Predicting Crime using Deep Learning

BRAC
UNIVERSITY

Inspiring Excellence

Muhammad Nafees Shihab (13301097)

Anupam Chowdhury (13301091)

SK. Belayet Mahmood (13301100)

Department of Computer Science and Engineering

Submission Date: 24 December, 2017

Supervisor:

Amitabha Chakrabarty, Ph.D.
Assistant Professor
Department of Computer Science and Engineering

Co-Supervisor:

Rubayat Ahmed Khan
Lecturer
Department of Computer Science and Engineering

Declaration

We hereby declare that, this thesis report is our own work and has not been submitted for any other degree or professional qualifications. All sections of the paper that use quotes or describe an argument or concept developed by another author, have been referenced in the reference section.

Signature of Supervisor

Amitabha Chakrabarty, Ph.D
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Signature of Co-Supervisor

Rubayat Ahmed Khan
Lecturer
Department of Computer Science and Engineering
BRAC University

Signature of Authors

Muhammad Nafees Shihab
(13301097)

Anupam Chowdhury
(13301091)

SK. Belayet Mahmood
(13301100)

Abstract

Criminal activities are available in every region of the world influencing social life and financial improvement. As such, it is a major concern of numerous legislatures who are utilizing distinctive advanced innovation to handle such issues. Crime Analysis, a sub branch of criminology, considers the behavioral example of criminal activities and tries to recognize the pointers of such events. Distinguishing the patterns of criminal activity of a place is vital in order to prevent it. Law enforcement organizations can work effectively and respond more rapidly if they have better knowledge about crime patterns in different geological points of a city. Deep learning agents work with data and utilize distinctive systems to discover patterns in data making it exceptionally helpful for predictive analysis. Law enforcement agencies utilize diverse patrolling techniques in light of the data they get the chance to keep a region secure. The aim of this paper is to use deep learning models to predict and classify a criminal incident by type, depending on its occurrence at a given location. The experimentation is conducted on a dataset containing crime records. For this supervised classification problem, we used a new approach - LSTM (Long Short Term Memory) and was able to classify crimes with 64.2% accuracy. CNN (Convolutional Neural Network) & Shallow dense model were used also. Solving the imbalanced class problem, the deep learning agent was able to classify crimes.

Index Term: Deep learning, Criminal incident, Supervised classification, LSTM, CNN, Shallow dense model.

Acknowledgement

First of all, we are so much grateful to ALLAH and to our parents.

Then, we are grateful to our supervisor, Dr. Amitabha Chakrabarty, for his immense support throughout the work. He made us to get out of our comfort zones and to push the limit. Without his guidance, it would have been impossible to get introduced to the amazing research prospect of Deep Learning.

We are also hugely indebted to our co-advisor Rubayat Ahmed Khan & Dipankar Chaki, lecturer of BRAC University, for sharing their ideas and experiences, which helped us to make right choices in the implementation phase.

Nevertheless, we would like to express our gratitude to Department of Computer Science and Engineering, BRAC University and our teachers for helping us with all the necessary support.

Contents

CHAPTER 1: Introduction

CHAPTER 3: Algorithm and Working Principles

CHAPTER 4: Dataset and Attributes

CHAPTER 5: Data Analysis & Result

LIST OF FIGURES

CHAPTER 4: Dataset and Attributes

CHAPTER 5: Data Analysis & Result

LIST OF TABLES

CHAPTER 4: Dataset and Attributes

CHAPTER 5: Data Analysis & Result

CHAPTER 1

Introduction

Deep learning have become a vital part of crime detection and prevention. A Deep learning agent can learn and analyze the pattern of occurrence of a crime based on the reports of previous criminal activities and can find hotspots based on time, type or any other factor. This technique is known as classification and it allows to predict nominal class labels. Classification has been used on many different domains such as financial market, business intelligence, healthcare, weather forecasting etc.

In this research, we used a dataset, Chicago's crime records from 2012 – 2017 [1] which contains the reported criminal activities in the neighborhoods of the city Chicago for a duration of 6 years. We used different deep learning models like LSTM, CNN and Shallow dense model to find hotspots of criminal activities based on the location. Results of different models have been compared and most the effective approach has also been documented.

1.1 Motivation

Criminal law and sociology specialists have also been studying the underlying pattern of crime and its relation to a region or an area's social or conomic development, the characteristics of different groups of people living there, family structure, level of education among other things. Different studies and researches have shown that significant concentration of crime happens at micro level of a region. This clustering is often called - hotspot. Research shows that a good neighborhood often has few streets or locations which have a higher concentration of criminal activities compared to others. It is also true in the opposite case, as a bad neighborhood often has many places which have relatively lower rates of crime.

David Weisburd, a renowned criminologist proposed to change people-centric paradigm of strategy to place-centric paradigm. He identified that geographical topology and microstructures are more important in dealing with hotspots [2]. Deep learning can harmonize the same concept by taking a data driven approach to identify hotspots. The agent can also incorporate social and economic standards at a micro level to find the indications.

The intention of this research has been to explore different deep learning techniques to find clusters and analyze the reported criminal activities to find underlying patterns.

1.2 Research Goal

Criminal activities take place all over the world and law enforcement agencies have to deal with them successfully. If enforcement agencies have an earlier assumption of the class of the crime, it would give them tactical advantages and help to resolve cases quicker. Also, a complete study of criminal activity in a geographic area helps to understand the underlying pattern of the crime.

With our research we hope to find answers to the following questions:

I. Does a criminal database that contains the geographical location & basic details of the criminal activity have plenty pointers to predict a type of crime?

II. Given just a geographic location, how accurately can we classify the crime?

1.3 Thesis Contribution

The objective of this thesis was to analysis and of crime data, to predict the crime type and find out the classification of crime using location data and to find out which deep learning model has better effectiveness for classifying the crime types according to the given location. We used a new approach – LSTM (Long Short Term Memory) to predict the crime type more accurately according to their location rather than CNN and Shallow dense model, which are widely used by other researchers. The analysis can be used by other researchers who are willing to work on innovating ways for predicting crime. This thesis gives an insight of use of deep learning models like LSTM, CNN and Shallow dense model for predicting crime.

1.4 Problem Statement

There is no digitalized criminal record available in Bangladesh. All records are hand written and those data are not well managed. They could not provide us any digital version of criminal records. We need a huge dataset with detail information. We attempted to collect crime incident data from Dhaka Metropolitan Police and also from newspapers manually for our research purpose but we could not find any source that provides crime data. We therefore had to use the data from Kaggle.com.

1.5 Methodology

We used a dataset, Chicago's crime records from 2012 – 2017 [1] which contains the reported criminal activities of Chicago state with so many detail information. Then we reduce the noise from the dataset and used it to three deep learning models – LSTM, CNN and Shallow dense model. We split the dataset into training and testing data and find out the loss function and accuracy rate of the models by using location data as input and to predict the crime type as output. Finally, we analyze the accuracy rates of those deep learning models.

1.6 Tools

1. Language – Python
2. Libraries – Keras, Pandas
3. IDE – Anaconda

1.7 Data

For our research, there was no data available. That is why we used Chicago's crime records from 2012 – 2017 [1]. This dataset contains huge amount of crime data.

1.8 Thesis outline

Chapter 1 gives a brief overview of our research, our estimated goal and what we have gained.

Chapter 2 discusses the literature review and background study of our thesis, what properties we considered and how they work and more.

Chapter 3 discusses about the algorithms and working principle.

Chapter 4 discusses about the dataset and its attributes. It also shows the data visualization.

Chapter 5 discusses the data analysis and the comparative analysis of the resulting output we found.

Chapter 6 ends our paper with conclusion and proposed future work for the system.

CHAPTER 2

Literature Review

To eradicate the criminal activity has always been a priority for governments around the world, many researches has been done to effectively find countermeasures and indicators of crime prior to happening. Criminologists have been pursuing to identify hotspots that need major attention from law enforcement agencies. Researches have been done to study the relation between criminal activities and socio-economic variables like unemployment [3], income level [4], race [5] and level of education [6].

