An Overview of Heuristic Based Crime Prediction and Analysis Using Social Media Data

[1]Boppuru Rudra Prathap, [2]Ramehsa K
[1]Research Scholar, [2]Professor
[1]Computer Science and Engineering,
[1]CHRIST(Deemed to be University), Bangalore, India

Abstract: Social media has grown to become one amongst the foremost powerful communication channels in human history and this is where individuals are sharing their perspectives, thoughts, suppositions, and feelings. The adoption of social media data for crime analysis is increasing day by day. In recent days, the collection of crime data from different heterogeneous sources becomes a primary step for the crime analysis and prediction. In this paper, a survey on Heuristic algorithms on different types of crime attributes considered for analysis and various techniques are used for prediction and analysis. Also various sources of social media used for analysis and prediction are also reviewed in detail. This media information can be considered as a one of the prominent asset, and we can use for crime investigation analysis and prediction and also we had identified the working model of a particular algorithm, research gaps in algorithms, Research challenges for crime analysis and prediction.

IndexTerms **- Crime, Crime Analysis, Crime Prediction, Prediction, Social Media, Predictive Analysis**

I. INTRODUCTION

Social media is a platform where people are sharing their views, ideas, news related to the event happened in particular geographical area. Crime analysis is one of the most crucial activities in solving the criminal cases. Modern technology and other advancements have enabled analysis of a large number of crimes. Unregulated migration and population growth have contributed to the high magnitude of crimes happening in cities. For many days researchers in the field of crime analysis and prediction used toward the justification of crime and public safety. Intelligence organizations and law enforcement organizations collect large amounts of crime data to predict the future occurrence. Since this involves analyzing a large number 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.

Several studies showed that crime analysis and forecasting could do with social media data. Such studies can apply for crime prediction and public safety in a particular geographic location. In addition, the rapidly advancing technologies can help address such issues. However, crime patterns are changing and proliferating. In this survey paper, we are discussing various crime analysis techniques and type of social media used by researchers. Few researchers had done the analysis based on models of the crimes happened in particular geographic location. In addition to that few researchers used the geographic location for identifying the crime prediction and based on this police resources can be allotted according to the crime type. In this paper we are mainly discussing about types of types of data sources what they had used for the analysis and prediction.

II. SOCIAL MEDIA USAGE FOR CRIME ANALYSIS

In this section we are going to discuss about different social media and algorithms usages for crime analysis and prediction based on different source of social media data. Table-1 and Table-2 shows the observations of different researchers involved in the crime analysis and prediction.

A. News feeds:

In this section, we are discussing the different researchers used news feeds as a source for crime analysis. I. Jayaweera et al.[1] focused mainly focused on the crime analysis through newspaper articles from different sources like Daily mirror, The Island, and Ceylon Today. Author extracted crime-related news from sources of the newspapers by using focused crawler, and then the news is classified by using SVM based classifier. Crime entries are extracted and duplicate detection is identified. By using pre-processed and classified data, crime analysis operation has been performed and the result has been shown in web-based GUI.

Piek Vossen [6] Author summarises about news items that is revealed in Europe. Daily news streams in four languages, extracting what happened, when, wherever and WHO was concerned. It compares the results across sources like Media Monitor, Yahoo Finance, Google News, Google Finance, Google Trends, and Reuters. what is more, it merges news of today with previous news, making a semipermanent history instead of separate events.

Roya Hassanian-Esfahani [16] has experimented with four main steps. The steps as a news retrieval, content analysis, and visualization. The author had used the different newspapers for analysis. This article thoroughly explained about various methods available how newspaper headlines can be retrieved and can use for analysis purpose.

Khmael Rakm Rahem et al. [30] used the news articles as a source of data. Author extracted the drug crime information from online newspapers. Based on this extraction the classification as follows: where the drug traffickers hide the drugs, nationalities of drug dealers, type of drug, and quality and market price of that drug.

Shiju Sathyadevan [35] has taken after strides in doing Crime Analysis: 1) Data gathering (gathers the information from various sources) 2.Classification (Naive Bayes Classifier: in Algorithm creator made a model via preparing Crime information identified with vandalism, kill, theft, thievery, sex mishandle, group assault, illegal conflagration, outfitted burglary, outrageous overpricing, grabbing and so on 3) Pattern Identification (Apriori calculation) 4) Prediction (choice tree idea). 5) Visualization. Creator utilizes the News destinations, sites, online networking, and RSS channels as a wellspring of information.

Masnizah Mohd and Nazlena Mohamad Ali [49] explains about news extraction system for Malaysian context. The author created a framework for crime news retrieval system (i-JEN).The author identified different crime based events and investigate on that events by using classification and clustering algorithms and finally developed the interactive system for visualization to evaluate the usability and system performance. The author used the Bernama news agency to get the data related to crime news.

Chung-Hsien Yu et all [46] has done the examination on computerized histories taken from twenty two years of news reports from the New York Times. He had made a corpus and done the examination and demonstrated on generous increments in the likelihood of sickness episodes, passing, and mobs ahead of time of the event of these occasions on the planet. Creator had portrayed how we can figure out how to anticipate the future in view of arrangement of detailed news occasions considering the illustrations. This paper utilized the Clustering and Entropy techniques as approach for investigation and expectation.

