Geo-Spatial Crime Analysis Using Newsfeed Data in Indian Context

Article *in* International Journal of Web-Based Learning and Teaching Technologies · October 2019

DOI: 10.4018/IJWLTT.2019100103

2 authors:

Prathap Rudra
Christ University, Bangalore
9 PUBLICATIONS **32** CITATIONS

SEE PROFILE

Ramesha K.
Dr. Ambedkar Institute of Technology
18 PUBLICATIONS **168** CITATIONS

SEE PROFILE

International Journal of Web-Based Learning and Teaching Technologies
Volume 14 • Issue 4 • October-December 2019

Geo-Spatial Crime Analysis Using Newsfeed Data in Indian Context

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

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

ABSTRACT

Social media is the platforms where users communicate, interact, share ideas, career interest, pictures, video, etc. Social media gives an opportunity to analyze the human behavior. Crime analysis using data from social media such as Newsfeeds, Facebook, Twitter, etc., is becoming one of the emerging areas of research for law enforcement organizations across the world. The intelligence gathered through data is used for identifying future attacks and plan for reinforcements. This article focuses on the implementation of textual data analytics by collecting the data from different newsfeeds and provides an optimized visualization. This article establishes a framework for better prediction of 16 types of crime in India and the Bangalore area by providing the coordinates of the crime area, along with the crime which might happen there.

KEYWORDS

Crime Analysis, Crime Prediction, Crime, Law Enforcement, Newsfeed Crime Data, Predictive Analysis, Social Media

1. INTRODUCTION

Crime imposes a price on economies because it acts as a tax on what's wrongfully created within the society. To forestall crime, the govt will invest in exploitation completely different tools. One among these is education. Our Tools can identify the relation between the crime rate India and Bangalore where more educated people are more in number. Obtaining a lot of individuals educated will increase the returns of legal activities against outlawed ones. However, education features a positive crime, because it will increase the quantity of being taken. The connection between education and crime is so ambiguous. A recent Local study known the links between the behavior patterns of youngsters and their later anti-social conduct and noted that each will most frequently be foreseen from school days.

The national security concern is that the primary goal of any nation. Criminology studies specialize in distinctive criminal characteristics. The applying of information mining techniques will facilitate with this identification. Crime analysis, a part of criminologMike any, is a law enforcement preforms that entails the systematic analysis of deciding on and analyzing both patterns and tendencies in crime and disorder. Crime analysis is one of the most critical activities in solving criminal cases. Modern technology and other advancements have enabled analysis of a considerable number of crimes. Unregulated migration and population growth have contributed to the high magnitude of crimes happening in cities. Intelligence organizations and law enforcement organizations collect massive amounts of crime data to predict future occurrence. Since this involves analyzing a large number

DOI: 10.4018/IJWLTT.2019100103

of data, manual methods for analyzing such data with huge variations have proven to be stressful and unproductive. Therefore, crime analysis has become one of the prominent problems in all law enforcement organizations and Intelligence organizations.

This paper presents a crime analytics system based on social media data RSS feeds from different newspapers. With the help of crime analysis, crime events that share common characteristics such as location, frequency, etc. and victim patterns can be identified. By combining various available data in social media, it is possible to create a profile of social misconduct based on a spatial and temporal analysis. With the correct processing of social media data, it is possible to take corrective actions. With the construction of an efficient data pipeline, it is possible to identify indicators and utilize them for predicting social mishaps.

This paper focuses on development on a software system that can analyze the news feeds data and shows the spatial visualization of different types of crimes with Indian and Karnataka-Bangalore context. The remainder of work followed with 2. Literature review 3. Methodology Used & Implementation of the System, Section 4. Results obtained and discussion. The last, section 5 concludes our work and suggests future research scope.

2. LITERATURE REVIEW

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.

Research by urban activist Jane Jacobs (2012) emphasizes that natural surveillance, i.e., the presence of high density of visitors and high diversity enhance the safety of the target area and in effect reduces crime. With the help of data-driven and place-centric approach, it is possible to determine whether a particular geographic area can be identified as a future crime area proposed by X. Wang et al. (2012). Crime models based on spatial analysis created from newsfeed proposed by Jayaweera et al. (2015) can help us in understanding criminal activity better. Also, this anonymized data has the advantage of limited to no privacy risks. Combining anonymized and aggregated user data with demographics can help in identifying whether specific locations are more prone to future crimes or not.

Other studies include Chaolun Xia et al. (2014) that proposes the development of systems that use hyper-local social media data to create a real-time visualization for crimes in cities. This visualization data is sourced from various platforms like Fourspace, Twitter, and Instagram. However, there are challenges in forming robust hyper-local event detection because of the scale and noise of social media data. Irrelevant content can dilute the information available to the system. Along with the screening, real-time updates are required in applications for users such as journalists and city officials. Spatial, Meta, Textual, and Historical (SMTH) Features are extracted from the real-time data. Twitter data has been a focal point in research in which the DataStream is used to detect events by grouping tweets into clusters.

According to research from M. Rospocher et al. (2016), European project news reader processes news-stream in 4 languages and extracts relevant data such as what happened, who was involved, when and where it happened, etc. taking sources from various news sources, the system compares information for contradictions and correlations. A study was done on J. H. Ratcliffe et al. (2006) population have used spatial and temporal patterns of historical crime to identify clusters of criminal activity. Research done by Pease (1998) indicates that clusters of crimes occur due to the encouragement from the success of previous attempts in the surroundings. Kennedy et al. (2011) also found that crime clusters contribute to socio-economic consequences such as depreciating housing prices, increased fear, etc.

K. R. Rahem et al. (2014) identified that crime happens in specific areas due to attractors such as drug supply, Ellen G. Cohn (1990) founds that cold season gives rise to violent crimes and summer season has high occurrences of non-violent crime. Data-driven crime analysis studies done in London have obtained nearly 70% accuracy in identifying whether a specific locality can be a crime hotspot or not. S. Chainey et al. (2008) also found that crime is concentrated at micro levels of geography. Clustering of crime in specific geographical areas does not align with the trends in the broader community.

When compared to using rich statistical data such as household census, demographics, language, employment, etc., usage of human behavioral information significantly improves the prediction accuracy proposed by X. Chen et al. (2015). News content analysis is another area of research of K. Kaur and V. Gupta (2012) in which semi-structured information from news articles are used to perform sentiment analysis, data mining, prediction of societal issues, etc. Interesting Patterns Mining is one of the popular tasks in news content analysis. Using NER (Named Entity Recognition), it is possible to extract location data and identify whether they are a crime location or not proposed by S. Huang et al. (2011). By combining temporal proximity, event similarity, and temporal relationships, identified it is possible to analyze the relationship between different events based on Ratcliffe, J et al. (2004).

