Chicago Crime Analysis and Prediction by Deep Neural Nets

Chandan Ramanna

Abstract

With the proliferation of technology and automation over the years, it is easier to determine the extent of vulnerability an individual is subjected to, at a specific geographic area on any occasion. The main objective of our project is to anticipate if a particular neighborhood in the city, at a given duration of the day will be a crime hotspot or not, with an acceptable rate of accuracy. The research aims to exploit background criminal knowledge procured from Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. The CLEAR system is an ideal and reliable source for assimilating criminal occurrences in Chicago, where criminal investigation records are preserved since 2001. The research, with the combination of demographic information, establishes an approach of detecting crimes in a particular geographic area by analyzing and studying the criminal occurrences of the area and thus making deductions by employing Deep Neural Networks.

1 Objective

According to the Chicago Crime report records, "Crime in Chicago has been tracked by the Chicago Police Department's Bureau of Records since the beginning of the 20th century. The city's overall crime rate, especially the violent crime rate, is substantially higher than the US average. Chicago was responsible for nearly half of 2016's increase in homicides in the US"(Exploring Chicago Crimes 2012-2016 , 2017). Chicago's homicide rate is higher the larger American cities of New York and Los Angeles, the reasons for the higher numbers in Chicago remain unclear(Crime in Chicago, 2017). However, "the Chicago police department tallies data differently than police in other cities, the FBI often does not accept their crime statistics."
In order to record the crime in Chicago, the Chicago police department developed a tool to assist city residents in problem-solving and combating crime and disorder in their neighborhoods, all thing shows Chicago police department has a long history of using data" (Jeanne Clery Disclosure Act, 2015). The report will cover the number of crime, the type of crime and the times series development of crime, etc.
Chicago Crime Report (2012-2016).From: amazonaws.com: Crime in Chicago has been tracked by the Chicago Police Department's Bureau of Records since the beginning of the 20th century. The city's overall crime rate, especially the violent crime rate, is substantially higher than the

US average. Chicago was responsible for nearly half of 2016's increase in homicides in the US. Keeping these concerns in mind, we have leveraged the dataset to analyze and perform prediction on crime events (except the murder cases) that occurred in the City of Chicago from 2001 to present. The data resource is from Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. For our project, We have more than 6280882 millions of the data records. The research aim to delve deeper into this statistic, properly detect outliers and deal with missing values in our dataset. Using 10 professional graphs to analysis the dataset. and using predictive analytics to solve some classification and prediction problems on our dataset.

Business Scenario Why is this dataset more important now? Because the number of crime in Chicago increase, more and more people care about their safety, and the government leader wants to build a good environment to the citizens. Therefore, to predict the occurrence of a crime at a location at a specific time of a day. The report uses data visualization to tell an interesting story to reader. Furthermore, the crime data is superimposed with housing information, literacy rate, employment and socioeconomic status of various community areas of Chicago, to generate an elaborate model to extend the research to a business/commercial model.

1.1 Existing solutions

Several crime analysis techniques have been established over the years for identifying and analyzing patterns and trends in crime and disorder. These techniques can be deployed by both the general public and the law enforcing bodies to observe vigilance and caution. The following section briefly describes the works proposed till date. The approach in [3] predicts crime hotspots based on human behavioral data derived from mobile network activity, in combination with demographic information thus considering both people- and place- centric perspectives. The people-centric approach involves individual and collective profiling and further determining behavioral patterns in crimes that are committed by the same offender or group of offenders. While place-centric perspective adopts a distinct modus operandi crime hotspot detection by analysis and the consequent derivation of useful insights leveraging mobile network activity as a source of human behavioral data. [3] establishes an approach called the Series Finder to tackle the problem of detecting precise patterns in crimes that are committed by the same offender or group of offenders. In this work, the authors take a machine learning approach to the problem of detecting specific patterns of crime that are committed by the same offender or group. The learning algorithm processes information similarly to how crime analysts process information instinctively: the algorithm searches through the database looking for similarities between crimes in a growing pattern and in the rest of the database, and tries to identify the modus operandi (M.O.) of the particular offender. The M.O. is the set of habits that the offender follows, and is a type of motif used to characterize the pattern. As more crimes are added to the set, the M.O. becomes more well-defined. The approach to pattern discovery captures several important aspects of patterns:

- Each M.O. is different. Criminals are somewhat self-consistent in the way they commit crimes. However, different criminals can have very different M.O.'s. Consider the problem of predicting housebreaks (break-ins): Some offenders operate during weekdays while the residents are at work; some operate stealthily at night, while the residents are sleeping. Some offenders favor large apartment buildings, where they can break into multiple units in one day; others favor single-family houses, where they might be able to steal more valuable items. Different combinations of crime attributes can be more important than others for characterizing different M.O's.
- General commonalities in M.O. do exist. Each pattern is different but, for instance, similarity in time and space are often important to any pattern and should generally by weighted highly. Our method incorporates both general trends in M.O. and also pattern-specific trends.
- Patterns can be dynamic. Sometimes the M.O. shifts during a pattern. For instance, a novice burglar might initially use bodily force to open a door. As he gains experience, he might bring a tool with him to pry the door open. Occasionally, offenders switch entirely from one neighborhood to another. Methods that consider an M.O. as stationary would not naturally be able to capture these dynamics.