Woo Kang and Bong Kang have worked on the multi-modal data using deep learning to predict the crime occurrence. They proposed a feature-level data fusion method with environmental context based on a deep neural network (DNN). Experimental performance results show that their DNN model is more accurate in predicting crime occurrence than other prediction models. [7]

Wang et al proposed the Series Finder, a Deep learning agent that tried to find patterns in crime committed by same criminal or groups of criminals. Clustering has also been used to study patterns of criminal behavior and terrestrial criminal history. [8]

A group of researchers were able to predict if particular areas of London would become a criminal hotspot by analyzing the usage of mobile network infrastructure and demographic information of people living in different areas of London. They have implied that data collected by mobile networks contain indicators for predicting crime levels. [9]

Iqbal, Murad, Mustapha, Panahy, & Khanahmadliravi combined two datasets - 1990 US LEMAS and crime data 1995 FBI UCR and applying classification techniques like Decision Tree and Naive Bayesian algorithm, 83.95% accuracy have been achieved when asked to predict a crime category for different states of USA. [10].

However, this paper does not disclose if there were any imbalanced classes of crime category. The same databases were also explored by Shojaee, Mustapha, Sidi & Jabar who employed a number of Deep learning algorithms, where k-Nearest Neighbor algorithm performed better than other algorithms by having an accuracy of 89.50%. They also used the Chi-square feature to improve the feature selection. [11]

Remond and Baveja have worked on the data noise problem and studied how some police reports or cases are idiosyncratic and do not contain good indicative matrices. Their proposed system called Case-Based Reasoning (CBR) filtered out these cases, and using this system, they were able to predict better compared to not having any filters on the data. [12]

Sadhana and Sangareddy have used twitter data and sentiment analysis to predict crime in real time. They also used this data to map the concentration of crime occurrences and find large scale hotspots. [13]

Machine learning is such type of strategy which evaluate the data and computerizes analytical model building. It is a branch of manmade intelligence in light of machines ought to have the capacity to learn and adjust through understanding [17].

The essential preface of machine learning is to construct algorithm which can extract or get input information and use statistical evaluation to predict an output value within a suitable range. Usually Machine learning algorithms can be divided into two categories, one is supervised and another is unsupervised. Supervised algorithms require people to provide each input and preferred output, similarly to furnishing comments about the accuracy of predictions throughout training. The algorithm will be applied to the new data when the training is finished [18].

A common use of it is to analysis historical data to predict future event. In the case of unsupervised algorithm deep learning approach is used to review data to reach to the conclusion instead of training data with desired outcome [18].

The main intention of unsupervised algorithm is to find out the hidden pattern from a large dataset [19].

Most of the practical machine learning uses supervised learning. In supervised learning we have to give input variables (X) and an output variable (Y) and use an algorithm to learn the mapping function from the input to the output.

$Y = f(X)$

The main idea is to approximate the mapping function so well that when the new input data is given (X) that we can predict the output variables (Y) for that data.

It is called supervised learning on the grounds that the procedure of an algorithm gaining knowledge from the collection of trained dataset can be thought of as an instructor regulating the learning procedure. We know the right answers, the calculation intelligently makes forecasts on the trained dataset and is remedied by the instructor. Learning stops when the calculation accomplishes an adequate level of execution [20].

Deep learning is a machine learning method that instructs machines to do what comes naturally to humans: learn by illustration. Deep learning is a key innovation behind driver less autos, empowering them to perceive a stop sign, or to recognize a person on foot from a lamppost. It is the way to voice control in buyer gadgets like telephones, tablets, TVs, and without hands speakers. Deep learning is getting loads of consideration recently and in light of current circumstances. It's accomplishing comes about that were impractical some time recently.

In deep learning, a PC demonstrate figures out how to perform classification errands straightforwardly from pictures, messages, or sound. Deep learning models can accomplish best in class precision, sometimes surpassing human-level execution. Models are prepared by utilizing an expansive arrangement of labeled data and neural network architectures that contain many layers.

While deep learning was first estimated in the 1980s, there are two primary reasons it has just as of late turned out to be helpful:

1. Deep learning requires a lot of labeled data. For instance, driver less vehicles improvement requires a large number of pictures and many hours of video.

2. Deep learning requires significant computing power. Superior GPUs have a parallel design that is proficient for deep learning. At the point when joined with groups or distributed computing, this empowers advancement groups to reduce training time for a deep learning system from weeks to hours or less.

Most deep learning methods use neural network architectures, which is why deep learning models are often referred to as deep neural networks.

The term “deep” usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150.

Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction [21].

CHAPTER 3

Algorithm and Working Principles

3.1 Introduction

Deep learning is a machine learning method that instructs machines to do what comes naturally to humans: learn by illustration. The term “deep” usually refers to the number of hidden layers in the neural network. Traditional neural networks only contain 2-3 hidden layers, while deep networks can have as many as 150. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction [21].

3.2 Dropout

Dropout is a recently introduced algorithm for preparing neural networks by arbitrarily dropping units amid preparing to keep their co-adjustment. A numerical examination of a portion of the static and dynamic properties of dropout is given utilizing Bernoulli gating variables, sufficiently general to oblige dropout on units or associations, and with variable rates. The system permits an entire examination of the group averaging properties of dropout in linear networks, which is helpful to comprehend the non-linear case. The gathering averaging properties of dropout in non-linear strategic networks result from three principal conditions: (1) the approximation of the expectations of calculated capacities by standardized geometric means, for which limits and estimates are inferred; (2) the algebraic equity between standardized geometric means of logistic functions with the logistic of the means, which mathematically characterizes logistic functions; and (3) the linearity of the methods as for aggregates, and in addition results of autonomous factors. The outcomes are also extended to other classes of transfer functions, including rectified linear functions. Estimate blunders tend to scratch off each other and don't gather. Dropout can likewise be associated with stochastic neurons and used to foresee terminating rates, and to back propagation by review the regressive spread as ensemble averaging in a dropout linear system. Additionally, the merging properties of dropout can be comprehended as far as stochastic angle plunge. At long last, for the regularization properties of dropout, the desire of the dropout slope is the angle of the corresponding approximation ensemble, regularized by an adaptive weight decay term with an affinity for self-predictable change minimization and sparse representations [22].

3.3 Activation Function

Activation functions are an extremely important feature of the artificial neural networks. They basically decide whether a neuron should be activated or not. Whether the information that the neuron is receiving is relevant for the given information or should it be ignored.

Y = Activation (∑ (weight * input) + bias) ……………………………….……….. (1)

The activation function is the nonlinear transformation that we do over the input signal. This transformed output is then observed to the following layer of neurons as input. When we don't have the activation function the weights and predisposition would just do a linear transformation. A linear condition is easy to solve however is constrained in its ability to take care of complex issues. A neural network without an activation function is basically only a linear regression model. The activation function does the non-linear change to the information making it competent to learn and perform more intricate errands. We would need our neural networks to take a shot at confused assignments like dialect interpretations and image classifications. Linear changes could never have the capacity to perform such assignments. Activation functions make the back-propagation possible since the gradients are supplied along with the error to update the weights and biases. Without the differentiable nonlinear function, this would not be possible.

3.4 Optimization

The interchange between optimization and machine learning is a standout amongst the most essential advancements in present day computational science.

Optimization formulae and strategies are turned out to be crucial in designing algorithms to extract fundamental learning from tremendous volumes of data. Machine learning, nonetheless, isn't just a tool of optimization innovation yet a quickly developing field that is itself creating new optimization thoughts. Optimization approaches have delighted in prominence in machine learning due to their wide pertinence and appealing theoretical properties. The expanding multifaceted nature, size, and assortment of the present machine learning models require the reassessment of existing presumptions. [23]

3.5 Adam optimizer

Adam is an advancement algorithm that can utilized rather than the traditional stochastic gradient plummet system to update network weights iterative situated in training data. Adam was presented by Diederik Kingma from OpenAI and Jimmy Ba from the University of Toronto in their 2015 ICLR paper (poster) titled “Adam: A Method for Stochastic Optimization “. The algorithm is called Adam. The authors pointed out some attractive benefits of using Adam on non-convex optimization problems, as follows:

- Simple to implement.
- Very efficient in computation.
- Uses little memory.
- Invariant to diagonal rescale of the gradients.

