**CRIME ANALYSIS IN CHICAGO FINAL REPORT - GROUP 16 - DATA ANALYTICS – SPRING 2017**

**Devyani Deshmukh, Shivanvitha Konda, Sneha Alkuru, Vaishnavi Kankanala**

**1. Abstract**

There are various crimes taking place in the Chicago city. The database extracted from Chicago Police Department’s CLEAR (Citizen Law Enforcement Analysis and Reporting) system contains data of about 15 years collected in this domain. Predicting the most crime-zone prone areas and analyzing which type of crime is common can have an important impact on issues like social quality of life and moreover the economic growth of society. Our outputs include

* There are various crimes taking place in the city so we want to analysis the crime type that occurred most in 2007 to 2016 and find if it is increased or decreased, and during which month it is very high.
* Visualize 5 most crime prone zones in Chicago city. We have analyzed them based on place where more than 40 crimes took place.
* To check the trend analysis of crime types in different months.

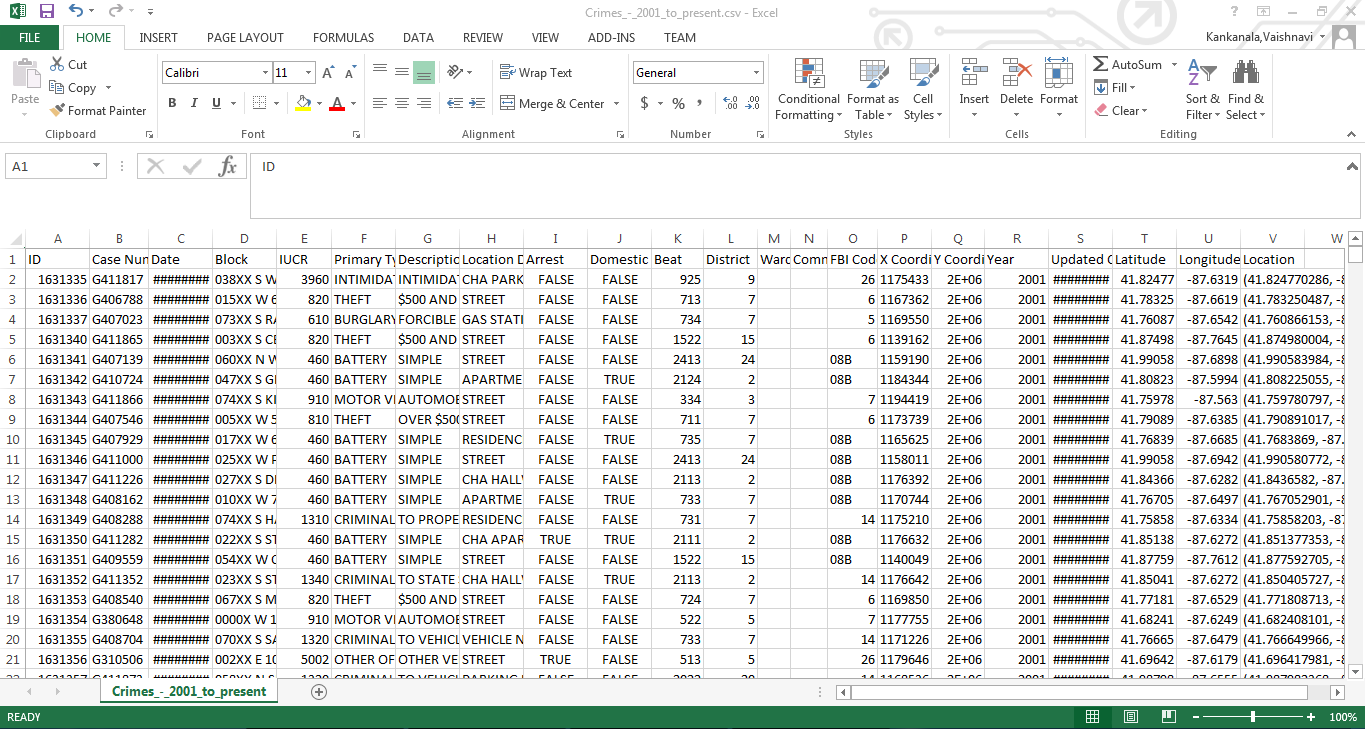
**2. Introduction/Background – Crime Analysis in Chicago**

Crime Analysis, can help us to take an important decision for people to move to crime prone areas or not. With the increase of crimes law enforcement agencies are demanding advanced geographic studies and data mining to improve results to protect communities.