[3] exploited data procured from the Datathon for Social Good - organized by Telef′onica Digital, at Open Data Institute and MIT during the Campus Party Europe 2013, London September 2013. The data provided by participants were categorized into two sections (i)smartstep data: smart anonymized and aggregated human behavioral data computed from mobile network activity in the London Metropolitan Area (ii)geo-localised open data: which included reported criminal cases, residential property sales, transportation, weather, etc. For each Smartsteps cell (geographic lo- cation with precise lat, log values), a prediction was made whether that particular cell will be a crime hotspot or not in the next month. The paper employed a variety of trained classifiers on the training data following a 5-fold cross validation strategy: logistic regression, support vector machines, neural networks, decision trees, and different implementations of ensembles of tree classifiers with different parameters. The decision tree classifier based on the Breiman's Random Forest (RF) algorithm was employed in order to yield the best performance when compared to all other classifiers. However, our model aims to adopt a machine learning approach and develop a concise learning algorithm to anticipate the criminal activity of a geographic location guaranteeing an acceptable rate of precision. Recent works in the domain are influenced by the proliferation of social media which has sparked an interest in using data from software applications to anticipate a variety of variables, including electoral outcomes and market trends. Following the trend, Wang et al. [4] proposed the ap-

plicability of social media to predict criminal incidents. Their approach relies on a semantic analysis of tweets using natural language processing along with spatio-temporal information derived from neighborhood demographic data and the tweets meta-data. The authors in [1] proposed a model named Series Finder which formulates predictions if an individual or a group of individuals will commit a crime in the near future by studying the past records and analyzing the pattern. Series Finder first uses one of the crimes as "seed" and then links it with the other crimes the criminal is involved using similarities such as type, location, etc., which leads to a pattern. These seed patterns need not be geographically closer which helps extend the search space. To be more thorough, the crimes that are grouped together on the basis of similar patterns, have some attributes that are almost identical and are calculated by using learned weight of each crime type and pattern based weights. The attributes considered are of various types: (i) occurrence of the crime with the presence and absence of people at the crime site (ii) exact time or time window (iii) the area of occurrence (iv) day of the week. Here, pattern building is done by using set building techniques where they start with an initial seed and then iteratively add the most similar crime to the set in a sequential manner until a cutoff value, which was predetermined. Validation of the model was carried out by using three different patterns (i) an existing(i.e., original) pattern (ii) a predicted pattern (iii) a verified pattern and based on these results success and the failures are recorded. Recent years have witnessed an augmented expansion of social media with millions of users. One such example being the twitter, which is considered to be one of the most abundant resources in the field of varied data. Statistical topic modeling and linguistic analysis of twitter-specific data are used to identify the major discussions, including crime, across various cities. Major attempts are taken to incorporate these techniques in extracting the crime related discussion across Chicago city in creation of a crime prediction model [2]. Gerber collected all the information between January 1, 2013 and March 31, 2013 documented by the Chicago Crime Department. The research considered the time-stamp of occurrence, latitude/longitude coordinates of the crime at the city-block level, and one among the 27 crime types provided by the Crime department. For the same period they considered twitter data, from official Twitter Streaming API, tagged with GPS coordinates of the city of Chicago [2]. The density of the tweets were correlated with different crime types documented with the crime department to predict the crime occurrence.

Drawbacks: There are many challenges to using Twitter as an information source for crime prediction. Tweets are notorious for (un)intentional misspellings, on-the-fly word invention, symbol use, and syntactic structures that often defy even the simplest computational treatments (e.g., word boundary identification). To make matters worse, Twitter imposes a 140-character limit on the length of each tweet, encouraging the use of these and other message shortening devices. Lastly, authors of [2] are interested in predicting crime at a city-block resolution or finer, and it is not clear how tweets should be aggregated to support such analyses (prior work has investigated broader resolutions, for example, at the city or country levels). These factors conspire to produce a data source that is not only attractive – owing to its real time, personalized content – but also difficult to process. Thus, despite recent advances in all stages of the automatic text processing pipeline (e.g., word boundary identification through semantic analysis) as well as advances in crime prediction techniques (e.g., hot-spot mapping), the answer to the primary research question in [2] has remained unclear.