Spengler et al. (2009) have developed a web-based tool that can extract qualitative information from unstructured data regarding the modus operandi of drug criminal groups. This method has revealed insights about Mexican drug cartel, their strategies to sell drugs and their evolution over the decades. Using conditional random field, Rexy Arulanandam et al. (2014) have extracted crime location from newspaper articles. The research was done by Bao et. Al (2013) utilizes Concentration Driven Model (CDM) to microscopic analysis of crime locations and Centre Based Model (CM) to macroscopically predict the hotspots of future criminal activity.

Other researchers S. Lee et al. (2008) have used NLP techniques to improve the efficiency of the system by defining keywords extracted from the news body as representative of the news articles. Matthew S. Gerber (2015) has proved Kernel Density Estimation (KDE) to perform well in predicting future crime patterns in geography compared to other hotspot mapping techniques. KDE is used to identify the density estimation of a particular activity in a region.

S. Chainey et al. (2008) Hotspot mapping is a method that is used in identifying, measuring and analyzing areas in which a particular activity is happening at high frequency. This technique is used in identifying crime patterns in geographical perspective. Xiangyu Zhao et al (2018) witnessed advanced ways in which to gather and integrate fine-grained urban, mobile, and public service information that contains varied crime-related sources also has made environmental and social info. Author has done the survey on different data mining algorithms can be used for crime analysis and prediction. Syeda Ambreen Zahra et al. (2018) put on the various utilities of the hotspot in GIS to acknowledge the crime in addition to encourage the advancement of investigation inclination strategy for policing.

. According to Messing and Westwood (2014), social media provides users with information from politically heterogeneous sources and what users consume depends on the social endorsements of the news. Some of the methods used in exploring the attributes relate to spatial and temporal attributes include hotspot detection, retrospective forecasting, spatial regression, machine learning, risk terrain analysis and near repeat concept. Alina Ristea, Ourania Kounadi and Michael Leitner (2018) used geosocial media posts to forecast hotspots of crime in Portland, Oregon. Twitter data is used for deriving the geographically weighted regression (GWR). GWR is a modeling approach for spatially heterogeneous processes. The two predictors used are crime related tweets and the population at the risk of being a victim of street crime. Geosocial media data are free and more accessible to obtain compared to authoritative data. It is found that with GWR, the prediction efficiency has improved than that of a baseline approach.

Based on the literature identified that Crime analysis is one of the emerging areas in Indian and world context using social media like Twitter, RSS News feeds, Instagram etc. Very few researchers have taken the Newsfeed as a source of data even though it's more authentic than other social media in

the present trend. In this paper, aim to do the Geospatial analysis to detect the crimes and the patterns evolving from data to efficiently identify crimes. Indian Newspaper headlines and description are identified as the source for analyzing crime data with the time span of 1st Jan 2017 to 1st Jan 2018.

2.1. Problem Identification Based on Literature

RSS feeds and social media data are considered to be good datasets for predicting crime. They are attractive as they have near real-time information with personalized content. However, using newspaper articles (RSS Feeds) as the primary source of crime prediction has many challenges. Untrained journalists write most newspaper articles; hence the information present in the articles are often incomplete and lack context. For predicting crime in a specific geographical location, it is necessary to have that information in the news articles. As there are no standards for newspaper articles, it becomes difficult to process for analytics algorithms. Despite the recent advancements in automatic text processing architectures, such as word boundary identification through semantic analysis, the research area is still an emerging one.

2.2. Novelty of Proposed Work

Automatic segregation and analysis of crime data is a new area of research. In this system, language parsing is done to extract spatial and temporal information from the news feed data. This research presents a new approach to crime hotspot identification in which crime hotspot locations, point assignments and other factors of interest are parameterized for further analysis. The system is able to predict both spatial and temporal aspects of the street level crime by incorporating predictive elements such as localization of distinct types of crime to geographic regions, tracking regional trajectories and assessing likelihood of the type of crime occurring in a particular region.

3. METHODOLOGY AND IMPLEMENTATION

3.1. Methodology

The research aims to create a framework for crime analysis through News feeds data. The research design includes the use of data mining to automate the extractive of predictive insights of space and time from the criminal records. The research done in this paper aims to use spatial analysis to help detect the crimes and the patterns evolving from data to efficiently identify crimes. Newspaper headlines and description of news are identified as the source for analyzing crime data.

The research in this paper focuses on the development of an Information Technology architecture that collects different RSS Newsfeeds from the internet. The feeds are then filtered for relevancy using criteria such as location, type of crime, etc. The RSS feed data is converted into a machine-readable format and then preprocessing is done. This step converts the raw data into rich information. With the help of this step, real-world data which is incomplete, inconsistent and lacking in trends or behavior can be processed. In these preprocessing steps, the data is cleaned, and features are extracted. The system then classifies the data based on the feature set. The analysis is then converted to visual hot-spots for easy understanding of patterns.

The architecture aims to find the data distribution and the pattern information. Our paper attempts to combine computer algorithms with statistical methods and introduce statistical approaches such as factor analysis, analysis of functional data into textual data mining. Data mining is the ideal technique to process massive crime dataset extracting useful information to support the police in crime analysis. The Flowchart of the proposed methodology is illustrated in Figure 1 which explains in seven stages of work.

In the first and second levels, the architecture paper has used RSS feeds for collecting crime-related news from different newsfeeds from 2017 Jan to 2018 Jan (27782 News related to crime). RSS known as Rich Site summary is a type of website format that allows users to get updates from

a particular website in a standard form. The third level the standard preprocessing algorithm is used for cleaning the data and converted it into a data frame. Preprocessing is done using Stop words, Lemmatizing, POS, and Stemming (Remove noisy data and cleaning process) using tools such as plyr & diplyr in R Language. A standard database is used to store the processed data (.rdat). The database can retrieve data with the help of a query language. The data is stored in a file .csv format.

The fourth level collected data is then sent for post-processing in which the date and description of news are extracted and cleaned further and converted into hash code. It translates the headline and details about the parent agency, in a unique hash code to reduce redundancy among data. Fifth level the preprocessing part, all the information is collected such as exact location, crime happening, etc. The process is repeated until sufficient data is processed. The data is then classified into 16 types of crimes like Arson, Assault, Burglary, Drunkenness, Fraud, Gambling, Harassment, Hurt, Kill, Molested, Murder, Robbery, Suicide, Trespass, Vandalism, Warrants. The algorithm has associated all the synonyms to conclude the crimes in 16 types as per the standard dictionary names. This is done using Lemmatization and Stemming process, which can get the familiar dictionary words advanced data analysis is done using Kernel density estimation and ARIMA time series model. Finally, in the seventh level paper presents the visualization of crime with respect places and type of crime.

