DATA MINING PRACTICES FOR EFFECTIVE INVESTIGATION OF CRIME
Prof. Hanmant N. Renushe#, Prof. Prasanna R. Rasal##, Prof. Abhijit S. Desai###
BVDU, Yashwantrao Mohite Institute of Management, Karad [M.S.], India#1, 2, 3

Abstract: This research paper highlights the importance of data mining technology to design proactive services to reduce crime incidences in the police stations jurisdiction. Crime investigation has very significant role of police system in any country. Almost all police stations use the CIPA system to store and retrieve the crimes and criminal data and subsequent reporting. It become useful for getting the criminal information but it does not help for the purpose of designing an action to prevent the crime. It has become a major challenge for police system to detect and prevent crimes and criminals. There is no any kind of information is available before happening of such criminal acts and it result into increasing crime rate. The presented paper highlights the use of data mining techniques for effective investigation of crimes.

Keywords: Crime, CIPA, CCIS, NCRB, Investigation, CrPC.

I. INTRODUCTION
Police plays an important role in civil administration in India. The Constitution of India assigns a responsibility to maintain the law and order in the country. In 1986 Govt. of India created National Crime Record Bureau (NCRB). Under NCRB the state crime record bureau [SCRB] for state and District crime Record Bureau [DCRB] for districts has been created. In order to making use of information technology, The Government of India designed Crime Criminal Information System [CCIS] to store and retrieve crime and criminal records. To provide the input to CCIS, the Common Integrated Police Application [CIPA] was also designed. In order to help the investigation officer [IO] the system needs to be designed in such a way that the information required by the IO should get on the figure tips.

II. POLICE DEPARTMENT IT INFRASTRUCTURE CURRENT SCENARIO
To understand current scenario of crime investigation, we need to know technological usage by the police force of the state. The Head of state police is Director General of Police [DGP]. The state is divided into administrative units called as Districts. A group of districts called as a Region and Head for each region is Deputy Inspector General of Police [DIGP]. Superintendent of Police [SP] is head for district and is assisted by Additional Superintendent of Police [Addl. SP] and Deputy Superintendent of Police [DySP] in each district.

Maharashtra, a highly industrialized State with large urban conglomerates, has adopted Commissionerates system for policing its large cities. The State has 10 Commissionerates and 35 district police units.

In order to make use of information technology Maharashtra police implemented the computerized system called CIPA at police station and CCIS as districts.

A. Common Integrated Police Application [CIPA]
CIPA is aimed at building the basic infrastructure and mechanisms for the Crime and Criminal Information System, based on CrPC, which is uniform across the country, from Police Station level onwards. CIPA being a National project is to be implemented in a time-bound manner from police station level onwards for computerization of police records and

ISSN: 2229-6093
The use of IT in their functioning on a uniform basis throughout the country.

The national level Central CIPA Implementation Committee comprising of Director, NCRB and representatives from the Ministry of Home Affairs (Police Modernization and Union Territories Divisions), NIC, National Institute of Criminology and Forensic Science and States, has been constituted to monitor the implementation.

State Crime Records Bureau and State Police Training Academies are conducting State Specific courses in this connection with the assistance of NIC. NCRB has introduced two advanced courses on CIPA in its training calendar for resource persons, who in turn will impart training and attend to trouble-shooting in the States.

B. Crime Criminal Information System [CCIS]

In order to make use of Information Technology the Government of India has designed Crime Criminal Information System [CCIS] to store and retrieve crime and criminal records. This system has been upgraded to CCIS Multi-Lingual web-enabled (CCIS MLe) in the year 2005 with facility for 5 regional languages i.e. Marathi, Gujarati, Tamil, Kannada and Gurmukhi, besides English and Hindi. Feature of crime analysis through data warehousing has also been added. The application has been web-enabled so that the field level investigating and supervisory officers can access the CCIS MLe database at National and State Levels through internet; anywhere - anytime.

C. Crime and Criminal Tracking Network System [CCTNS]

The CCTNS is under implementation as a “Mission Mode Project” [MMP] and adopted the guidelines of the National e-Governance Plan [NeGP].

The CCTNS aims at creating a comprehensive and integrated system for enhancing the efficiency and effectiveness of policing at all levels and especially at the PS level through adoption of principles of e-Governance. CCTNS operates through the creation of a nationwide networked infrastructure for evolution of IT enabled state-of-the-art tracking system around “investigation of crime and detection of criminals” in real time, which is a critical requirement in the context of the present day internal security scenario. The scope of CCTNS spans all 35 States and Union Territories and covers all PSs (14,000+) and all Higher Police Offices (6,000+) in the country. The CCTNS project includes connectivity of police units at various levels within the States – police stations, district police offices, state headquarters, SCRB and other police formations and States, through state headquarters and SCRB, to NCRB at GOI level.