The CLEAR system is an ideal and reliable source for assimilating criminal occurrences in Chicago, where criminal investigation records are preserved since 2001. The research, with the combination of demographic information, establishes an approach of detecting crimes in a particular geographic area by analyzing and studying the criminal occurrences of the area and thus making deductions by employing a well-founded and reliable learning algorithm. The analysis is further extended to incorporate the impact of housing and inhabitation, literacy rate, employment and socioeconomic status on the crime occurrence rate.
The rest of the report is divided into six sections. Section 2 elaborates on the system model chosen for our research. Section 3 comprises of the sub-problem statements and their analysis, each of which are demonstrated by graphs. Section 4 describes the algorithmic approach to our problem in detail. Section 5 and 6 consist of experimentation results and conclusion.

2 Adopted System Model

Chicago, Illinois ranks third in the United States in population (2.7 million), second in the categories of total murders, robberies, aggravated assaults, property crimes, and burglaries, and first in total motor vehicle thefts [1]. In addition to its large population and high crime rates, Chicago maintains a rich data portal containing, among other things, a complete listing of crimes documented by the Chicago Police Department. Crime analysis techniques play a vital role in devising solutions to crime problems, and contriving crime prevention strategies. Analysts study crime re- ports, arrests reports, and police calls for service to identify emerging patterns, series, and trends as quickly as possible. With the adoption of automation and machine learning techniques, these activities can be accomplished at a more accelerated and efficient rate. [4] derives automatic techniques to interpret and detect patterns on analyzing a previously existing criminal record of an individual or a group of individuals. This work is briefly elaborated in the next section. Our objective however is to determine patterns in the type of criminal activity irrespective of the perpetrator. We have procured criminal records from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. The CLEAR system is an ideal and reliable source for assimilating criminal occurrences in Chicago, where criminal investigation records are preserved since 2001. Table 1 depicts the attributes considered by our system model and is referred to as the dataset.

To accomplish our objective of recognizing crime patterns across the city based on geo-

Variable	Description
ID	Unique identifier for the record.
Case Number	The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
Date	Date when the incident occurred. this is sometimes a best estimate.
Block	The partially redacted address where the incident occurred, placing it on the same block as the actual address.
IUCR	The Illinois Uniform Crime Reporting code. This is directly linked to the Primary Type and Description.
Primary Type	The primary description of the IUCR code.
Description	The secondary description of the IUCR code, a subcategory of the primary description.
Location Description	Description of the location where the incident occurred
Arrest	Indicates whether an arrest was made.
Domestic	Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
Beat	Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts.
District	Indicates the police district where the incident occurred.
Ward	The ward (City Council district) where the incident occurred.
Community Area	Indicates the community area where the incident occurred. Chicago has 77 community areas.
FBI Code	Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
Y Coordinate	The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
Year	Year the incident occurred.
Updated On	Date and time the record was last updated.
Latitude	The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
Longitude	The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

Table 1: Attributes of crime-data provided by the Chicago Police Department

graphical locations, our first measure is dividing the entire city of Chicago into smaller units called cells, each district of Chicago is evaluated as a cell. In the meta-data obtained from the CLEAR system of Chicago Police Department, each criminal record is characterized by several attributes that includes crime description, location, longitudes and latitudes, etc as elaborated in Table 1. These attributes comprise the dataset of the system model adopted by our project and will be conducive while plotting the exact locations of the crimes. In addition, the CLEAR system classifies the crimes into 32 different categories as depicted in Table 2. With all the attributes, we expect to depict the pattern of each crime-type across the City of Chicago for an entire year. Since there has been a radical decrease in the crime rate over the past decade, it is detrimental to consider early time and location attributes for each criminal record. For a sound prediction of the occurrence of a crime at any location and any hour of a day, it is required to consider the data that is consistent and out of exemptions. Therefore in order to abstain from false conjectures and guarantee a reliable prediction model, we plan on considering the criminal records of past 6 years to train our algorithm. The derived prediction model is then tested against the records from recent years for validation and determining the accuracy rate of our model. To increase the reliability of the prediction we intent to compare the accuracy rate among various machine learning algorithms such as logistic regression, support vector machine, gradient boosting machine, neural networks, and decision trees.

Theft	Battery
Robbery	Criminal Damage
Deceptive Practice	Narcotics
Domestic Violence	Non-Criminal (Subject Specified)
Assault	Criminal Trespass
Gambling	Arson
Burglary	Prostitution
Concealed Carry License Violation	Human Trafficking
Motor Vehicle Theft	Weapons Violation
Homicide	Offense involving Children
Crime Sexual Assault	Sex Offense
Obscenity	Non-Criminal
Liquor Law Violation	Interference with Public Officer
Kidnapping	Public Peace Violation
Intimidation	Stalking
Public Indecency	Ritualism

Table 2: Classification of crimes

3 Sub-problem statements and their Analysis

The crime data extracted from the CLEAR of Chicago police department is cleansed and normalized to eliminate the error data sets and to remove N/As. Following the path of analyzing and predicting the crime and vulnerability of the location to the crimes, we manipulated the data set to extract time of a day, month and year attributes. Also, with the occurrence rates and similarities in the crime types, we have merged some of the data.

Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	...	Community Area	FBI Code	X Coordinate	Y Coordinate	Year	Updated On	Latitude	Longitude
2003-11-26	062XX S JUSTINE ST	0610	BURGLARY	FORCIBLE ENTRY	APARTMENT	False	...	67.0	05	1167071.0	1863359.0	2003	04/15/2016 08:55:02 AM	41.780614	-87.663033
2003-11-26	067XX S HONORE ST	0610	BURGLARY	FORCIBLE ENTRY	APARTMENT	False	...	67.0	05	1165179.0	1859961.0	2003	04/15/2016 08:55:02 AM	41.771330	-87.670066
2003-11-27	053XX W NORTH AVE	1320	CRIMINAL DAMAGE	TO VEHICLE	OTHER	False	...	25.0	14	1140231.0	1910112.0	2003	04/15/2016 08:55:02 AM	41.909443	-87.760292

Figure 1: Sample Data Set

3.1 Determining training and testing set from the data

The line graph below (Figure 2) shows the trend of the number of crimes that have occurred from the year 2001- 2017. The line graph was chosen to visualize the above condition as line charts help to clearly visualize the trend over a period. From the above graph we could understand that the year 2003 has the highest number of crimes which gives us a branch to analyze and drill deep into the crime data set. Since there has been a radical decrease in the crime rate over the past decade, it is detrimental to consider early time and location attributes for each criminal record. For a sound prediction of the occurrence of a crime at any location and any hour of a day, it is required to consider the data that is consistent and out of exemptions. Consequentially, we designate crimes occurred in the years 2011-2015 as the training data set and 2015-2017 as our test case.

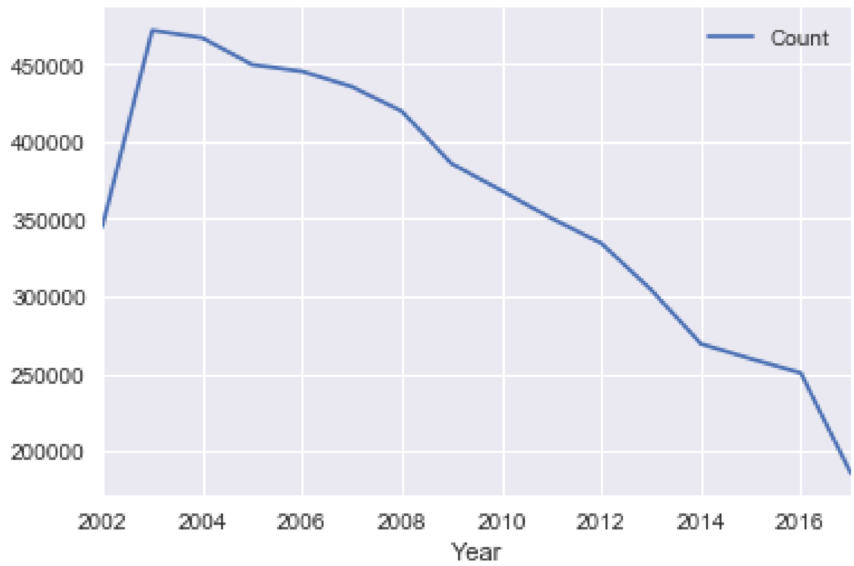

Figure 2: Occurence of Crime over the years (2001 - 2017)

3.2 Determining top 5 most dangerous communities w.r.t crime occurrence

Figure 3 & 4 shows the top ten most dangerous community areas in the City of Chicago. The City of Chicago is divided into various community areas. To understand the rate of crimes based on the community areas, we plotted a column chart which shows the distribution of crime rates based on the community areas. On mapping the community area number to the area on the map of Chicago we understood that the Community Area 25 which has the highest rate of crimes from 2001- Present.

