# Criminal Intent: Analysis and Prediction

Dandekar, Apurwa Kamat, Deepali Fortunato, Eric

## ABSTRACT

Ensuring the safety of a person is one of the primary duties of any government. Globalization with all its benefits has also brought us several detrimental effects like inequality and unequal distribution of wealth. Due to all these changes, there has been an upsurge in several types of crimes. Criminal behavior can be curbed if timely countermeasures can be taken and implemented properly. Criminal data patterns get influenced by various factors that can be used to extract valuable information like the location of the crime, the time and date that it happened, and use it to calculate the crime rate in that area. The study of these factors and the circumstances under which they occur can help us in predicting some of these crimes and tackle the sociological issues responsibly. Keeping all these factors in mind, we aim at designing an application that can predict the possibility of manifestation of the crime so that concerned legal authorities can take necessary steps to prevent them and save loss of life and property.

## 1. OVERVIEW

The plan includes identifying influential factors in crimes that have occurred in the United States of America over a period of time and calculating the crime rate. We used the criminal data sets from the UCI website [3] to study the factors that affected the occurrence of crime. Most crimes have multiple factors affecting them like the per Capita Income, number of members in the family, the time of the crime, among others. So depending upon these factors and their effect the crime rate vary from different areas. This led to the deduction that most crimes are subjective to the areas they transpire in. Hence, we narrowed down our studies to the data set of one particular region; Chicago [1] and studied the relation of crime with respect to the time of the year it happened. The occurrence of a crime cannot be subjected to one particular spot so in-order to classify any location as either safe or risky we need to consider the entire region around it. Hence, we have considered the crime rates of the ward that the location is a part of.

By applying predictive analysis to this data, we have developed our model to succeed in challenging scenarios such as how any particular attribute in a data set would affect the crime prediction pattern. Studies have shown that there is a direct co-relation between the weather conditions and the crime rate in an area[2]. Predictive algorithms within data mining techniques help us infer information that is normally hidden. After working with the criminal data-sets, we were able to discover the behaviour of crime pattern over the period of ten years based on the attributes of the available data. We have designed and developed an application that uses data mining analysis such that the key attributes of the data sets could be used to study the crime patterns and the locations to classify them into high risk and safe regions based on the crime type and the month they occurred in.

