

# Criminal Investigation Tracker with Suspect Prediction

Dr. C. S. Shinde<sup>1</sup> Miss. Priyanka More<sup>2</sup> Miss. Sandhya Sonkamble<sup>3</sup> Miss. Nayana Patil<sup>4</sup>  
Miss. Krutika Kulkarni<sup>5</sup>

<sup>1,2,3,4,5</sup>Department of Information Technology

<sup>1,2,3,4,5</sup>Dr. J. J. Magdum College of Engineering, Jaysingpur. (Shivaji University, Kolhapur.), India

**Abstract**— Crime is a violation against the humanity that is often accused and punishable by the law. Criminology is a study of crime and it is interdisciplinary sciences that collects and investigate data on crime and crime performance. The crime activities have been increased now-a-days and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are voluminous data of crime exist. So, we need methodologies to predict crime. Data Mining provides classification technique for this purpose. Classification in crime is mainly used to identify the patterns in crime and also used to predict the crime. Classification is a technique of data analysis that is used to extract and predicts future trends in data based on similarity measures.

**Keywords:** Criminal Investigation Tracker, KNN, Decision Tree

## I. INTRODUCTION

Crimes are the significant threat to the humankind. There are many crimes that happens regular interval of time. Perhaps it is increasing and spreading at a fast and vast rate. Crimes happen from small village, town to big cities. Crimes are of different type – robbery, murder, rape, assault, battery, false imprisonment, kidnapping, homicide. Since crimes are increasing there is a need to solve the cases in a much faster way. The crime activities have been increased at a faster rate and it is the responsibility of police department to control and reduce the crime activities. Crime prediction and criminal identification are the major problems to the police department as there are tremendous amount of crime data that exist. There is a need of technology through which the case solving could be faster.

In recent years, security intelligence agencies have placed great emphasis on big data and its applications. In the field of criminal investigation criminal investigation experts pay special attention to the collection and analysis of criminals and network information related data. Analyzing of a large number of such data can help criminal investigation departments to make effective decisions on the prevention of crime. According to whether the investigation uses information means, the investigation can be divided into traditional investigation and information-led investigation. A crime is an action which constitutes an offence and is punishable by law. analyzing and identifying the hidden crime patterns are the major problems to the police department as there are voluminous data of crime exist.

## II. RELATED WORK

Shiju Sathyadevan, Devan M. S et al., (2014) [1] predicted the regions which have high probability for crime occurrence and visualized crime prone areas. The authors classified the data using the Naive Bayes classifiers algorithm which is a

supervised learning as well as a statistical method for classification and has provided 90% accuracy.

Ryan Heart field, George Loukas et al., (2016) [2] predicts the rate of crimes that occurs due to Semantic Social Engineering Attacks and explores the feasibility of predicting user susceptibility to deception-based attacks. The authors have predicted using logistic regression and a random forest prediction model, with the accuracy rates of .68 and .71, respectively.

S. Sivaranjani, S. Sivakumari et al., (2016) [3] used various clustering approaches like the K-Means clustering, Agglomerative clustering and Density Based Spatial Clustering with Noise (DBSCAN) algorithms are used to cluster crime activities in Tamil Nadu. The performance of each clustering algorithms is evaluated using the metrics such as precision, recall and F-measure, and the results are compared. Based on the above metrics, DBSCAN algorithm gave the best results compared to the other two selected algorithms.

## III. DATASET

In data collection step we are collecting data from different web sites like news sites, blogs, social media, RSS feeds etc. The collected data is stored into database for further process. Crime data is an unstructured data since the no of field, content, and size of the document can differ from one and other. The crime dataset consists of crime information. The crime dataset consists of crime instances and attributes. The attribute crime type consists of different crime types like murder, rape, robbery, burglary and so on.

## IV. IMPLEMENTATION

### A. KNN

The KNN classification searches through the dataset to find the similar or most similar instance when an input is given to it. The input to KNN is the query i.e. the attribute values of crime dataset. Based on that query KNN gives result that assist to analyze the large crime database and also helps in predicting the crime future in various cities. It draws crime patterns for various cities.