3.2 Implementation of system

Implementation started with the RSS Feeds and then, as you can see in the below flow chart, divided the whole process into three parts.

Figure 2 Shows the system architecture of the crime Analysis software is divided into two parts: Part 1. Data mining, cleaning, and exploratory data analytics and Part 2. Preprocessing and classification.

3.2.1. Data Mining, Cleaning, and Exploratory Data Analytics

Figure 3 explains the RSS feeds from various websites are collected and then sent to an RSS mining algorithm. The algorithm then creates a hash code to remove duplicates in the data. The data is merged with historical data, and the duplicates are removed once again. Information scrapping of time, location, and crime types are done on the collected data. It is then stored in a database. Text extraction from news feeds (XML format) is done using an XML Parser. Preprocessing of the news feeds by using Stop words, Lemmatizing, POS, and Stemming (Remove noisy data and cleaning process) is done by tools such as plyr & diplyr in R language. The data is then stored in a local server (Excel) .CSV format.

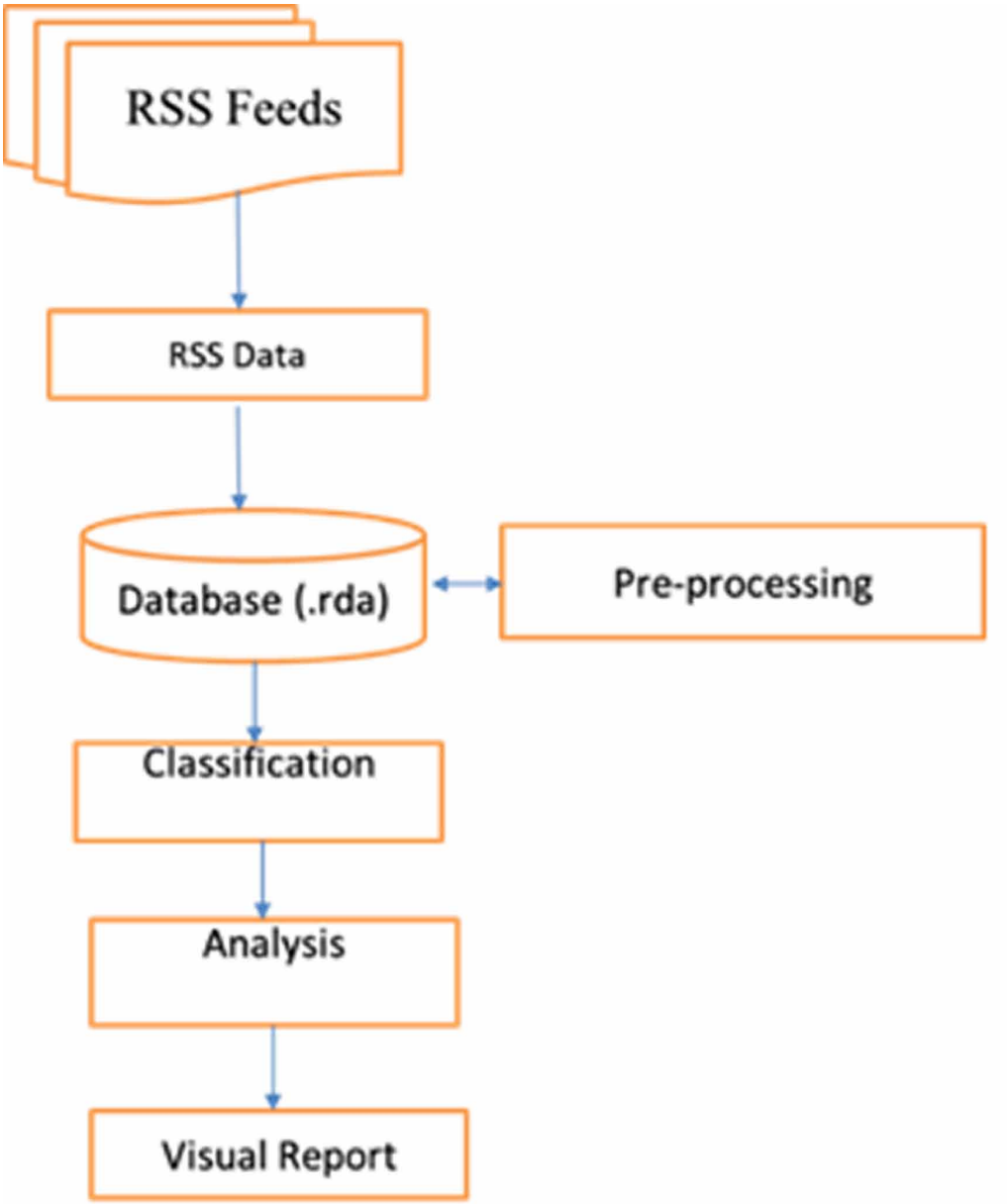
3.2.2. Preprocessing and Classification Part

Figure 4 Post processing (Duplicate detection can be done with hash code) is done using digest package. It is then checked whether the data is sufficient for analysis (minimum 3000 Trial and error). Then locating hot spots is done using Ggmap. The classification model is prepared for processing the data. The extracted data is then stored in both csv and xml format and sent to the database. Bayesian classification model is used for classifying the Crime related news.

4. RESULTS AND DISCUSSION

The research has produced the output in the form of hotspots of crime. Cartographic work and crime maps for various geographic regions of Bangalore and India are created as the results. Comparative analysis of the shift in criminal activity over a period of one year (2017 Jan to 2018 Jan) is calculated. It also shows the clustering of various crime types with respect to commercial areas. The user can choose the crime type from the drop-down menu and get the corresponding visualization. The results can be utilized for the recommendation of police action for analysis of crime in geographic locations

Figure 1. Proposed architecture framework



identified clusters. It is observed from the experimental results that the model can perform with reasonable accuracy. The visual output of the results is as follows.