Khmael Rakm Rahem [29] Author work aims to extract offered drug crime info from on-line newspaper articles. Author work has the subsequent subtasks: assess wherever and the way drug traffickers hide drugs, determine the nationalities of drug dealers, identify the kinds (names) of drugs, and assess the amount and costs of drugs within the local market.

B. Twitter

Matthew S. Gerber [2] articles mainly concentrates on the prediction of crime throw twitter tweets by considering the data from Chicago crime data and identified 25 types of crimes. The author did this using the official Twitter Streaming API and defining the coordinates [-87.94011, 41.64454] (lower-left corner) and [-87.52413,42.02303] (upper-right corner) to collect the tweets from different users. The author used KDE (Kernel density estimation) method for prediction of crime. This methodology can be utilized for the analyst to fast visualization and identify the locations of crime rates. Author published software as open source code name of asymmetric threat tracker.

Anthony J. Corso et al. [3] explains the predictions mechanism of crime by using NLP combined with data mining techniques. Author has done extensive research on data and their proficiencies when pre-processing social media's noisy knowledge, government knowledge, and ambiguously collected subject knowledge to be used in a predictive GIS artifact. To implement this artifact design author has done the data collection and discussion result embedded as an investigative study. Finally, author proved the results indicates the connection between social media and domain-specific datasets exits.

Maximilian Walther and Michael Kaisser [7] explains about two phases. The first phase is algorithm on geospatial event detection which observes about all tweets about Twitter issued in geolocation. In the second phase of the article explain about the analysis of resulted spatiotemporal clustered post with machine learning component, in order to detect whether its consist the real world event or not. Finally, the detected events displayed on the map with a location where they happen.

Tony H. Grubesic [9] Propose the usage of statistical measures for identifying and comparing the spatiotemporal based information which is related to robbery, burglary, and assault. For this study author used the Real-time social media data from Twitter as a data source.

Mingjun Wang [13] article focuses on two problems First investigate the prediction of social media users' spatial trajectories. Second, authors experimented the correlation between the occurrence of crimes in next-place prediction with prediction mechanism.The author could able to predict crime rate main cities in the United States. For research author used the Twitter real-time data as a data source.

Nikhil Dhavase and A. M. Bagade [27] article proposes Geoparsing which identifies the location in the text automatically. In this experiment, an author uses the real-time data from Twitter. He collects the Twitter data related to crisis situation and location information. By extracting of location information like street address and building can be helpful to identify the exact place of an event, this has been implemented with NLP Methods.

Fabrizio Albertetti [42] proposed a mechanized strategy for extensive information Crime linkage, in view of a fluffy different criteria basic leadership way to deal with recognize the situational, behavioural and measurable data. Trials are directed with arrangement in robberies from genuine information and contrasted with master comes about.

Xiaofeng Wang et al. [51] Explains regarding the investigation of Twitter-based criminal incident prediction. The author principally targeting the tactic that will the automated linguistics analysis and understanding of linguistic communication Twitter posts, combined with spatiality reduction via latent Dirichlet allocation and prediction via linear modelling. The author tested our model on the task of predicting future hit-and-run crimes. analysis results indicate that the model well outperforms a baseline model that predicts hit-and-run incidents uniformly across all days.

Nyalleng Moorosi and Vukosi Marivate [23] In this paper discusses concerning privacy problems associated with mining South African crime and public safety incidents from social media posts. The paper touches on matters associated with possession of social media data, privacy preservation challenges once many sorts of knowledge from totally different sources is integrated further as legal protection of the process of non-public info.

C. Mobile data

Andrey Bogomolov and Bruno Lepri [4] experimented on one month anonymised mobile data with demographic information to predict whether a particular neighborhood in London will be a crime hotspot or not. Author has implemented a model by using

random forest algorithm, and he achieved 68.37 % accuracy on only in mobile plus demographic information, and 69.54% accuracy when adding census data.

Andrey Bogomolov et all [10] present a new approach to predict crime in a geo-location based on specific mobile phone and demographic data. The primary work of author lies on human behavioural data gathered from mobile network activity. This activity can be used to tackle the crime prediction problems. The author considered the police data from London for experimental analysis, and he obtains an accurateness of almost 70% when predicting particular area in the city will be a crime hotspot or not.

Izyana Ariffin et all [28] done a comparative study on following six mobile applications which use crime crowdsourcing apps to report crime-related information: 1. Enforce Crime Map 2. Crime Watch Mobile 3.Community against Crime 4. Malaysia Crime 5. Community Alert. 6.MyDistress. The author examined these apps based on the seven criteria points list as follows: map viewing, list of crimes, sharing of incidents, reports of authority, sharing data in media, tutorial and finally system support. Based on all application author concluded that all apps mainly useful for community alert.

Teddy Mantoro [27] mainly explains the crime information assistance or accident information system through a mobile app. The author proposes a framework for crime assistance to help victim and police. By using this framework, the victim can send the crime-related information like location, where the accident happened by using Google API. This mobile also capable of sending and receiving of information about crime incidents and can inform nearest police station automatically. For this work author used a built-in database with the combination of Google map APIs, which will allow the police to find the location of accident or crime.

