

# A Case for Data Mining in Police Administration

Singh Dilbag

Associate Professor

Department of Computer Sc.& Applications,  
Chaudhary Devi Lal University, Sirsa, Haryana, India

Kumar Omda

Research Scholar

Department of Computer Sc.& Applications  
Chaudhary Devi Lal University, Sirsa, Haryana, India

## ABSTRACT

Application of data mining in police administration may help the crime prevention and criminal tracking. The present study was aimed to assess the effectiveness of data mining in crime detection. Data of vehicle theft in Sirsa district has been analyzed to draw the inference and pinpoint the hotspots. Classification and Clustering techniques of data mining have been used through Weka 3.7.7 to inference the useful knowledge for the police administration.

## General Terms

Crime prevention and criminal tracking, effectiveness of data mining in crime detection, data mining techniques, electronic data collection.

## Keywords

Data Mining, Weka, KDD, Police administration, datasets.

## 1. INTRODUCTION

Computing technology has significantly influenced lives, business data processing as well as scientific computing. In the beginning, computer professionals were concerned with designing files to store data so that information could be explored for business and other requirements. Storage size and speed of accessing the data were the limitations. Then the era of database systems simplified to storage the data. Responsibility like declarative aspects of programs was passed on to the database administrators and the users were facilitated to queries in simpler languages such as SQL. On the hardware side, the storage cost and machine cost have been reduced to affordable amount by the time. Organizations are now accumulating a large amount of data in archival files. As a result, huge volume of data are heaping up in the electronic domes of organizations, research institutions and government departments. Before 1990's such data were dumped into archival files, but in recent years, such collection of data is being used with the help of tools of knowledge discovery and data mining. Data mining is the exploration and analysis of large data sets for discovering meaningful patterns and rules[1]. All business enterprises collect a variety of information electronically. In recent years, storage of electronic information has been grown due to automatic and computerized operations like barcoding of products and decline in data storage costs. Telecommunication houses, banks, scientific and government agencies also collects huge amount of data. The data so gathered is being treated as gold mine of information and there is potential business intelligence hidden in the large volume of data. With the help of data mining, this large data could be used for success of the business [2].

Crime detection also closely associates with data mining. In order to apply data mining techniques by police, the case should be filed electronically and contains descriptive information about the culprits. It also contains a description of modus operandi. Any technique may be used to examine a situation where a group of similar physical descriptions

coincide with a group of similar modus operandi. If there is a good match and the offenders are known for one or more of the offences, then each of the unsolved cases could have well been committed by the same criminal/group. By matching unsolved cases with known culprits, it would be possible to clear up old cases and determine patterns of behavior. Alternatively, if the criminal is unknown but a large cluster of cases seem to point to the same offenders, then these frequent offenders may require for careful examination [1].

## 2. DATA MINING IN POLICE ADMINISTRATION

Till now police in India is using ICT and computers for electronic data collection and computerization of various processes through its software like CIPA, CCIS, CCTNS and others. These software have helped a lot to the police for searching relevant record of the crime and criminals and other relevant information required time to time and hence, time and efforts are being saved. But with the help of data mining, police can benefited by extracting beneficial trends and patterns and behavior among similar cases for making future planning. In this paper, an experiment has been conducted on the dataset relating to vehicle thefts happened in the district Sirsa from 1<sup>st</sup> January, 2012 to 30<sup>th</sup> may, 2012. Efforts have been made to extract the useful knowledge to improve decision making for preventing vehicle theft in particular areas. Copies of FIRs of vehicle theft in Sirsa district have been retrieved from [http://haryanapolice.nic.in/police/fir\\_query\\_Stolen.asp\[3\]](http://haryanapolice.nic.in/police/fir_query_Stolen.asp[3]) and normalized in Microsoft Excel worksheet. The data so normalized were converted to “.arff” file. Abstract of the dataset and its structure have been given in table 1 & 2, respectively.