Figure 2. Flow of data

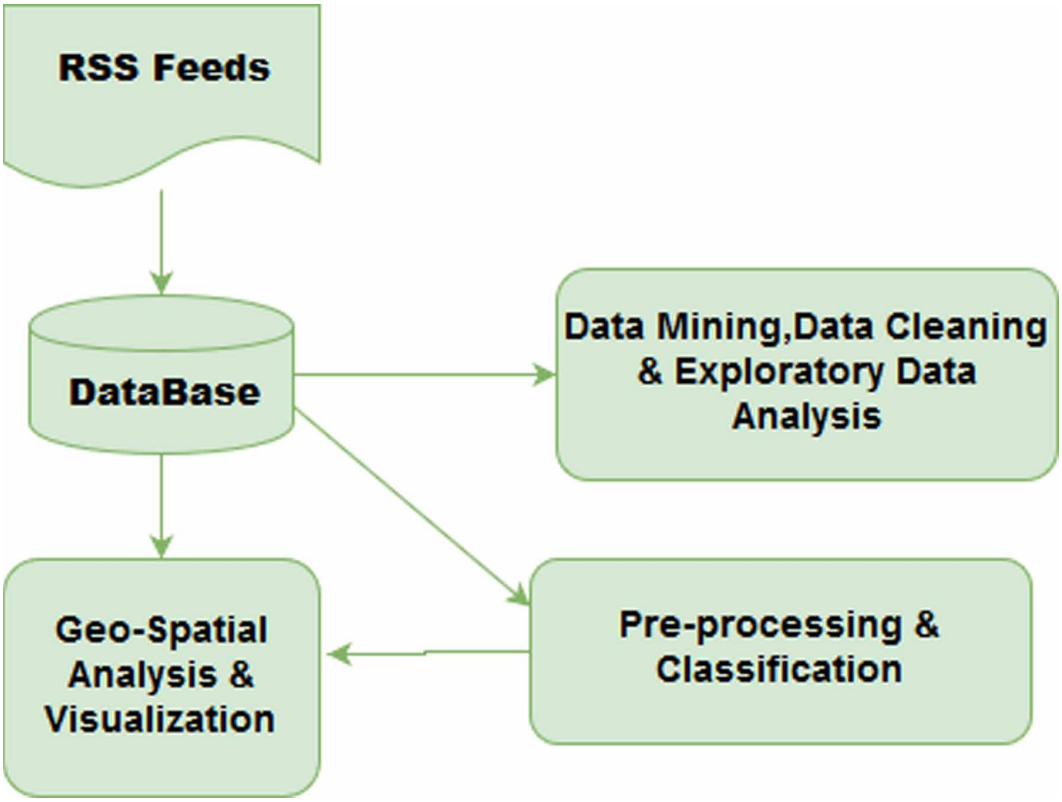


Figure 3. Data mining, cleaning, and exploratory data analytics

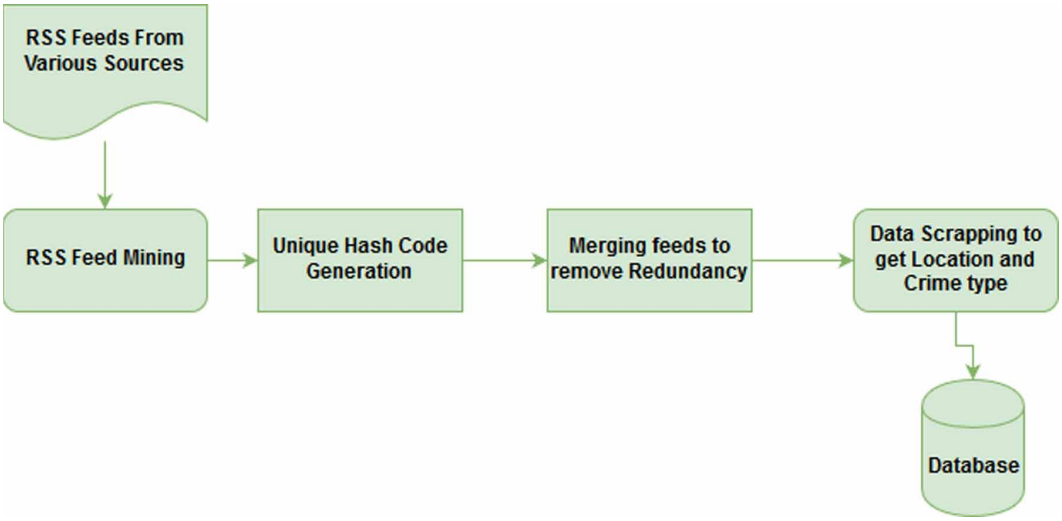
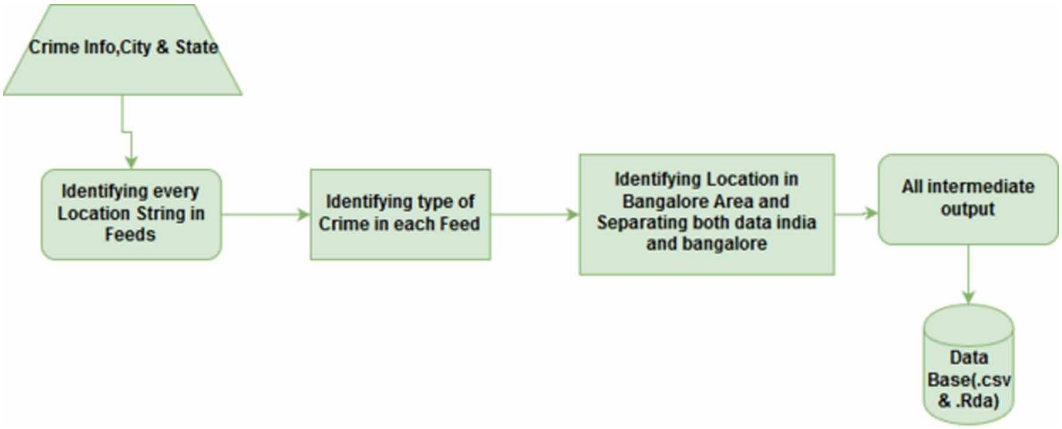


Figure 4. Preprocessing and classification



4.1. Data Preparation (Indian Crime Data)

The crime related data from newspaper articles related to all the cities in India have been collected as shown in Figure 5. Mining Data having the Fields like s.no, Description, Date, Link, Hashcode, webSourceTitle, Crime, State, latitude, longitude. Hashcode generation is also done in this case. Total 27782 News, mined related to crime in the Indian context.

4.2. Data Preparation (Bangalore Crime data)

Figure 6 having a fields like s.no, Description, Date, Link, Hash, webSourceTitle, Crime, State, latitude, longitude. Hash code generation is also done in this case. With the help of spatial algorithms, the data pertaining to Bangalore is selected and the data related to other locations are filtered out. In experiment Filtered News associated with crimes 27782 to 472 on Bangalore context. No-crime field has taken as missing values and removed from database.

