

Crime analysis, detection and prediction using big data via cloud computing

¹ Prof. Pranalini A. Joshi, ² Prof. Rahul Bhole, ³ G. Maneeshwari, ⁴ Aditi Khade,
⁵ Dipali Bhosale, ⁶ Vishakha Bichkar,

^{1, 2, 3, 4, 5, 6} Department Of Information Technology,

^{1, 2, 3, 4, 5, 6} Zeal College of Engineering And Research, Pune, India

Email - ¹ pranalini.ketkar@zealeducation.com, ² rahul.bhole@zealeducation.com, ³ maneeshwari@gmail.com,
⁴ aditiganpatkhade1999@gmail.com, ⁵ deelepaliBhosale2841999@gmail.com, ⁶ vishakhabichakar@gmail.com

Abstract: *Crime is most important factor in smart city. Crimes now days are expanding step by step and with various degrees of force and flexibility. The result is great loss to society in terms of monetary loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. Crime analysis will be to generate the crime hot-spots that will help in deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources. To overcome this problem the computing era can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. The crime rate expectation methodologies can be connected on chronicled information accessible in the police records by looking at the information at different points like reason of crime, recurrence of similar kind of crimes at specific location with other parameters to prepare model the crime prediction also in proposed system Fake crime detection, Shows Shortest Distance route from current place to police station on map. We have developed this system for monitoring crime in smart city as well as try to reduce crime in particular area.*

Key Words: *Crime analysis, Cloud computing, Naive Bayes, Police Station.*

1. INTRODUCTION:

For a long time, inquiry in the region of crime investigation has been utilized towards the moderation of crime and open well-being. In the previous decade, with the coming of Big Data, open information accessibility, and e-administration there has been a practically exponential increment in the formation of information examination and representation devices for the policing and treatment of open security frameworks. Stop word Removal algorithm is used for searching purpose, Naive Bayes algorithm is used for classification of crime and K-Nearest Neighbour (KNN) for finding shortest distance on map. Law enforcement agencies store information about reported crimes in many cities and this information is made publicly available in the spirit of open-data. This data typically has the type of crime (e.g., arson, assault, burglary, robbery, theft, and vandalism), as well as the time and location of the crime. Area wise crime detection system, identify fake crime. In our system we show the Shortest Distance route from current place to police station on map, Patterns in crime and the effect of law enforcement policies on the amount of crime in a district can be examined utilizing this information with the objective of lessening crime.

2. LITERATURE REVIEW:

Ayidh alqahtani et.al [1] states that security has reliably been one of the most imperative concerns. Government and security associations are endeavouring to turn away violations and guarantee their family. Regardless, the challenge of overseeing a tremendous proportion of data has transformed into an imperative issue for all affiliations. Thus, a crime information system that jars strategy with an immense proportion of data in a brief time span is required for specialists to know infringement hotspots, wrongdoing plans and to predict future ones. This system gives a plan of a Crime Data Information System. Data pre-preparing is done in the Crime Database and two systems for crime assessment are performed. These two philosophies are examined, and results are certified with ground truth. Andrew J. Park et.al [2] Crime examiners and law requirement offices can detail powerful systems using current wrongdoing patterns and examples. Present day computational advancements, especially enormous information investigation and representation can be applied to accessible wrongdoing information and different assets. Visual examination, a logical thinking procedure with intelligent visual interfaces, is an incredible asset that serves to enlarge the pattern and example acknowledgment abilities of specialists. This paper presents a three-dimensional (3D) visual examination system that intelligently imagines wrongdoing information and other pertinent datasets on an exceptionally precise 3D model of the City of Vancouver, Canada. This 3D perception advances wrongdoing investigation exercises through an increasingly exact presentation of the region under examination. The 3D visual investigation can improve ID of fleeting and spatial

criminal regions and give vital and powerful designs to ace dynamic police organization. The open information list from the City of Vancouver, Canada and the Vancouver Police Department open source property related misconduct information were utilized right now.

