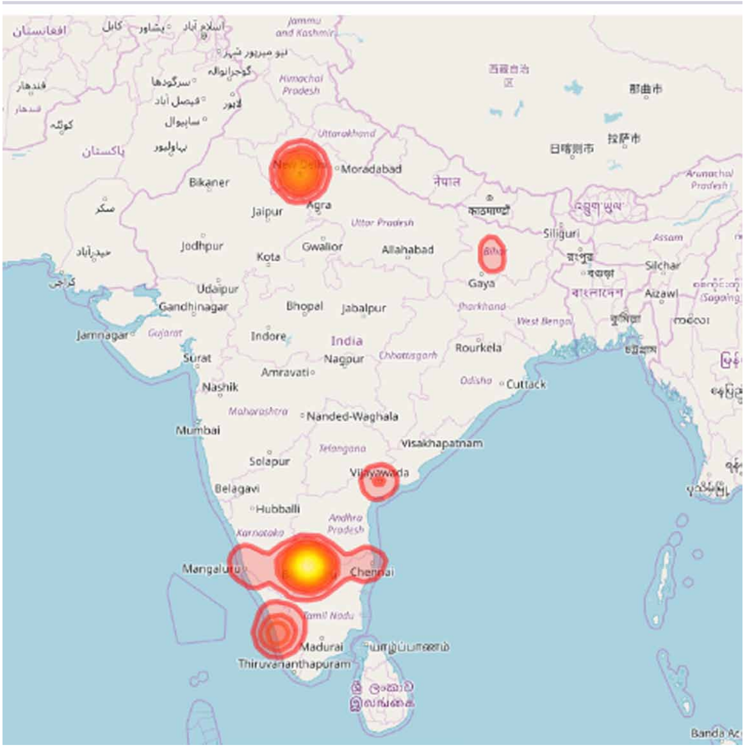


Figure 7. Density of individual crimes using KDE algorithm-India (violent crimes)

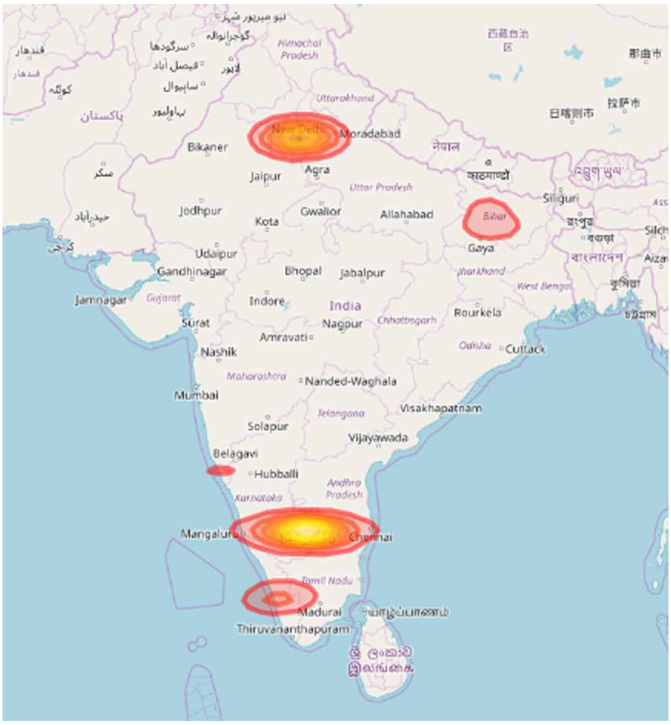


Figure 8. Density of individual crimes using KDE algorithm-India (commercial crimes)

