



Criminal Intent: Analysis and Prediction

Apurwa Dandekar Deepali Kamat Eric Fortunato
Rochester Institute Of Technology



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. 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.

Presently there are applications like PredPol and CommandCentral Predictive, whose purpose is to predict the occurrence of a crime at a place by taking into consideration the time, place and type of the crime. These applications, however, do not provide us with the risk percentage involved with each region and the percentage of all the crime types.

Goal

- To design and develop an application that uses data mining analysis such that the key attributes of the data sets can be used to study the crime patterns of each particular location
- Classify locations into high risk and safe regions based on the crime type and the time of occurrence of crime
- Use application to obtain the crime rates and the predict risk factor associated with various wards of Chicago

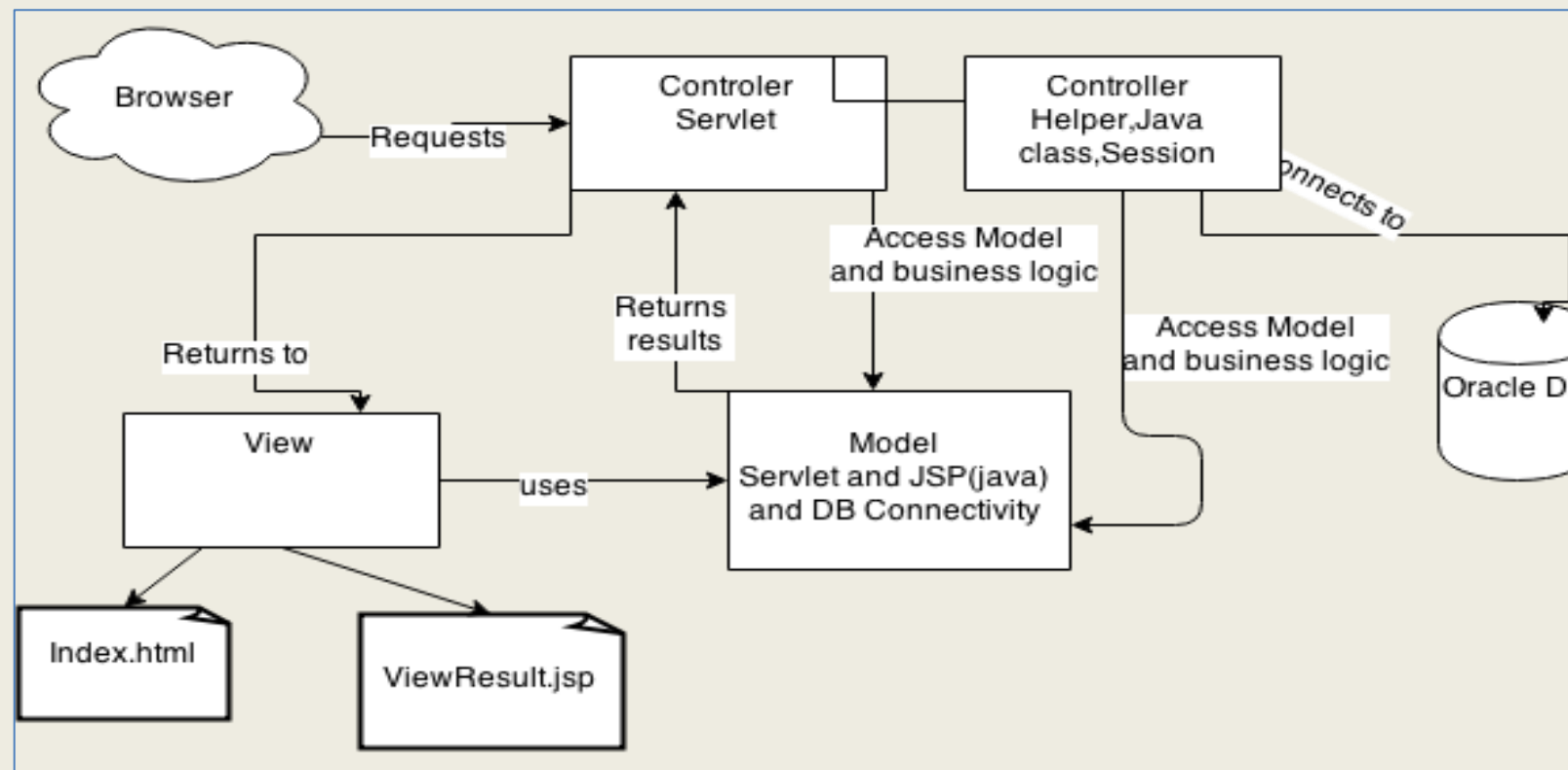


Figure 1. Block Diagram Of the Architecture of the Application

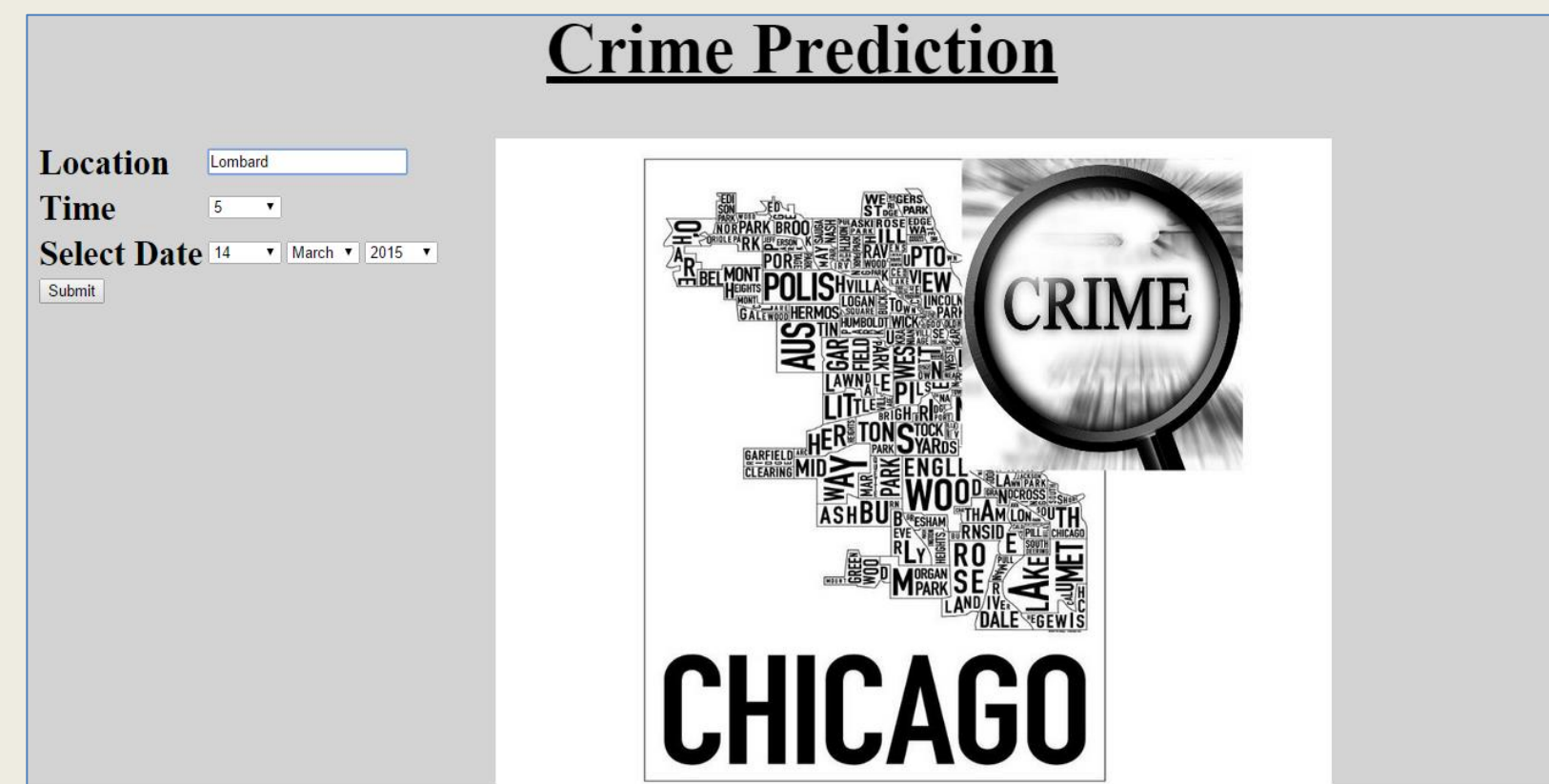


Figure 2. User Interface.