Mark Petty et.al [3] proposed an improved procedure to help measurable agents in recognizing positional difference in protests because of wrongdoing scene tainting. Either deliberately or on the other hand coincidentally, wrongdoing scene sullyng can happen during the examination and documentation process. This new proposed strategy uses an ASIFT-based element identification calculation that analyses pre-and post-tainted pictures of a similar scene, taken from various perspectives. The dispute is that the ASIFT enrolment strategy is more qualified to genuine world wrongdoing scene photography, being increasingly hearty to relative mutilation that happens when catching pictures from various perspectives. The proposed philosophy was tried with both the Filter and ASIFT enrolments strategies to show that (1) it could distinguish missing, planted and dislodged objects utilizing both Filter and ASIFT and (2) ASIFT is better than SIFT as far as mistake in uprooting estimation, particularly for bigger perspective disparities between the pre-and post-pollution pictures. Peng Chen and Justin Kurland [4] proposed given the snappy paced nature of current police work, the improvement and use of bleeding edge data burrowing contraptions for wrongdoing examination can play a fundamental factor in easing future insidiousness and helping with wrongdoing evasion. This paper intends to deal with the issue of perceiving potential successive chargeable models using as of now underutilized characteristics from police recorded wrongdoing data. To achieve this wrongdoing data taking care of methodology is recommended that gathers three factors in police recorded wrongdoing event data: (1) time; (2) setting; and (3) modus operandi of doing things method for getting things done. Each wrongdoing event quality is shown using the Apriori estimation, ordinarily used for customary thing set mining and association standard picking up from complex datasets. Results from the model suggest that Apriori can perceive basic affiliations and consequently can highlight wrongdoing model examples settled inside progressively broad police-recorded wrongdoing databases, which could incite more fruitful police responses than by and by methods for regular indicative techniques. Sharmistha dutta et.al [5] proposed the Character Crime Detection using the possibility of adaptability is a multi-layered data mining based approach to manage verifies the private information of applicants applying for charge cards. This assessment relies upon three thoughts: quality by giving various security layers, quality data by slaughtering steady goofs and adaptively by perceiving legitimate and coercion directly. Beside these power drivers, this application furthermore has certain confinements. Before long, the CD and SD counts can look on a gigantic moving window, greater number of qualities and number of association types, thus eating up basically a greater proportion of time for the structure to deliver results. Likewise, even after standard update of the figuring's, the evaluation is unquestionably not a certifiable one as aggressors don't get time to change their technique due to the counts as it would have been at whatever point sent logically. Starting now and into the foreseeable future, the possibility of adaptively isn't properly shown. Subsequently, the future augmentation can be loosened up in reducing these obstructions. Sunil yadav et.al state that crime is a social irritation and costs our overall population significantly in a couple of various ways. Any assessment that can help in clarifying wrongdoings quickly will pay for itself. About 10% of the crooks do about a portion of the crime. The system is set up by supporting before year's record of wrongdoings taken from certifiable online passage of India posting various infringement, for instance, murder, commandeering and capturing, dacoits, robbery, burglary, ambush and other such infringement. As per data of Indian bits of knowledge, which gives data of various wrongdoing of late years (2001-2014) backslide model is made and the crime rate for the following a long time in various states can be envisioned. We have used managed, semi-directed and solo learning procedures on the wrongdoing records for data disclosure and to help in growing the farsighted precision of the crime. This work will be valuable to the close by police central station in crime camouflage. [6]

Sheikh, J. et.al introducing the most recent year's crime examination has transformed into a wide range term that needs a lot of research on crime assessment and crime mapping. Crime mapping and spatial examination reinforce all of them and accept an indispensable occupation in the naturally new sort of crime depiction, recognition and to respond alluringly to the issue of fault. This exploration blends factual strategies (bunch investigation) and spatial models made with GIS, developed on police wrongdoing reports. This framework puts on the varying utilities of GIS to see the problem areas of crime notwithstanding bolster the movement of assessment inclination strategy for policing. The valuable methodology in the present assessment for crime mapping can be viably associated for improvements of UIs organize the movement of safe city strategies. [7] Chamath sajeewa et.al states that Crime examination is one of the most huge activities of the vast majority of the savvy and law necessity affiliations wherever all through the world. Generally they assemble private and outside Crime related data (information) to prevent future attacks and utilize a foreordained number of law prerequisite resources in a perfect manner. An important test looked by most of the law prerequisite and learning affiliations is viably and definitely researching the creating volumes of crime related data. The huge land arranged assortment and the multifaceted idea of crime models have made the examining and recording of crime data logically problematic. Data mining is an indispensable resource that can be used satisfactorily for researching tremendous databases and deciding huge illustrative results. This framework displays a sharp crime examination system which is planned to beat the recently referenced issues. The proposed structure is an online system which contains crime examination frameworks, for instance, hotspot distinguishing proof, crime assessment and crime plan portrayal. The

proposed system contains a rich and revised condition that can be used effectively for strategies of crime examination. [8]