- Well applicable for problems that consists large number of data and/or parameters.
- Appropriate for non-stationary objectives.
- Appropriate for problems with very noisy/or sparse gradients.
- Hyper-parameters have intuitive interpretation and typically require little tuning.

3.5.1 How Does Adam Work

Adam is distinctive to traditional stochastic gradient descent. Stochastic gradient descent keeps up a solitary learning rate (named alpha) for all weight update and the learning rate does not change during preparing. The Adam can be describe as combining the advantages of two other extensions of stochastic gradient descent. Specifically:

- Adaptive Gradient Algorithm (AdaGrad) that keeps up a parameter learning rate that enhances execution on issues with inadequate gradients (e.g. common dialect and PC vision issues).
- Root Mean Square Propagation (RMSProp) that likewise keeps up per-parameter learning rates that are adjusted in light of the average of recent magnitudes of the gradients for the weight (e.g. how rapidly it is evolving). This implies the calculation does well on the web and non-stationary issues (e.g. uproarious).[24]

3.6 Models

We approached a new model for predicting crime – LSTM and then we used two widely used model, CNN and Shallow dense model.

3.6.1 Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) systems are a sort of intermittent neural system fit for learning request reliance in sequence prediction problems. This is a conduct required in complex issue areas like machine interpretation, speech recognition, and more. LSTMs are a perplexing zone of deep learning. It can be difficult to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field. There are few that are better at obviously and accurately articulating both the guarantee of LSTMs and how they function than the specialists that created them [25].

3.6.2 Convolutional Neural Network (CNN)

The field of machine learning has taken a dramatic twist in recent times, with the rise of the Artificial Neural Network (ANN). These biologically inspired computational models are able to far exceed the performance of previous forms of artificial intelligence in common machine learning tasks. One of the most impressive forms of ANN architecture is that of the Convolutional Neural Network (CNN). CNNs are primarily used to solve difficult image-driven pattern recognition tasks and with their precise yet simple architecture, offers a simplified method of getting started with ANNs. [26].

3.6.3 Shallow Dense Model

Shallow dense model is just a single hidden layer. The reason, of course, is the ability of deep nets to build up a complex hierarchy of concepts. It's a bit like the way conventional programming languages use modular design and ideas about abstraction to enable the creation of complex computer programs. Comparing a deep network to a shallow network is a bit like comparing a programming language with the ability to make function calls to a stripped down language with no ability to make such calls. Abstraction takes a different form in neural networks than it does in conventional programming, but it's just as important. [27]

3.7 Performance Metrics

The subsequent stage executing a machine learning algorithm is to discover how compelling the model in light of metric and datasets is. Diverse performance metrics are utilized to assess distinctive Machine Learning Algorithms. For instance a classifier used to recognize mages of various objects; we can utilize classification performance metrics, for example, Log-Loss, Average Accuracy, AUC, and so on. In the event that the machine learning model is attempting to predict a stock price, at that point RMSE (rot mean squared error) can be utilized to compute the effectiveness of the model. Another case of metric for assessment of machine learning algorithms is accuracy review or NDCG, which can be utilized for sorting algorithms essentially utilized by search engines. Along these lines, we see that distinctive metrics are required to quantify the productivity of various algorithms, likewise relying on the current dataset. It is imperative to pick appropriate metrics to assess how well an algorithm is performing. [28].

3.7.1 Accuracy

Accuracy is the most intuitive execution measure and it is essentially a proportion of effectively anticipated perception to the aggregate observation. One may imagine that, on the off chance that we have high exactness then our model is ideal. Truly, accuracy is a false positive and false negatives are practically same. Therefore, you have to look at other parameters to evaluate the performance of your model [29]. Accuracy measures how many predictions are matched exactly with the actual or true label of the testing dataset and returns the percentage of correct results.

3.7.2 Categorical Cross entropy

Categorical cross entropy is an alternative (and probably preferred, but more to that later) cost function for multinomial classification. If we have a set of classes C, and a bunch of samples x1,x2,x3,…,xN together with my predictions x^1,x^2,x^3,…,x^N for these samples, it’s defined as:

L(x) =H (p,p^)= –∑c∈Cp(x=c)logp^(x=c)…………………… (2) [30]

CHAPTER 4

Dataset and Attributes

4.1 Crime dataset and attributes

We did our experiment on a specific dataset. This dataset reflects reported incidents of crime that occurred in the City of Chicago from 2001 to present. The dataset contains more than 6,000,000 records/rows of data. We took the data from Kaggle.com. There are good number of previous analysis on this dataset. There are some good data representations, mapping and forecasting that really helped us to explore more about this dataset. The attributes of the dataset is given below.

Table 4.1.1: Attributes of Dataset

ID	Unique identifier for the record.
Case Number	The Chicago Police Department RD Number
Date	Date when the incident occurred
Block	The partially redacted address where the incident occurred
IUCR	The Illinois Uniform Crime Reporting code
Primary Type	The primary description of the IUCR code
Description	The secondary description of the IUCR code

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
District	Indicates the police district where the incident occurred
Ward	The ward (City Council district) where the incident occurred
X Coordinate	The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection
Y Coordinate	Coordinate - The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection
Year	Year the incident occurred
Latitude	The latitude of the location where the incident occurred
Longitude	The longitude of the location where the incident occurred
Location	The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal

So, there are a lot of information integrated in the dataset. This dataset is available in different formats. We worked on CSV file format. Here we will try to predict the crime type over the years. We choose a limited number (2012-2017) of dataset for our research purpose. Now, some data visualizations are given below.

4.2 Classification of crimes

There are 34 types of crime in this dataset. Some of them are very frequent and some of the types are very less in number. So, from this table we can say that not every types of crime is very serious. Moreover, different types of crimes makes it a multiclass problem.

Table 4.2.1: Crime Frequency

Crime type	Frequency
Theft	326987
Battery	281728
Criminal damage	154150
Narcotics	134829
Assault	90515
Other offense	87000
Burglary	82730
Deceptive practice	74843

Motor vehicle theft	60426
Robbery	58700
Criminal trespass	38595
Weapons violation	17058
Public place violation	13068
Offense involving children	11802
Prostitution	7601
Crime Sexual Assault	6759
Interference with public officer	6159
Sex offense	4885
Homicide	2618
Gambling	2212
Arson	2204
Liquor law violation	1950
Kidnapping	1098
Stalking	821
Intimidation	658
Obscenity	187

