

# CRIME ANALYSIS, DETECTION AND PREDICTION IN SMART CITY

Vivek Amrutkar<sup>1</sup>      Ajay Sanas<sup>2</sup>      Prasad Umbarkar<sup>3</sup>      Shubham Sonawane<sup>4</sup>      Prof. S. N. Uke<sup>5</sup>  
UG Student              UG Student              UG Student              UG Student              Ass.Professor

*Bachelor of Engineering(IT)*

*<sup>1</sup>Department of Information Technology,*

*<sup>1</sup>SKN Sinhgad Institute of Technology & Science Lonavala, Maharashtra, India*

**Abstract-** Crime is most important factor in smart city. Crimes now days are expanding step by step and with various degree 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 developed this system for monitoring crime in smart city as well as try to reduce crime in particular area.

**Keywords-** *Crime analysis, Cloud computing, Naive Bayes, Police Station*

## I INTRODUCTION

For a long time, inquire about 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, shows 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.

## II LITERATURE SURVEY

Ayidh alqahtani et.al [1] states that security has reliably been one of the most critical concerns. Government and security associations are endeavouring to turn away crimes and guarantee their kinfolk. Regardless, challenge of overseeing tremendous proportion of data has transformed into a significant issue for all affiliations. Thusly, a wrongdoing information structure that jars strategy tremendous proportion of data in a brief time span is required for operators to know infringement hotspots, wrongdoing plans and to predict future ones. This paper gives plan of Crime Data Information System. Data pre-handling is done in Crime Database and two approaches for crime assessment is performed. These two procedures are dissected, and results are attested with ground truth.

Neetu Narayan and sharmistha dutta [2] proposed that data mining based approach to manage verifies the private information of up and comers applying for charge cards. This assessment relies upon three thoughts quality by giving various security layers, quality data by slaughtering consistent goofs and adaptively by perceiving legitimate and coercion direct. Beside these force drivers, this application moreover has certain limitations. Before long, the CD and SD figuring's can look on tremendous moving window, greater number of qualities and number of association types, thusly eating up basically greater proportion of time for the system to create results. Additionally, much after standard update of the figuring's, the appraisal is unquestionably not a certified one as aggressors don't get time to change their system due to the estimations as it would have been at whatever point sent continuously. Starting now and into the foreseeable future, the

possibility of adaptivity isn't suitably shown. Henceforth, the future augmentation can be loosened up in decreasing these hindrances.

Peng chen and Justin Kurland introducing a fast paced nature of current police work, the improvement and usage of bleeding edge data burrowing devices for crime examination can play an essential factor in easing future insidiousness and helping with wrongdoing shirking. This paper intends to deal with the issue of perceiving potential successive at fault models using as of now underutilized characteristics from police recorded wrongdoing data. To achieve this wrongdoing data dealing with system is recommended that moves three factors in police recorded wrongdoing event data: (1) time; (2) setting; and (3) modus operandi as usual method for getting things done. Each crime event quality is shown using the Apriori count, typically used for customary thing set mining and connection standard picking up from complex datasets. Results from the model recommend that Apriori can perceive basic affiliations and subsequently can include crime model examples settled inside progressively broad police-recorded wrong doing databases, which could incite more fruitful police responses than by and by offered by methods for regular symptomatic techniques.[3]

Rohit Vishwakarma et.al states that crime is a social irritation and cost our overall population significantly in a couple of various ways. Any assessment that can help in clarifying violations quickly will pay for itself. About 10% of the culprits do about portion of the violations. The system is set up by continuing before year's record of crimes taken from certifiable online passage of India posting various infringement, for instance, murder, seizing and abducting, dacoits, burglary, robbery, ambush and other such infringement. As per data of Indian bits of knowledge, which gives data of various crime of late years (2001-2014) backslide model is made and the crime percentage for the following a long time in various states can be envisioned. We have used regulated, semi-administered and unaided learning method 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 command in crime covering.[4]

Indika Perer et.al[5] introducing a crime examination is one of the most huge activities of the greater part of the astute and law prerequisite affiliations wherever all through the world. Generally they accumulate private and outside Crime related data (information) to thwart future ambushes and utilize a foreordained number of law prerequisite resources in a perfect manner. An important test looked by most of the law prerequisite and information affiliations are successfully and correctly researching the creating volumes of wrongdoing related data. The colossal land grouped assortment and the multifaceted idea of crime models have made the examining and recording of wrongdoing data continuously inconvenient.

