# Criminal Shortlisting and Crime Forecasting based on Modus Operandi

Malith Munasinghe[#1], Shanika Udeshini[#2],.Harsha Perera[#3],Ruvan Weerasinghe[#4]

[#]*University of Colombo School of Computing, No.35, Reid Avenue, Colombo 07, Sri Lanka.*

[1]mpmunasinghe@gmail.com
[2]udeshini14@gmail.com
[3]harsha.perera91@gmail.com
[4]arw@ucsc.lk

***Keywords*—Modus Operandi, Machine Learning, Criminal Profiling**

## INTRODUCTION

Crime detection and prevention is a very crucial wok which is in the hands of police, law enforcement agencies and local government. Experts in crime analyzing use crime scene evidences to capture unique ways a criminal has acted during a crime, which is also called as *Modus Operandi(MO)*. Using *MO* as the main focus, the efforts taken in this research is to shortlist and predict criminals and criminal activities with the support of machine learning based algorithms.

## BACKGROUND

According to the published data from Sri Lanka police for the first quarter of 2012 out of 13173 true cases reported 9423 cases are under pending status. Inability to directly identify possible suspects, lack of efficiency in looking through past records and other physical and technological problems are incorporated with producing such a low rate of success. Although fingerprints can be used, it is not always possible to trace fingerprints left at the crime scene. Therefore the need of a supportive property to direct the search on identifying possible suspects is there.

As in criminological terms *Modus Operandi* is a key supportive factor which is used by criminologists to identify potential suspects based on how the crime was organized, executed and gotten away. Even though, Local police stations follow a method to forecast possible crimes based on a map and clock approach a centralized or an interconnected approach is not there for this. Thus it fails to capture many of the important MO attributes that could have lead into solving the case.

## METHODOLOGY

During the first phase of the methodology, we studied available literature on MO classifications and feature extraction and carried out interviews with crimes analyzing experts from Police College, Kaluthara and Criminal Records Division in Sri Lanka to incorporate their expertise in M.O feature extraction. Iteratively while evaluating the degree of efficiency of the MO extraction features we created the final revised MO feature extraction form which is capable of extracting 20 attributes in the domain of *house breaking and theft* crimes in Sri Lanka.

Data collection was done in two stages. Stage 1 was collecting crime data sample using selective sampling technique by collecting past 10 years GCR records data from Pettah and Cinnamon Gardens police stations. Stage 2 was collecting data of the most active criminals in Sri Lanka. Altogether there were 395 crime records under the house breaking and theft category from 60 different police stations by 201 criminals after cleaning the data there were 322 crimes from 57 different police stations by 105 distinct criminals for the model implementation.

## MODEL IMPLEMENTATION

Since there were no predefined MOs we had to go with partitioning or clustering. We chose the clustering algorithm and number of clusters we want. For the evaluation we used the clusters as black box when inserting a new crime to see whether the criminal of that crime is matched with the criminals in the cluster which it belongs to.

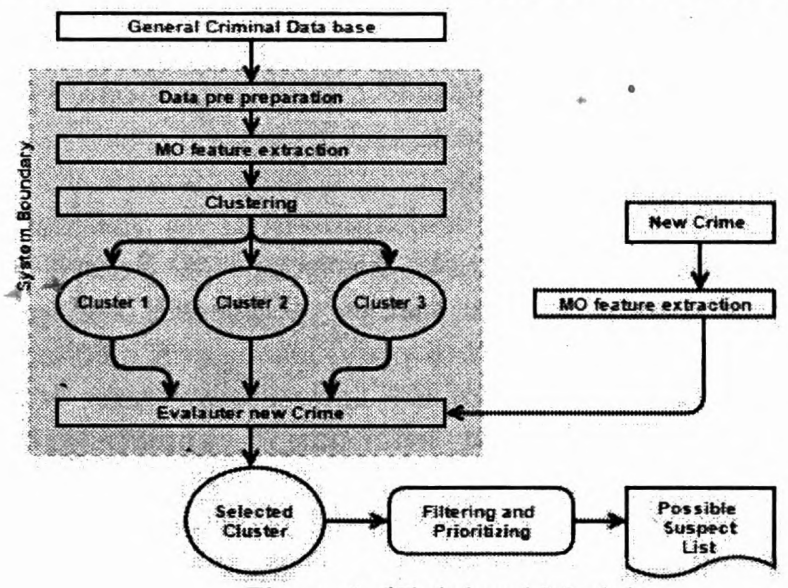


Figure 1: High level Model

In this model first we separate all 322 crimes into 5 sets. We selectively separated 58 crimes for the first set called world set and randomly assigned 66 crimes to the other remaining four sets. Purpose of the world set was to keep a separate untouched data set for the final evaluation. Folds acted as iterative training and evaluation data sets. We had four iterations for a single evaluation where this evaluation model suggests as cross validation.

## RESULT AND DISCUSSION

Based on the evaluation results we choose Hierarchical clustering with Average linking over K means and Hierarchical with mean linking as the best optimal clustering algorithm in the range of 55 to 65 clusters. Based on those two criterions we evaluate the world set and the results are 79.31034 of accuracy for 55 clusters, 77.58621of accuracy for 60 clusters and 75.86207 of accuracy for 65 cluster.

If tailored carefully this methodology can be applied to to Latin American, Asia Pacific region and other related countries that are using similar procedures to forecast crime and criminal short listing.