Non-criminal	90
Concealed carry license violation	85
Public Indecency	61

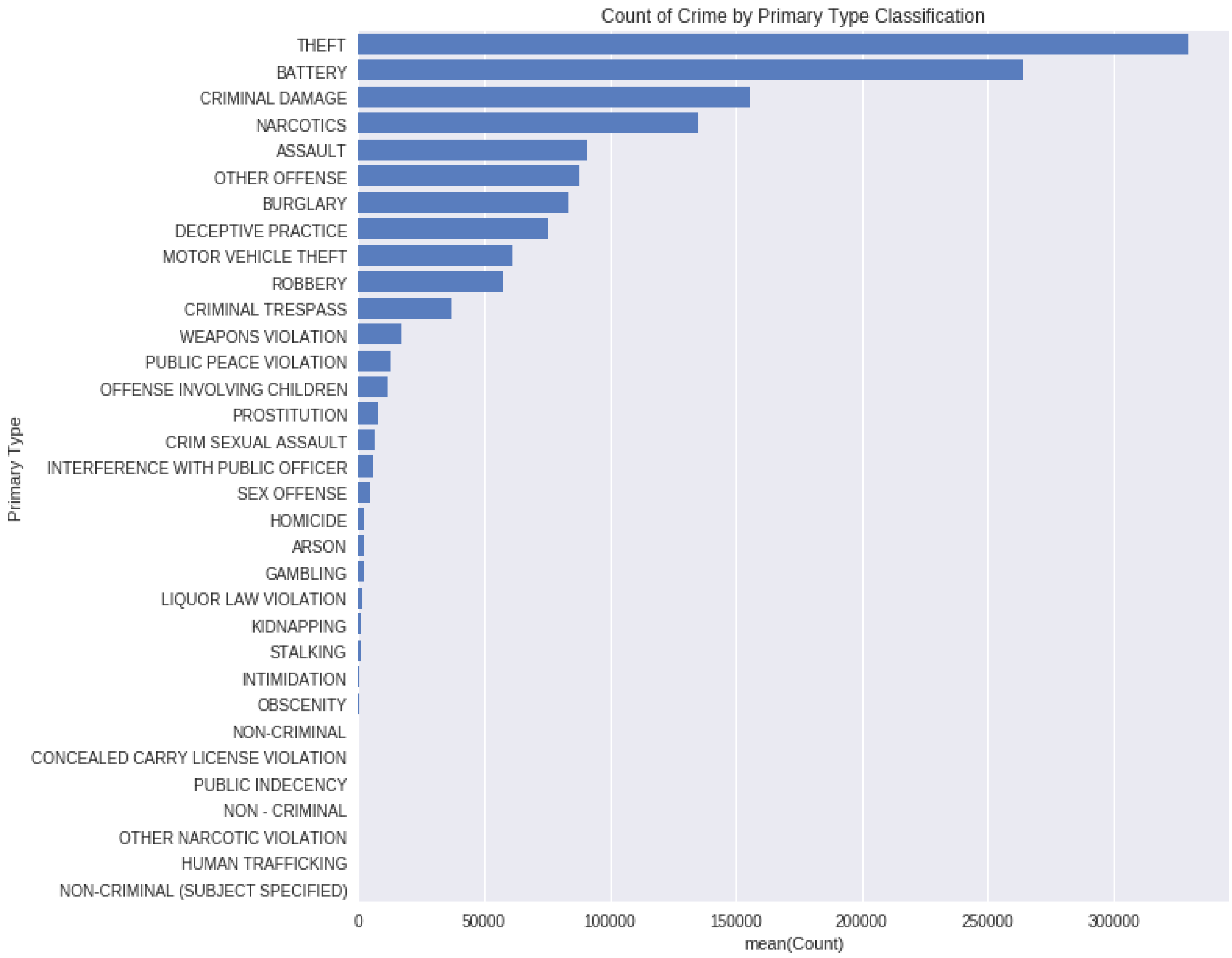


Fig. 4.2.1: Frequency of crime categories

In this graph there are a few number of crime types are occur very often and some of them are very less in number. For an example, theft is occurred most in that city which is 326987 times. Moreover, battery, criminal damage, narcotics, assault are happened more than 100000 times. On the other hand, more than half of the crime types are happened less than 5000 times and some of them are less than 1000 times. Also a few number of them are in between 50 to 500. So, it can be said that the classes are not equally distributed.

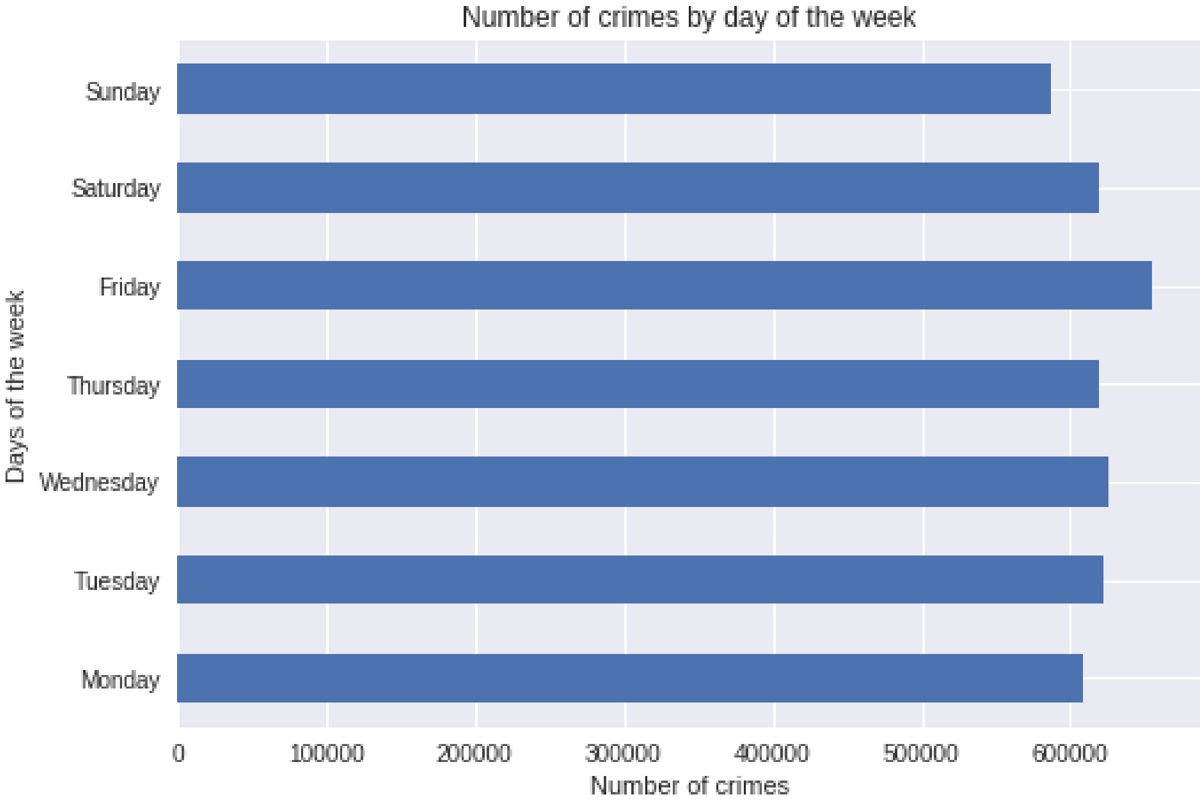


Fig. 4.2.2: Crimes occurring in different weeks

This graphs shows the weekly committed crime over the years in total. From this figure it is clear that crime occurrence happened every day of a week in a large number. Friday has the maximum number of crime occurrence. From Saturday to Wednesday the occurrence increase a bit.

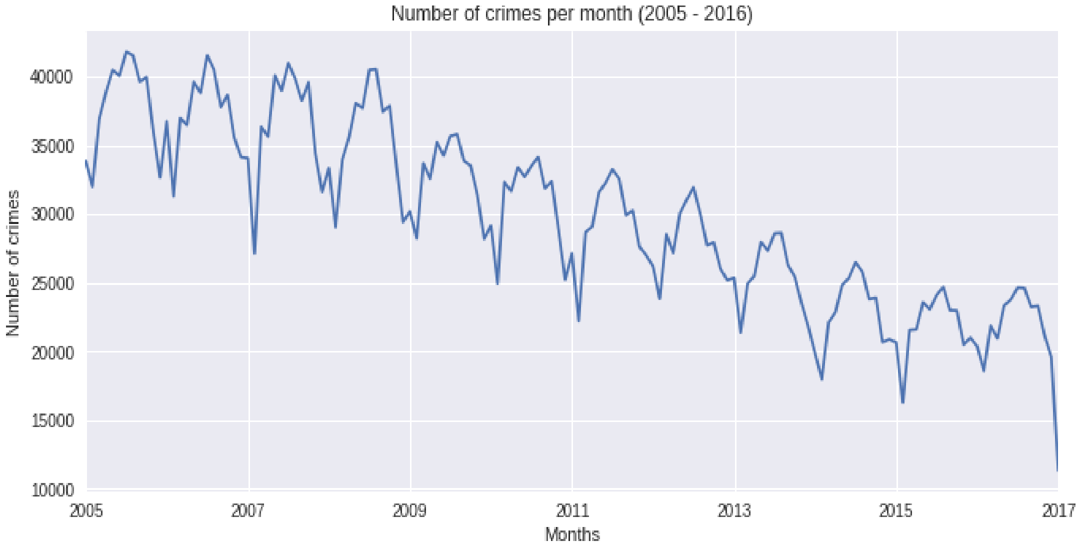


Fig. 4.2.3: Crimes occurring in different months

From Fig. 4.2.3, it is clear that the occurrence of criminal activity increase gradually through half of the year and decrease gradually through last half of the year. So, the criminal activity is at its highest pick at the middle of the year which is June, July and August. From that information it can be said that, Chicago had less number of crime in the winter season and maximum crimes in summer season.

