

Importance of Data Mining Time Series Technique in Crime and Criminal Investigation: A Case Study of Pune Rural Police Stations

Hanmant N Renushe
BVDU YMIM Karad
Karad [MS], India

Dr. Milind J. Joshi
Shivaji University, Kolhapur
Kolhapur [MS], India

Prasanna R. Rasal
BVDU YMIM Karad
Karad [MS], India

ABSTRACT

This research paper highlights the importance of data mining technology in crime and criminal investigation to reduce crime incidences in Pune Rural Police Stations. Crime investigation has very significant role of police system in any country. Pune Rural Police Stations Crime and criminal record is stored and retrieved using CIPA and CCIS and 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 Time series technique to forecast crime in a specific area and seasonal crime pattern which can be useful for crime prevention.

Keywords

Time Series, Crime, NCRB, Investigation, CCIS, CIPA, CrPC.

1. 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. Police force is in insufficient ratio as compare to population. It leads to increasing graph of crime in nation; another reason is lack of use of information technology in investigation and in prevention. Keeping these things in mind govt. of India designed G2G model. 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. CIPA software install in every police station, CIPA is only limited to the informative purposes and doesn't forecast or shows any seasonal crime in specific police station region, therefore it need to be advancement in existing system such as use of data mining technology in CIPA as well as CCIS system.

2. PUNE POLICE CURRENT SCENARIO

To understand current scenario of crime detection, we need to know police structure and hierarchy, technological usage of police force.

2.1 Pune Police System

The motto of Maharashtra Police is. 'सद्रक्षणाय खलनिग्रहणाय'

It means that Maharashtra Police is committed to PROTECTING THE RIGHTEOUS AND CONTROLLING & ANNIHILATING THE EVIL. 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 Rural Police Stations system for policing its large cities. The State has 10 Rural Police Stations and 35 district police units. Pune Rural has comes under Rural Police Stations system, Pune police is headed by Commissioner supported by joint Commissioner and addl. Commissioner, and the 7 DCP headed by each zone, special branches such as head quarter, Traffic, Crime and Spl. branch. Along with no. of Police Inspector, Asst. Police Inspector, Police Sub Inspector and Constable are working.

2.2 Common Integrated Police Application

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 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.

2.3 Crime Criminal Information System [CCIS]

In 1986 Government of India created National Crime Record Bureau [NCRB]. Under NCRB the State Crime Record Bureau [SCRB] for States and District Crime Record Bureau [DCRB] for Districts has been created. 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.

Information Technology audit of computerization in Police Department noticed that Crime and Criminal Information System and Common Integrated Police Application meant for crime data storage and retrieval did not deliver the desired output. The deficiencies in input controls and supervisory checks have resulted in incomplete and incorrect database, making the systems unreliable and thus not useful. No tangible benefits have thus accrued so far from the computerization.(CAG Report 2009)

3. TREND ANALYSIS FOR CRIME FORECASTING

Understanding and predicting Crime Incidences is vital to police officer to maintain law and order. While forecasts are never perfect, they are necessary to prepare for actual crime incidences. In order to maintain law and order and effective control in the police station area, accurate crime incidences forecasts are imperative.

Business Intelligence [BI] is a concept and method to improve business decision making by using fact-based support systems. Business Intelligence often aims to support better business decision-making.

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 approached to the implementation of data mining, first is to copy data from data warehouse 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 nonchurners 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.

The Microsoft Time Series Algorithm is a novel forecasting algorithm it is a hybrid of auto regression and decision tree technique.