3. METHOD:

Crime is one of the dangerous factors for any country; analysis of crime is the action in which analysis is done on crime activities. Today culprits have most extreme utilization of every single current innovation and logical techniques in perpetrating crimes. The law authorities need to successfully address out difficulties of crime control and upkeep of open request. In a proposed system mainly includes Admin, User, Crime reporter and Police module. In Proposed system authenticate crime reporters add different types of crime according to category. According to the collection of data of crime, analyze the data. Admin can check if the crime is fake or not. If any crime reporter adds any fake crime the first time the admin sends a warning to that crime reporter and if again adds fake crime by that reporter then the system can block that crime reporter. Users can search crime wise as well as area wise. Users can view the shortest distance route on a map for the police station. Police also view the crimes and do investigation. Using a cloud system we can access anywhere and anytime. In the proposed system, Stop word Removal algorithm is used for searching purpose, Naive Bayes algorithm is used for classification of crime and K-Nearest Neighbours (KNN) for finding shortest distance on map. A stop word is a commonly used word that (the, is, a, about, more etc.) a search engine has been programmed to ignore, both when indexing entries for searching.

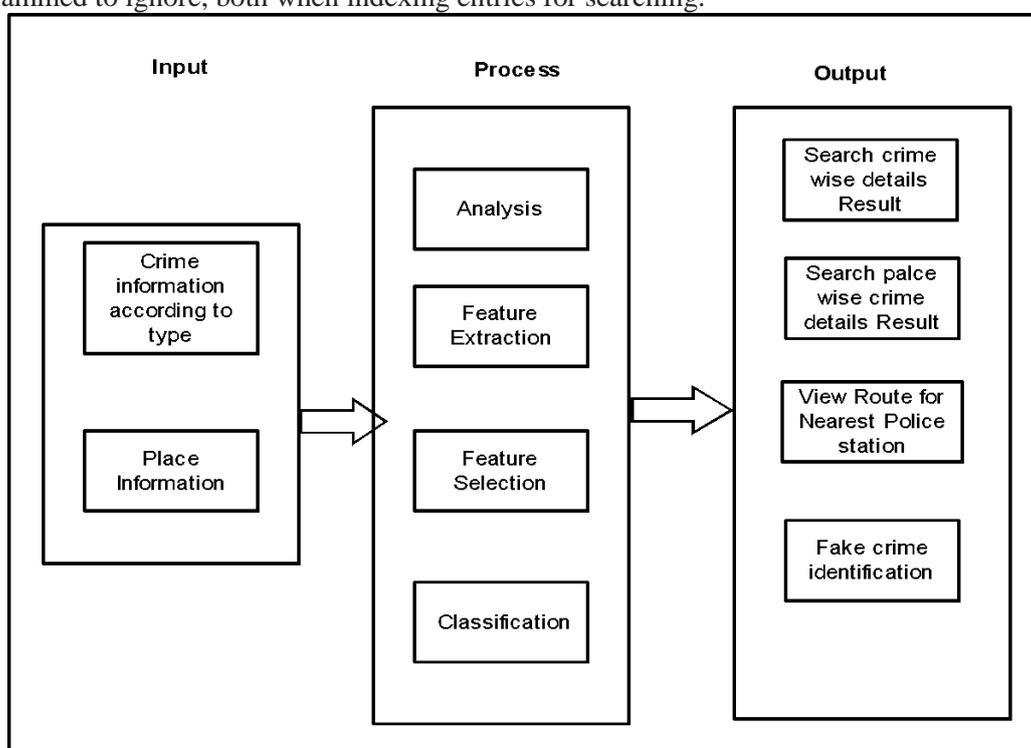


Figure 1. Block Diagram of Proposed System

This algorithm is used in search engine, Natural language processing (NLP). Using stop word removal algorithm user search the crime or user search place wise crime. Using this algorithm we can classify the crime according to category after classification we predict crime category. Euclidean distance is separation is the straight line separation between two points. Euclidean space becomes a metric space. This algorithm is used for finding the optimal distance on a map.

Working of User:- User login with proper authentication, Search crime area wise, Search crime category wise, View Crime, View Nearest police station route on map.

Working of Admin:- Admin login with proper authentication, View crime reporter information, View Crime information, Send warning to crime reporter add Police station dataset, Block crime reporter.

Working of Crime Reporter:- Crime Reporter login with proper authentication, Add crime, View Crime and View Warning

Working of Police:- Police login with proper authentication, View Crime and do investigation

4. DISCUSSION:

Critical resources in these smart cities will be more rapidly deployed to regions in need, and those regions predicted to have an imminent or prospective need. For example, crime data analytics may be used to optimize the

distribution of police, medical, and emergency services. Existing systems work on crime analysis of data in USA cities with the help of different parameters. This system does not do any work on Fake crime identification and blocks the crime reporter. To find out shortest distance route on map for police station for crime investigations

5. ANALYSIS:

5.1. Stop words Removal Algorithm:

A stop word is a commonly used word that (the, is, a, about, more etc.) a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. This algorithm is used in search engine, Natural language processing (NLP)

Input: - Place name or Crime Type

Output: - View all the crime details according to place name or crime type.