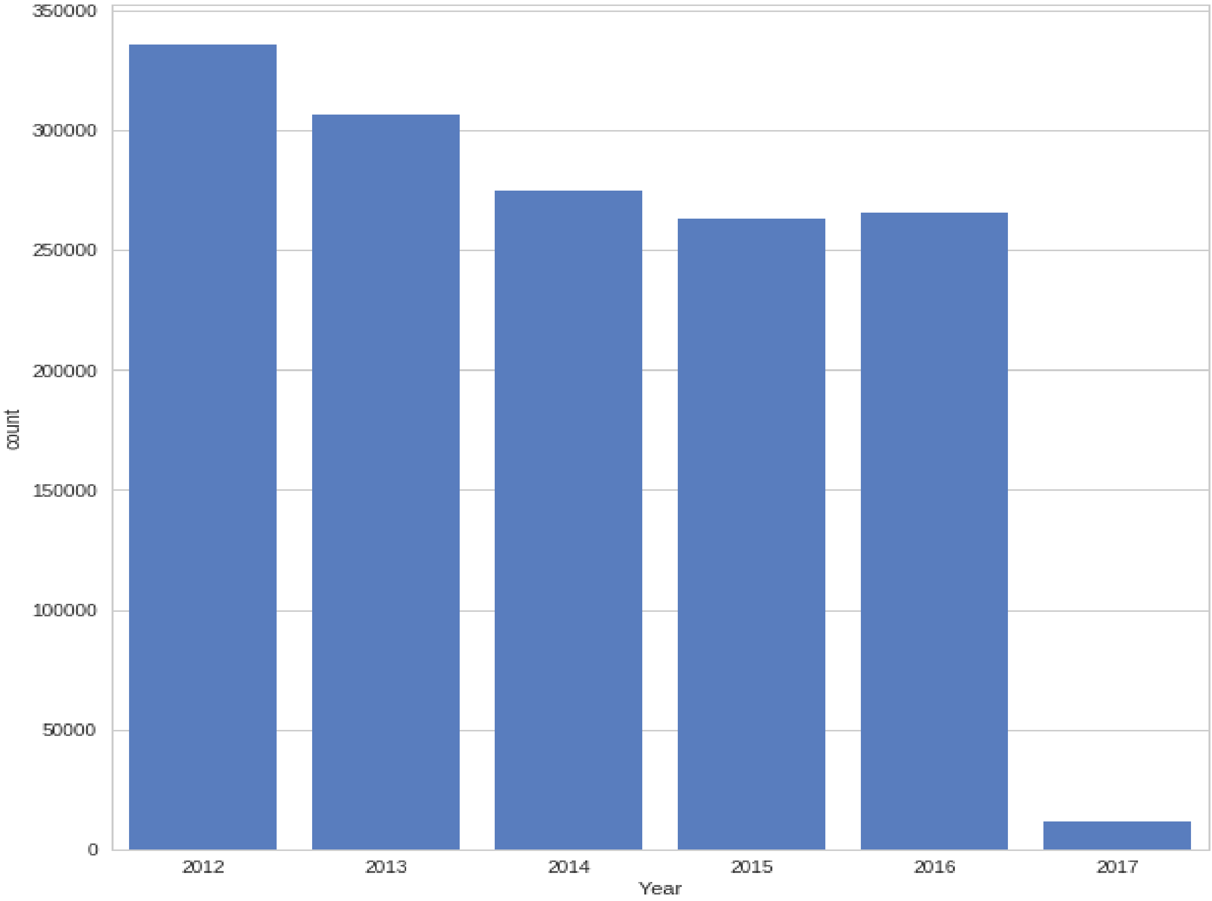


Fig. 4.2.4: Crimes occurring in different years

In this figure, the yearly occurred crimes are presented. Day by day the density of crimes are decreasing. We do not have the full crime report of 2017 in the dataset that's why we got a very less number of crimes in that year.

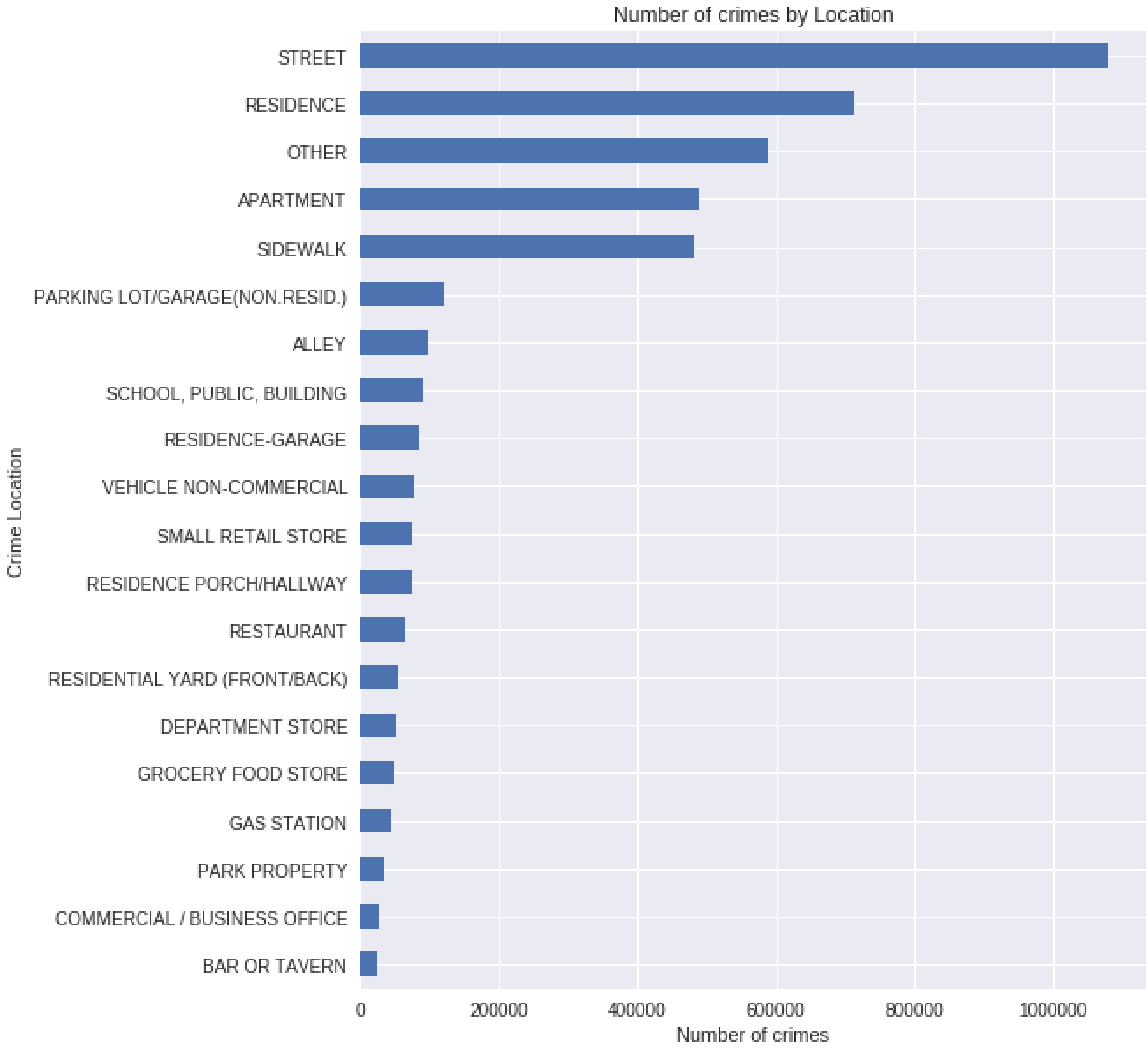


Fig. 4.2.5: Crimes occurred in different locations

We have different location types in our dataset. After plotting according to locations we have seen that the maximum number of crimes happened on the streets. A big amount of crimes occurred in residence, apartment, sidewalks and other places.