KNN method stores all available objects and classifies new objects based on the similarity measure. It is used for criminal identification by considering the past crimes and discovering similar crimes that match the current crime based on number of nearest neighbors matched. The attributes crime type and crime city are considered as crime input attributes for KNN. All the attributes in the crime dataset can be considered but for the sake of brevity only two attributes are considered here. KNN method searches for nearest neighbor of the input values and filter those values from the dataset. The result of this method helps to retrieve data from the database and also assists in understanding the crimes.

1) *Algorithm: KNN classification*

Input:

- 1) A finite set D of points to be classified,
- 2) A finite set T of points,
- 3) A function  $c: T \rightarrow \{1, \dots, m\}$ ,
- 4) A natural number k.

Output:

A function  $r: D \rightarrow \{1, \dots, m\}$

Method:

- 1) Begin
- 2) For each x in D do
- 3) Let  $U \leftarrow \{\}$
- 4) For each t in T add the pair  $(d(x, t), c(t))$  to U;
- 5) Sort the pairs in U using the first components;
- 6) Count the class labels from the first k elements from U;
- 7) Let  $r(x)$  be the class with the highest number of occurrences;
- 8) End For each
- 9) Return r
- 10) End

B. *Decision Tree*

For prediction we are using the decision tree concept. A decision tree is similar to a graph in which internal node represents test on an attribute, and each branch represents outcome of a test. The main advantage of using decision tree is that it is simple to understand and interpret. The other advantages include its robust nature and also it works well with large data sets. This feature helps the algorithms to make better decisions about variables.

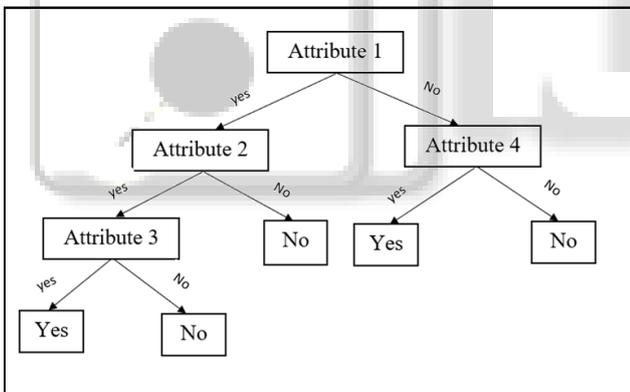


Fig. 1: Decision Tree

The working of decision tree seems to be little confusing but it's really easy. If first condition is satisfied then we check the next case and if the first condition itself is not satisfied then there is no need to check the rest. So, the series of questions and their answers can be organized in the form of a decision tree. The tree has three types of nodes:

- A Root node, that has incoming edges and zero or more outgoing edges.
- Internal nodes, each of which has one incoming edge and two or more outgoing edges.
- Leaf node or end node, each of which has exactly one incoming edge and no outgoing edges.

This supervised machine learning technique builds a decision tree from a set of class labelled training samples and by using this tree, tests the new samples. It is a predictive model which uses a set of binary rules to calculate the class value. The tree determines:

- Which variable to split at a node.
- Decision to stop or split.
- Assign terminal nodes.

V. CONCLUSION

An endeavor has been made in this paper to propose a tracker system framework which sufficient to recognize the criminal through the dialog with the victims to help the law enforcement activities. Precision and accuracy of the proposed system is highly dependent on the perception of the eye witnesses.

REFERENCES

- [1] Shiju Sathyadevan, Surya Gangadharan. S, (2014), "Crime analysis and prediction using Data Mining", First International Conference on networks and Soft Computing, Kerala, India.
- [2] Heartfield, Ryan, George Loukas, and Diane Gan, (2016), "You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks." IEEE Access 4, University of Greenwich, London.
- [3] Sivaranjani, S., S. Sivakumari, and M. Aasha, (2016), "Crime prediction and forecasting in TamilNadu using clustering approaches." Emerging Technological Trends (ICETT), International Conference on. IEEE, University of Coimbatore, India.
- [4] Mary Shermila A, Amrith Basil Bellarmine, Nirmala Santiago, (2018) "Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach", IEEE 2nd International Conference Trends in Electronics and information (ICOEI 2018), Chennai, Tamil Nadu.
- [5] Lawrence McClendon and Natarajan Meghanathan (2015), "Using machine learning algorithms to analyse crime data." Machine Learning and Applications: An International Journal (MLAIJ).