Figure 5. Data scrapping from different news feeds Indian context

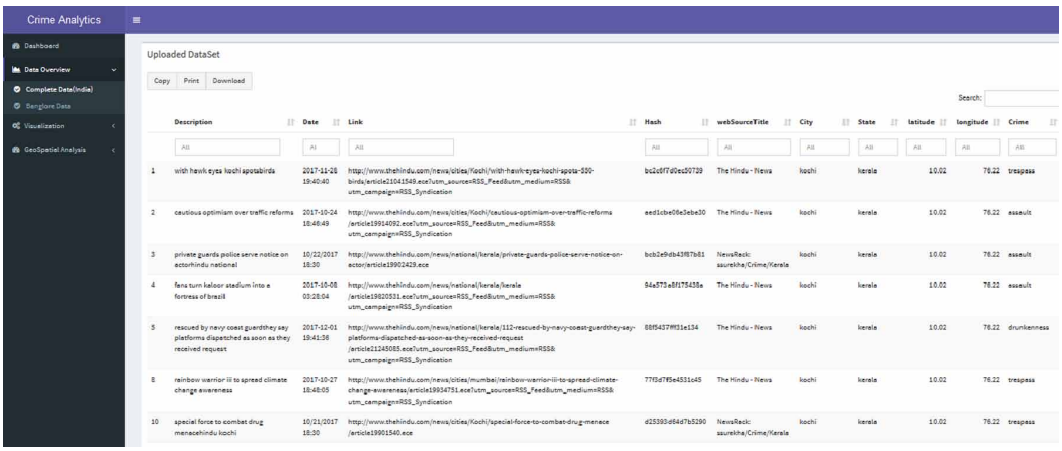
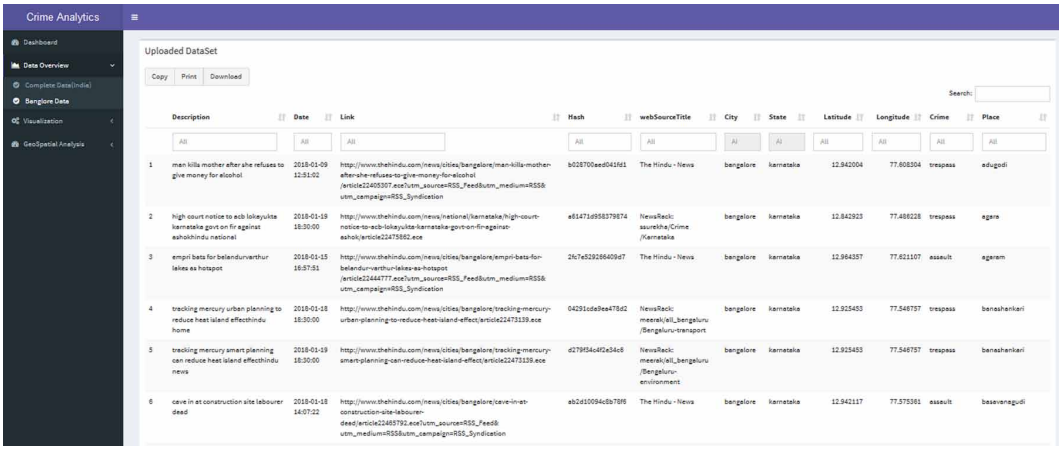


Figure 6. Data scrapping from different news feeds Bangalore context



4.3 Visualization of Crime Anlaysis

4.3.1 India Crime Statistics

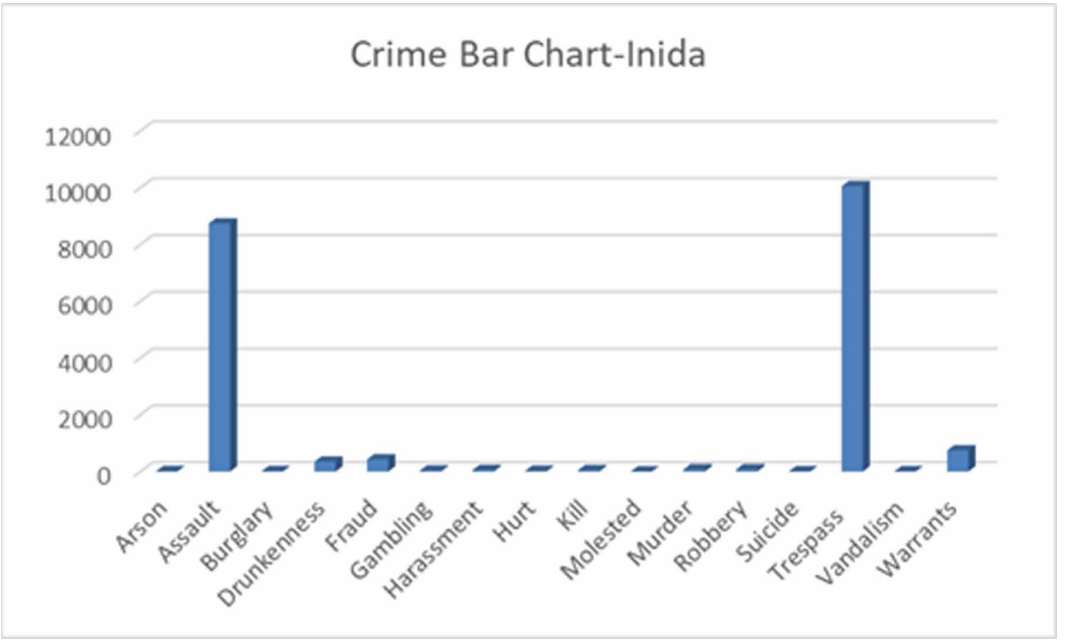
Table 1 gives the statistics from 2017 Jan to 2018 Jan with 16 types of crimes in Indian context.

Figure 7 shows the Bar-chart of 16 types of crimes in Indian context. According to the result Assault (8744) and Trespass (10054) more in number of news identified. Figure 8 shows the processed data of India with classification done on the various crime types in the form of a Pie chart. According

Table 1. Crime types India statistics

S. No.	Crime	Count
1.	Arson	15
2.	Assault	8744
3.	Burglary	18
4.	Drunkenness	351
5.	Fraud	433
6.	Gambling	38
7.	Harassment	55
8.	Hurt	32
9.	Kill	51
10.	Molested	8
11.	Murder	78
12.	Robbery	82
13.	Suicide	9
14.	Trespass	10054
15.	Vandalism	9
16.	Warrants	748

Figure 7. Bar-chart of crimes India



to the result more assault (42%) and trespass (49%) is the first two high level of crime is recorded in newsfeeds in Indian context.

Figure 9 shows the hotspot mapping of different types of crime at the country level is depicted. With the help of this visualization, the crime density can be identified easily. Based on retrieval of crime information from different newsfeeds it shows the density of crime events in all states India.

