



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 8      Issue: V      Month of publication: May 2020**

**DOI: <http://doi.org/10.22214/ijraset.2020.5038>**

**[www.ijraset.com](http://www.ijraset.com)**

**Call:  08813907089**

**E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)**

# Crime Investigation using Data Mining

Abhishek Wagh<sup>1</sup>, Tejas Zambre<sup>2</sup>, Ronak Sharma<sup>3</sup>, Harsh Badera<sup>4</sup>, Rakesh Shirsath<sup>5</sup>

<sup>1, 2, 3, 4</sup>UG Scholar, <sup>5</sup>Professor, Computer Engineering Department, Sandip Institute of Technology and Research Centre, Nashik, Maharashtra, India

**Abstract:** *The Crime rate around the world is very large in number and increasing dramatically. In the present investigation techniques and approaches, police officers have to put their time and manpower to solve these cases by analysing collected information, figuring suspects & finding criminal(s). As information collected throughout crime investigation is enormous, data processing is the technique that can prove very helpful. Data Mining is a technique that will be used to extract useful information from a large amount of crime information so that possible suspects of the crime will be known with efficiency. The best playacting algorithmic program is used against sample crime and criminal information to spot the possible suspects of the crime.*

**Keywords:** *Application, Cloud Technology, Crime Analysis, Criminal Investigation, Criminology, Data Mining, Mobile Prediction Algorithms, Suspect Prediction*

## I. INTRODUCTION

In India, the police department preventing the crimes; maintaining law, rules, and regulations in society; peace throughout the country at large scale. The problem with the Indian or any other police department in the world is that they're still using the manual paper documentation method for maintaining records and filing the First Information Report (FIR) to analyse the records of crime and criminals. On the other hand, criminals are taking benefits from advanced and latest technologies to commit the crime. Every year crime rate is increasing dramatically. The total number of crime cases recorded every year is very large in number. Crime rate is increasing in no time in India and that we have become unsafe. If we tend to perform necessary calculations on the statistics, the local department should handle close to regarding 6400 cases per day. To stop crimes, a police officer needs to establish evidence for those cases. We need some investigation techniques using which our police department can get some sort of idea to solve the problem. We have designed a crime investigation mobile application that is used for suspect prediction using different prediction algorithms, data mining techniques, and based on data available in the database. All the data entered by the user is stored on the cloud using Cloud technologies. Finally, crime analysis is performed using all these techniques and algorithms.

## II. LITERATURE SURVEY

Crime data mining is utilized to convey entire crime measurements of a selected scene or territory that offer an advantage to the overall public by striking the administration and law implementation organizations to comprehend expansion causes that grow the crime rates. The administration and law demand offices can make higher decisions for the more senior living of the nationals that ordinarily mean a parcel of lives. The decision tree algorithmic rule and Apriori algorithmic rule that identified the information and classified the information consequently. Decision Tree Pseudocode

- 1) *Step 1* – Initialize the algorithm with Set S as the root node
- 2) *Step 2* – Iterate this Set S continuously. Where on each iteration it will iterate through only unused attributes in the Set and will calculate Information gain and Entropy of each attribute.
- 3) *Step 3* – Select the attribute with highest information gain or lowest entropy.
- 4) *Step 4* – Split the set by using the attribute selected in the previous step to form new subsets.
- 5) *Step 5* – Use the above steps/algorithm on each formed subset while considering unused attributes only. [1]

Frequent set mining plays an important role in information mining. The Apriori development calculations are the most known calculations which can be utilized for Frequent data Pattern mining. This paper proposes the study of different Pattern Mining and Data Mining calculations, and can be applied to crime pattern mining. The survey analysis provides the data that would tell about what has been done in the particular area, what is the present status and what are the other related attributes. This paper assists the working researchers to get a clear idea towards the application of frequency related data mining algorithms in various possibilities. [2] In existence, crime keeps increasing and threatens the lives of the people publicly. The accuracy and time of tracing are robust, whereas the data processing technique is indulged. The hurdle within the process starts from choosing the connected variable for analysis and their sensitiveness. The data mining techniques like classification, clustering, and rule mining is well supporting the

purpose of handling any type of data. The effective management of crime-related data along with the optimized algorithms is the potential area of research. Profiling the offender information then applying the data mining paves the way to understand the behavioral pattern of the criminals. [3] This paper shows data mining techniques can be used on Crime details that impacts the high properties. Supervised machine learning uses prebuilt data sets as an input to teach, testing, and produce desired outputs using them, whereas unsupervised machine learning divides a piece of unstructured information into different categories. The algorithms of Decision tree and Naïve Bayes algorithm are well known examples of the supervised machine learning methods in data mining and machine learning on present data. Using such algorithms, we can make prediction or classify data in more useful form. [4]