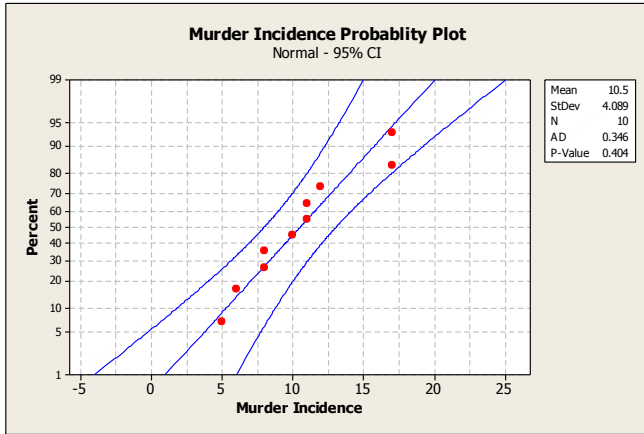
TABLE 1
TOTAL CRIME HAPPENS DURING THE JANUARY
2010- OCTOBER 2010

Month / Crime Type	Murder	Att. To Murder	RIOT	Hurt	Rape	Dacoits	Robbery	Theft
Jan	5	7	27	52	4	2	11	139
Feb	10	8	38	76	5	5	7	118
Mar	6	4	37	81	3	0	20	157
Apr	8	11	51	85	6	1	14	165
May	17	7	36	76	9	2	9	137
Jun	12	9	30	69	4	1	17	126
Jul	11	8	29	41	7	3	15	143
Aug	8	6	42	77	5	7	11	122
Sep	11	5	29	70	6	5	13	116
Oct	17	12	45	72	7	2	13	132

Probability Plot is used to determine how well a specific distribution fits your data and to obtain estimated population parameters and percentiles. Compare the fitted line to the actual data in order to determine how well the distribution fits. An Anderson-Darling [AD] statistic and a p-value are calculated to evaluate the distribution fit of your data. Choose value for AD is 0.10 and p-value is 0.05. In the Probability plot, the curve in the tails indicates a skewed distribution and point far away from the line indicate outlier.

Probability plots typically contain: Y-axis representing a transformed percentile scale, X-axis representing Murder Incidences, Plot Symbols representing the actual cumulative distribution observed in the sample. As the number of observations is 10 only, the probability plot may show even greater variation and nonlinearity therefore the AD statistics and p-values is used.

**GRAPH 1:
PROBABILITY PLOT FOR MURDER CATEGORY**



The probability plot for the Murder incidences shows that the data points for distribution fall close to the fitted normal line and within the confidence interval. Further, the AD value 0.345 is low and the p-value 0.404 is above any reasonable significance level. Together, these values suggest that the normal distribution fits both sets of the Murder Incidences well.

The Minitab 16 generate the fitted line for Crime Type Murder using the equation

$$Y_t = 6.20 + (0.78 * t).$$

The *t* represents the month during which each data point was collected.

Forecast accuracy in the crime incidences can be measured using the Mean Absolute Percent Error [MAPE].^[10]

Fitted Value is $Y_t = a + (b * t)$

Value of *a* and *b* can be calculated using

$$\sum X, \sum Y, \sum XY, \sum X^2$$

$$\sum X = 1 + 2 + \dots + 8 = 36$$

$$\sum Y = 5 + 10 + \dots + 17 = 105$$

$$\sum XY = 5 + 20 + \dots + 170 = 642$$

$$\sum X^2 = 1 + 4 + \dots + 100 = 385$$

$$\sum Y = Na + b \sum X$$

$$\sum XY = a \sum X + b \sum X^2$$

In order to calculate the value for *a* and *b* we have line equation

$$\sum Y = Na + b \sum X \text{ -----1}$$

$$\sum XY = a \sum X + b \sum X^2 \text{ -----2}$$

From equation 1 and 2

$$105 = 10a + 55b \text{ -----3}$$

$$642 = 55a + 385b \text{ -----4}$$

Multiplying equation 3 by 55 and equation 4 by 10 we get

$$5775 = 550a + 3025b \text{ -----5}$$

$$6420 = 550a + 3850b \text{ -----6}$$

Equation 6 - equation 5 we get

$$645 = 825b$$

$$b = 0.78$$

Substituting value *b* into equation 3 we get

$$105 = 10a + (55 * 0.78)$$

$$A = 6.20$$

$$a = 6.20$$

$$b = 0.78$$

Absolute Deviation = Volume [Y] – Fitted Value

Squared Deviation = Square of Absolute Deviation

MAD = Sum of Absolute Deviation / Number of Observation

MSD = Sum of Squared Deviation / Number of Observation

MAPE = Mean Absolute Deviation [MAD] / Mean Ratio

Or

MAPE = Sum of Absolute Deviation / Total Crime Incidences

Hence the for the month of Nov 2010 forecast value for Murder is as below

$$Y_{11} = 6.20 + (0.78 * 11)$$

$$= 14.78$$

$$= 15$$

This show that November 2010 may have murder incidences 15 in the Pune area

The fitted value, Absolute Deviation, Squared Deviation and Absolute Percent Error for the category Murder is shown in below