Teddy Mantoro[26] Author proposed Mobile app on Crime data help or accident system and that they proposes a framework of mobile crime data help to assist the users (victim) from locus (location wherever accident or crime happened) with location aware capabilities. For their analysis author used the Google API.

Izyana Ariffin[28] during this paper author measure six mobile applications that use crowdsourcing to report crime connected incidents. The analysis is finished supported seven criteria's. These embody map viewing, crime listing, incident sharing, authority coverage, media sharing, tutorial and supported system.

D. Instagram

Ke Xie et al. [8] proposed a method for event detection from Instagram photos using two mathematical steps. The first step using the time series data to analyze and detect the irregular signals in small regions. After identifying the irregular signals use a classifier to decide if the identified activity belongs to an actual event or not. In this experiment, the author has collected the 906235 photos from Instagram by giving geo region boundaries. As a second step author uses the Gaussian process Regression to identify the time series prediction, candidate event classification for spatial data.

E. Police data

Christoffer Gahlin and Erik Johanson [11] researched two aspects first, calculation of accuracy and performance of Kernal density estimation algorithm using a small data sets, and second step as a study on the amount of crime data needed to compute accurate and reliable hotspots. Author used the three geographical areas in Sweden, including Stockholm, Gothenburg and Malm¨o are analyzed and evaluated over a one year. The police department provides data for the experiment.

Shoaib Khalid et al. [12] analyzed location and time analysis for crime hot spots. Author considered the data from Faisalabad city of Pakistan Crime data from the Police department. To identify the hotspot KDE algorithm has been used and this complete research had done in 4 phases as follows. In the first phase. Data collection, Graphical positioning survey, Zone separation, digitization of data and final interview of police officials.In second phase crime report geocoding, locating on the map, density analysis of crime and analysis of network has been done. In phase three COMSTAT model is proposed based on CRIMEGEOGRFIX. In Final phase Operational Analysis, Strategic Plans, Revised Duty Rosters, New locations of Police Check Points has been observed. The result of experiment identified following crimes.1.Theft of bike and car were observed at parking place and road sides. 2. Snatching is happing in the populated areas of city which is lacking of street lights.3.Robberites were observed in the posh areas of cites.

Arunima S. Kumar and Raju K. Gopal [14] survey deal with the study of data mining based systems for analyzing crime information and thus automates the crime investigation procedure of the police officers in the Indian context. Author has explained 3 Different frameworks available in Crime analysis. 1. Regional crime analysis program (recap), 2.Data mining framework for crime pattern identification. 3. Narcotics network in tucson police department as a part of Indian initiatives. Based on above frame works National Crime Records Bureau (NCRB) proposed following system in order to cater the ever growing need of technology, NCRB keeps itself abreast of latest technology in software development. 1. Crime and Criminal Information System (CCIS) 2. Common Integrated Police Application (CIPA) 3. Talash Information System (TIS) 4. Crime and Criminal Tracking Network System (CCTNS). Author used raw data from different police stations in Indian context.

Somayeh Nezami and Ehsan Khoramshahi [18] done their research in 3 phases. As in 1st part crime knowledge associated with drug importing in Khorasan Province of the Islamic Republic of Iran is studied. The second part author examined the distribution of crimes and its relation with every of special and descriptive variables. Finally within the third part crime distribution was modelled in keeping with variables like population, distance from town, distance to the closest police office and illiteracy by exploitation Geographically Weighted Regression. Author used the Southern Khorasan Province police knowledge for experimental purpose.

Mohammad A. Tayebi et al. [34] has drawn attention on Spatial crime analysis, firstly concentrates on crime hotspots, areas of higher crime density. The author proposed the CRIMETRACER which gives the probabilistic model of spatial space related to crime. The author used the Real-time data from police records for experimental purpose.

Omowunmi E. Isafiade and Antoine B. Bagula [38] analyzed various crime classes across thirty-two locations among the Western Cape of Good Hope Province of the Republic of an African nation. Author planned the Citisafe algorithmic rule by mistreatment FP-Growth algorithmic rule. The visual image techniques presenting the crime patterns, through authors algorithmic rule, would assist enforcement agencies and public safety organizations to channelize their resources consequently to realize a targeted and efficient crime hindrance strategy. Author thought of Raw-data from Police Records as a knowledge supply.

Zhanhong Wang, Jianping Wu, Bailang Yu [50] study explained about the monthly hotspots of thefts and robberies in Shanghai in 2009 are analyzed and mapped by using the hotspot analysis tool of ArcGIS 9.3. Two types of crimes identified with spatiotemporal hotspots identified. The author used PCA(Principal component analysis) for investigating the 18 types of indicators involved in the crime distribution. The author considered the Shanghai police data for the experiment.

L. S. Thota, M. Alalyan, A. O. A. Khalid, F. Fathima, S. B. Changalasetty and M. Shiblee [51]. This paper creator demonstrated that Crime profiling and zoning can be displayed with use of information mining calculations. Creator utilized the NCRB (National Crime Records Bureau) dataset which is kept up by administration of India. Creator utilized the K-Means grouping calculation on criminal informational collection and makes a custom India outline bunch zones of the expressed. This guide indicates general Crime profiles in Indian guide with states.