Architecture

- The user interface model was developed using the MVC framework.
- The Oracle 11g database server is used to store the criminal incidence data.
- JDBC is used to connect the Oracle database server using Java.
- The entire Web application is deployed on Apache Tomcat Server.
- HTML/CSS is used to design the user interface and the client side programming.
- Field validation is done using JavaScript.
- Server side programming is done in Java using Servlet and JSP, where business logic is incorporated and data extraction involves fetching of accurate data from the database.
- HttpServlet, HttpRequest and HttpResponse classes from Java are used in design of the client-server interaction

Data Analysis

- Data is classified based on the crime types and the month of occurrence of the crime for each ward
- R and Rattle were used for data classification
- Further classification involved creating 12 data subsets (January to December) consisting of the crime count for each month for the 10 year(2005-2014) span per ward.
- Study of the box plots led to the conclusion that the position of each ward in the plot gives us the nature of the risk involved in that area.
- Plotted a histogram representing the relationship between the crime count for the fifty wards over the ten year span.
- Used ggplot2 package in R to generate a map of Chicago which displays the varying intensities of crime across the city
- prediction of crime related risk using ggplot was same as that obtained by using box plots and histogram, thereby confirming accuracy of the predictive analysis

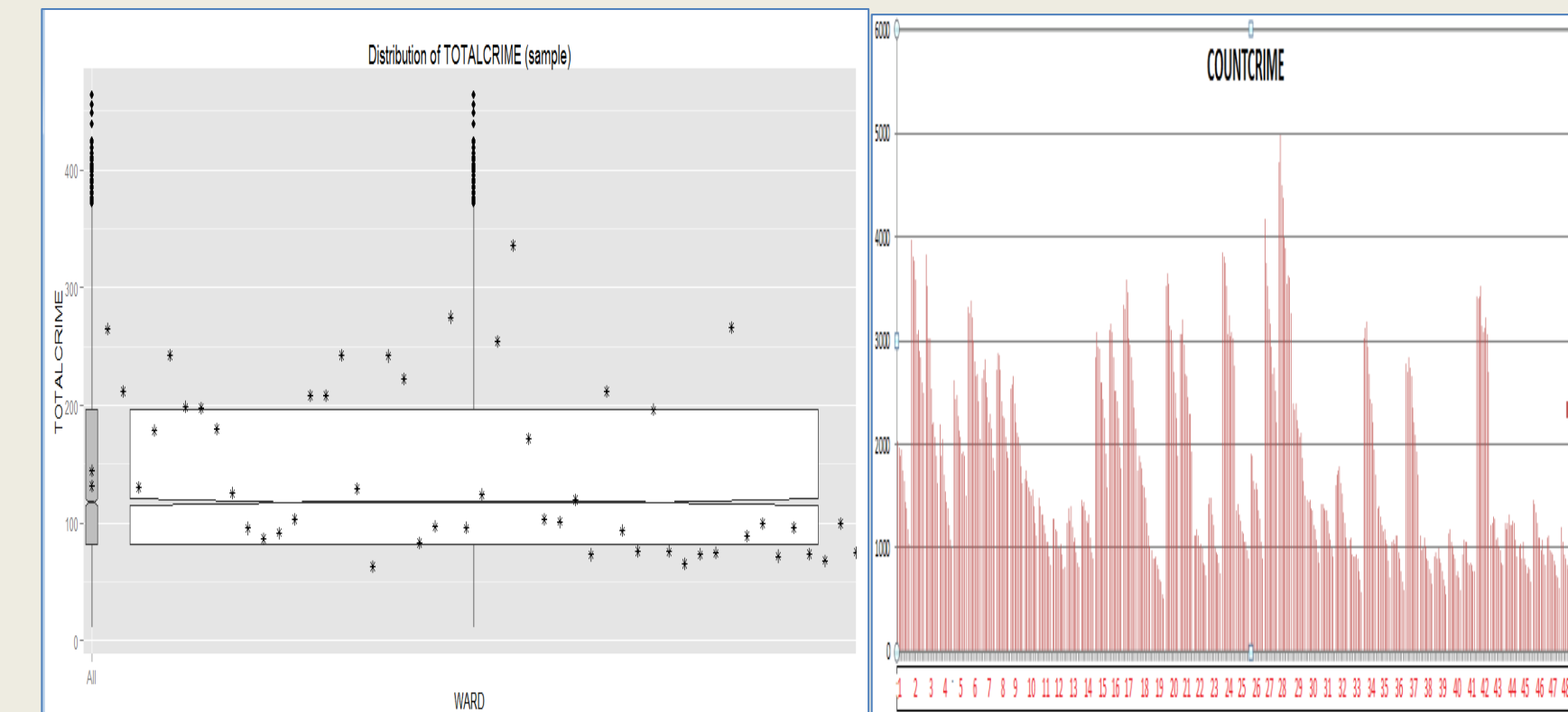


Figure 3: Box Plot Representation of Crime per Ward

Figure 4: Histogram Representing Crime Data over 10 years

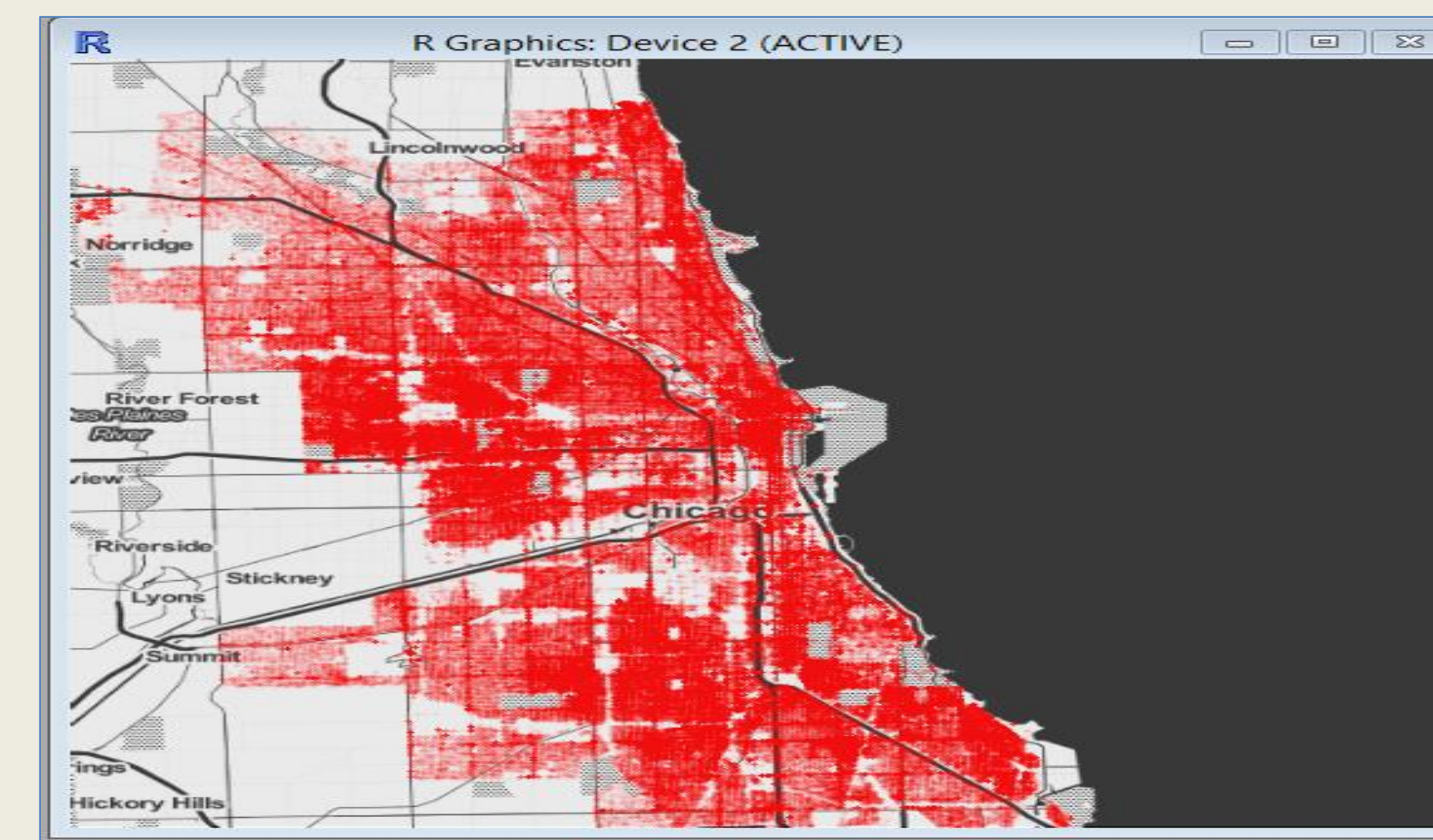


Figure 5. Map of Chicago with Crime Density Display

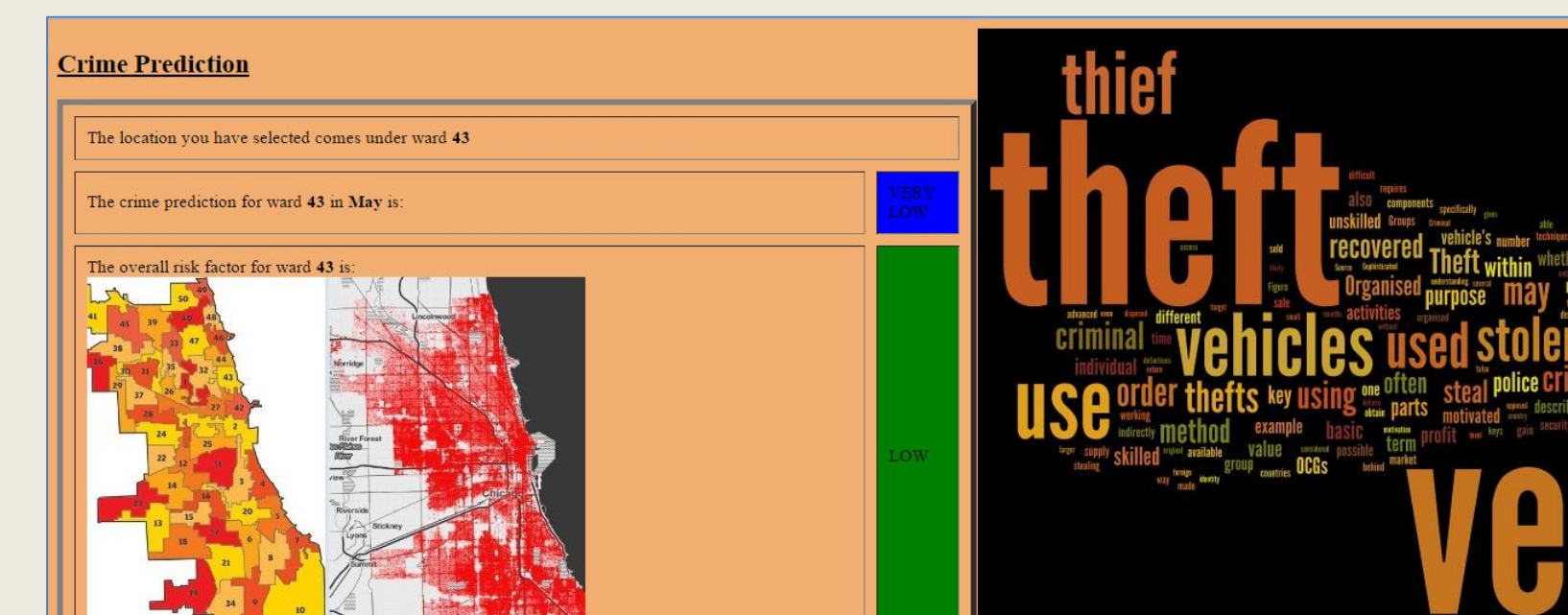


Figure 5. User Interface: Crime Prediction Result

Current Status & Future Work

- Developed an application that takes in a location and month as input from the user
- The application matches the entries with those in the database and then displays the risk factor of the ward that the location belongs to and the crime percentage for the different types of crime for that ward

Future Work

- Incorporating the same principles as these and predicting risk factor for regions outside of Chicago.
- Making a smartphone application based on current project
- Taking other attributes that influence crime rate like the economic and weather conditions and studying the variation of crime pattern with respect to the same

Lessons Learned

- Big data and the tools used allow us to identify risky areas of the city of Chicago easily which would have been difficult to catalog using other conventional methods
- Feature selection and decision trees help in understanding influential factors involved in rise and fall of crime rate
- Data mining and big data analysis is affected by missing values if the data has not been cleaned and prepared before hand.
- The removal of irrelevant elements helps in the efficient classification of risk.
- Statistical tools are key to perform efficient analysis and understanding the results
- R and Rattle provide multiple functions to extract features, plot results and make visualizations easy to comprehend.

REFERENCES

1. PredPol (<https://www.predpol.com/>)
2. CommandCentral Predictive
<http://www.publicengines.com/products/commandcentral-predictive.php>
3. Brittany Suszan, D. E. Crimes - 2001 to present, 2011.
2. Murataya, R., Scholar, P. D. M., and Gutie' rrez, 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.
5. Soumen Chakrabarti, Martin Ester, U. F. J. G. J. H. S. M. G. P.-S. W. W. Data mining curriculum, 2006.