As we know that Data mining is the process in which useful information is extracted. This useful extracted information can be used in more effective and various way like for prediction of possible suspects and/or criminal in this case. There are a number of data mining strategies available. Using specific data mining techniques has a more significant influence on the results. There are various data processing techniques and algorithms like Naïve Bayes, JRip and J48. These algorithms will be tested for their performance against sample data. Output from each technique or algorithm then will be compared against each other and then the algorithm with best results will be selected. Data mining is a process of extracting data from an immense quantity of knowledge stored in databases, information warehouses, and information repositories. Clustering on the opposite is the method of combining feature objects into groups. The data objects present in Groups (Clusters) are similar and they are dissimilar compared to other Groups (Clusters) [5]

This research proposes, the most significant features of data mining strategies were used through a multi-utility model for advanced crime investigation. This frame uncovered some strategies and approached for using SOM and MLP Neural networks for classification of Criminal information and clustering. Challenges of using partition clustering techniques in crime data clustering were also proposed. We hope we can make crime spatial-temporal data involved in the analysis besides behavioral crime variables. We also intend to implement this framework as integrated enterprise software. [6]

### III. ALGORITHM

#### A. Naive Bayes Classifier

Naive Bayes is Supervised Learning based and an option classifier strategy based on Bayes's Theorem. Naive Bayes classifier is well known for its performance and accuracy. Naive Bayes is reliable algorithm. Naive Bayes classifiers can perform efficiently and effectively on large datasets. This algorithm assumes that the effect of a feature in a particular class is independent of other properties of class. Even if these features are dependent on each other, these features are still considered as independent ones. This assumption from Naïve Bayes simplifies computation and known as class conditional independence.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- 1)  $P(h)$ : This is known as the prior probability of h.
- 2)  $P(D)$ : This is called as the prior probability.
- 3)  $P(h/D)$ : This is known as posterior probability.
- 4)  $P(D/h)$ : This is known as the posterior probability.

Naive Bayes classifier calculates the probability of an input in the following steps:

- a) *Step 1*: Using above equation, for the given feature calculate the probability
- b) *Step 2*: Find Likelihood probability with each attribute for each class
- c) *Step 3*: Calculate posterior probability putting these values in Bayes Formula
- d) *Step 4*: Check the class with a higher probability, given the input belongs to the higher probability class

#### B. Decision Tree

Decision Tree is used for Classification. Decision Tree algorithms are very easy to use compared to other algorithms for the same applications. It is one of the Supervised Learning approaches for problem solving. Problems of Regression and Classifications are solved by using Decision Tree.

Using Decision tree Sample model will be created for classification or predicting the target variable value.

As the name suggests, the algorithm uses tree like structure for classification. In this tree representation, internal nodes of the tree represent an attribute and leaf nodes of the tree represents class label or value of target variable.

1) *Decision Tree Algorithm Pseudocode*

- a) *Step 1* – Initialize the algorithm with Set S as the root node
- b) *Step 2* – Iterate this Set S continuously. Where on each iteration it will iterate through only unused attributes in the Set and will calculate Information gain and Entropy of each attribute.
- c) *Step 3* – Select the attribute with highest information gain or lowest entropy.
- d) *Step 4* – Split the set by using the attribute selected in previous step to form new subsets.
- e) *Step 5* – Use above steps/algorithm on each formed subset while considering unused attributes only.

**IV. ARCHITECTURE**

Figure 1 describes about the system architecture of the main application. It describes about all the components and sub-components which the system is consists of. Here, the system is designed using user interface, server and it's consists of a database. User Interface service is provided to the user when he/she opens the application present on their device. The User Interface is directly connected to the server where the main algorithmic tasks, data mining and data analysis algorithms are performed. All the data entered or the files attached by the user on the UI are stored in the database. So, this is the overall architecture of the system where all the platforms are linked together to perform a given task (i.e. suspect prediction).