H.P. Kavya, Veena Karjigi[31] This work investigates the applying of keyword recognizing to notice crime and it is used at the side of telephone sound and audio observation devices by security organization. Author worked on completely different speeches from thirty-nine phones.

F. Hardware related

Chaolun Xia et all [5] proposed CityBeat. The work is to give a city data about ongoing perception of Social media information to writers. In this article writer has particular equipment setup with cameras settled in various geo-loacitons.Author separated the photographs by utilizing cameras as information, utilizes the time arrangement examination to recognize the area and time of occasion happened. Creator utilized the SVM Algorithm to characterize the information. Creator principally considered the geo-tweeted.

G. Dataset

Nidhi Tomar and Amit Kumar Manjhvar[15] proposed ACO calculation to enhance the K-Means calculation, which recognizes the distinctive Crime groups in the ruff dataset. In this article writer considered the ruff informational collection for test reason.

Martin Lnenicka et all [19] proposed web application, which filters the chose broad communications servers for the words associated with the Crime sort and the solid region. All web wellsprings of Czech Republic had been be utilized as a wellspring of information.

H. JIŘÍ, I. IGOR [20] clarifies how the OLAP can utilize as Multidimensional information base for Crime examination. Creator for the most part clarifies the plan of multidimensional database for Crime avoidance, which contains the information from various Crime registers and sources. The measurements in the work are populace, joblessness, properties, wellbeing, offices and so on. The Fact tables in the database are connected with two shared measurements like space and time.

B.V.S. Varma and Valli Kumari [38] proposed algorithmic administer to mine boundless Crime designs from uncommon databases. The well known Crime designs with totally unique quality esteems in an exceedingly Crime data territory unit found. The framework continued with the advancement of PCrimeSpatial-tree with the help of H-table. Later decide the prevalent Crime designs from the made tree. Once the advancement of tree take the anticipated information so it is less demanding to search out the examples from the distinctive Crime episodes taken. Creator utilized ruff dataset for trial reason.

Xifan Zheng, Yang Cao, Zhiyu Ma [49] proposed acquainted a model with foresee the future serial Crime area with help of criminal geographic profiling and the time and areas of past Crime scenes. This model, produce the results of separation rot and the nearby geographic highlights into thought and define a likelihood thickness capacity without bounds serial Crime site to understand the forecast. Creator utilizes the Sample dataset to demonstrate his model work.

Alice Hutchings [21] Research examines the takedown of websites used for criminal functions. These boxed-in the achievement of money mules, faux websites, moreover as those used for phishing, malware dissemination, the sale of counterfeit and illicit product, and child regulatory offence content. Authors conducted drawn-out interviews with a selection of people actively engaged in data processor takedown, moreover as business corporations that offer specialist services, organisations targeted by criminals, uk social control and repair suppliers United Nations agency answer takedown requests.

Ubon Thongsatapornwatana [22]Paper reviews the literatures on various data mining applications, especially applications that applied to solve the crimes. Author has done the survey on Association, classification, Clustering algorithms with respect to crime data and Databases, Web Sites, Sensors has be considered for their survey.

H. Organizations

Qiang Zhang and Pingmei Yuan [17] proposed a Hot spot expectation demonstrate by utilizing spatial fleeting qualities and LDA-KNN to anticipate the elevated amounts of each sub-range, accordingly separation the Hot spots. Straight Discriminant investigation utilized as dimensionality diminishment strategy and KNN is utilized as a Classification technique. This work is upheld by National Natural Science establishment of China. This work basically break down the Crime occurrences of burglary, robbery and other usurpation in the city zone of Nanchang in the vicinity of 2014 and 2015.The aggregate length of the time arrangement of the Crime episodes directed is two months. Creator Mainly centered around vital occasions for Crime problem area and expectation like Chinese Spring Festival, Valentine's Day, Lantern Festival, Tomb Sweeping Festival, May Day, Dragon Boat Festival, Double Seventh Festival, Mid-Autumn Festival, National Day, Christmas and New Year's Day.

Mehmet Sait Vural[39] Authors motivation for this approach is that unattended approaches for crime analysis do not would like completely realistic info set therefore on develop deciding algorithms. The model depends on GIS by approximating the

characteristics of the population in real-life. Then, results of various GIS connected queries area unit incontestable on the GIS map to change the visual analysis of the incidents.

I. Email

Mugdha Sharma [33] proposes an apparatus which applies an improved Decision Tree Algorithm to identify the suspicious messages about the criminal exercises. The goal is to distinguish the suspicious criminal exercises and limit them.That's the reason the instrument is named as "Z-Crime" delineating the "Zero Crime" in the general public. Creator utilized E-Mail information as information source.

J. Telephone Phone Call information

M. Kumar, M. Hanumanthappa and T. V. S. Kumar [52] Author done the criminal examination by utilizing Mobile telephone call information and dissected the call detail record (CDR) to track the suspects/criminal for examination. Creator utilized CDR Database with the energy of diagram investigation instruments, which provides for creator solid piece of information to split the case.