Data mining is a fundamental resource that can be used satisfactorily for exploring tremendous databases and deciding huge illustrative results. This framework displays a sharp wrong doing examination structure which is planned to beat the recently referenced issues. The proposed structure is an online structure which contains crime examination frameworks, for instance, hotspot distinguishing proof, crime assessment and wrongdoing plan portrayal. The proposed structure includes a rich and modified condition that can be used effectively for systems of crime examination.

Sheikh, J et.al proposed that most recent year's crime examination has transformed into a wide range term that needs a lot of research on wrongdoing assessment and wrongdoing mapping. Wrongdoing mapping and spatial examination reinforcements all of them and expect a crucial activity in the characteristically new kind of wrongdoing depiction, recognition and to respond alluringly to the issue of offense. This exploration blends mixes factual strategies (group investigation) and spatial models made with GIS, developed on police wrong doing reports. This framework puts on the various utilities of GIS to see the problem areas of wrong doing not with standing bolster the movement of assessment inclination method for policing. The helpful methodology in the present assessment for crime mapping can be viably associated for advancement of UIs organizes the movement of safe city systems.

Shyh Wei Teng et.al proposed an improved technique to help quantifiable specialists in perceiving positional contrast in fights on account of bad behavior scene spoiling. Either intentionally or then again unintentionally, bad behavior scene sully can occur during the assessment and documentation process. This new proposed methodology utilizes an ASIFT-based component recognizable proof figuring that investigations pre-and post-spoiled photos of a comparative scene, taken from different points of view. The debate is that the ASIFT enrolment technique is progressively able to certified world bad behavior scene photography, being progressively healthy to relative mutilation that happens when getting pictures from different points of view. The proposed way of thinking was attempted with both the Filter and ASIFT enrolments techniques to show that (1) it could recognize missing, planted and removed articles using both Filter and ASIFT and (2) ASIFT is superior to SIFT similar to both in evacuating estimation, especially for greater point of view inconsistencies between the pre-and post-contamination pictures.

Andrew J. Park et.al introducing that[8] crime analysts and law prerequisite workplaces can detail amazing frameworks utilizing current bad behavior examples and models. Present day computational progressions, particularly tremendous data examination and portrayal can be applied to open bad behavior

data and various resources. Visual assessment, a legitimate intuition strategy with insightful visual interfaces, is a staggering resource that serves to amplify the example and model affirmation capacities of experts. This paper exhibits a three-dimensional (3D) visual assessment framework that keenly envisions bad behaviour data and other appropriate datasets on an extraordinarily exact 3D model of the City of Vancouver, Canada. This 3D observation advances bad behaviour examination practices through an undeniably precise introduction of the area under assessment. The 3D visual examination can improve ID of transitory and spatial criminal areas and give fundamental and amazing structures to pro powerful police association. The open data list from the City of Vancouver, Canada and the Vancouver Police Department open source property related unfortunate behaviours data were used at the present time.

### III PROBLEM STATEMENT

It is the major challenge to understand the versatile data available with us then model it to predict the future incidence with acceptable accuracy and further to reduce the crime rate.

**Goal:** - Using this system we monitoring crime in smart city also prevent fake crime news and also help for maintains all crime report to the police station.

#### Objectives

- To search area wise crime.
- Analysis the crime with the help of existing data.
- To find out shortest distance route on map for police station.
- Fake crime identification and block the crime reporter

### IV EXISTING SYSTEM APPROACH

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 system work on crime analysis of data in USA city with the help of different parameters. In this system does not any work on Fake crime identification and block the crime reporter. To find out shortest distance route on map for police station for crime investigations.

### V ALGORITHM APPROACH

#### 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 is been utilized to remove stopwords from document. A generic stopword list containing 400 stopwords created using hybrid approach is used

**The algorithm is implemented as below given steps.**

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

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

**Step 3:** The stop word is compared to target text in form of 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 stopword completely, another stopword is read from stopword list and again 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.

#### 2. NAIVE BAYES

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

**Input:**-crime related words

**Output:**-Crimes accord to category.

Naive Bayes is a basic system for developing classifiers: models that appoint class names to issue occasions, spoke 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.

**This algorithm, summarized as follows.**

$P(W_k/class) = (n_k + 1) / (n + Vocabulary)$

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

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

**Algorithm Steps:**

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

#### 3.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 optimal distance on 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.

1. Load the data.
2. Initialise the value of k.
3. For getting the predicted class, iterate from 1 to total number of training data points.
4. 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 cosine matrix, etc.
5. Sort the calculated distances in ascending order based on distance values.
6. Get top k rows from the sorted array.
7. Get the result.