Table 1. Abstract of dataset of vehicle theft in Sirsa

Description	Value
Geographical Area of dataset	Sirsa district
Period	1 <sup>st</sup> Jan., 2012 to 30 <sup>th</sup> may, 2012
Total No. of instances	147
Missing values	No

**Table 2. Structure of dataset of vehicle theft in Sirsa**

Name of Attribute	Category	Range/Value
POLICESTATION	Numeric	1-12{RORI, CITYMANDIDABWALI, SIRSACITY, RANIA, SADARSIRSA, KALANWALI, ELLENABAD, DING, BARAGUDHA, MANDIDABWALISADAR, NSCHOPTA, ODHAN}
TIME	Numeric	1-24 {31-130 = 1, 131-230 = 2, 231-330 = 3, ..... 2331-0030 = 24,}
DAY	Numeric	1-7{SUNDAY, MONDAY, TUESDAY, WEDNESDAY, THURSDAY, FRIDAY, SATURDAY}
BEAT	Numeric	21-1018{Code of Police Station mentioned before the beat No.}
PROPERTY STOLEN	Numeric	1-8{MOTORCYCLE, SCOOTER, CAR, JEEP, TRACTOR, TRUCK, VAN, MINITRUCK}
MAKE	Numeric	1-15{MAHINDRA, SUZUKI, HONDA, HERO, TVS, BAJAJ, MARUTI, HUNDAI, SWARAJ, JOHNDEAR, ESCORT, TATA, LML, FORD, YAMAHA}
MODEL	Numeric	1-36{SPORTS, CT100, PASSION, DISCOVER, CDDELUXE, SPLENDER, JEEP, VESPA, BOLERO, SAFARI, PLATINA, J5310, BULLET, MARUTI800, ACTIVA, SWARAJ855, INDICA, CHETAK, SWIFT, VERNA, SWIFTDEZIRE, COMMANDER, SWARAJ724, CALIBER, PULSAR, CDDAWN, GLAMOUR, CD100, BOXERAT, VICTOR, AVENGER, CENTRA, ESTEEM, MAXPICKUP, T1210}
YEAROFMANUFACT	Numeric	1947-2012

(Source:[http://haryanapolice.nic.in/police/fir\\_query\\_Stolen.asp](http://haryanapolice.nic.in/police/fir_query_Stolen.asp) accessed on 23.11.2012) [3]

### 3. RESULTS & DISCUSSION

In data mining data are transformed into information and then into knowledge by using application of particular data mining methods. The above mentioned data have been analyzed by

using Classification and Clustering techniques of data mining through Weka 3.7.7<sup>1</sup>.

#### 3.1 Classification

Results generated by using classification technique of data mining are as under:-

==== Run information ====

Scheme: weka.classifiers.rules.ZeroR

Relation: firvehicle3

Instances: 147

Attributes: 7

PS

OccuDay

OccuTimeNormalized

BeatCode

PropertyStolenInvolved

Make

ManufacYear

Test mode: 10-fold cross-validation

==== Classifier model (full training set) ====

ZeroR predicts class value: SIRSACITY

Time taken to build model: 0 seconds

==== Stratified cross-validation ====

== Summary ==

Correctly Classified Instances 97 65.9864 %

Incorrectly Classified Instances 50 34.0136 %

Kappa statistic 0

Mean absolute error 0.1111

Root mean squared error 0.2307

Relative absolute error 100 %

Root relative squared error 100 %

Coverage of cases (0.95 level) 95.9184 %

Mean rel. region size (0.95 level) 70 %

Total Number of Instances 147

<sup>1</sup>Weka 3.7.7 is data mining tool developed by the University of Waikato, Hamilton, New Zealand.