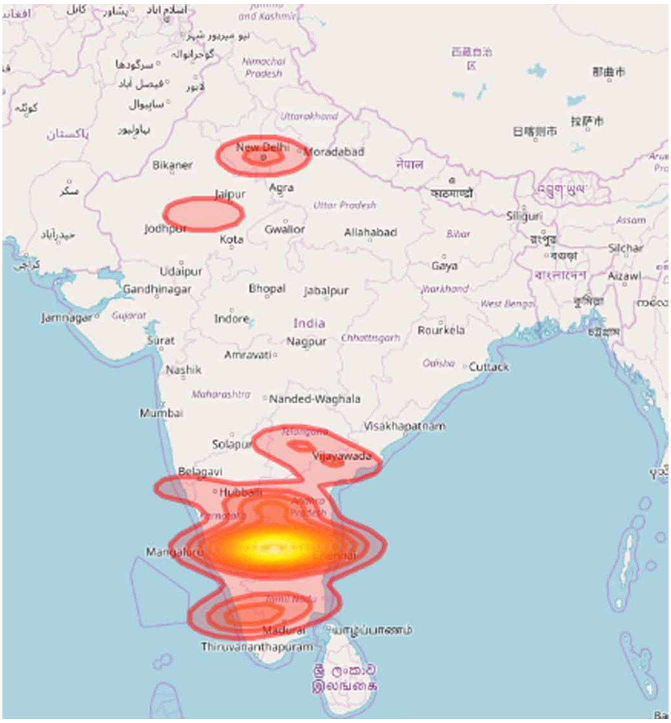


Figure 9. Density of individual crimes using KDE algorithm-India (property related crimes)

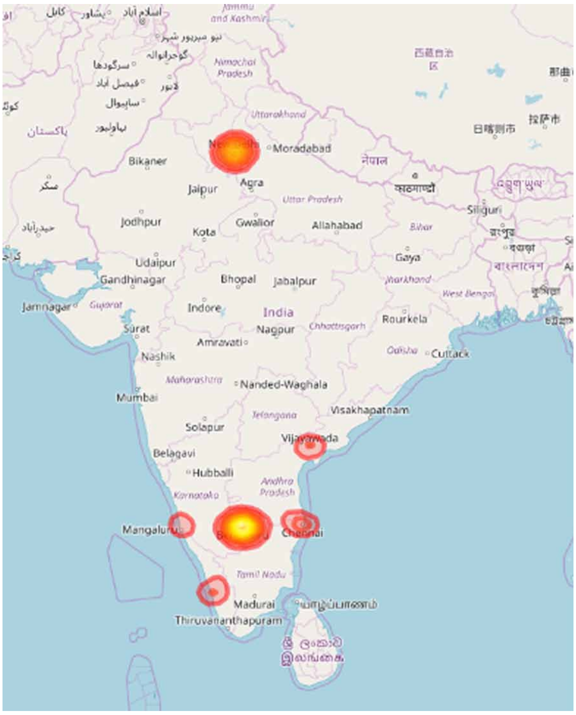
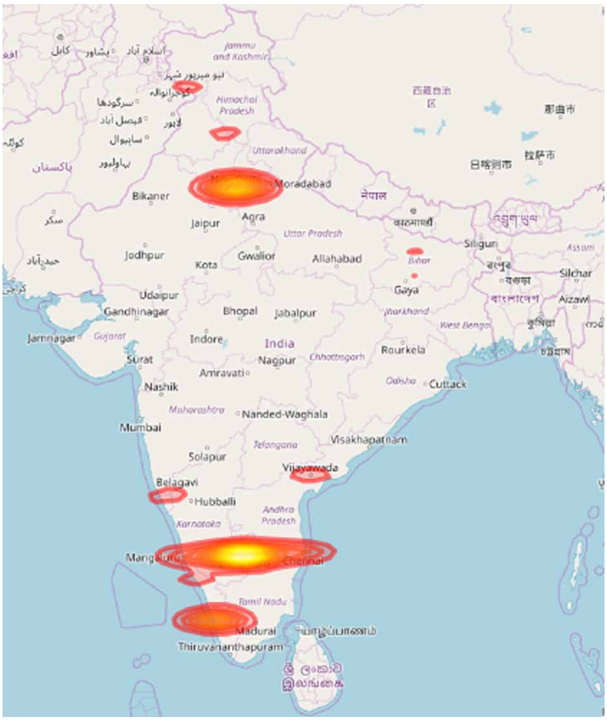


Figure 10. Density of individual crimes using KDE algorithm-India (drug related crimes)



KDE algorithm shows the density of all crimes happening in Bangalore in Figure 11. It can be found that the crime hotspots are concentrated around the Major places in city like Corporation circle, kempegowda majestic, and white field etc.