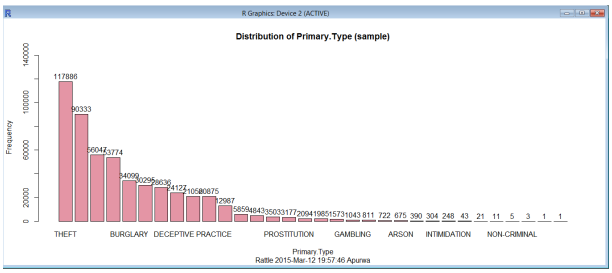



**Figure 1: Bar plot: Data Analysis**

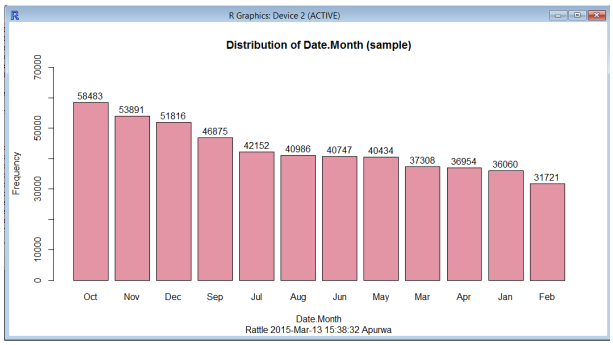



**Figure 2: Bar plot: Month vs Crime**

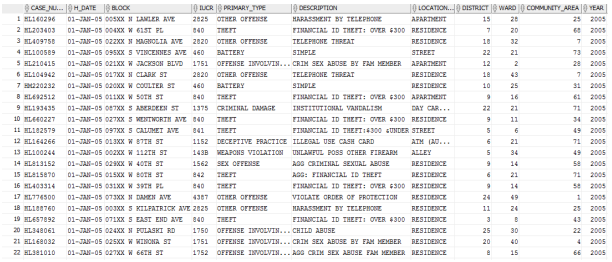
## 2. IMPLEMENTATION

To develop an appropriate prediction of the scenario,events it is required to have done a thorough analysis of data which can lead to the proper conclusion and prediction of similar events.After the data analysis is done the appropriate predictions methods are applied which will give the desired results.We have adapted the same methodologies to develop our predictive model.

## 2.1 Data Extraction

Criminal data sets were taken from the UCI website in the csv format[4]. After merging the data with its description, the empty and missing values from the data were deleted. Certain unwanted, inappropriate attributes were removed. The data cleaning and data processing were done with the help of the language R and Rattle. Decision trees in R/rattle were used to determine the features present and how each attribute affects the overall the crime rate. The observations of the trees were used to conclude that location was a common attribute among all the types of crimes. This led us to performing a similar kind of analysis on a data set pertaining to only one region, Chicago.

We got a real time data in the form of csv which has the data describing the crime w.r.t the date, time , area , crime type etc. for the past 10 years in the city of chicago.There were many missing entries and unwanted columns were present. Therefore the data cleaning was performed with the help of R and Rattle to delete the missing entries and unwanted column.The cleaned data was exported as a csv file which is further used for all the data extraction and analysis purpose.Then this data was imported into an Oracle 11g database with the help of SQLDeveloper tool which is used for all the data related task like creating the tables, views , exporting the sub data sets.