==== Confusion Matrix ====

a b c d e f g h i j <-- classified as

0 0 2 0 0 0 0 0 0 0 | a = RORI

0 0 26 0 0 0 0 0 0 0 | b = CITYMANDIDABWALI

0 0 97 0 0 0 0 0 0 0 | c = SIRSACITY

0 0 2 0 0 0 0 0 0 0 | d = RANIA

0 0 4 0 0 0 0 0 0 0 | e = SADARSIRSA

0 0 4 0 0 0 0 0 0 0 | f = KALANWALI

0 0 7 0 0 0 0 0 0 0 | g = ELLENABAD

0 0 3 0 0 0 0 0 0 0 | h = DING

0 0 1 0 0 0 0 0 0 0 | i = BARAGUDHA

0 0 1 0 0 0 0 0 0 0 | j = DABWALISADAR

The above results depict that in district Sirsa maximum 97 theft of vehicles has been registered in Sirsa City police station in the year 2012(1<sup>st</sup> January to 30<sup>th</sup> May, 2012) followed by City MandiDabwali police station (26) and Ellenabad police station(7). On the other hand Rori, Rania, Sirsa Sadar, Kalanwali and Ding police stations registered 1 to 4 cases of vehicle theft, while single FIR has been registered in Baragudha and DabwaliSadar police stations in the above period. It is also noted that no case of vehicle theft has been registered in NathusariChopta and Odhan police stations during the above period.

### 3.2 Clustering

Results generated by using clustering technique of data mining have been derived as under:-

==== Run information ====

Scheme: weka.clusterers.EM -I 100 -N -1 -M 1.0E-6 -S 100

Relation: firvehicle3

Instances: 147

Attributes: 7

OccuDay

OccuTimeNormalized

BeatCode

PropertyStolenInvolved

Make

ManufacYear

Ignored:

PS

Test mode: Classes to clusters evaluation on training data

==== Clustering model (full training set) ====

EM

==

Number of clusters selected by cross validation: 2

Attribute	Cluster	
	0	1
	(0.2)	(0.8)

=====

#### OccuDay

WEDNESDAY	5.0067	22.9933
MONDAY	9.2267	15.7733
THURSDAY	3.9977	21.0023
TUESDAY	5.748	18.252
SUNDAY	2.9959	20.0041
SATURDAY	5.9972	14.0028
FRIDAY	2.9699	13.0301
[total]	35.9422	125.0578

#### OccuTimeNormalized

20	7.5835	25.4165
21	4.765	19.235
19	3.9526	11.0474
.	.	.
.	.	.
.	.	.
2	1.9998	1.0002
[total]	48.9422	138.0578

#### BeatCode

21	2.9736	22.0264
35	1.9994	18.0006
314	9.5675	10.4325
.	.	.
.	.	.
.	.	.
34	1.0001	1.9999

[total]	60.9422	150.0578
<b>PropertyStolenInvolved</b>		
MOTORCYCLE	1.0751	117.9249
JEEP	11.9864	1.0136
SCOOTER	3.9831	2.0169
TRACTOR	3.9563	1.0437
CAR	10.9731	1.0269
TRUCK	1.9925	1.0075
VAN	1.9759	1.0241
[total]	35.9422	125.0578

<b>Make</b>		
HEROHONDA	1.0106	76.9894
BAJAJ	2.3839	31.6161
MAHINDRA	10.9647	1.0353
TVS	1.0104	8.9896
.	.	.
.	.	.
.	.	.
YAMAHA	1.0052	1.9948
[total]	41.9422	131.0578

<b>ManufacYear</b>		
2011	6.9557	19.0443
2010	6.9979	18.0021
2008	4.0019	14.9981
.	.	.
.	.	.
.	.	.
1978	1.0083	1.9917
[total]	48.9422	138.0578

Time taken to build model (full training data) : 4.07 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	30 (20%)
1	117 (80%)

Log likelihood: -11.82311

Class attribute: PS

Classes to Clusters:

0 1<-- assigned to cluster

0 2 | RORI

3 23 | CITYMANDIDABWALI

23 74 | SIRSACITY

0 2 | RANIA

1 3 | SADARSIRSA

1 3 | KALANWALI

0 7 | ELLENABAD

1 2 | DING

1 0 | BARAGUDHA

0 1 | DABWALISADAR

Cluster 0 <-- CITYMANDIDABWALI

Cluster 1 <-- SIRSACITY

Incorrectly clustered instances : 70.0 47.619 %

Results reveal that in Sirsa district out of total 147 cases of vehicle theft, which were registered from 1<sup>st</sup> January, 2012 to 30<sup>th</sup> May, 2012, maximum thefts occurred between 19:31 to 20:30 hours, followed between 20:31 to 21:30 hours and between 18:31 to 19:30 hours. Beatwise maximum theft of vehicles has been occurred in Beat No. 1 of City MandiDabwali police station followed by Beat No. 5 & 14 of Sirsa City police station and then Beat No. 11 of Sirsa City police station. Motorcycles have been stolen maximum followed by Jeep, Car, Scooter, Tractor, Truck and Van. It is also noted that no case of vehicle theft has been registered w.r.t. other remaining type of vehicles during the above period. Theft of products manufactured by Hero is on maximum followed by Bajaj and then by Mahindra, TVS, Maruti, Tata, Hundai, Swaraj, Escorts, Honda, LML and Yamaha. Theft of Hero Splender remained maximum followed by CDDeluxe and then by Bajaj Platina, Sports, Passion, Bolero, Pulsar, Discover, CT100, Jeep, Swift Dezire, Boxer AT, Safari, Bullet, Maruti800, Chetak, Verna, Victor. There is no major difference of number of thefts in various week days, however, maximum theft cases occurred on Wednesday followed by Monday and Thursday and then Tuesday, Sunday, Saturday, Friday in the Sirsa district during the above period.

#### 4. CONCLUSION

At present, data are being actively collected and stored in large data-warehouses in most of the government and social organizations. It is the necessity of the time to recognize the potential value of these data stores as an information source for making useful decisions. Data mining can cater this requirement of decision support in every area of the enforcement agencies. The present research has discussed

how data mining can help the police administration to explore the available datasets and draw relationship and patterns between occurrences. The results have underlined the usefulness of data mining to predict and at preemptively about the events of thefts etc. Thus, need is to understand the recognized fact to explore and mine the available data-heaps to draw valuable information for decision making for police administration.

## **5. REFERENCES**

- [1] Pujari, A.K. 2010. Data Mining Techniques. Universities Press, Second Edition, 2010.
- [2] Gupta, G.K. 2008. Introduction to Data Mining with Case Studies. Prentice-Hall of India Private Limited, New Delhi, Second Printing, August, 2008.
- [3] [http://haryanapolice.nic.in/police/fir\\_query\\_Stolen.asp](http://haryanapolice.nic.in/police/fir_query_Stolen.asp)
- [4] Dunham, M.H. 2011. Data Mining – Introductory and Advanced Topics. Pearson Education, Eighth Impression, 2011.
- [5] Dwivedi, H. and Bhargava, D 2008 “An Uphill Journey in the Police Stations from Traditional System to Modern System : CIPA in Rajasthan”, Poddar Institute of Management, University of Rajasthan, Jaipur, India.
- [6] <http://scr.b.haryanapolice.gov.in/CW.aspx> accessed on 23.11.2012
- [7] <http://ncrb.gov.in> accessed on 23.11.2012
- [8] <http://ncrb.nic.in/ccismle.htm> accessed on 23.11.2012
- [9] Singh, V. and Kumar, R. 2009 “E-Police: An Information and Communication Technology Enabled Police,” INFOCOM’09, Department of Computer Science & Applications, Chaudhary Devi Lal University, Sirsa, February 19, 2009.
- [10] Adriaans, P. and Zantinge, D. 2003. Data Mining. Pearson Education, Seventh Indian Reprint, 2003.