**3. Data Description and Preprocessing**

We extracted our data from the website: <https://catalog.data.gov/dataset/crimes-2001-to-present-398a4>

Due to the immense size of the data set, as well as the fact that the most crimes occurred was from 2011 and it’s very recent. When we began the study, we decided to focus our tests on the last five years (2007-2016). Initial raw data files has 22 dimensions and around 10, 48,576 instances. Fortunately, the raw data was available to download in .csv file format with minimal wait time. Figure 1(a) shows a snapshot of one of the .csv files, and Figure 1(b) shows the expanded labels of the initial 22 attributes:



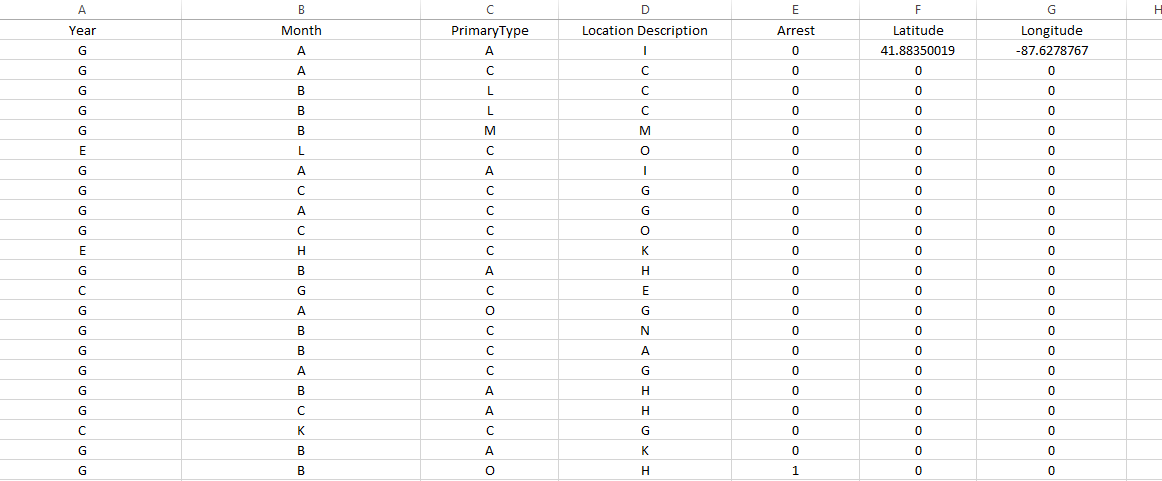
**Figure 1(a) shows the raw data of .csv file**

|  |  |  |
| --- | --- | --- |
| **S.No** | **Attribute Name** | **Description** |
| **1** | **ID** | **Unique ID** |
| **2** | **Case Number** | **Criminal activity number** |
| **3** | **Date** | **Crime date** |
| **4** | **Block** | **Place where crime happened** |
| **5** | **IUCR** | **Illinois Uniform Crime Reporting (IUCR) codes are four digit codes that law enforcement agencies use to classify criminal** |
| **6** | **Primary Type** | **Crime Type** |
| **7** | **Description** | **Crime type description** |
| **8** | **Location Description** | **Place description** |
| **9** | **Arrest** | **Criminal arrested or not** |
| **10** | **Domestic** | **Domestic violence** |
| **11** | **Beat** | **A beat is smallest police geographical area** |
| **12** | **District** | **Police district where incident occurred** |
| **13** | **Ward** | **Ward place where it is occurred** |
| **14** | **Community** | **Community area where Chicago has 77** |
| **15** | **FBI code** | **Crime classification as outlined in FBI’s NIBRS** |
| **16** | **X Coordinate** | **X coordinate of location** |
| **17** | **Y Coordinate** | **Y coordinate of location** |
| **18** | **Year** | **Case occurred year** |
| **19** | **Updated On** | **Record was last updated** |
| **20** | **Latitude** | **Location** |
| **219** | **Longitude** | **Location** |
| **22** | **Location** | **Location (Latitude, Longitude)** |

**Figure 1(b) snapshot gives description of all the 22 attributes**

Our first step in preprocessing was to eliminate the attributes of the raw data to a selective one which provides more valuable information for the study. Since we were concerned with numerical prediction, we first eliminated all dimensions that were not of numeric type. We also eliminated those columns that were particularly sparse, and those that were less familiar crime attributes (such as IUCR, Beat, FBI code etc.)