Figure 8. Pie-chart of crimes India

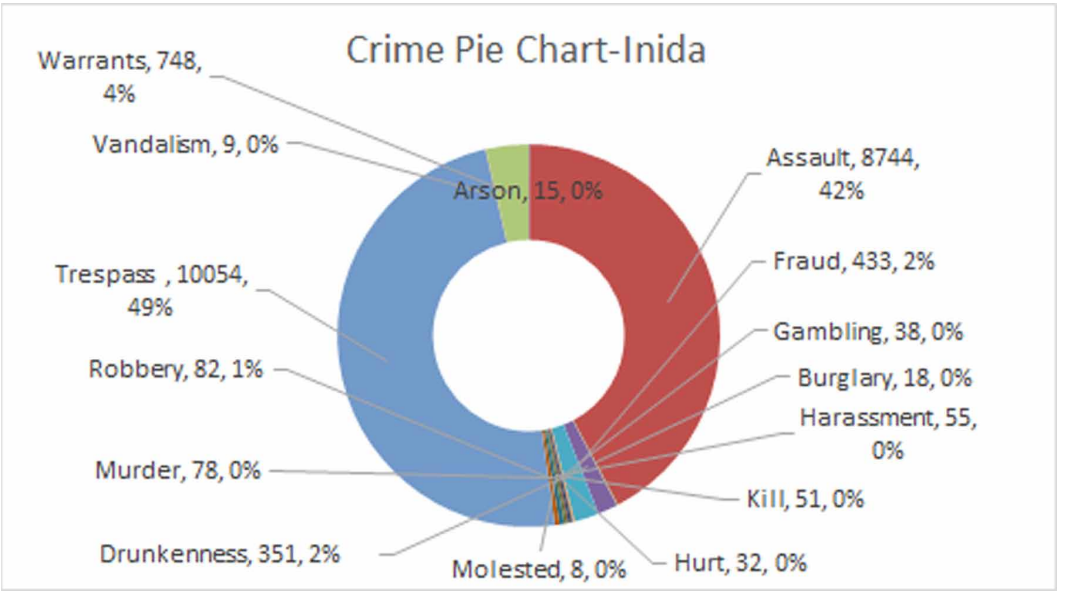
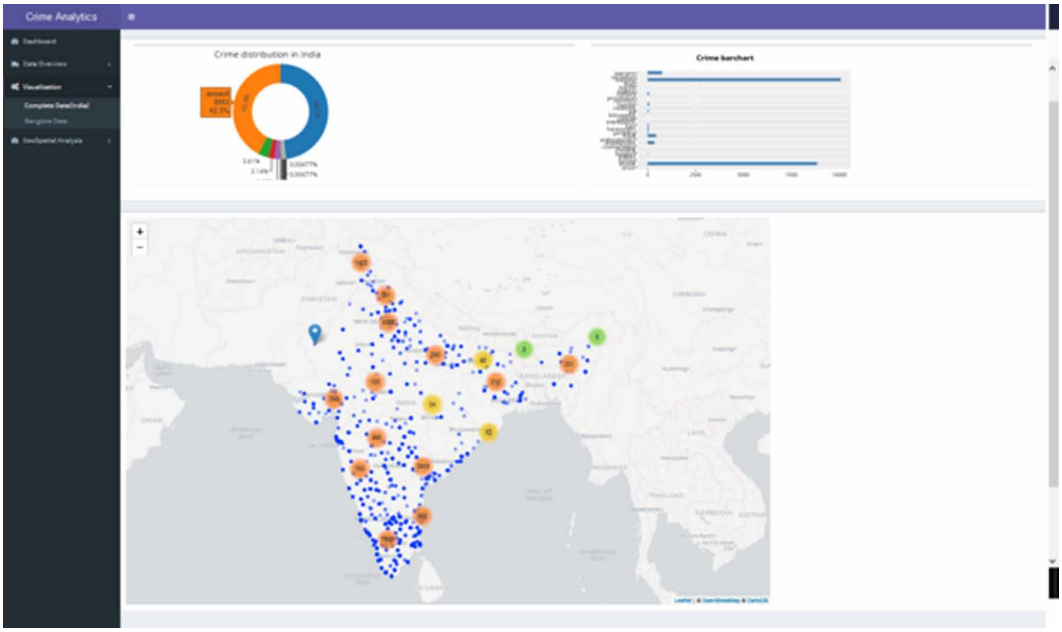
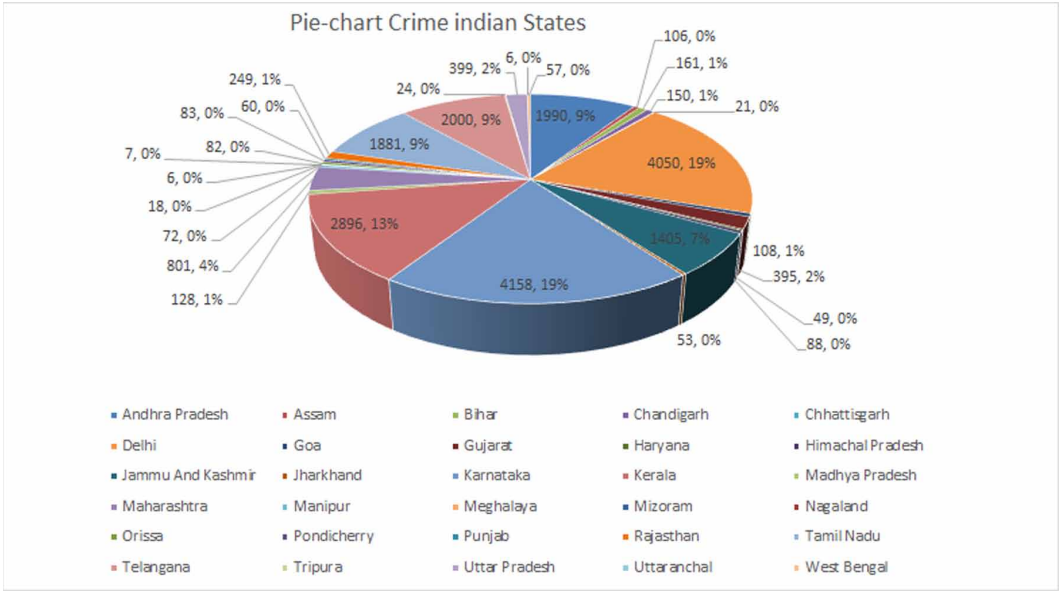


Figure 9. Geo graphical locations crimes India



According to our retrieval of data from news feeds crime rate has identified in different states in India identified as follows and the data shown here crimes with the timespan of 1st Jan2017 to 1st Jan 2018. Figure 10 shows the crime statistics in indian states. According to the statistics Andhra Pradesh- 1990, Delhi-4050, Jammu And Kashmir-1405, Karnataka-4158, Kerala-2896,Tamil Nadu-1881,Telangana-2000 etc.