K. Software's related to crime analysis and prediction

My Neighborhood [50]Update is the group entryway for neighbour's to know who is close to client and furthermore about the occasions what's going on close-by, which made by Corona Solutions. This task for the most part for law authorization group to distinguish the occasions happening and encourage the stream of data. This task enables the client to perceive what's going on and happened history in the clients range, city or nation. This venture is financed by neighbourhood police offices and sheriffs' workplaces. The concentration of the undertaking is distinguishing proof of occasions occurring in as for neighbourhood.

Community oriented Online Social Media Observatory (COSMOS) [51], Social Media and Data Mining. Universe is programming stage that decreases the specialized and methodological hindrances to get to and investigate the online networking. This task is an Economic and Social Research Council (ESRC) vital. The primary concentration of the activities is " Big Data investigation" which incorporates the distinctive application spaces like social, PC, political, wellbeing, factual and scientific researchers to contemplate the methodological, hypothetical, observational and specialized measurements of web-based social networking information in social and arrangement settings. The task is open source, code and dataset which and be utilized for trial reason.

NewsReader [52] is venture, which gathers the constant information from news encourages and stores the information in volumes and break down for basic leadership. The test of this venture is volume of information is expanding step by step and covering billions of filed archives with a large number of reports included day by day. These archives are interconnected with various sources like life stories and friends databases. Newsreader constructed a framework that concentrates the information, investigate, and recognize the occasions on the base of what occasion, whom had done that and when that occasion occurred by utilizing on labels of the newsfeed and put away the whole information in organized database. The Newsreader bolsters diverse dialects like English, Spanish, Italian and Dutch.

Malaysia Crime [53] is a site that shows Crime reports in Malaysia on a guide. Client can investigate the distinctive Crime regions or occasions in the Google delineate. By tapping on area or stick client can perceive what sort of Crime is better known on that specific place. By clicking that stick the clients can see the entire data about the Crime which clarify about Crime sort, area, nitty gritty depiction of Crime. The Project is open source, which utilizes the twitter informational index to do the test think about.

Crime Reports and Public Engines are to enable the Crime ID, to avert programming in light of cloud condition. It gives simple client condition the item incorporate Crime reports like most prevalent Crime in a specific place in maps, order focal which demonstrates the information perception motor. Open Engines have more than 2500 clients in around the world.

III. LITERATURE ON SPATIO-TEMPORAL SOCIAL MEDIA DATA FOR CRIME ANALYSIS

A. Spatio - temporal analysis

Using data distribution of crime across space and time can be analyzed in several ways. One of the traditional methods for visualization is the use of hot-spot maps. It utilizes the kernel density estimation (KDE) methodology to fit two-dimensional spatial probability density function to historical crime record data. Hotspot maps are widely used as they enable the analyst to plot the data and create intuitive visualizations for identifying areas with high crime concentrations [2].

B. Crime prediction

Title Crime prediction is based on the fact that future crimes occur in the vicinity of past crimes. Hot-spot maps are useful in identifying crimes. There are advanced techniques such as self-exciting point process models that capture the spatiotemporal clustering of crimes [3]. However, they come with various limitations. These include locally descriptive models which cannot be scalable to different geographical areas, inability to construct models independent of historical data, inability to extract insights from rich social media data.

Various researchers have addressed the first two limitations of hot-spot maps by correlating the data with a feature space such as police headquarters, local railways, etc. this approach makes use of historical and spatial variables when making predictions. Also, the inclusion of spatial information enables forecast for geographic areas for which no historical data is available [6].

C. Prediction using social media

Social media based predictive modeling is a relatively new area with applications in prediction or detection of disease outbreaks, earthquakes, box office performance of movies, etc. The prediction modeling for crime data differs in the spatial resolution [5]. While the above said research areas could be done with a spatial resolution that encompasses the whole city with a single prediction, criminal activity varies significantly from one city block to other. Author's research takes reference from the works of Wang et. all [7] which have used tweets from news agencies to predict crime. However, the types of crime are limited to break-in and enter, vehicular hit and run crimes. Also, the research work does not include GPS information available with the tweets to give location-specific prediction.

The relationship between crime, people dynamic and urban environment have been studied in detail by the urban activist Jane Jacobs. She emphasizes that natural surveillance, i.e., the presence of high density of visitors and high diversity contribute to the safety of a particular area and in effect less crime. Criminologists have started to investigate crime concentrations at micro levels of geography [14]. Using place-centric and data-driven approach it is possible to determine whether a particular geographic area can become a scene of crime [1].

News content analysis is another area of research 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 [13]. Using NER (Named Entity Recognition), it is possible to extract location data and identify whether they are a crime location or not [9]. By combining temporal proximity, event similarity, and temporal relationships, it is possible to analyze the relationship between different events [4]. Spengler et. all 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 [12]. Using conditional random field, Arulanandam, Savarimuthu, and Purvis have extracted crime location from newspaper articles. Research done by Bao et. al [11] utilizes Concentration Driven Model (CDM) to microscopic analysis of crime locations and Centro-Based Model (CM) to macroscopically predict the sites of future crime. Other researchers 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 [8].

Based on the survey on different research papers Table-1 listed about the different sources that can be considered for Crime analysis and prediction.Table-2 shows the heterogeneous algorithms and research gaps of that algorithms and challenges of algorithms Identified.