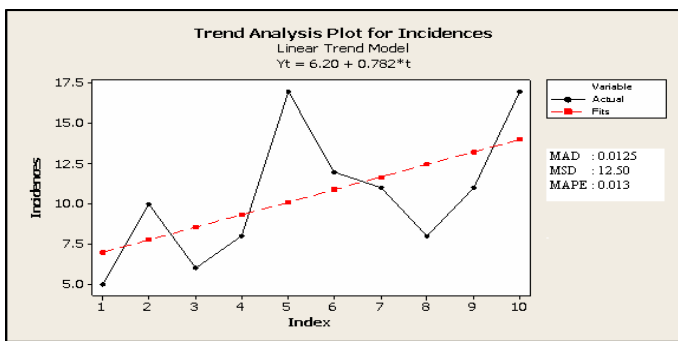
**TABLE 2:
STATISTICAL VALUE CALCULATED FOR MURDER
CATEGORY OF YEAR 2010**

Month / Crime Type	t	Murder	Fitted Value	Absolute Deviation	Squared Deviation	Absolute Percent Error
Jan	1	5	6.98	-1.98	3.9204	0.02
Feb	2	10	7.76	2.24	5.0176	0.01
Mar	3	6	8.54	-2.54	6.4516	0.01
Apr	4	8	9.32	-1.32	1.7424	0.01
May	5	17	10.1	6.9	47.61	0.00
Jun	6	12	10.8	1.12	1.2544	0.00

			8			
Jul	7	11	11.66	-0.66	0.4356	0.00
Aug	8	8	12.44	-4.44	19.7136	0.01
Sep	9	11	13.22	-2.22	4.9284	0.00
Oct	10	17	14	3	9	0.00
			SUM	0.1	100.034	0.109947
			N	8	8	8
			MA D	0.0125	12.50925	0.013743
					MSD	MAPE

MAD Stands for Mean Absolute Deviation, which is the average of absolute deviation. An absolute deviation is the absolute value of the actual crime minus the fitted value. The best fitted line should have zero MAD and Murder Category MAD value is approximately Zero value hence model is suited perfect. The trend analysis plot for murder is generated as below

**GRAPH 2:
TREND ANALYSIS FOR MURDER INCIDENCES**



Similarly Trend Analysis Plot for all Crime categories has been plotted.

**TABLE 3:
STATISTICAL VALUE CALCULATED FOR ALL CATEGORIES**

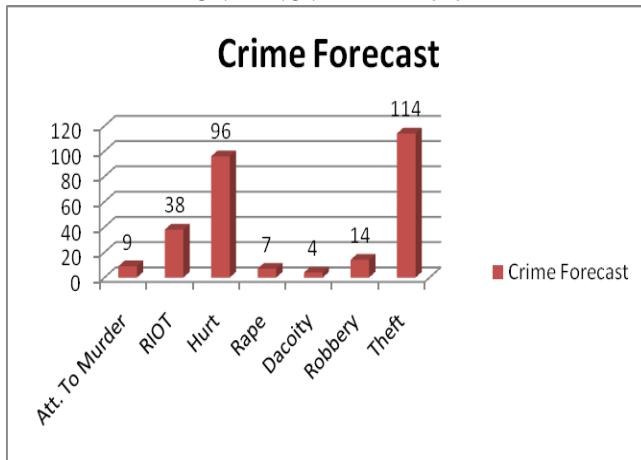
Variable/ Crime Type	ATT. To Murder	RIO T	Hurt	Rape	Dacoity	Robbery	Theft
A	6.82	34.66	43.13	4.2	1.46	12.13	135.49
B	0.16	0.31	4.72	0.25	0.24	0.15	-1.99
Crime Forecast	9	38	96	7	4	14	114
MAD	0	0.04	0.01	0.02	0.02	0.04	11.32
MSD	5.38	55.22	4	2.3	3.87	12.79	6
MAPE	0.00	0.01	0.00	0.02	0.09	0.04	0.82