Figure 10. Crime rate India states



4.3.2 Bangalore Crime Statistics

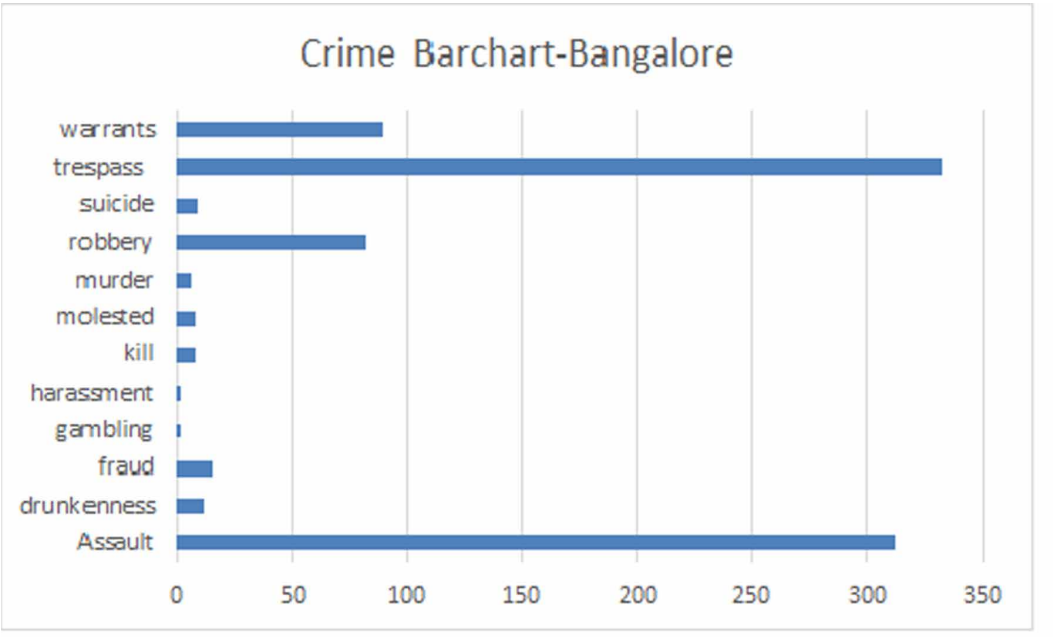
Table 2 gives the statistics from 2017 Jan to 2018 Jan with 14 types of crimes in Bengaluru context. Here identified 14 types of crimes in different Geographical locations in Bangalore.

Figure 11 shows the processed data of Bangalore with classification done on the various crime types in the form of a bar chart. According to the result identified the news related to crime as Assault

Table 2. Crime types India-Bangalore statistics

S. No.	Crime	Count
1.	Assault	312
3.	drunkenness	12
4.	fraud	16
5.	gambling	2
6.	harassment	2
8.	kill	8
9.	molested	8
10.	murder	6
11.	robbery	82
12.	suicide	9
13.	trespass	333
14.	warrants	90

Figure 11. Bar-chart of crimes Bangalore



(312), Drunkenness (12), Fraud (16), Gambling (2), Harassment (2), Kill (8), Molested (8), Murder (6), Robbery (82), Suicide (9), Trespass (333), Warrants (90). Based on this analysis Assault and Trespass more in number identified.

Figure 12 shows the processed data of Bangalore with classification done on the various crime types in the form of a pie chart. According to the result more assault (36%) and trespass (38%) is the first two high level of crime is recorded in newsfeeds in Bangalore context.

Figure 13 shows the hotspot mapping of different types of crimes at Bangalore is depicted. With the help of this visualization, the crime density can be identified easily. According to the result highest

Figure 12. Pie-chart of crimes Bangalore

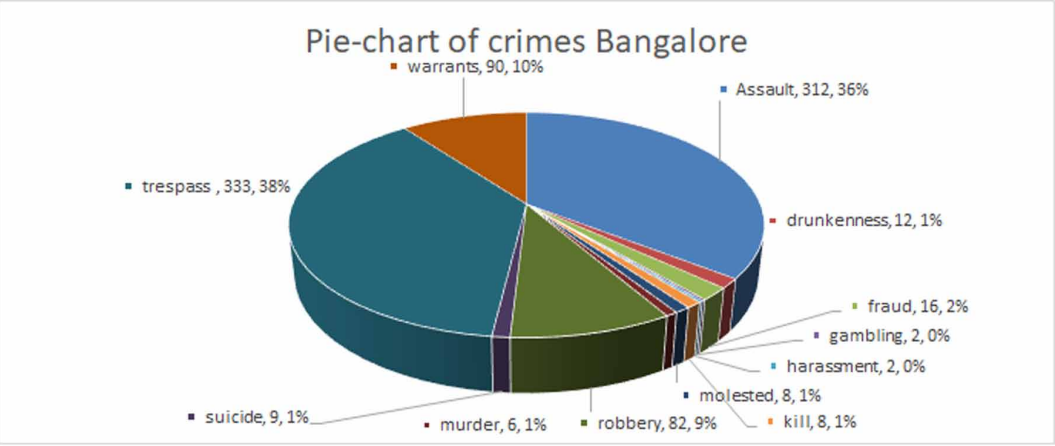
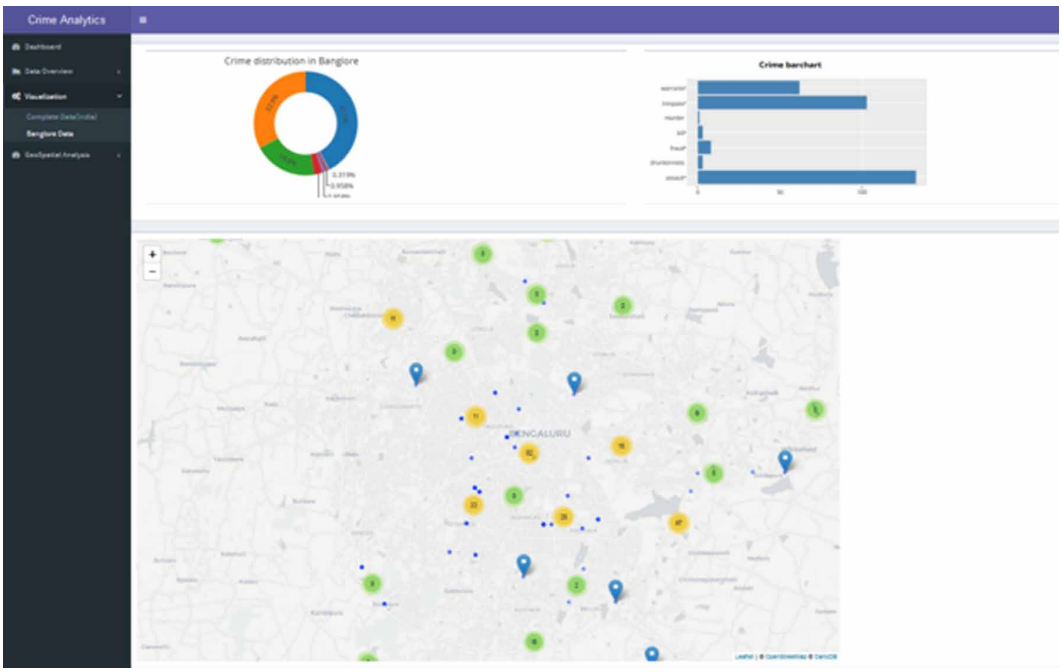


Figure 13. Geo spatial crime locations Bangalore



rate of crimes in Corporation Circle (Lat-12.960779, Lon-77.58424) and Ballandur (Lat-12.926031, Lon-77.676246) are more density of crime identified. Based on our result identified the crime hotspots and density of crime based on Geo location by using web-based techniques like RSS feed mining.