CHAPTER 5

Data Analysis & Result

5.1 Preprocessing

As we are working on a large dataset we have got 32 types different information and values or attributes for a particular crime. We used the CSV format of the dataset for our research. The attributes for a particular crime types are in different types of file format. Like, case id, latitude & longitude are in numeric values but Arrest is in Boolean and primary type or location description are in string type. In order to use these dataset in deep learning models all of our input set must be converted into numeric value.

Now, to predict primary crime type we do not need to use all of the attributes. We have to use the attributes which are very relevant to predict a crime type. So we selected some of the attributes which are given below.

Table 5.1.1: Input and Output

Input (X)	Location_Description
	Arrest
	Beat
	District

	Ward
	Community area
	Latitude
	Longitude
Prediction (Y)	Primary_type

We used different python libraries to do the work. For an example, we used numpy to label our data, we used pandas for out visual representation. Also, we used read_csv from pandas to take input from our CSV file.

5.2 Training and Testing Dataset

We took a recent portion of the total dataset which is 2012 to 2017 time zone to train our model. We divided our data into two parts, training dataset and another is testing dataset. We keep 60% of the data for training purpose. We have used three deep learning models for the training. They are LSTM (Long short-term memory), CNN (Convolutional Neural Networks), and Shallow dense layer model. All of them came up with decent accuracy. Firstly, we have tried to predict the total 32 primary crime types.

Supervised classification models are applied on Chicago Crime Dataset to predict the category of a crime incident. In the following part, performance of different classification models are discussed.

5.3 CNN (Convolutional Neural Networks)

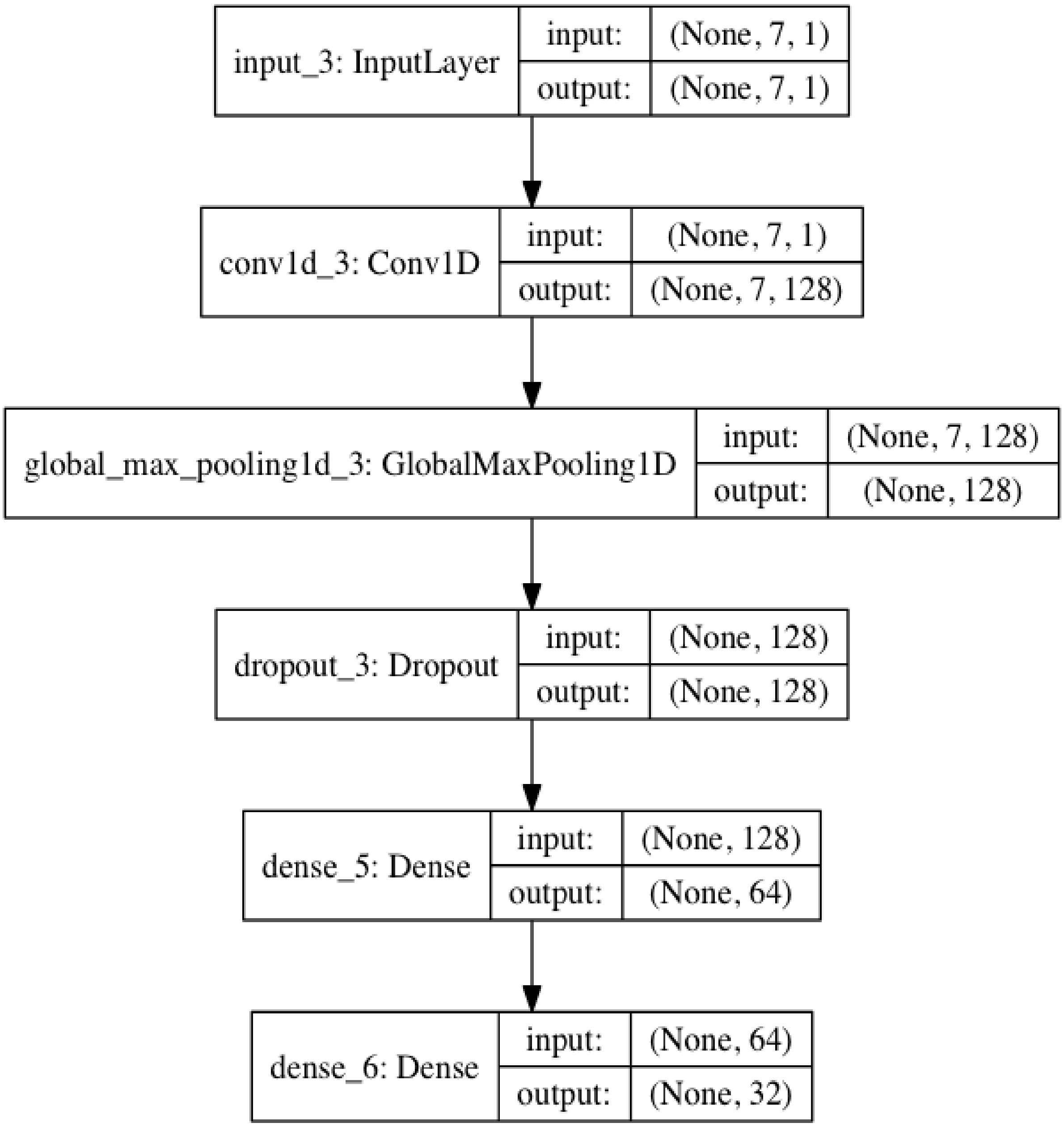
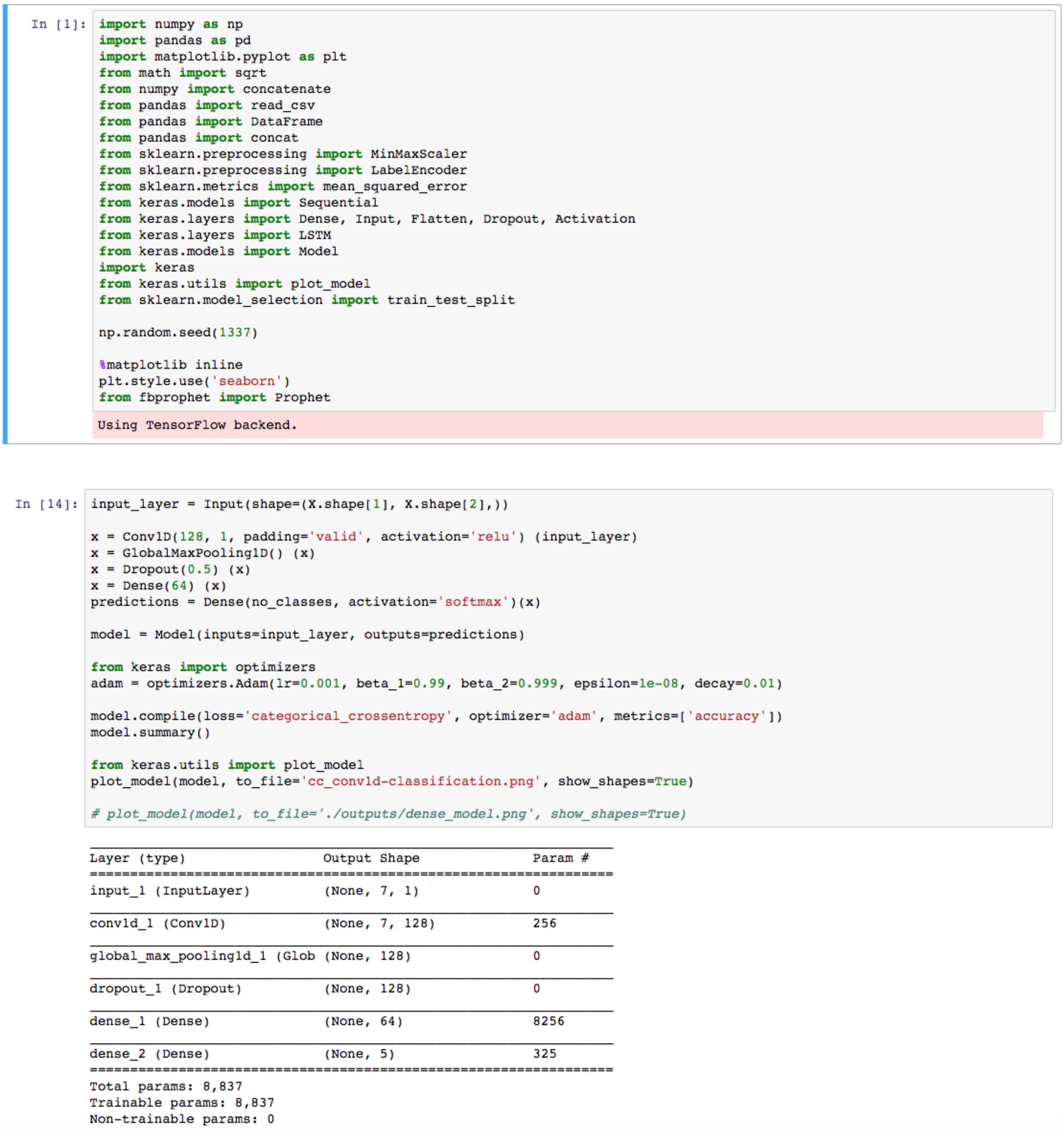