Figure 12 and 13 Show the density of Property related and Drug related crimes using KDE algorithm. It can be found that the crime hotspots are concentrated around the Major areas as Corporation circle, kempegowda majestic, Koramangala, Bellandur, Chikkaballapur, White field etc.

Figure 14 shows the Time series graph using ARIMA Model. For experimental purpose here considered the 1-year newsfeeds data i.e. 2017 Jan to 2018 Jan. Black colored line in graph is represented 365 days of crime frequency and blue color line represented 15days forecasting identified using exiting of 1 year.

Figure 11. Density of all crimes using KDE algorithm- Bangalore

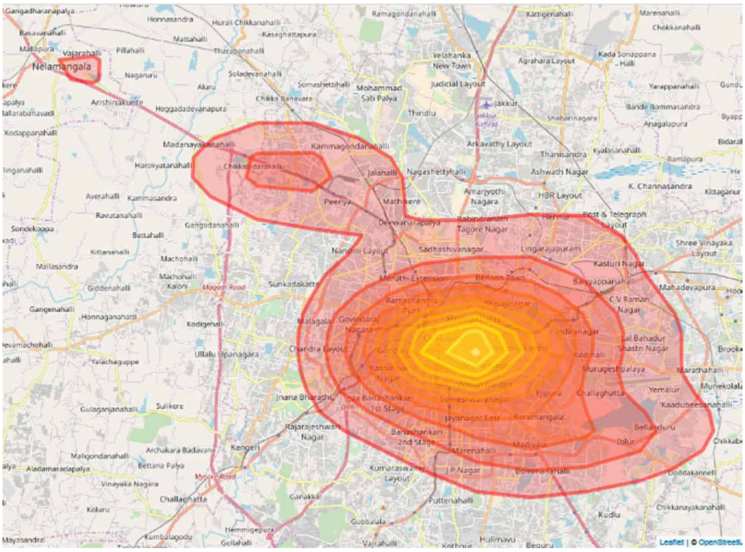


Figure 12. Density of crimes using KDE algorithm – Bangalore (property related crimes)

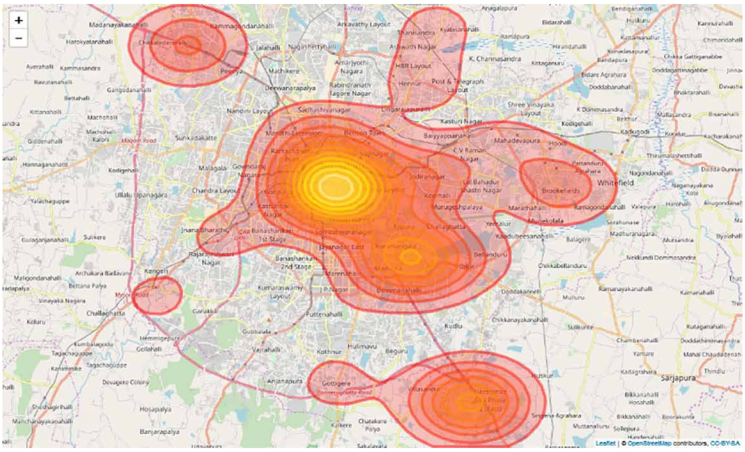


Figure 13. Density of crimes using KDE algorithm – Bangalore (drug related crimes)