5. CONCLUSION

The proposed system can perform crime analysis operations such as hotspot detection, crime comparison and crime pattern visualization using R language. Before this research, use of RSS news feed data for crime Analysis and prediction is relatively unknown. Also, the combination of textual analysis with spatial and temporal information for crime analysis and prediction is a relatively new area that needs exploration based on web-based learning techniques. There are challenges in identifying and classifying the different types of crime. The performance of the analytics algorithms with standard hotspots has not been studied intensely. Our results have shown that news feed data with spatiotemporal information can increase the analysis and prediction performance for 16 types of crimes in Indian and Bangalore context. This indicates a potential to enhance the efficiency of police by better allocation of scarce resources such as police patrols. This application leads to a reduction in effort and improvement in crime response rates.

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

Xia, C., Schwartz, R., Xie, K., Krebs, A., Langdon, A., Ting, J., & Naaman, M. (2014, April). CityBeat: real-time social media visualization of hyper-local city data. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 167-170). ACM.

Shahaf, D., & Guestrin, C. (2010, July). Connecting the dots between news articles. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 623-632). ACM. doi:10.1145/1835804.1835884

Das, P., & Das, A. K. (2018). *Application of Classification Techniques for Prediction and Analysis of Crime in India. In Advances in Intelligent Systems and Computing* (Vol. 711). Springer.

Cohn, E. G. (1990, January 1). Weather And Crime. *British Journal of Criminology*, *30*(1), 51–64. doi:10.1093/oxfordjournals.bjc.a047980

Ratcliffe, J. H. (2006). A temporal constraint theory to explain opportunity-based spatial offending patterns. *Journal of Research in Crime and Delinquency*, *43*(3), 261–291. doi:10.1177/0022427806286566

Jayaweera, I., Sajeewa, C., Liyanage, S., Wijewardane, T., Perera, I., & Wijayasiri, A. (2015, April). Crime analytics: Analysis of crimes through newspaper articles. In *2015 Moratuwa Engineering Research Conference (MERCon)* (pp. 277-282). IEEE.

Kaur, K., & Gupta, V. (2012). A survey of topic tracking techniques. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*, 384–393.

Rahem, K. R., & Omar, N. (2014). Drug-related crime information extraction and analysis. In *Proceedings of the 6th International Conference on Information Technology and Multimedia*, Putrajaya, India (pp. 250-254). doi:10.1109/ICIMU.2014.7066639

Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology*, *27*(3), 339–362. doi:10.1007/s10940-010-9126-2

Gerber, M. S. (2014, May). Predicting Crime Using Twitter and Kernel Density. *Journal of Decision Support Systems(Elsevier)*, *61*, 115–125. doi:10.1016/j.dss.2014.02.003

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, *41*(8), 1042–1063. doi:10.1177/0093650212466406

Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., & Bogaard, T. et al. (2016). Building event-centric knowledge graphs from news. *Journal of Web Semantics*, *37*, 132–151.

NewsReader Grant agreement no.: 316404 – Co-funded by the EU - FP7 Work Programme Research theme 'Information and Communication Technologies.' (n.d.). Retrieved from http://www.newsreader-project.eu/

Ratcliffe, J. H. (2004). The hotspot matrix: A framework for the spatiotemporal targeting of crime reduction. *Police Practice and Research*, *5*(1), 5–23. doi:10.1080/1561426042000191305

Fletcher, R., & Nielsen, R. K. (2018). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. doi:10.1177/1461444817724170

Ristea, A., Kounadi, O., & Leitner, M. (2018). Geosocial Media Data as Predictors in a GWR Application to Forecast Crime Hotspots (Short Paper). In LIPIcs-Leibniz International Proceedings in Informatics (Vol. 114). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, *21*(1-2), 4–28. doi:10.1057/palgrave.sj.8350066

Spengler, A., & Gallinari, P. (2009, May). Learning to extract content from news webpages. In *2009 International Conference on Advanced Information Networking and Applications Workshops* (pp. 709-714). IEEE.

Bao, P., Shen, H. W., Chen, W., & Cheng, X. Q. (2013). Cumulative effect in information diffusion: Empirical study on a microblogging network. *PLoS One*, *8*(10), e76027.

Huang, S., Peng, X., Niu, Z., & Wang, K. (2011). News topic detection based on hierarchical clustering and named entity. In *7th IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)* (pp. 280-284). doi:10.1109/NLPKE.2011.6138209

Lee, S., & Kim, H.-J. (2008). News keyword extraction for topic tracking. In *Fourth IEEE International Conference on Networked Computing and Advanced Information Management, NCM'08* (pp. 554-559).

Chen, X., Cho, Y., & Jang, S. Y. (2015). Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*, Charlottesville, VA (pp. 63-68). doi:10.1109/SIEDS.2015.7117012

Wang, X., Brown, D., & Gerber, M. (2012). Spatiotemporal modeling of criminal incidents using geographic, demographic, and Twitter-derived information. In *Intelligence and Security Informatics*. IEEE Press.

Zhao, X., & Tang, J. (2018, May). Crime in Urban Areas: A Data Mining Perspective. *SIGKDD Explorations*, *20*(1), 1–12. doi:10.1145/3229329.3229331

Zahra, S. A. (2018). Crime Mapping in GIS by Using Hotspot. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *2*(1), 13–19.

Prathap Rudra Boppuru is an Assistant Professor and 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 is a Professor in Dr. Ambedkar Institute of Technology Bangalore-India. He has completed his PhD in JNTU Hyderabad-India in 2013. His research interests include machine learning, image processing, VLSI, digital signal processing, and communication. He had presented research papers 7 international conferences and published more than 20 research papers in reputed international journals in the area of Image processing and digital signal processing. Currently, he is working on the machine learning algorithms for crime detection and prevention.