D. Organized Crime Intelligence Systems [OCIS]

The growth and spurt in the Organized Crime, especially Mafia activities and activities by the terrorists and insurgent groups in the recent times have attained considerable magnitude that need to be tackled with greater professional expertise and effectiveness which requires inter-agency co-ordination in sharing criminal intelligence. For achieving this, NCRB has developed an Organized Crime Intelligence System [OCIS] for collecting, storage and retrieval of information on organized crime and criminals and provide co-ordination amongst different police forces at State and Central level.

NCRB is running the OCIS since November 2005. NCRB has created a database on Organized Criminal Gangs on criminal activities which is being refined. Data Security Systems are also in place. Nodal officers of the rank of IG or equivalent have been designated for organized crime units in all the states and UT’s and OCUs are functioning in 8 pilot states.
A pilot project has been taken up in states of Haryana, Punjab, Jammu & Kashmir, Uttar Pradesh, West Bengal and Delhi on the activity “Theft of automobiles” and Andhra Pradesh and Maharashtra for “Sale and Purchase of Women and Children on Prostitution and maid services”.

E. Motor Vehicle Coordination System [MVCS]

The MVCS has been implemented in all States/UTs. The main objective of the software is to provide information to public, police and other agencies regarding the recovered/lost motor vehicle. The modified version has already been released to all States/UTs/MV counters in the month of December 2006. As and when request received from any States/District/UT for opening of MV Counter,

F. POLNET

The POLNET is a Govt. of India project, being implemented through the Directorate of Co-ordination of Police Wireless (D.C.P.W.), New Delhi and Directorate of Police Wireless Maharashtra State, Pune. Under this scheme, Maharashtra Police has been allotted 33 units of V-SAT and 850 Remote Subscriber Units. When fully operational, POLNET will be utilized for the transmission of Crime & Criminal data, Fax transmission, Voice Communication, Image-Photographs & Finger print transmission throughout the country.

I. Finger Print Bureau [FPB]

In Maharashtra, Finger Print Bureau was established in 1899 at Pune under the control of Inspector General of Police, Bombay Presidency. At present there are 3 regional Bureaux at Mumbai, Nagpur and Aurangabad, which are engaged in recording and searching finger impression slips of arrestee and retrieval of chance prints found at scenes of crime. Advance Computerized Finger Print Analysis and Criminal Tracing System [FACTS] have been installed and have become functional from July 2004. A central server installed at Pune is linked with 41 Police Units. Finger Print Data of 2, 00,000 criminals is updated on the FACTS. More than two hundred persons are working at 4 bureau and 41 police Units, continuously with modern equipments to handle Finger Print work for investigation.

J. Talash Information System [TIS]

The TIS is an integrated system for linking Arrested, Wanted, Missing, Traced, Kidnapped, Deserters, Escapee, Unidentified dead bodies and unidentified persons based on attribute data available from various police agencies in the country. It is proposed to further integrate attribute-based search with photograph based facial search. To keep pace with the changing time the package is redesigned using windows platform so as to enable users to use the software more effectively.

This new version of the package is with value addition features like photograph scanning and export and import of data electronically etc.

K. Telephone Call Interception System [TCIS]

Ministry of Home Affairs has invited bids for setting up Communications Monitoring facility at all the State capitals, on turnkey basis. Communications Monitoring facility will be setup at all the State capitals with facilities to monitor Voice Calls, SMS & MMS, GPRS and FAX communications on Landlines (PSTN), CDMA and GSM networks. The system across all the States will be compatible and interoperable. The TCIS can not only listen to phone conversation, but can also track down precise location on a map and match voice with known suspects before the call is complete. The system, to be set up by next April, will also be able to analyze the
calling pattern of a target to identify correlated calls and record the locations of all mobile devices. It will also have the capability to integrate the data on a digital map including satellite imagery and software — that can analyze millions of calls and their locations and spot the possible hideouts being used by a suspect. An integrated voice recognition system will enable intelligence officers to identify the voices of people the target is talking to.

### III. DATA MINING FOR INVESTIGATION OF CRIME

Data mining is basically used to find out unknown patterns from a large amount of data. There are popular tools of data mining to rub data mining algorithms. There are two approaches for the implementation of data mining, first is to copy data from data warehouse or source and mine it. Other approach is to mine the data within a data warehouse. There are various data mining techniques available as follows:

**Classification** is used to classify database records into number of predefined classes on criteria. The data with sharing common properties are specified into predefined classes.

**Clustering and segmentation** is used to segment a database into subsets, or clusters based on set of attributes. It is a method to group data into classes with identical characteristics in which the similarity of intra-class is maximized or minimized.

