

## Crime pattern detection using historical data

### Problem statement

Crimes now days are increasing day by day and with different level of intensity and versatility. 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 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 prediction strategies can be applied on historical data available in the police records by examining the data at various angles like reason of crime, frequency of similar kind of crimes at specific location with other parameters to prepare model the crime prediction. 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.

### Background

Many researchers have worked on the same techniques to predict the future crime based on certain input parameters. The approaches used by them to evaluate the prediction models include: artificial neural network, support vector and time series analysis etc. We with help of crime data modeling the crime pattern can be detected or worked upon. Crime data analysis can even speed up the resolution to the court cases with ease and timely. The previously unknown information can be better extracted from the historical data. Some researcher used clustering algorithms for crime analysis. Clustering approach can work well in this case as the crime is totally random in nature and the available data cannot effectively predict the future incidences? In this case the clustering process can find similarities between crimes in increasing size pattern and in remaining database. Such types of crime prediction approach work on pattern of crimes in terms of similarity of attack or crime such as time of incidence, the methodology of crime is same etc.

### Methodology

Crime can be better predicted if we use clustering approach rather than classification due random nature of crimes. Clustering approach may work better.

Following steps may apply for the crime analysis:

1. **Collect the data:** In this phase the data from various sources is collected from various government sources, social media platform about the incidence like facebook, blogs etc. It may expect that the data received may be in unstructured form with different types and size. So oops approach may be better to extract the information. Explicitly NoSql can be used as it run without schema detail.
2. **Classification:** The standard algorithm useful for classification which may be applied is Naïve bayes Classifier. This classifier will the probability of falling into different classes of crimes and the crime predicted to belong to specific class the probability is highest. Naïve bayes classifier is more efficient as compared to regression in terms of convergence rate.
3. **Identify pattern:**  
Next phase in the methodology is to find the sequence of crimes which are similar in nature and belongs to same class. Such pattern may be identified as suggested in literature through the apriori algorithm. Apriori algorithm work on principle of association rules. Association rules highlight the trend the crime pattern database. Pattern identification outcome may be

location based pattern detection. Pattern information may help the government or police department to effectively speedily work on resolving the criminal cases.

4. Prediction the Crime: For the prediction outcome the decision trees are used. Here in decision tree each internal node has test for occurrence on an incidence and outcome is normally yes or no and based on the outcome again the next level of internal node has another question test and so on to reach to final decision for prediction.

For example the decision tree table is given below.

Area sensitivity	Notable event	VIP presence	Criminal group	crime
Yes	Yes	Yes	No	yes
Yes	Yes	No	Yes	no
No	No	No	Yes	no
Yes	No	No	No	no
Yes	Yes	Yes	yes	yes
No	Yes	No	No	no

From this table the decision tree can be easily drawn and hence the prediction can be very easily done. In Such a way supervised machine learning approach can be exploited to build the decision tree from labeled training data set.

5. Visualization or presentation of result: The results can be visualized using appropriate graphs or maps showing sensitive areas of having high probability of crimes.

## Experimental Design

The concept to co reference resolution is suggested through the literature to find the referenced entity in a given crime text. Coreference is used to describe the entity or person which is referred by two different crime texts. For example: Reema said she will come i.e. here "she" refers to person "Reema". Likewise all referenced entities in a text are extracted. Below example shows the working of Coreference concept

Input: A pillion bike rider snatched away a gold mangalsutra worth Rs 85,000 of a 60-year-old woman pedestrian in sector 19, Kharghar on Friday. The victim, Shakuntala Mande, was walking towards a vegetable outlet around 9.40am, when a bike came close to her and the pillion rider snatched her mangalsutra. A robbery case has been registered at Kharghar police station.

Sample output of Coreference Resolution

```
[ "The victim" -> "Shakuntala Mande"
  "her mangalsutra" -> "a gold mangalsutra"
  "Kharghar" -> "Kharghar on Friday"
  "her" -> "Shakuntala Mande"
  "the pillion rider" -> "A pillion bike rider"]
```

The figure-1 shows the regions that has more possibility for crime. Heat maps are effectively used here to show the intensity level of crime occurrence. Heat maps give category based color images with Gradient range. We can analyze data of our interest. So after identifying the probable regions the prevention mechanism can appropriately applied.



Figure-1

This representation is outcome of analysis on the existing crime records. These crimes are majorly related to robbery, murder, burglary which is having higher probability where proper security is not implemented.

This type of analysis on historical data can also help in determining the criminal profile based on the characteristic behavior inferred from the data. It can help the investigator to accurately predict the profile of unknown criminals.