#### VI MATHEMATICAL MODULE

Symbol

- $TN_c$  = Total number of crime.
- $TN_t$  = Total number count of crime type
- $CA_1$  = Crime in first area.
- $CA_2$  = Crime in second area.
- $TCP_a$  = Total crime in particular area.
- $TRC_a$  = Total remaining crime in area.
- $TCPt$  = Total crime of particular type
- $TRCt$  = Total remaining crime type.

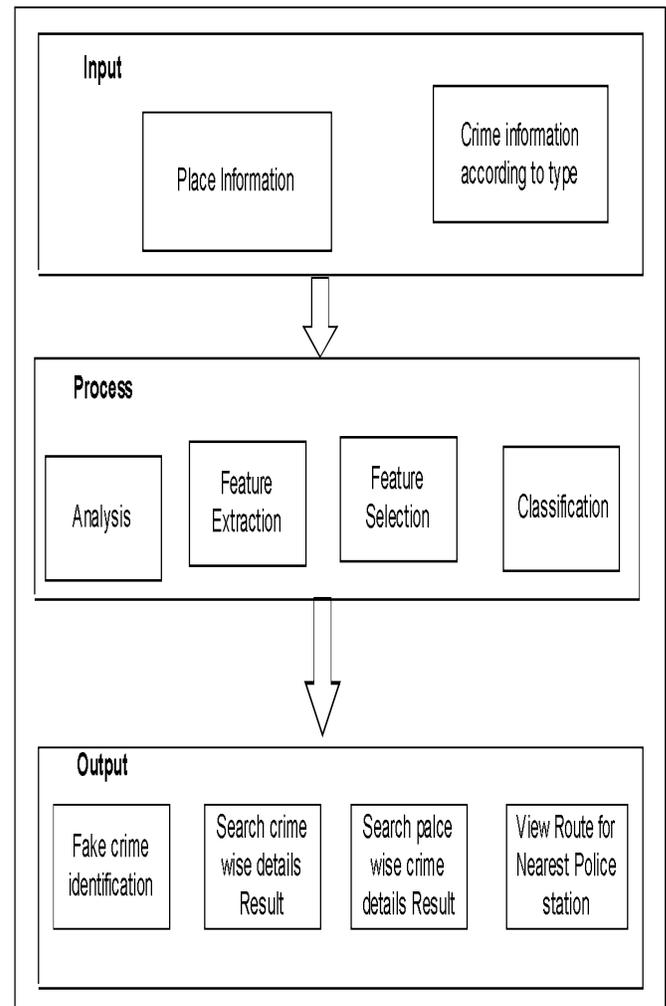
Equation

1. Total number of crime = Total crime of first area + Total number of crime second area... + Total number of crime N area
- $TN_c = CA_1 + CA_2 + \dots + CA_N$
2. Crime in particular area = Total number of crime - Total Remaining area in crime.
- $TCP_a = TN_c - TRC_a$
3. Total count of crime of particular type = Total count of crime of particular type - Total Remaining count of crime of particular type.
- $TCPt = TN_t - TRCt$

#### VI. SYSTEM ARCHITECTURE

Investigation of crime is the activity wherein examination is done on crime exercises. Today offenders have most extraordinary use of each and every present development and sensible procedures in executing crimes The law specialists need to effectively address out challenges of crime control and up keep of open solicitation .In a proposed framework for the most part incorporates Admin, User, Crime Reporter and

Police module. In Proposed framework confirm crime reporter include various kinds of crime as indicated by class as per collection of information of crime, investigation the information. Administrator can check the crime is fake or not. On the off chance that any crime correspondent include any fake crime first time administrator send warning to that crime reporter and in the event that again include block to the crime reporter by that admin, at that point framework can obstruct that crime reporter. Users can look through crime astute just as territory savvy. User can see most limited separation course on map for police headquarters. Police additionally see the violations and do examination Using cloud system we can access anywhere and anytime.



**Figure 1 System Architecture**

In system architecture, We can collect the information from various place of pune like hadapasar, narhe, Shivajinagar and baner. 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. 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 classified 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 optimal distance on map.

### VII COMPARATIVE RESULTS

In our experimental setup, as shown in table 1, the total numbers of crime according to crime type were 600. These crimes were then divided into five main categories; among which 145 were Robbery, 120 were Murder, 105 were Decoity, 95 were Kidnapping and and rest 135 were Rape.