| | CASE_NU... | H_DATE | BLOCK | IUCR | PRIMARY_TYPE | DESCRIPTION | LOCATION... | DISTRICT | WARD | COMMUNITY_AREA | YEAR |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | HL160296 | 01-JAN-05 | 005XX N LAMLER AVE | 2825 | OTHER OFFENSE | HARASSMENT BY TELEPHONE | APARTMENT | 15 | 28 | 25 | 2005 |
| 2 | HL203403 | 01-JAN-05 | 004XX W 61ST PL | 840 | THEFT | FINANCIAL ID THEFT: OVER $300 | RESIDENCE | 7 | 20 | 60 | 2005 |
| 3 | HL409758 | 01-JAN-05 | 022XX N MAGNOLIA AVE | 2820 | OTHER OFFENSE | TELEPHONE THREAT | RESIDENCE | 18 | 32 | 7 | 2005 |
| 4 | HL100589 | 01-JAN-05 | 095XX S VINCENNES AVE | 460 | BATTERY | SIMPLE | STREET | 22 | 21 | 73 | 2005 |
| 5 | HL210415 | 01-JAN-05 | 021XX W JACKSON BLVD | 1751 | OFFENSE INVOLVIN... | CRIM SEX ABUSE BY FAM MEMBER | APARTMENT | 12 | 2 | 28 | 2005 |
| 6 | HL104942 | 01-JAN-05 | 017XX N CLARK ST | 2820 | OTHER OFFENSE | TELEPHONE THREAT | RESIDENCE | 18 | 43 | 7 | 2005 |
| 7 | HH220232 | 01-JAN-05 | 020XX W COULTER ST | 460 | BATTERY | SIMPLE | RESIDENCE | 10 | 25 | 31 | 2005 |
| 8 | HL492512 | 01-JAN-05 | 011XX W 50TH ST | 840 | THEFT | FINANCIAL ID THEFT: OVER $300 | APARTMENT | 9 | 16 | 61 | 2005 |
| 9 | HL193435 | 01-JAN-05 | 087XX S ABERDEEN ST | 1375 | CRIMINAL DAMAGE | INSTITUTIONAL VANDALISM | DAY CAR... | 22 | 21 | 71 | 2005 |
| 10 | HL460227 | 01-JAN-05 | 027XX S WENTWORTH AVE | 840 | THEFT | FINANCIAL ID THEFT: OVER $300 | RESIDENCE | 9 | 11 | 34 | 2005 |
| 11 | HL162579 | 01-JAN-05 | 097XX S CALUMET AVE | 841 | THEFT | FINANCIAL ID THEFT:$300 &UNDER | STREET | 5 | 6 | 49 | 2005 |
| 12 | HL164266 | 01-JAN-05 | 013XX W 87TH ST | 1152 | DECEPTIVE PRACTICE | ILLEGAL USE CASH CARD | ATM (AU... | 6 | 21 | 71 | 2005 |
| 13 | HL102244 | 01-JAN-05 | 092XX W 112TH ST | 1488 | WEAPONS VIOLATION | UNLAWFUL POSS OTHER FIREARM | ALLEY | 5 | 34 | 49 | 2005 |
| 14 | HL613152 | 01-JAN-05 | 029XX W 40TH ST | 1562 | SEX OFFENSE | AGG CRIMINAL SEXUAL ABUSE | RESIDENCE | 9 | 14 | 58 | 2005 |
| 15 | HL615870 | 01-JAN-05 | 015XX W 80TH ST | 842 | THEFT | AGG: FINANCIAL ID THEFT | RESIDENCE | 6 | 21 | 71 | 2005 |
| 16 | HL403314 | 01-JAN-05 | 031XX W 39TH PL | 840 | THEFT | FINANCIAL ID THEFT: OVER $300 | RESIDENCE | 9 | 14 | 58 | 2005 |
| 17 | HL776500 | 01-JAN-05 | 073XX N DAMEN AVE | 4387 | OTHER OFFENSE | VIOLATE ORDER OF PROTECTION | RESIDENCE | 24 | 49 | 1 | 2005 |
| 18 | HL158760 | 01-JAN-05 | 093XX S KILPATRICK AVE | 2825 | OTHER OFFENSE | HARASSMENT BY TELEPHONE | RESIDENCE | 11 | 24 | 25 | 2005 |
| 19 | HL657892 | 01-JAN-05 | 071XX S EAST END AVE | 840 | THEFT | FINANCIAL ID THEFT: OVER $300 | RESIDENCE | 3 | 8 | 43 | 2005 |
| 20 | HL348061 | 01-JAN-05 | 024XX N POLASKI RD | 1750 | OFFENSE INVOLVIN... | CHILD ABUSE | RESIDENCE | 25 | 30 | 22 | 2005 |
| 21 | HL168032 | 01-JAN-05 | 025XX W WINONA ST | 1751 | OFFENSE INVOLVIN... | CRIM SEX ABUSE BY FAM MEMBER | RESIDENCE | 20 | 40 | 4 | 2005 |
| 22 | HL581010 | 01-JAN-05 | 027XX W 66TH ST | 1752 | OFFENSE INVOLVIN... | AGG CRIM SEX ABUSE FAM MEMBER | RESIDENCE | 8 | 15 | 66 | 2005 |

**Figure 3: Oracle 11g Database**

## 2.2 Data Analysis

Once the data has been imported into the database we started with the data mining functionality.The application we have designed is aimed at predicting the crime for following scenarios,

1. Prediction of the risk for all the wards w.r.t all the months
2. Predicting the overall risk for the ward
3. Provide the statistics of different types of crime for that particular area. . . .

different wards for all the months depending upon the previous 10 years crime occurrence.

## 2.3 Prediction

We have created a web application that involves taking in the location and month from the user. The application connects to the database and matches the location of the block to the ward in which it is present and returns the probability of a risk for the selected month in that ward along with the overall probability of the risk. It also provides the statistics of the different crime types that has occurred in that area. Our application involves the use of Oracle for database connectivity and HTML/CSS, servlet,jsp and JavaScript for client-server interaction.