TABLE I

Survey on Crime analysis and prediction Data Sources

Source of data	Performed researchers	Algorithms used for Analysis & Prediction
News feeds	[1],[16],[30],[36],[49],[46]	SVM [1,49],Naive Bayes Classifier[30,36,46]
Twitter	[2],[3],[7],[9],[13],[27],[42]	KDE[2,13], Risk Terrain Modeling[3], Naïve Bayes[7,27] ,Multilayer perception[7,27], Pruned C4.5[7,27], decision tree[7,27], ZeroR[7,27], K-means[9],
Mobile	[4],[10],[28],[27]	Random forest algorithm [4,10]
Instagram	[5],[8]	SVM[5], Gaussian Processes Regression(Time series prediction)[8]
Police Record	[11],[12],[14],[18],[34],[38]	KDE[11,12], Geographically Weighted Regression
Ruff datasets	[15],[19]	Ant Colony Optimization (ACO)[15] algorithm to improve the K-Means algorithm, Decision Tree.
Organizations	[17]	LDA((Linear Discriminant Analysis)-KNN(K Nearest Neighbor) [17]
Emails	[34]	Decision Tree Algorithm[34]
Phone calls	[52]	CDR Database with the power of graph analysis tools

TABLE II
Summary of Researches in Crime Analysis and Prediction

Performed Researches	Techniques	Tasks	Research Gaps	Research Challenges
[8]	Apriori algorithm	Extracting elements from data sources and analyzing the crime patterns	False detection	Improve precise detection
[1,49,5]	SVM algorithm	Analyzing the crime patterns	No flexibility of crime model and visualization	Model the crime future attacks and improve visualization
[2,13,11,12]	KDE(Kernel Density Estimation)	Learn from experience and Learn in which part of the space actions will be successful	"Maybe" even find the spot with the highest success probability	Nonparametric kernel density estimation
[3]	Risk Terrain Modeling	Empirical assessment of micro level places that are at an increased risk of illegal behavior and crime.	Risk factors identification and Operationalization, Spatial Influence	Relative Risk Value
[7,27,19]	Decision tree	Classification accuracy with related to crime data.	Classification rules can be used to predict the suspect of the crime.	Accuracy of crime related data.
[4,10,28,27]	Random forest algorithm	Variable selection for crime identification. RF do not require parameter tuning, and do not require specification of a feature space, as support vector machines.	Robust to overfitting problem, overfitting	Over fitting occurs when there are too many features to considered resulting poor performance in predictions or classifications.
[8]	Gaussian Processes Regression(Time series prediction)	geo-referenced information of public services (e.g., shopping centers, banks) and criminal incidents to approximate the prior risk function as a Gaussian Mixture Model	optimum number of centroids for each Gaussian kernel	Loses the efficiency with high dimensional spaces
[30,36,46,7,27]	Naive Bayes Classifier	Predict the probability of a given tuple dependency to a specific class.	Rate of population increases, then the accuracy rate increases otherwise with less data no proper accuracy.	higher sensitivity

IV. Conclusion

All In light of this examination on Crime investigation and expectation we give a review of past and flow scientists who required on Crime examination from Social media information. In view of our review we recognized that Social media is one of the well-known information for doing of Crime examination and forecast. In view of this review we can state that specialists has researched on the extraction designs in light of numerous different of properties like news, tweets, pictures, ruff informational indexes and police ongoing information. Underneath table clarify that diverse specialists engaged with the Crime examination and expectation.

REFERENCES

1. **[1]** Jayaweera, C. Sajeewa, S. Liyanage, T. Wijewardane, I. Perera and A. Wijayasiri, "Crime analytics: Analysis of crimes through newspaper articles," 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, , pp. 277-282, 2015.
2. Matthew S. Gerber, "Predicting Crime Using Twitter and Kernel Density",Journal of Decision Support Systems(Elsevier), Vol. 61, pp. 115-125, May 2014.
3. Anthony J. Corso, Gondy Leroy and Abdulkareem Alsusdais "Toward Predictive Crime Analysis via Social Media, Big Data, and GIS" Proceedings of In iConference 2015.
4. Andrey Bogomolov and Bruno Lepri "Predicting Crime Hotspots Using Aggregated and Anonymized Data on People Dynamics" Proceedings in conference on the scientific analysis of mobile phone datasets, MIT Media Lab, pp. 7-10 April 2015.
5. Chaolun Xia , Raz Schwartz ,Ke Xie ,Adam Krebs, Andrew Langdon , Jeremy Ting and Mor Naaman, "CityBeat: real-time social media visualization of hyper-local city data", International World Wide Web Conferences Steering Committee Republic and Canton of Geneva, Switzerland, pp. 167-170 April 2014.
6. Piek Vossen, German Rigau, Luciano Serafini Pim Stouten, Francis Irving and Willem Van Hage, "NewsReader: recording history from daily news streams", Proceeding in the European Union as project ICT- 316404 pp. 200-208 January 2013.
7. Maximilian Walther and Michael Kaisser, "Geo-spatial event detection in the twitter stream". ECIR'13 Proceedings of the 35th European conference on Advances in Information Retrieval, Springer-Verlag Berlin Heidelberg, pp. 356-367, Novemeber 2013.