Fig. 5.3.1: Convolutional Neural Networks Model

We are proposing Convolutional Neural Network (CNN) as one of our classification model. The model takes input in the (7, 1) shape and returns the same size output. In this model it could extract up to 128 features. We have used 100 epoch for every model. This model uses Global Maxpooling after every convolution. Then we used dropout as our data was over fitted while running the algorithm. Then we flattened the output of last convolution layer and feed into two consecutive dense layers. In the last layer this model got 32 classes which is the number of the unique crime types in the dataset.



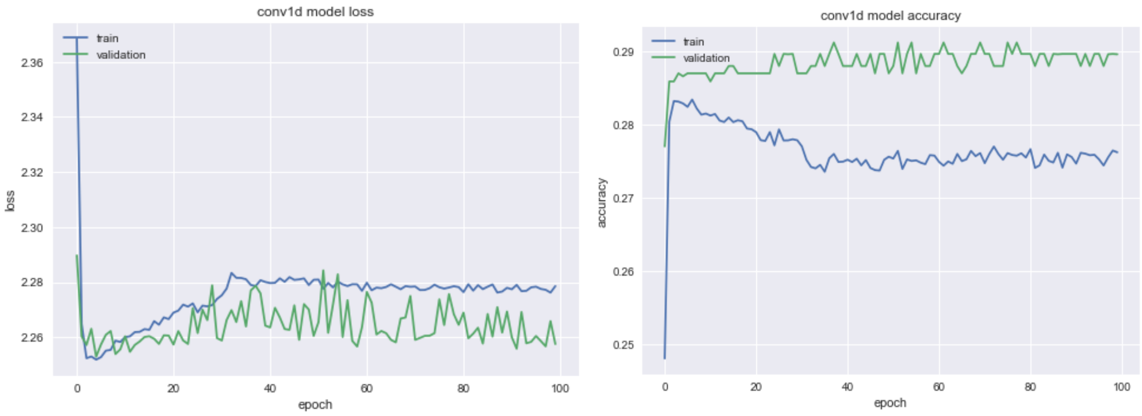
```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        from math import sqrt
        from numpy import concatenate
        from pandas import read_csv
        from pandas import DataFrame
        from pandas import concat
        from sklearn.preprocessing import MinMaxScaler
        from sklearn.preprocessing import LabelEncoder
        from sklearn.metrics import mean_squared_error
        from keras.models import Sequential
        from keras.layers import Dense, Input, Flatten, Dropout, Activation
        from keras.layers import LSTM
        from keras.models import Model
        import keras
        from keras.utils import plot_model
        from sklearn.model_selection import train_test_split

        np.random.seed(1337)

        %matplotlib inline
        plt.style.use('seaborn')
        from fbprophet import Prophet

        Using TensorFlow backend.
```

```
In [14]: input_layer = Input(shape=(X.shape[1], X.shape[2],))

         x = Conv1D(128, 1, padding='valid', activation='relu') (input_layer)
         x = GlobalMaxPooling1D() (x)
         x = Dropout(0.5) (x)
         x = Dense(64) (x)
         predictions = Dense(no_classes, activation='softmax')(x)

         model = Model(inputs=input_layer, outputs=predictions)

         from keras import optimizers
         adam = optimizers.Adam(lr=0.001, beta_1=0.99, beta_2=0.999, epsilon=1e-08, decay=0.01)

         model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
         model.summary()

         from keras.utils import plot_model
         plot_model(model, to_file='cc_conv1d-classification.png', show_shapes=True)

         # plot_model(model, to_file='./outputs/dense_model.png', show_shapes=True)
```