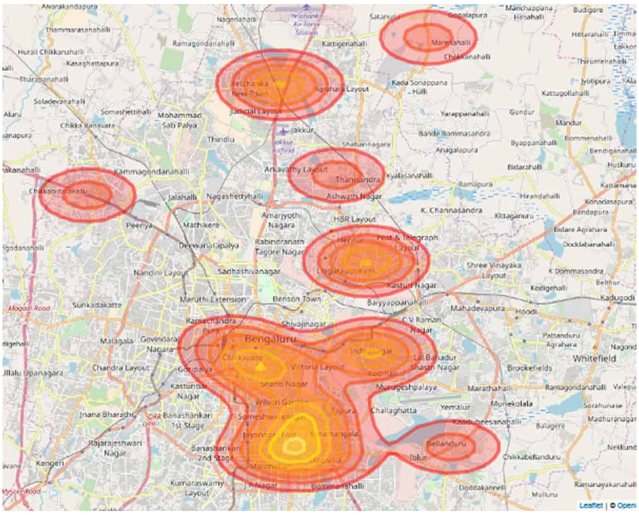


Figure 14. Forecasting using ARIMA Logarithmic scale model (time series analysis- 1 day period)-complete data-India- ALL crimes

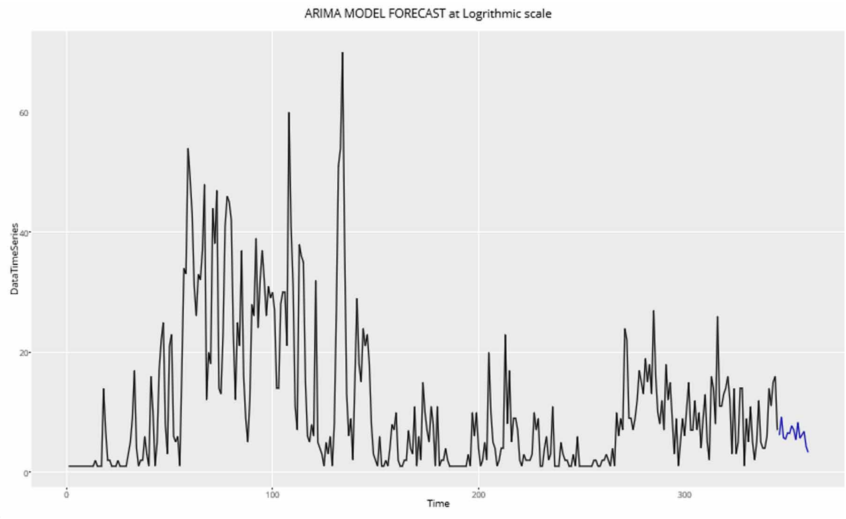


Table 2 divided into 2 parts first part is Random values (High or low crime rate change compared to other days) in the 365days and second part shows the crime rate 365 to 400 days' time period. In part one From S number 1 to 14 is the values considered form the crimes happening in the period of Jan 2017 to Jan 2018. S no 15 to 18 forecasting of crimes identified using on ARIMA Model. The ARIMA model is applied for 365 days of data and identified next 15 days of crimes in the Indian context. Day 108 and 134 having more crime rate because of some specific events in the area. Here 40 is the highest crime Value we considered for crime rate for forecasting crime. To validate the forecast values considered the real-time data of news feed from 366 day to 400 day and compared with our Forecasting values. The forecasting values is matched with 75% Accuracy with real time data. On specific dates.

Table 2. Forecasting using ARIMA Logarithmic scale model (time series analysis- 1-day period)-complete data-India-ALL crimes