Sr. No	Crime Type	Number of Crime
1	Robbery	145
2	Murder	120
3	Decoity	105
4	Kidnapping	95
5	Rape	135

Table1 1: Classification of crime wise category

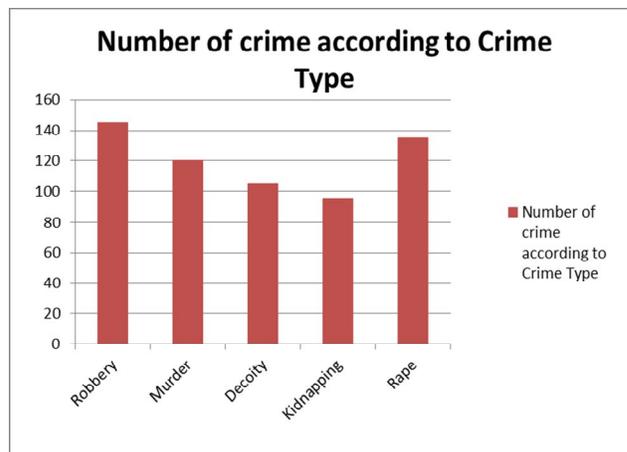
In our experimental setup, as shown in table 2, the total numbers of crime according to place were 600. These crimes were then divided into five main categories; among which 163 were Housing, 120 were Hostel, 137 were Market, 89 were Street and rest 91 were Store.

Sr. No	Place Type	Number of Crime
1	Housing	163
2	Hostel	120
3	Market	137
4	Street	89
5	Store	91

Table1 2: Classification of place wise category

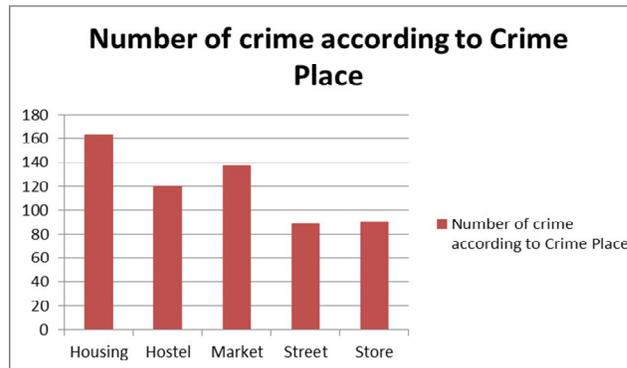
### VIII RESULTS

Graph 1: Number of crime according to crime type From above data, as shown 600 total crime in particular crime type in graph 7.1, the numbers of crime found to be Robbery were 285, Murder crime were 140, Decoity crime were 105, kidnapping crime were 95 and rest Rape crime were 135.



Graph 2: Number of crime according to crime place.

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



### IX CONCLUSION

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 monitory loss, social loss and further it enhances the level of threat against the smooth livelihood in the society. 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 applied on authentic information accessible in the police records by analyzing the information at different edges like explanation of crime, recurrence of comparable kind of crimes at specific location with other parameters to prepare model the crime prediction also in proposed system Shows Shortest Distance route from current place to police station on map as well Fake crime detection system. With the help of cloud we can easily access the system form anywhere in the world.

#### ACKNOWLEDGMENT

This work is supported in a crime analysis in smart city of any state in india. Authors are thankful to Faculty of Engineering and Technology (FET), SavitribaiPhule Pune University,Pune for providing the facility to carry out the research work.

#### 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] Neetu narayan ,sharmistha dutta ; “Identity crime detection using data mining” 2017 3rd international conference on computational intelligence and networks (cine)
- [3] Peng chen ; 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)
- [4] Rohit vishwakarma rohit vishwakarma ,Sunil yadav ; meet timbadia ; ajit yadav ; ; nikhilesh yadav “crime pattern detection, analysis & prediction” 2017 international conference of electronics, communication and aerospace technology (iceca)
- [5] Indika perer;Isuru jayaweera ; chamath sajeewa ; sampath liyanage ; tharindu wijewardane ;adeesha wijayasiri “Crime analytics: analysis of crimes through newspaper articles” 2015 moratuwa engineering research conference (mercon)
- [6] Sharif, M., Zahra ,Sheikh, J., Shafique, I, S. A., & Farid, T. (2017). IST: Role of GIS in crime mapping and analysis. 2017 International Conference on Communication Technologies (ComTech).
- [7] Shyh Wei Teng ,Mark Petty ; Manzur Murshed “Improved Image Analysis Methodology for Detecting Changes in Evidence Positioning at Crime Scenes” 2019 Digital Image Computing: Techniques and Applications (DICTA)
- [8] 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)