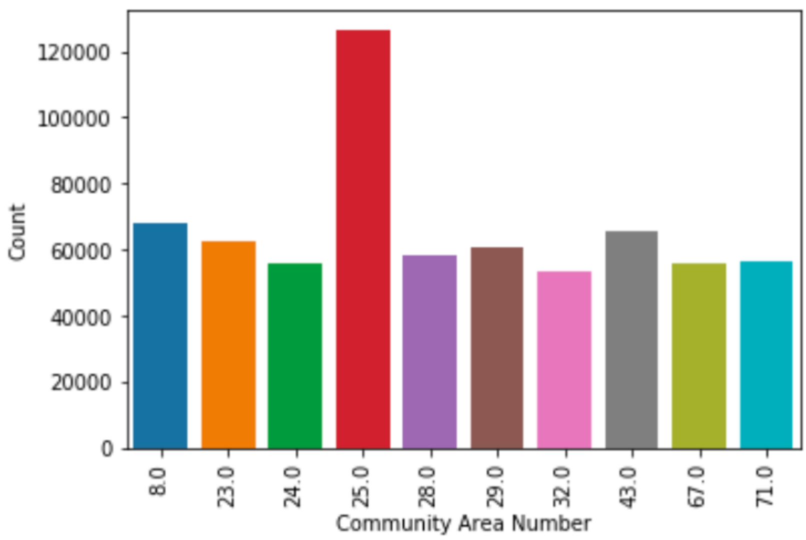

Figure 3: Top 10 most dangerous communities w.r.t crime occurrence

In order to validate our previous interpretation of identifying the community area with most number of crimes, we analyze the past five year trend of the crime rates in community areas. On plotting a trend line graph we understand that Community Area 25 which is Austin has the highest crime rate from 2001 – Present.

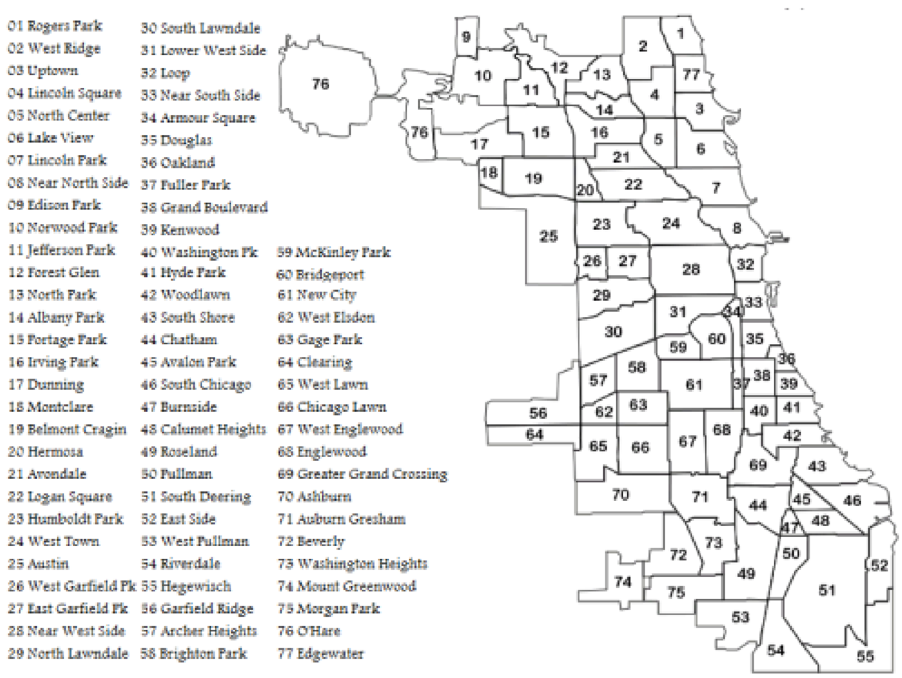

Figure 4: Geographical illustration of top 5 most dangerous communities w.r.t crime occurrence

The density of crime occurrence is demonstrated using a heat map (Figure 5). The graph (heat map) shows the geographical distribution of the crimes in the city of Chicago. The intensity of the color shows the number of crimes. The higher the intensity and higher is the number of crime in that particular area of City of Chicago. This could give a quick insight on which region in the City of Chicago ranks the highest in Crime Rates. The graph is plotted based on the latitude and longitude of the crimes.