Number of attributes are ultimately reduced from 22 to 7 and the instances are reducedfrom 10, 48,576 to 63,326.



**Fig 1 (c).**

The symbolic representation of above characters in the pre-processed data can be explained with the help of below screenshots:

|  |  |
| --- | --- |
| **Letter** | **Month** |
| **A** | January |
| **B** | February |
| **C** | March |
| **D** | April |
| **E** | May |
| **F** | June |
| **G** | July |
| **H** | August |
| **I** | September |
| **J** | October |
| **K** | November |
| **L** | December |

**Fig 1 (d).**

|  |  |
| --- | --- |
| **Letter** | **Year** |
| **A** | **2010** |
| **B** | **2011** |
| **C** | **2012** |
| **D** | **2013** |
| **E** | **2014** |
| **F** | **2015** |
| **G** | **2016** |
| **X** | **2007** |
| **Y** | **2008** |
| **Z** | **2009** |

**Fig 1 (e).**

|  |  |
| --- | --- |
| **Value** | **Arrest** |
| **False** | 0 |
| **True** | 1 |

**Fig 1 (f).**

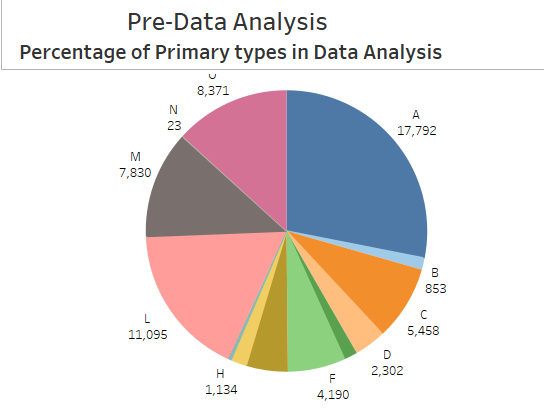
|  |  |
| --- | --- |
| **Letter** | **Crime Type** |
| **A** | theft,burglary,robbery |
| **B** | kidnapping |
| **C** | deceptive practice |
| **D** | motor vehicle |
| **E** | sexual |
| **F** | drugs |
| **G** | homicide |
| **H** | rules violation |
| **I** | gambling |
| **J** | interference with public |
| **K** | intimidation |
| **L** | battery |
| **M** | criminal damage |
| **N** | non-criminal |
| **O** | others |

**Fig 1 (g).**

|  |  |
| --- | --- |
| **Letter** | **Crime Type** |
| **A** | abandoned building |
| **B** | airport/aircraft |
| **C** | apartment |
| **D** | ATM |
| **E** | Bank |
| **F** | vehicle |
| **G** | residence |
| **H** | street, sidewalk |
| **I** | stores, restaurant |
| **J** | gas stations |
| **K** | Public places |
| **L** | parking lot |
| **M** | schools, colleges |
| **N** | cab, currency exchange |
| **O** | others |

**Fig 1 (h).**

Below screenshot of pie chart describes the primary crime type analysis. For example crime type A took place 17,792 times out of 63,326 instances of crimes.



**Fig 2 (a).**

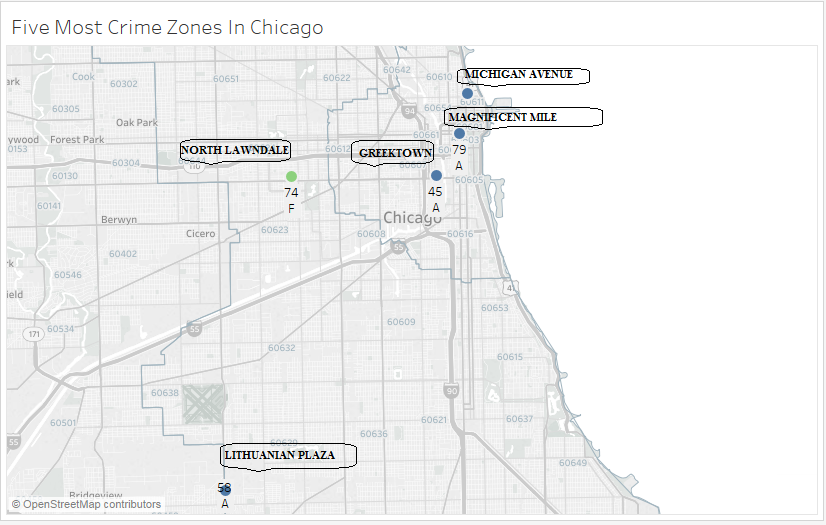
**4. Algorithms and results for all the outcomes:**