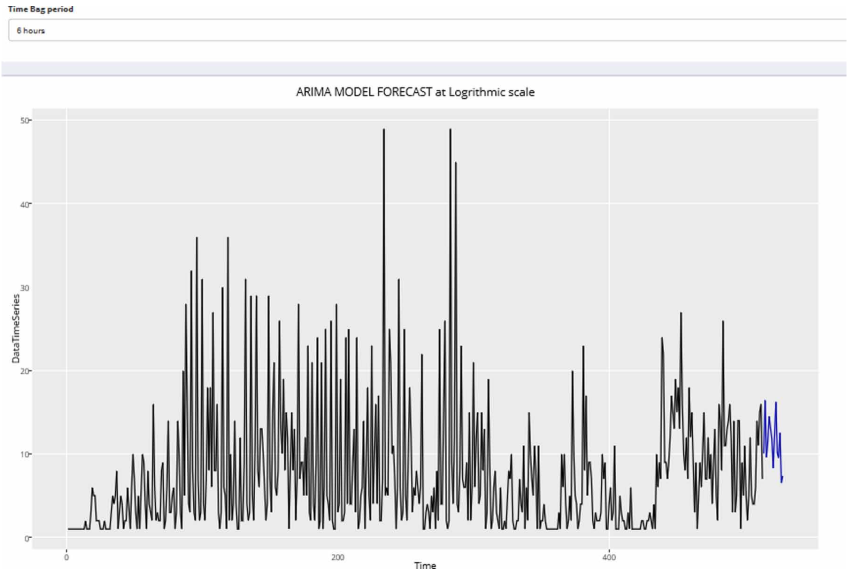
Day 1 to 365			Day 365 to 400		
S.No	Time	Crime Frequency/ Forecast	S.No	Time	Crime Frequency/Forecast
1.	18	14	1.	365	13
2.	33	17	2.	368	16
3.	41	16	3.	370	15
4.	51	23	4.	371	15
5.	59	54	5.	373	19
6.	108	60	6.	375	8
7.	134	70	7.	377	8
8.	114	29	8.	380	25
9.	190	25	9.	390	17
10.	205	20	10.	400	32
11.	229	9			
12.	285	27			
13.	316	26			
14.	350	32			
15.	368	8.30			
16.	389	12.60			
17.	373	10.20			
18.	379	5.79			

Figure 15 shows the Time series graph using ARIMA Model. For experimental purpose here considered the 1 year newsfeeds data i.e. 2017 Jan to 2018 Jan. Black colored line in graph is represented 530-6 hour time bag crime frequency and blue color line represented 15-6 hour time bag forecasting identified using exiting 563- 6 hour time bag. The forecasting values is matched with 78% Accuracy with real time data. On specific 6-hour time bag.

Table 3 divided in to 2 parts first part shows the Random values in the 563 time slots in a period of Jan 2017 to Jan 2018. From S number 1 to 16 is the values considered form the crimes happening in the period of Jan 2017 to Jan 2018 every 6 Hours' time bag. S no 17 to 20 forecasting of crimes identified using on ARIMA Model. In Part2 to validate the forecast values considered the real time data of news feed from 532 to 580 6 hours' time bag data considered and compared with our Forecasting values. The forecasting values is matched with 78% Accuracy with real time data. On specific time slots.

Table 4 Shows the Random values in the 52 time slots in a period of Jan 2017 to Jan 2018. From S number 1 to 6 is the values considered form the crimes happening in the period of Jan 2017 to Jan 2018 every 1week time bag. S no 7 to 8 forecasting of crimes identified using on ARIMA Model. Similarly, we had identified the 6 classes (Violent crimes, Property crimes, Drug related crimes etc.) of crimes time series analysis with same time bag period based on our analysis we identified Violent crimes is more possible crime in Indian context and Commercial Crimes more in Bangalore context (Figure 16).

Figure 15. Forecasting using ARIMA Logarithmic scale model (time series analysis- 6 Hours period)-complete data-India- ALL crimes