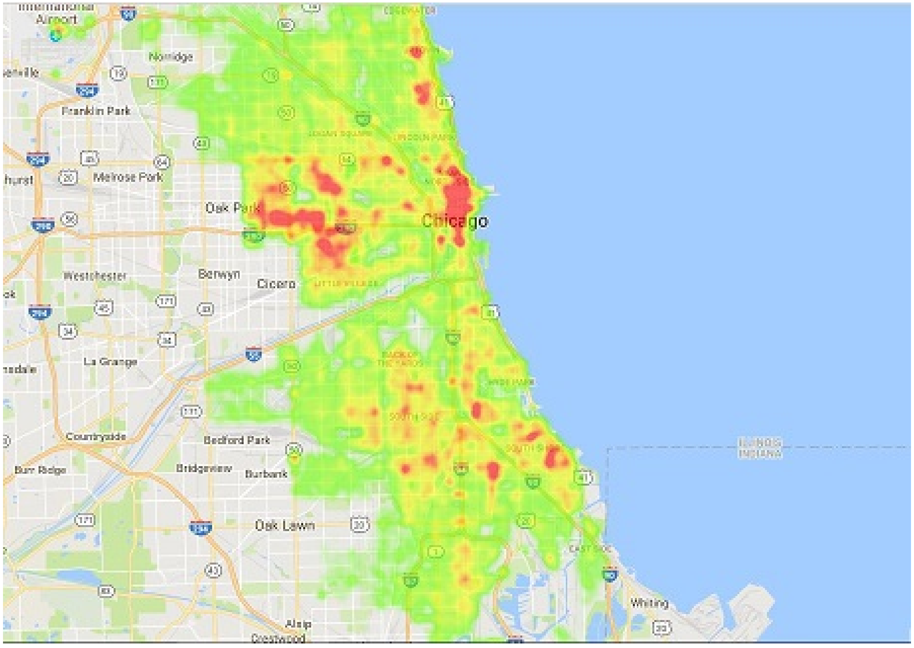

Figure 5: Geographical illustration of top 5 most dangerous communities w.r.t crime occurrence

3.3 Crime Vs. Time of Occurrence

In the heat map plotted below (6) we try to visualize and analyze the distribution of crime rate based on the time of the day. We had a column named date which had the time elements in our dataset. We split the time elements and added it to a separate column named time and grouped it into four categories 00 AM-06 AM, 06 AM-12 PM. 12 PM – 18 PM and 18 PM-00 AM. We plotted a heat map based on the rate of crime, based on the time of the day and the type of the crime. This could give answer the question of which crime happens the most and at what time. So the intensity of the color in the heat map shows the higher number in crime rate. By observing the heat map we understand that the theft has the higher intensity and its from 12:00PM – 18:00PM.

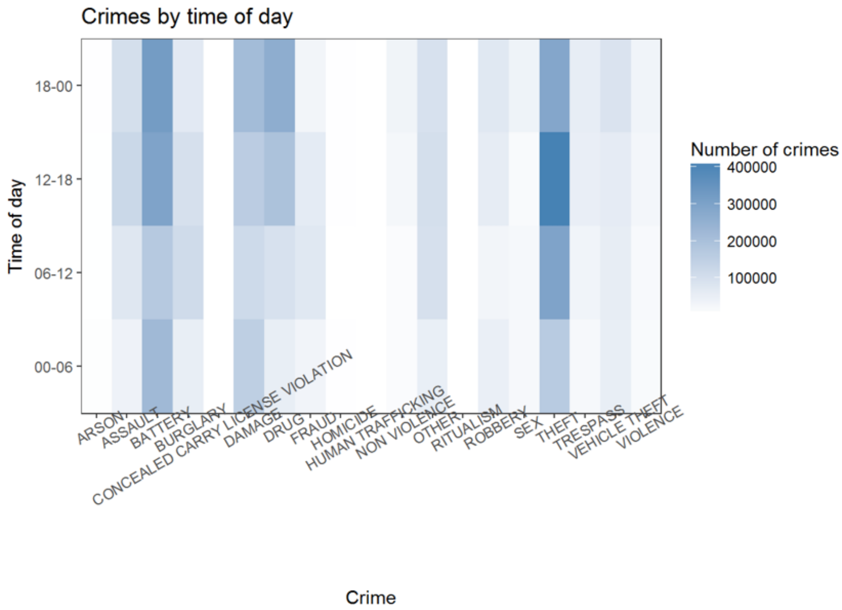

Figure 6: The above graph illustrates the nature of occurrence of the crime

3.4 Determining top 10 crime types for prediction

From Figure 2 we were able to analyze the distribution of crime (primary) types based on the percentage of their crime rates. This gives us a better understanding of the major crimes that occur. The graph in Figure depicts different crime types with respect to number of times of occurrence. We interpret that theft has the highest percentage and is the crime type with the highest crime rate.

Thus extracting the top 10 crime types on the basis of occurrence we obtain the figure below (Figure 7)