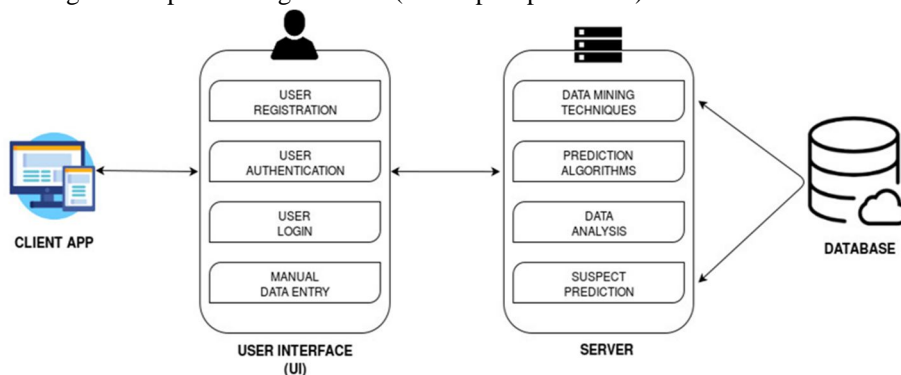


Fig. 1 Crime Investigation architectural diagram

Here in the above figure 2 describes about the activities performed by user from the start step (i.e. opening of the application) to the end step (i.e. closing of the application). Opening of the application is very first step performed by the user then the second step is the authentication step which consists of sign up and login processes. Then the user (i.e. officer) choose the required feature and enters the related information. The next activity performed is to choose service and analysing the data using certain data mining techniques and after that suspect prediction task is performed using certain predictive algorithms. At last the result of the suspects is displayed on the main user interface (UI). So, these are all the activities which are performed by the user on the user interface (UI).

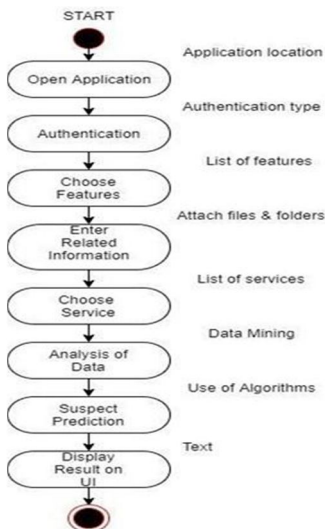


Fig. 2 Crime Investigation flow diagram



## V. CONCLUSION

Traditional crime investigation processes need a lot of talented workforce and work. There's a lack in the use of technology for a sensitive domain like a crime investigation. Therefore, crime investigation has become a time-consuming method. Data mining is the process of extracting valuable data or knowledge from massive data sources. A great deal of experience is collected throughout the crime investigation method, and solely useful information is needed for analysis. Therefore, data mining can be used for this purpose. The selection of specific data mining techniques has a more considerable influence on the results obtained. This is often the main reason behind the performance comparison and range of best performing data mining algorithm.

## REFERENCES

- [1] Dr. Ritu Bhargava, Pramod Singh, Rameshwar Singh Sangwa, "ANALYSIS OF CRIME DATA USING DATA MINING ALGORITHM"
- [2] K.S.N.Murthy, A.V.S.Pavan Kumar, Gangu Dharmaraju, "ANALYSIS OF CRIME DATA USING DATA MINING"
- [3] R.Sujatha, Dr.Ezhilmaran, "A COMPARATIVE STUDY ON PREDICTION OF CRIME PATTERNS"
- [4] Prajakta Yerpude, Vaishnavi Gudur, "PREDICTIVE MODELLING OF CRIME DATASET USING DATA MINING"
- [5] S.R.Deshmukh, Arun S. Dalvi, Tushar J .Bhalerao, Ajinkya A. Dahale, Rahul S. Bharati, Chaitali R. Kadam, "CRIME INVESTIGATION USING DATA MINING."
- [6] Mohammad Reza Keyvanpoura, Mostafa Javidehb, Mohammad Reza Ebrahimi, "DETECTING AND INVESTIGATING CRIME employing DATA MINING: A GENERAL CRIME MATCHING FRAMEWORK."



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)