8. Ke Xie, Chaolun Xia, Nir Grinberg, Raz Schwartz and Mor Naaman, "Robust detection of hyper-local events from geotagged social media data", Proceedings of the Thirteenth International Workshop on Multimedia Data Mining, pp. 1-9, August 2013.
Tony H. Grubesic and Elizabeth A, "Spatio-Temporal Interaction of Urban Crime", Journal of Quantitative Criminology, Vol 24, Iss 3, pp. 285-306, March 2008.
9. Andrey Bogomolov and Nuria Oliver, "Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data" ICMI '14 Proceedings of the ACM 16th International Conference on Multimodal Interaction pp. 427-434 September 2014.
10. E. Johansson, C. Gåhlin and A. Borg, "Crime Hotspots: An Evaluation of the KDE Spatial Mapping Technique," 2015 European Intelligence and Security Informatics Conference, Manchester, pp. 69-74, 2015.
11. S. Khalid, Jiechen Wang, M. Shakeel and Xia Nan, "Spatio-temporal Analysis of the Street Crime hotspots in faisalabad city of Pakistan," 2015 23rd International Conference on Geoinformatics, Wuhan, pp. 1-4, 2015.
12. M. Wang and M. S. Gerber, "Using Twitter for Next-Place Prediction, with an Application to Crime Prediction," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, pp. 941-948, 2015.
13. S. Kumar and R. K. Gopal, "Data mining based crime investigation systems: Taxonomy and relevance," 2015 Global Conference on Communication Technologies (GCCT), Thuckalay, pp. 850-853, 2015.
14. N. Tomar and A. K. Manjhvar, "An improved optimized clustering technique for crime detection," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, pp. 1-5, 2016,.
15. R. Hassanian-esfahani and M. j. Kargar, "A survey on web news retrieval and mining," 2016 Second International Conference on Web Research (ICWR), Tehran, pp. 90-101, 2016,.
16. Q. Zhang, P. Yuan, Q. Zhou and Z. Yang, "Mixed spatial-temporal characteristics based Crime Hot Spots Prediction," 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Nanchang, pp. 97-101, 2016,.
17. S. Nezami and E. Khoramshahi, "Spatial Modeling of Crime by Using of GWR Method," 2016 Baltic Geodetic Congress (BGC Geomatics), Gdansk, pp. 222-227, 2016,.
18. M. Lnenicka, J. Hovad, J. Komarkova and M. Pasler, "A proposal of web data mining application for mapping crime areas in the Czech Republic," 2015 10th International Joint Conference on Software Technologies (ICSOFT), Colmar, pp. 1-6, 2015,.
19. H. Jiří, I. Igor, D. Michala, H. Bronislava and B. Petr, "Multidimensional database for crime prevention," 2016 17th International Carpathian Control Conference (ICCC), Tatranska Lomnica, pp. 242-247, 2016.
20. Hutchings, R. Clayton and R. Anderson, "Taking down websites to prevent crime," 2016 APWG Symposium on Electronic Crime Research (eCrime), Toronto, ON, pp. 1-10, 2016,.
21. U. Thongsatapornwatana, "A survey of data mining techniques for analyzing crime patterns," 2016 Second Asian Conference on Defence Technology (ACDT), Chiang Mai, pp. 123-128, 2016,.
22. N. Moorosi and V. Marivate, "Privacy in mining crime data from social Media: A South African perspective," 2015 Second International Conference on Information Security and Cyber Forensics (InfoSec), Cape Town, pp. 171-175, 2015,.
23. F. Herrera, R. Sosa and T. Delgado, "GeoBI and Big VGI for Crime Analysis and Report," 2015 3rd International Conference on Future Internet of Things and Cloud, Rome, pp. 481-488, 2015,.
24. J. Azeez and D. J. Aravindhar, "Hybrid approach to crime prediction using deep learning," 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Kochi, pp. 1701-1710, 2015,.
25. N. Dhavase and A. M. Bagade, "Location identification for crime & disaster events by geoparsing Twitter," International Conference for Convergence for Technology-2014, Pune, pp. 1-3, 2014,.
26. T. Mantoro, Feriadi, N. Agani, M. A. Ayu and D. Jatikusumo, "Location-Aware Mobile Crime Information Framework for Fast Tracking Response to Accidents and Crimes in Big Cities," 2014 3rd International Conference on Advanced Computer Science Applications and Technologies, Amman, pp. 192-197, 2014,.
27. Ariffin, B. Solemon and W. M. L. Wan Abu Bakar, "An evaluative study on mobile crowdsourcing applications for crime watch," Proceedings of the 6th International Conference on Information Technology and Multimedia, Putrajaya, pp. 335-340, 2014,.
28. K. R. Rahem and N. Omar, "Drug-related crime information extraction and analysis," Proceedings of the 6th International Conference on Information Technology and Multimedia, Putrajaya, pp. 250-254, 2014.
29. Babakura, M. N. Sulaiman and M. A. Yusuf, "Improved method of classification algorithms for crime prediction," 2014 International Symposium on Biometrics and Security Technologies (ISBAST), Kuala Lumpur, pp. 250-255, 2014,.
30. H. P. Kavya and V. Karjigi, "Sensitive keyword spotting for crime analysis," 2014 IEEE National Conference on Communication, Signal Processing and Networking (NCCSN), Palakkad, pp. 1-6, 2014,.
31. G. Molcho, S. Maier, F. Melero and C. Aliprandi, "CAPER: Collaborative Information, Acquisition, Processing, Exploitation and Reporting for the Prevention of Organised Crime," 2014 IEEE Joint Intelligence and Security Informatics Conference, The Hague, pp. 316-316, 2014,.
32. M. Sharma, "Z - CRIME: A data mining tool for the detection of suspicious criminal activities based on decision tree," 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC), New Delhi, pp. 1-6, 2014.
33. M. A. Tayebi, M. Ester, U. Glässer and P. L. Brantingham, "CRIMETRACER: Activity space based crime location prediction," 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, pp. 472-480, 2014,.
34. S. Sathyadevan, D. M. S and S. G. S., "Crime analysis and prediction using data mining," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, pp. 406-412, 2014.
35. S. Sathyadevan, D. M. S and S. G. S., "Crime analysis and prediction using data mining," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, pp. 406-412, 2014.
36. Musdholifah, S. Z. B. M. Hashim and I. Wasito, "KNN-kernel based clustering for spatio-temporal database," Computer and Communication Engineering (ICCCE), 2010 International Conference on, Kuala Lumpur, pp. 1-6, 2010.
37. O. E. Isafiade and A. B. Bagula, "CitiSafe: Adaptive Spatial Pattern Knowledge Using Fp-Growth Algorithm for Crime Situation Recognition," 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, Vietri sul Mere, pp. 551-556, 2013,.
38. M. S. Vural, M. Gök and Z. Yetgin, "Generating incident-level artificial data using GIS based crime simulation," 2013 International Conference on Electronics, Computer and Computation (ICECCO), Ankara, pp. 239-242, 2013,.
39. P. Intani and T. Orachon, "Crime warning system using image and sound processing," 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), Gwangju, 2013, pp. 1751-1753.
40. F. Albertetti, P. Cotofrei, L. Grossrieder, O. Ribaux and K. Stoffel, "Crime linkage: A fuzzy MCDM approach," 2013 IEEE International Conference on Intelligence and Security Informatics, Seattle, WA, pp. 1-3, 2013,.
41. C. Yang, H. Liu, Y. Sun and A. Abraham, "Multi-knowledge extraction from violent crime datasets using swarm rough algorithm," 2012 12th International Conference on Hybrid Intelligent Systems (HIS), Pune, pp. 560-565, 2012.
42. Guiyun Zhou, Jiayuan Lin and Wenfeng Zheng, "A web-based geographical information system for crime mapping and decision support," 2012 International Conference on Computational Problem-Solving (ICCP), Leshan, pp. 147-150, 2012.
43. Canyu Wang, Xuebi Guo and Hao Han, "Crime detection using Latent Semantic Analysis and hierarchical structure," 2012 IEEE International Conference on Computer Science and Automation Engineering, Beijing, pp. 337-340, 2012.
44. C. H. Yu, M. W. Ward, M. Morabito and W. Ding, "Crime Forecasting Using Data Mining Techniques," 2011 IEEE 11th International Conference on Data Mining Workshops, Vancouver, BC, pp. 779-786, 2011.