**Association** identifies affinities/associations among the collection of data as reflected in the examined records. A result is patterns describing rules of association in data.

**Decision Tree** is predictive model that can be viewed as tree, each branch is a classification question and leaves of the tree are partitions of data set with their classification. It divides data on each branch point without losing any of the data. The number of churners and non churners is conserved as we move up or down the tree. ID 3, C4.5, CART and CHAID are some algorithms used in this technique.

**Neural Networks** are biological systems that detect patterns, make predictions and learn. The artificial neural networks are computer programs implementing sophisticated pattern detection and machine learning algorithms on a computer to build predictive models for historical databases.

In order to achieve the goal there are number of commercial software packages available and each offers a combination of relevant features. For the research paper we have used SPSS Clementine.

**SPSS Clementine**

SPSS Clementine utilizes a visual approach to data mining with an emphasis on a person with domain knowledge performing the analysis. It combines learning algorithms and statistical techniques with the facilities to manipulate, display and visualize the data.

Data mining tools in Clementine helps to solve a wide variety of business and organizational problems. The data and modeling tools in Clementine reside in palettes, the area below the stream canvas. Each tab contains groups of nodes that are a graphical representation of data mining tasks, such as accessing and filtering data, creating graphs, and building models.

Clementine includes a number of machine-learning and modeling technologies, which can be roughly grouped according to the types of problems they are intended to solve.

**Predictive model:** it includes decision trees, neural networks, and statistical models.

**Clustering model:** focus on identifying groups of similar records and labeling the records according to the group to which they belong. Clustering methods include Kohonen, k-means, and TwoStep.

**Association rules:** Associate a particular conclusion with a set of conditions.
Screening model: It can be used to screen data to locate fields and records that are most likely to be of interest in modeling and identify outliers that may not fit known patterns. Available methods include feature selection and anomaly detection.

The following image shows the SPSS Clementine environment with the model design.

Image1. SPSS Clementine 11.1 Environment for the study

For the research paper we have taken the incidences occurred under Faraskhana police station of Pune city. Following table shows the records

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Major Head</th>
<th>Accused Name</th>
<th>Offence Place</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>KURAV SATESH THAMBE</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>SACHIN DINESH KALOKE</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>SALIM EDILI SHIDAR</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>SANTWAN BHOOMIR BHUSNARE</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>BAKHET MAHDUR MAHIN</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>WUSAWAT PANCHU SHOLMA</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>AKOL NANDIR POYWAR</td>
<td>SAIWALI ROAD</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>KURANGIR LALI</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>RAJESH NAUSHIKH MORE</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>AKOL GHADE</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>AKBIR RAHIM BHINDI</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>VIJAY DHONDAI SOLKHAR</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>UNKNOW</td>
<td>RAWHAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>OTHER</td>
<td>VIJAY DHONDAI SOLKHAR</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>SHANKARUN,BHUSLE</td>
<td>BHUJAR PETH</td>
</tr>
<tr>
<td>PUNE CITY</td>
<td>MV THEFT</td>
<td>PROD NAVAPASA SHANDE</td>
<td>BHUJAR PETH</td>
</tr>
</tbody>
</table>

The above table shows various crimes committed over the time in the jurisdiction. The graphical representation is much more effective therefore we added distribution of crimes location wise in the model.

Graph1. Number of crime committed on the offence place

Network structure used to show the association between the variables. The network structure uses the link between the nodes, the links such as Strong link, Medium link and Weak link. Strong link shows the more association between nodes whereas Weak link shows less association between nodes.

Image3: offender structure of Major Head Vs Offence place

The above network structure shows 31 medium links and remaining weak links i.e. the occurrence of MOTOR VEHICLE THEFT in BUDHWAR PETH is 31 times therefore focused area for the MOTOR VEHICLE THEFT is BUDHWAR PETH and requires more patrolling in the shown area.
IV. CONCLUSION

Crime Investigation is one of the important tasks of police organization in the India. In today’s IT enabled era many techniques are available for crime prevention and investigation.

Data mining practices is one aspect of crime investigation, for which numerous technique are available. In the present study the researchers have used crimes occurred under Faraskhana police station of Pune city for the year 2011

There is huge gap between number cases registered and completion of investigation, due to many reasons which are stated below.

- Technology Usage: Police must use the intelligence technology for investigation. Presently there are many technological system are available but are not used effectively
- Innovative Practices Training [IPT] must be provided to the investigation personnel on regular basis.
- Common Platform must be formed for all available computerized systems for effective investigation and prevention of crime.

V. ACKNOWLEDGMENT

The researchers are grateful to the authors, writers and editors of the books and articles, which have been referred for preparing the presented research paper. It is the duty of the researchers to remember their parents whose blessings are always with them.

VI. REFERENCES