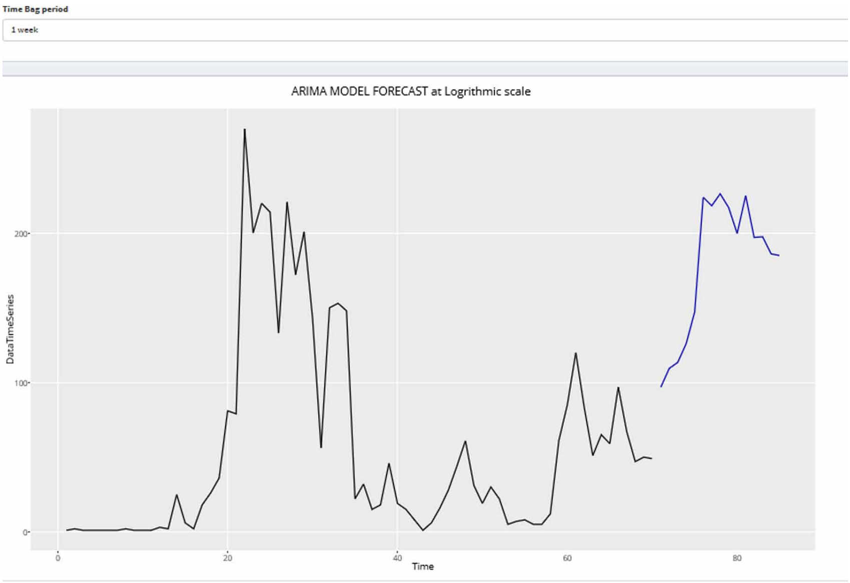


Time Bag period
6 hours

Table 3. Forecasting using ARIMA Logarithmic scale model (time series analysis-6 Hours period)-complete data-India-ALL crimes

1 to 532 (6 Hours' Time Bag)			532 to 580 (6 Hours' Time Bag)		
S. NO	**Time**	**Crime Frequency/ Forecast**	**S. NO**	**Time**	**Crime Frequency**
1.	20	5	1	532	14
2.	75	14	2	533	11
3.	96	36	3	534	12
4.	119	36	4	538	14
5.	132	31	5	540	15
6.	171	28	6	555	12
7.	199	28	7	560	15
8.	222	18	8	565	10
9.	234	49	9	568	11
10.	373	20	10	570	11
11.	383	17	11	580	12
12.	439	24			
13.	453	27			
14.	484	28			
15.	518	32			
16.	530	16			
17.	533	11.638			
18.	534	11.9			
19.	538	14.11			
20.	547	11.11			

Figure 16. Forecasting using Arima Logarithmic scale model (time series analysis- 1 Week period)-complete data-India- ALL crimes



Time Bag period: 1 week

Table 4. Forecasting using ARIMA Logarithmic scale model (time series analysis 1 Week- period)-complete data-India-ALL crimes

S.No	Time	Crime Forecast
1.	12	3
2.	22	150
3.	25	180
4.	33	153
5.	39	46
6.	48	61
7.	51	36.60
8.	52	20.20

6. CONCLUSION

This paper has utilized 1-year crime data (Jan 2017 to Jan 2018) that are specific to Indian conditions. The research findings in this paper have used Kernel density estimation (KDE) for crime hotspot identification and ARIMA model to predict the future crime behavior. This helps in identifying the probable factors responsible for causing the crime. The prediction rate accuracy is about 75%. With the help of these insights, regions with high levels of crime can be selected for intense observation as a preventative method for reducing crime rates. 68 types of crime are identified, and the data is classified based on them. Time series analysis enables prediction of crime rates in the same location in future. Along with the present scope of our project, which is prediction of the crime prone areas, we can also predict the estimated time for the crime to take place as a future scope. Along with this, one can try to predict the location of the crime. We will test the accuracy of frequent-item sets and prediction based on different test sets. So the system will automatically learn the changing patterns in crime by examining the crime patterns. Also the crime factors change over time. By shifting through the crime data we have to identify new factors that lead to crime. Since we are considering only some limited factors full accuracy cannot be achieved. For getting better results in prediction we have to find more crime attributes.