We tested our first outcome that is analyzing which crime took place highest in which month of the year based on various algorithms. All of these algorithms approximately gave 33% of accuracy. We ran below algorithms:

1. Multiclass classifier
2. SVM
3. Decision tree J48

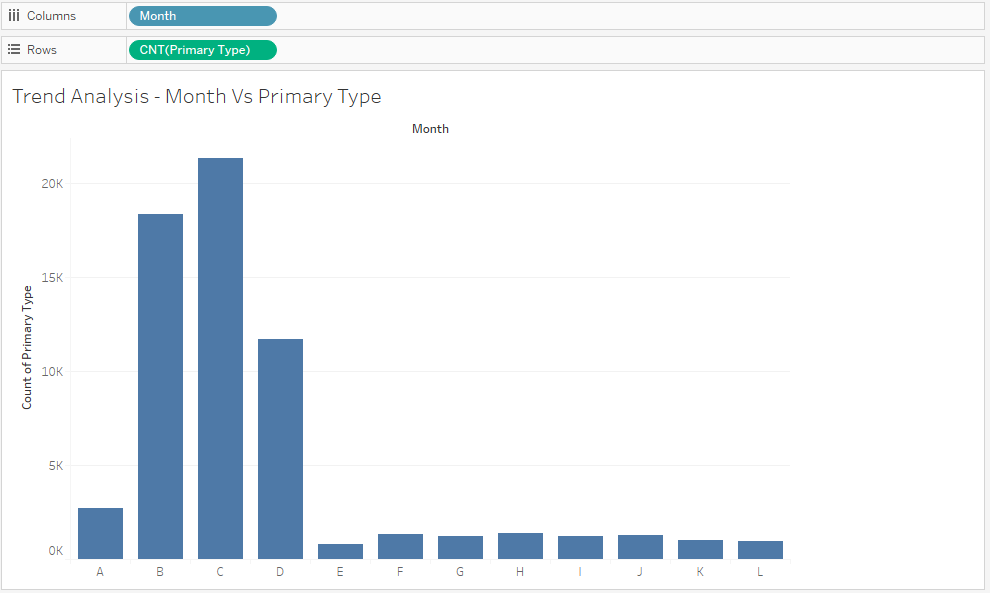
The output of the multiclass classifier is as below:

For our second output that is analyzing the five most crime prone zones in Chicago city we have used the visualization tool that is tableau. We decided our outcome based on the crimes which were more than 40 in a particular zone. We also added landmarks to that zone so as to spread the awareness among public. The screenshot of our analysis is as follows.



**Fig 4.**

For our third analysis that is analyzing the trends in crimes in different months we have used the tableau as visualization tool. The trend analysis graph is as below:



**Fig 5.**

We have analyzed the trends in crime based on months and count of primary crime types.

**5. Conclusion:**

For our first outcome after running the above mentioned algorithms we concluded that the crime type is robbery/theft is high in the year 2016 and in the month of March.

For our second outcome from the graph we analyzed that most crime prone zone is in East Chicago and the common crime types are theft/robbery.

From the trend analysis graph we conclude that crimes are mostly occurring in the month of March.

We hope this report will be helpful to the common public in Chicago so as to be aware and keep themselves safe from crimes taking place in the city.

**6. References**

**1. Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *International Journal of Data Mining & Knowledge Management Process*, Vol.15, No.4**

**2.** [**https://arxiv.org/ftp/arxiv/papers/1508/1508.02050.pdf**](https://arxiv.org/ftp/arxiv/papers/1508/1508.02050.pdf)

**3.** [**https://catalog.data.gov/dataset/crimes-2001-to-present-398a4**](https://catalog.data.gov/dataset/crimes-2001-to-present-398a4)

**4.** [**http://www.cs.waikato.ac.nz/ml/weka/downloading.html**](http://www.cs.waikato.ac.nz/ml/weka/downloading.html)

**5.** [**http://www.pentaho.com/download**](http://www.pentaho.com/download)