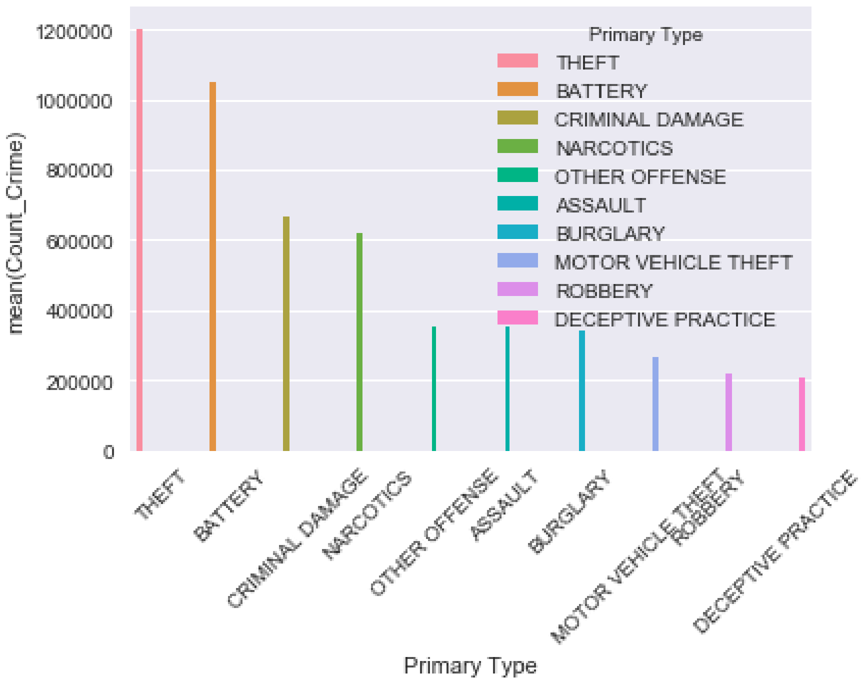

Figure 7: Top 10 crimes in the City of Chicago

4 DNN Classifier

The model, DNN Classifier, suggests the particular crime as the primary type in the prediction and the corresponding predictors include: time of a day, month, year, location, location description (latitude & longitude), beat, weekday and community area . With the help of the predictors we aim to anticipate the crime type that is most likely to occur with respect to the location description, the time of the day, month and year.

In this section, we describe the structure and learning method of our prediction model. We employed DNN Classifier to perform unsupervised learning and classify the different primary types based on various predictors as mentioned above.

For working on the neural network, the data was divided into training and testing sets, split by 70% for training the model and the remaining 30% is for evaluation. The predictor data is then fed into the neural network as the input. This data is then fed into the hidden layer along a path with some weights. ReLU (Rectified Linear Unit) activation function is used for the hidden layers. The output of the hidden layer is then given to the output node where the error is calculated with respect to the Testing data. This process is called the feed forward. Dropout is used for regularization to reduce the over-fitting of the data points.

This error is used to calculate a delta value which is then back propagated to modify the weights to get minimum error.

5 Implementation of the model

For the prediction, an estimator which has been modeled by DNN (Deep Neural Network) Classifier is used. The model is designed with 2 hidden layers with 20 nodes each. The hidden layers are provided with ReLU activation function. Also, the model also hosts, adamOptimzer with leaning rate of 0.001.

The data is label encoded before fed to the classifier. The output labels are one-hot encoded. Now the encoded training predictors are converted into training functions and then fed to the estimator. The output are then compared with the output functions for the training accuracy. Now to to find the generalization error, the test data which has never fed to the estimator is used and accuracy of the prediction is calculated.

However, by partitioning the available data into three sets, we drastically reduce the number of samples which can be used for learning the model, and the results can depend on a particular random choice for the pair of (train, validation) sets.

A solution to this problem is a procedure called cross-validation (CV for short). A test set should still be held out for final evaluation, but the validation set is no longer needed when doing CV. In the basic approach, called k-fold CV, the training set is split into k smaller sets (other approaches are described below, but generally follow the same principles). The following procedure is followed for each of the k "folds":

- A model is trained using k-1 of the folds as training data;
- the resulting model is validated on the remaining part of the data (i.e., it is used as a test set to compute a performance measure such as accuracy).

Below screenshots show the demo of the model (since the UI is not been created yet, only code demo is shown).

	Encoded_output	PrimaryTypes
0	0	BURGLARY
1	1	THEFT

Input Sample

	Year	Month	Hour	Primary Type	Weekday	Community Area
9	2016	4	14	THEFT	1	25.0

Prediction - Calculation

```
from sklearn.preprocessing import LabelEncoder, OneHotEncoder, StandardScaler

sample1 = sample.drop(['Primary Type'], axis=1)
```

```
for f in sample1.columns:
    if sample1[f].dtype == 'object':
        lbl_enc = LabelEncoder()
        # same as above encoding. it takes every object dtype from
        # pandas dataframe and converts to numerical labels
        sample1[f] = lbl_enc.fit_transform(sample1[f].values)

sample1
X = sample1.values
X
```

```
array([[  2.01600000e+03,   4.00000000e+00,   1.40000000e+01,
          1.00000000e+00,   2.50000000e+01]])
```

```
new_samples = np.array( X , dtype=np.float32)
predict_input_fn = tf.estimator.inputs.numpy_input_fn(
      x={"x": new_samples},
      num_epochs=1,
      shuffle=False)
```

Figure 8: Model Demo with 2 classes

Prediction - Output

```
list(predictions)
```

```
[1]
```

Figure 9: Model output with 2 classes

6 Evaluation

K-fold Cross-Validation method is used to increase the accuracy of the estimator that has be explained in the above section. Considering the huge data points and probability bounds distributed among the hidden layers, we have used 15-fold cross-validation. Here, instead of dividing the data set into training set, dedicated to train the estimator and the test data set, the entire data is randomly divided into 15 different sets and one random set is chosen to be a test data. The method helps in utilizing major portion to model the estimator.