A dictionary based approach has been utilized to remove stopwords from documents. A generic stopword list containing 400 stopwords created using hybrid approach is used.

5.2. The algorithm is implemented as below given steps:

Step 1: The target document text is tokenized and individual words are stored in an array.

Step 2: A single stop word is read from the stopword list.

Step 3: The stop word is compared to the target text in the form of an array using sequential search technique.

Step 4: If it matches, the word in array is removed, and the comparison is continued till length of array.

Step 5: After removal of the stopword completely, another stopword is read from the stopword list and again the algorithm follows step 2. The algorithm runs continuously until all the stopwords are compared.

Step 6: Resultant text devoid of stop words is displayed, also required statistics like stopword removed, no. of stopwords removed from target text, total count of words in target text, count of words in resultant text, individual stop word count found in target text is displayed.

5.3. NAIVE BAYES:

Using this algorithm we can classify the crime according to category after classification we predict crime category.

Input:-crime related words

Output:-Crimes according to category.

Naive Bayes is a basic system for developing classifiers: models that appoint class names to issue occasions, spoken to as vectors of highlight esteems, where the class marks are drawn from some limited set. There is certifiably not a solitary calculation for preparing such classifiers, yet a group of calculations dependent on a typical guideline: all innocent Bayes classifiers expect that the estimation of a specific element is autonomous of the estimation of some other component, given the class variable.

5.4. This algorithm, summarized as follows:

$$P(W_k/class) = \frac{n_k + 1}{(n + \text{Vocabulary})}$$

Where, n = total no. of words with specified class

n_k = no. of times word occurred with the specified class Vocabulary = size

5.5. Algorithm Steps:

- Calculate n_k i.e., no. of times the word occurred with class.
- Calculate n i.e., total no. of words for given class.
- Calculate $p(w_k/v_j) = n_k/n$ i.e. the probability of word for the given class .
- Calculate the probability of each class
- Calculate the vocabulary i.e., the total no. of unique words
- Repeat the process.

5.6. K-Nearest Neighbour:

Euclidean distance is the straight line distance between two points. Euclidean space becomes a metric space. This algorithm is used for finding the optimal distance on a map.

Input:-Source and destination location name.

Output:-Shortest path on map

In pattern recognition, the k-nearest neighbours' algorithm (k-NN) is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space. If $k = 1$, then the object is simply assigned to the class of that single nearest neighbour.

- Load the data
- Initialise the value of k
- For getting the predicted class, iterate from 1 to total number of training data points
- Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are cosine matrix, etc.
- Sort the calculated distances in ascending order based on distance values
- Get top k rows from the sorted array
- Get the result.

6. FINDINGS:

In our experimental setup, as shown in table 1, the total numbers of crime according to crime type were 1000. These data sets are formed by collecting data from various sources . These crimes were then divided into five main categories; among which 325 were Robbery, 165 were Murder, 210 were Decoity, 170 were Kidnapping and and rest 130 were Rape.

Sr. No	Crime Type	Number of Crime
1	Robbery	325
2	Murder	165
3	Decoity	210
4	Kidnapping	170
5	Rape	130

Table1: Classification of crime wise category

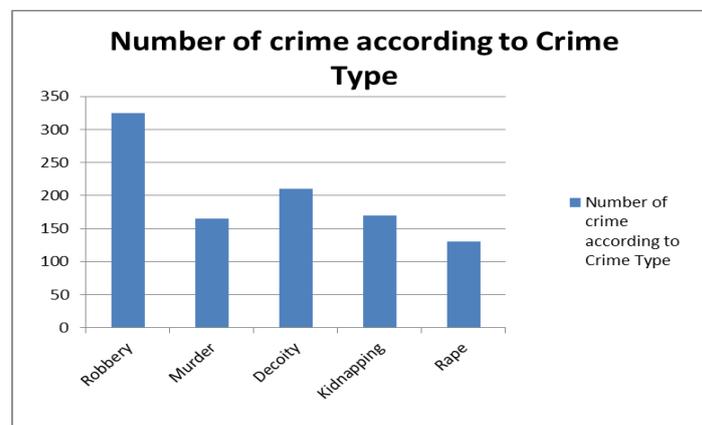
In our experimental setup, as shown in table 6.2, the total numbers of crime according to place were 1000. These crimes were then divided into five main categories; among which 285 were Housing, 140 were Hostel, 240 were Market, 145 were Street and rest 190 were Store.