Accordingly Crime Forecast has been plotted

**TABLE 4:
CRIME FORECAST OF ALL CATEGORIES FOR THE MONTH NOVEMBER 2010**

Variable/ Crime Type	ATT. To Murder	RIO T	Hurt	Rape	Dacoity	Robbery	Theft
Crime Forecast	9	38	96	7	4	14	114

GRAPH 3:
CRIME FORECAST OF ALL CATEGORY FOR THE
MONTH NOVEMBER 2010



4. CONCLUSIONS

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.

Crime forecasting is one aspect of crime investigation , for which numerous technique are available such as secular trend cyclic trend, irregular trend etc. , secular trend is used for short time forecasting. In the present study the researchers has used short time forecasting of crimes in Pune district for the month November 2010 using the crime data of the period Jan 2010 to October 2010.

From the study it has observed that actual no of crime reported during the November 2010 are approximately same as the forecasted crimes using short term forecasting. Hence the short term forecasting method is very useful in crime prevention and investigation Process.

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

- Integrated Mechanism for Investigation: The Common Integrated Crime Analysis Cell [CICAC] must be formed to help the investigation officer on requirement.
- Technology Usage: Police must use the intelligence technology for investigation. As on today they are mostly investigate with traditional way, on contrary criminals are using very sophisticated technology and often finds the loop hole, due to which, there is tremendous increase in crime ratio.
- Intelligence failure is an important issue, and it requires improving intelligence.
- Innovative Practices Training [IPT] must be provided

to the investigation personnel on regular basis.

5. 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

6. REFERENCES

- [1] <http://nitawriter.wordpress.com/2007/08/20/poor-people-to-police-ratio/>,23 December 2009 4:30 pm...
- [2] <http://ncrb.nic.in/origin.html> , 24 December 2009 2:20 pm
- [3] <http://mahapolice.gov.in/aboutus> 23 December 2009 4:50 pm
- [4] <http://punepolice.maharashtra.gov.in>
- [5] CIPA, <http://cipa.gov.in/cipa/pdf/cipabros.pdf>
- [6] SALIENT FEATURES OF CCIS MLe, <http://ncrb.nic.in/featuresMLe.pdf>
- [7] <http://www.cag.gov.in> Audit report of Govt. of India for the ended year 31 march 2009
- [8] Alex Berson, Stephan J. Smith (2004) Data Warehousing, Data Mining, & OLAP, TATA McGraw HILL Publications, New Delhi.
- [9] http://mahacid.com/agree_monthly_crime.htm
- [10] Hyndman, R.J., Koehler, A.B (2005)"Another look at measures of forecast accuracy", Monash University.
- [11] Brantingham, P. J., & Brantingham, P. L. (1984). *Patterns in crime*. New York: Macmillan
- [12] Sherman, L.W., Gartin, P. R., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology*, 27, 27–55.
- [13] Gorr, W. L., Olligschlaeger, A. M., & Thompson, Y. (2003). Short-term time series forecasting of crime. *International Journal of Forecasting*.
- [14] Liu, H., & Brown, D. E. (2003). Criminal incident prediction using a point-pattern based density model. *International Journal of Forecasting*
- [15] Gorr & Harries (2003) Introduction to Crime Forecasting. *International Journal of Forecasting*19-551-555.
- [16] Felson & Poulson 2003 Simple indicator of crime by time of day. *International Journal of Forecasting*19-595-601
- [17] Corcoran, Wilson & Ware 2003 Predicting the Geo-temporal variation of crime disorder. *International Journal of Forecasting*19-623-63