Figure below explains the accuracy of the training model and the upgradation of the activation function for each batch of data set fed to the estimator. It can easily be observed that the weights have been updated to have the activation function under equilibrium.

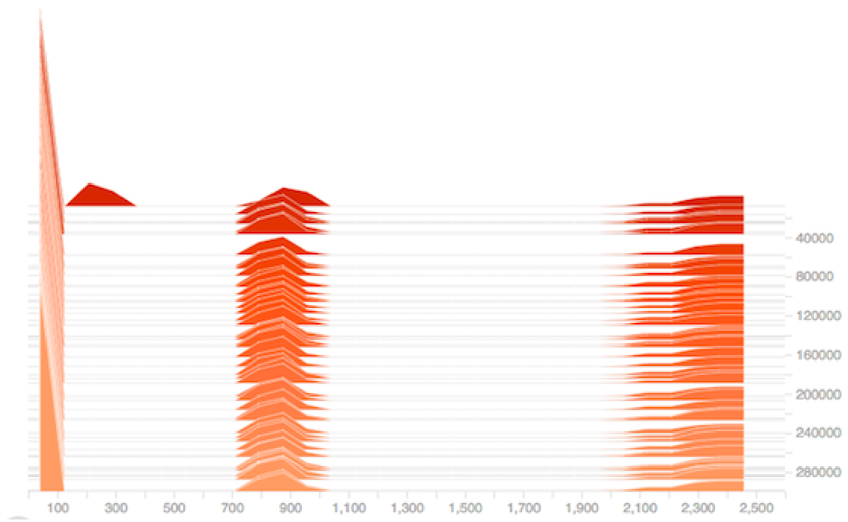

Figure 10: Mapping the output of activation function of the hiddenlayers for each batch of training sets

The model suggests the particular crime as the primary type in the prediction and the corresponding predictors include: time of a day, month, year, location, location description (latitude & longitude), beat, and community area . With the help of the predictors we aim to anticipate the crime type that is most likely to occur with respect to the location description, the time of the day, month and year.

The below figure summarizes the evaluation of the model with the SVM model.

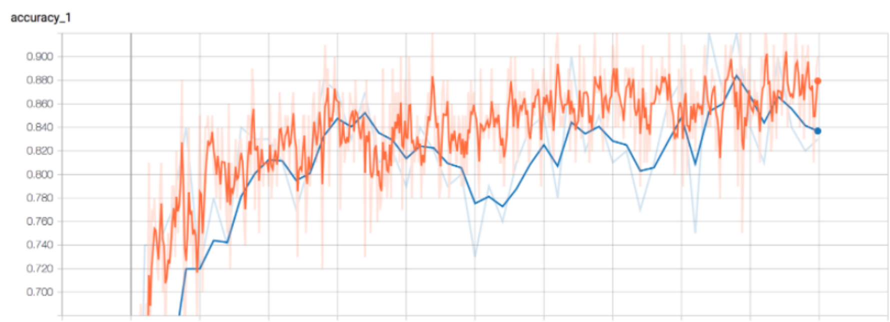

Figure 11: Accuracy graph of the DNN Classifier model

Estimator	Accuracy (%)	AUC
DNNClassifier	84.25	0.833
SVM	67.01	0.5052
6-fold Cross Validation on DNN estimator	79.03	--

Figure 12: Evaluation of the DNN model with the SVM model

7 Conclusion

In the first level of analysis, we tried to understand the overall trend of the Crimes in Chicago since 2001- Present under various factors. In section 3, the sub-problems along with the graphs we analyze the increase and decrease of the number of crimes based on the Community Area and the type of crimes and the time of the day. By understanding these trends, we try to drill down the data and focus on the comparison of the number of crimes based on the type of crime, community area and the time of the day. The City of Chicago is divided into different community areas which gives us a better understanding of which part of the city has more crime rates. The types of crimes give us a better understanding to analyze the different types of crimes that have been committed in the past in the City of Chicago. On analyzing the time of the day, we could drill deep into the particular time of the day a crime has been committed.

In summary, the research focused on the prediction of a particular crime in a location based at a particular time of the day of a moth using Deep Neural Networks.

It was understood that the neural network, with a single hidden layer will be insufficient for prediction of the crime if more inputs are given to it and a deep learning is required to process the amount of data given. The key to improving the prediction is increasing the number of hidden layers in the network. To avoid a local minimum every time it is required to start the weights at random values at the start. This way the minimum can be avoided and the optimum output can be obtained.

References

[1] Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang, Yi Qin, and Michael Chau. Crime data mining: a general framework and some examples. *computer*, 37(4):50–56, 2004.

[2] Matthew S Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.

[3] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. Learning to detect patterns of crime. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 515–530. Springer, 2013.

[4] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. Automatic crime prediction using events extracted from twitter posts. *SBP*, 12:231–238, 2012.