45. Z. Shiyao, "A research on dealing with serial crimes with mathematical method," 2011 IEEE 3rd International Conference on Communication Software and Networks, Xi'an, pp. 421-426, 2011.
46. J. Lu and G. a. Tang, "The spatial distribution cause analysis of theft crime rate based on GWR model," 2011 International Conference on Multimedia Technology, Hangzhou, pp. 3761-3764, 2011.
47. M. Mohd and N. M. Ali, "An Interactive Malaysia Crime News Retrieval System," 2011 International Conference on Semantic Technology and Information Retrieval, Putrajaya, pp. 220-223, 2011.
48. Xifan Zheng, Yang Cao and Zhiyu Ma, "A mathematical modeling approach for geographical profiling and crime prediction," 2011 IEEE 2nd International Conference on Software Engineering and Service Science, Beijing, pp. 500-503, 2011.
49. My neighborhood is a secure, private social meeting place for residents of your neighborhood. [Online] Available:http://myneighborhood.com/
50. Matthew S. Gerber "Predicting crime using Twitter and kernel density estimation" Decision Support Systems 61 (2014) 115–125.
51. S.J. Yang, A.M. Greenberg, and M. Endsley "Automatic Crime Prediction Using Events Extracted from Twitter Posts" Springer-Verlag Berlin Heidelberg 2012, (Eds.): SBP 2012, LNCS 7227, pp. 231–238, 2012.
52. COSMOS ESRC-funded Social Data Science Lab (that maintains the software) at Cardiff University.[Online] Available: https://www.cs.cf.ac.uk/cosmos/
53. NewsReader Grant agreement no.: 316404 – Co-funded by the EU - FP7 Work Programme Research theme 'Information and Communication Technologies', [online] Available: http://www.newsreader-project.eu/