REFERENCES

Angers, J., Biswas, A., & Maiti, R. (2016). Bayesian Forecasting for Time Series of Categorical Data. *Journal of Forecasting*, *36*(3), 217–229. doi:10.1002/for.2426

Behrens, K., & Robert-Nicoud, F. (2014). Survival of the Fittest in Cities: Urbanisation and Inequality. *Economic Journal (London)*, *124*(581), 1371–1400. doi:10.1111/ecoj.12099

Berestycki, H., Wei, J., & Winter, M. (2014). Existence of Symmetric and Asymmetric Spikes for a Crime Hotspot Model. *SIAM Journal on Mathematical Analysis*, *46*(1), 691–719. doi:10.1137/130922744

Bowers, K., & Newton, M. (2018). A gis-linked database for monitoring repeat domestic burglary. *Mapping And Analysing Crime Data-Lessons From Research And Practice, 5*(40), 120-137. Retrieved from https://link.springer.com/article/10.1057/palgrave.sj.8350066

Chainey, S., & Radcliffe, J. (2018). GIS and crime mapping. *IEEE Journal*, *45*(3), 115–118. doi:10.13140/RG.2.2.11064.14081

Chen, P., & Yuan, H. (2008). Forecasting crime using the arima model. *Fifth International Conference On Fuzzy Systems And Knowledge Discovery*, *5*(10), 627-630. doi:10.1109/FSKD.2008.222

Clancey, G., Kent, J., Lyons, A., & Westcott, H. (2017). Crime and crime prevention in an Australian growth centre. *Crime Prevention and Community Safety*, *19*(1), 17–30. doi:10.1057/s41300-016-0012-1

Eck, J., & Weisburd, D. (2018). Crime and place, crime prevention studies. Criminal Justice Press, 4.

Fayyad, U. (2012). Knowledge Discovery and Data Mining: Towards a unifying framework. *2Nd Int. Conf. On. Data Mining and Knowledge Discovery*, *45*(3), 112–115. Retrieved from https://www.aaai.org/Papers/KDD/1996/KDD96-014.pdf

Green, M. (2002). Crime and society. *International Journal of Social Economics*, *29*(6), 781–795. Retrieved from https://www.essayagents.com/blog/crime-and-society-essay-paper

Jiang, F., Yang, X., & Li, S. (2018). Comparison of Forecasting India's Energy Demand Using an MGM, ARIMA Model, MGM-ARIMA Model, and BP Neural Network Model. *Sustainability*, *10*(7), 2225. doi:10.3390/su10072225

Kumar, S., & Revathy, S. (2016). Crime Mapping Analysis: A GIS Implementation in Madurai City. *International Journal of Science and Research*, *5*(3). doi:10.21275/v5i3.nov162301

Marzan, C., & C. Baculo, M. (2018). Time Series Analysis and Crime Pattern Forecasting of City Crime Data. *ACM Digital Library, 40*(11), 113-118. doi: 10.1145/3127942.3127959

Wang, B., & Luo, X. (2018). Graph-Based Deep Modeling and Real Time Forecasting of. *ACM Digital Library*, *54*(12), 113–115. doi:10.1371/journal.pone.0176244

Zhang, S., & Wu, X. (2011). Fundamentals of association rules in data mining and knowledge discovery. *Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery*, *1*(2), 97–116. doi:10.1002/widm.10

Zhao, X., & Tang, J. (2018). Crime in Urban Areas. *ACM SIGKDD Explorations Newsletter*, *20*(1), 1–12. doi:10.1145/3229329.3229331

Prathap Rudra Boppuru is an Assistant Professor & Research Scholar in Computer Science and Engineering at CHRIST (Deemed to be University) India. He has completed his B.Tech in Andhra University and M.Tech in JNTUK University. He exhibits a solid commitment toward continued career development embracing every opportunity to achieve educational excellence. Prathap writes Research articles on crime and analytics, which, considering where you're reading this, makes perfect sense in an Indian context. He has published articles on Crime analysis in Indian context using Social media data.

Ramesha K. PhD is a Professor in Dr. Ambedkar Institute of Technology, Bangalore, India. He has completed his PhD in JNTU Hyderabad 2013. His research Interest includes Machine Learning, Image Processing, VLSI, Digital Signal Processing, and Communication. He had presented and published several research papers in the area of Image processing and Digital Signal Processing. He guided several research scholars in the area of Image processing. Currently, he is working on machine learning algorithms for crime detection and prevention.