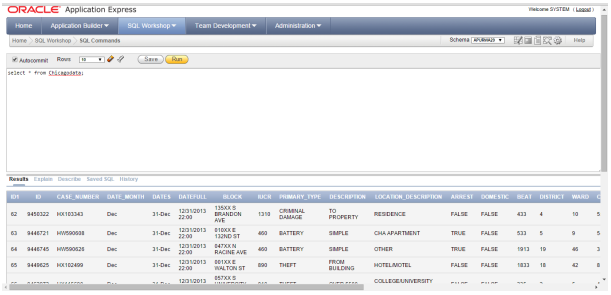



**Figure 4: Oracle 11g Database**

# 3. ARCHITECTURE

We have web based application using HTML/CSS where the client side programming will be done using JavaScript and server side programming will be done in Java using Servlet and JSP, where we will incorporate business logic and calculation onto the extracted data. We are using the Oracle 11g database server to store the criminal incidence data. JDBC is used to connect the Oracle database server with Java. The Web application will be deployed on Apache Tomcat Server.

# 4. DESIGN

We have created a web based application in HTML/CSS which enables the user to enter the Location and date for which he/she wishes to acquire the crime rate for the area around the location. We make use of JavaScript to validate whether the user has put in appropriate data, if not an alert message is displayed. After successful validation, the form data entered by the user is sent to the server side where the actual business logic is applied and SQL queries are performed to give appropriate idea about the severity of the crime and crime rate. The region we have taken into consideration is the ward number, that is, depending upon the location provided by the user, we are extracting the ward number for that location from the database. After that we are calculating the crime rates for the different types of crime. Data clustering and classification will be performed on the data at hand. In further stages we plan on applying prediction algorithms based on the data modeling and data mining procedures we have performed in the earlier stages.

# 5. CURRENT STATUS AND FUTURE WORK

We extracted the criminal dataset for the city of Chicago from the year 2011 from the internet[1] in the csv format.The

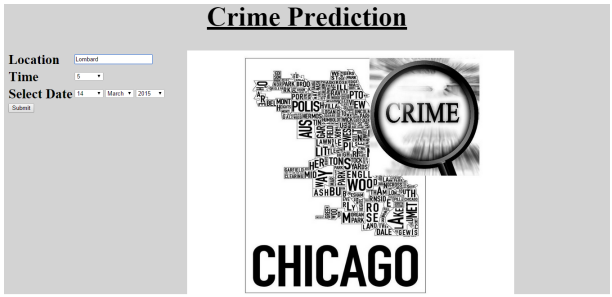



**Figure 5: User Interface**

data mining procedure has been applied in order to extract useful information and form an understanding from it. At the back end, Oracle is used to import the csv file into the Oracle database using the Oracle express Edition, which will be further used for the data extraction and analysis. Using this data, we run SQL queries which help us get the overall statistics of the crime depending upon the region, time and the weather.

We have created a web based application using the HTML/CSS format where a simple user interface has been produced. The validation of the data of the HTML page is done using JavaScript. We have tried to establish a database connection with Java using JDBC. In the future, we plan on developing an appropriate business logic depending upon the statistics and analysis of the result we obtained in data extraction from the Oracle database. We will be designing an algorithm to predict the accurate result for a specific application, that is the application will provide the user with the statistics for the different types of crime that occurred in that area and depending upon these statistics and business logic, the application will provide the information about the severity of the crime rates for that region. Going ahead we plan to finish the ER diagram for the data set and also the implementation of the relational database for it. This plan also includes steps for normalization and finishing the code for classification and clustering algorithms.

# 6. REFERENCES

[1] Brittany Suszan, D. E. Crimes - 2001 to present, 2011.

[2] Murataya, R., Scholar, P. D. M., and Gutiérrez, D. R. Effects of weather on crime.

[3] Redmond, M. Communities and crime data set normalized, 2009.

[4] Redmond, M. Communities and crime data set unnormalized, 2009.