Sr. No	Place Type	Number of Crime
1	Housing	285
2	Hostel	140
3	Market	240
4	Street	145
5	Store	190

Table2: Classification of place wise category

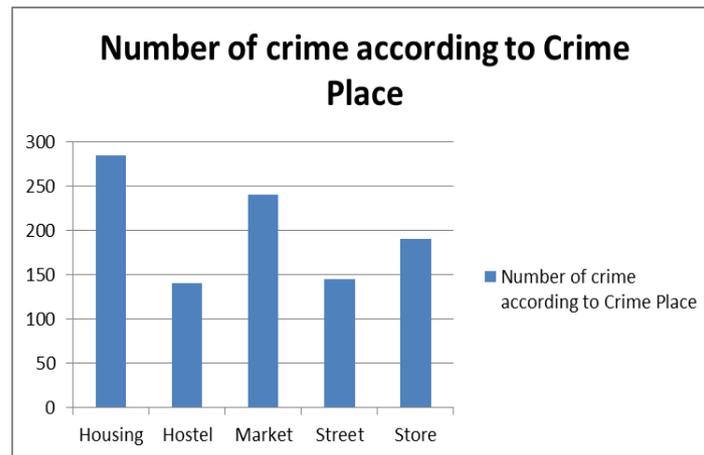
7. RESULT:

From above data, as shown 1000 total crime in particular crime type in graph 1, the numbers of crime found to be Robbery were 325, Murder crime were 165, Decoity crime were 210, kidnapping crime were 170 and rest Rape crime were 130.



Graph 1: Number of crime according to crime type

From above data, as shown 1000 total crimes in particular place in graph 2, the numbers of places found to be Housing were 285, Hostel crime were 140, Market crime were 240, Street crime were 145 and rest store crime were 190.



Graph 2: Number of crimes according to crime place.

8. CONCLUSION:

Crimes nowadays are expanding step by step and with various degrees of force and flexibility. The result is great loss to society in terms of monetary loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. To overcome this problem this project can help to reduce the crime or even may be helpful in predicting the crime so that sufficient measures can be taken to minimize the loss to property and life. With the help of the cloud we can easily access the system from anywhere in the world.

REFERENCES:

1. Ayidh Alqahtani, Andrey Bogomolov, Ajwani Garima, Ahmad Alaiad “CRIME ANALYSIS IN CHICAGO CITY” 2019 10th international conference on information and communication systems (icics)
2. Andrew J. Park ; Valerie Spicer ; Herbert H. Tsang ; Kyle Behiels ; Justin Song Discovering “Crime Trends and Patterns Using Three-Dimensional Visual Analytics” 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)
3. Mark Petty ; Shyh Wei Teng ; Manzur Murshed “Improved Image Analysis Methodology for Detecting Changes in Evidence Positioning at Crime Scenes” 2019 Digital Image Computing: Techniques and Applications (DICTA)
4. Peng Ahen ; Justin Kurland “Time, place, and modus operandi: a simple apriori algorithm experiment for crime pattern detection” 2018 9th international conference on information, intelligence, systems and applications (iisa)
5. Sharmistha dutta ; Neetu Narayan “Identity crime detection using data mining” 2017 3rd international conference on computational intelligence and networks (cine)
6. Sunil Yadav ; Meet Timbadia ; Ajit yadav ; Rohit vishwakarma ; Nikhilesh Yadav “crime pattern detection, analysis & prediction” 2017 international conference of electronics, communication and aerospace technology (iceca)
7. Sheikh, J., Shafique, I., Sharif, M., Zahra, S. A., & Farid, T. (2017). IST: Role of GIS in crime mapping and analysis. 2017 International Conference on Communication Technologies (ComTech).
8. Chamath sajeewa, Isuru jayaweera; Sampath liyanage ; Tharindu wijewardane ; Indika perera; Adeesha wijayasiri “Crime analytics: analysis of crimes through newspaper articles” 2015 moratuwa engineering research conference (mercon)
9. Dipali Bhosale : Vishakha Bichkar : G.Maneeshwari : Aditi Khade : Prof. Pranalini. A. Joshi : Prof. Rahul Bhole “SURVEY ON CRIME ANALYSIS, DETECTION AND PREDICTION USING BIG DATA VIA CLOUD COMPUTING”. 2019 International Journal For Research & Development In Technology.