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         (None, 7, 1)              0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 7, 128)            256
_________________________________________________________________
global_max_pooling1d_1 (Glob (None, 128)               0
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 64)                8256
_________________________________________________________________
dense_2 (Dense)              (None, 5)                 325
=================================================================
Total params: 8,837
Trainable params: 8,837
Non-trainable params: 0
```

Fig. 5.3.2: Code of Convolutional Neural Networks Model

5.3.1 Loss and Accuracy for CNN



Loss **Accuracy**

Fig. 5.3.1.1: Loss and accuracy for CNN model

Convolutional Neural Network (CNN) did not generate a very good result. It has got an accuracy of 29% and loss 2.23. In the loss graph the training and the validation loss are close enough but in the accuracy graph the training accuracy got a higher number than the validation accuracy. This Convolutional Neural Network (CNN) model has an unstable training accuracy and loss found from the graphs. This graphs also shows that with epoch the accuracy got lower for train data and got higher for test data.

5.4 Shallow Dense Classification

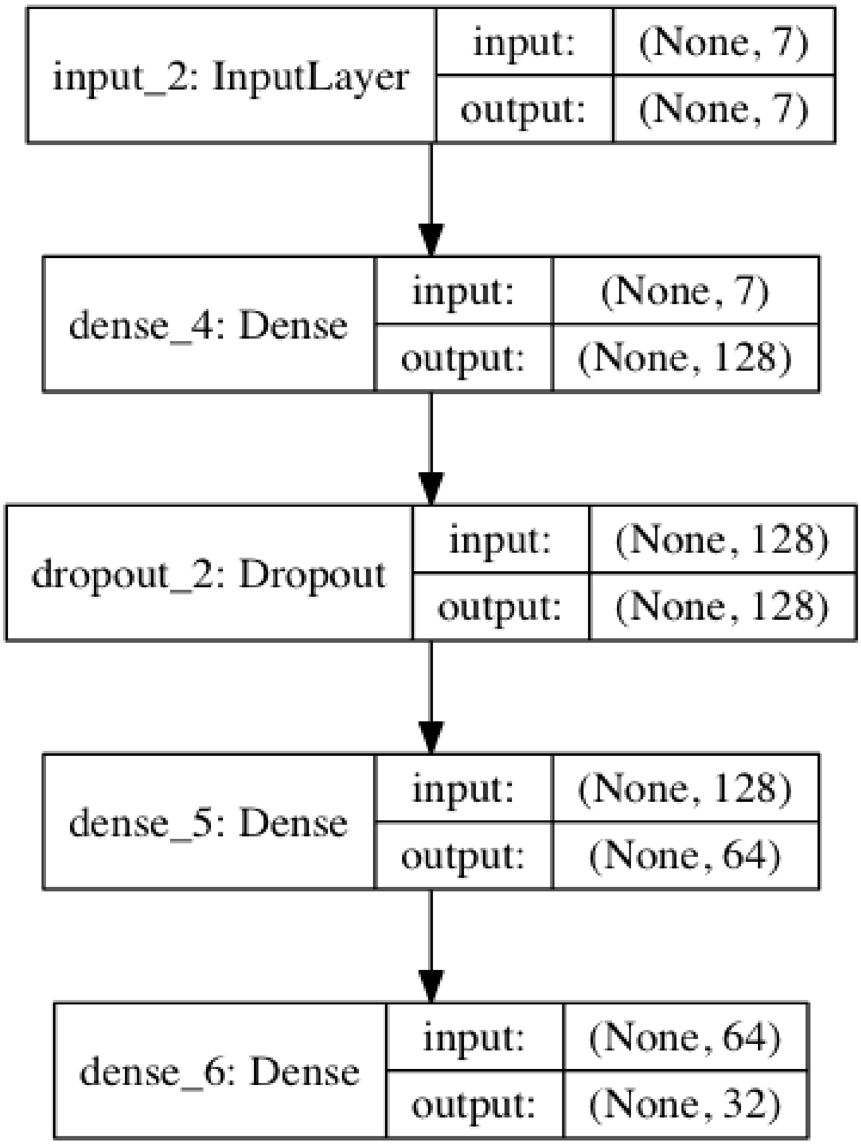
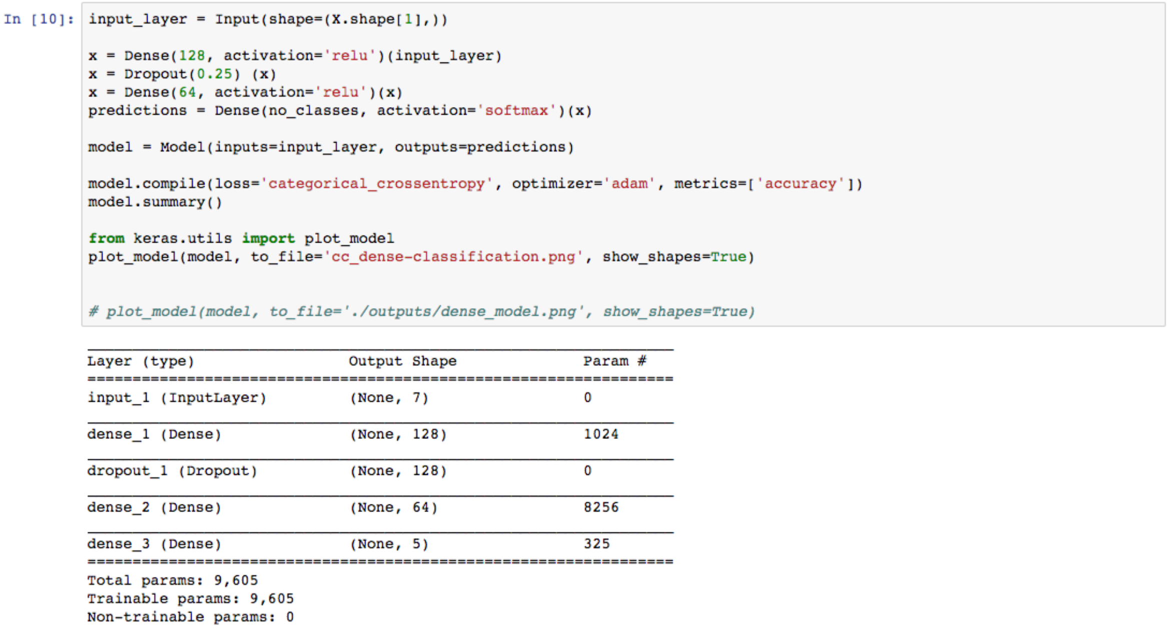


Fig. 5.4.1: Shallow Dense Classification Model

We also proposed Shallow Dense Classification model as one of our classification model. In this model it could extract up to 128 features. We have used 100 epoch for the model. Then we used dropout as our data was over fitted while running the algorithm. Then we flattened the output of last convolution layer and feed into two consecutive dense layers. In the last layer this model got 32 classes which is the total crime type.

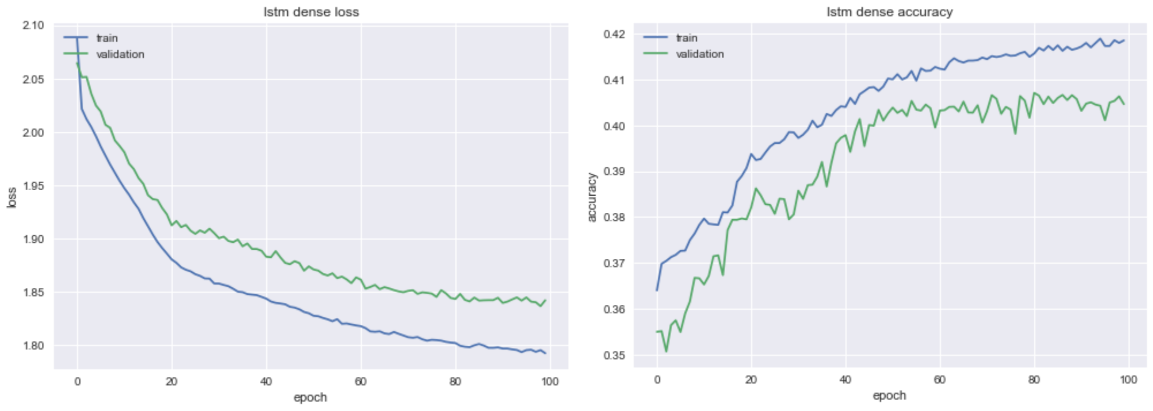


```
In [10]: input_layer = Input(shape=(X.shape[1],))

         x = Dense(128, activation='relu')(input_layer)
         x = Dropout(0.25) (x)
         x = Dense(64, activation='relu')(x)
         predictions = Dense(no_classes, activation='softmax')(x)

         model = Model(inputs=input_layer, outputs=predictions)

         model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
         model.summary()

         from keras.utils import plot_model
         plot_model(model, to_file='cc_dense-classification.png', show_shapes=True)


         # plot_model(model, to_file='./outputs/dense_model.png', show_shapes=True)
```

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_1 (InputLayer)         (None, 7)                 0         
_________________________________________________________________
dense_1 (Dense)              (None, 128)               1024      
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_2 (Dense)              (None, 64)                8256      
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 325       
=================================================================
Total params: 9,605
Trainable params: 9,605
Non-trainable params: 0
```

Fig. 5.4.2: Code of Shallow Dense Classification Model

5.4.1 Loss and Accuracy for Shallow Dense Classification Model



Loss **Accuracy**

Fig. 5.4.1.1: Loss and accuracy for Shallow Dense Classification

Shallow dense classification model did generate a good result. This model is far better than Convolutional Neural Network (CNN) model for our dataset. It has got an accuracy of 41.8% and loss 1.78. In this graphs green line indicates test result and green line indicates train result. The loss got lower and accuracy got higher with the higher number of epoch in terms of both train and test data.

5.5 LSTM (Long Short Term Memory)

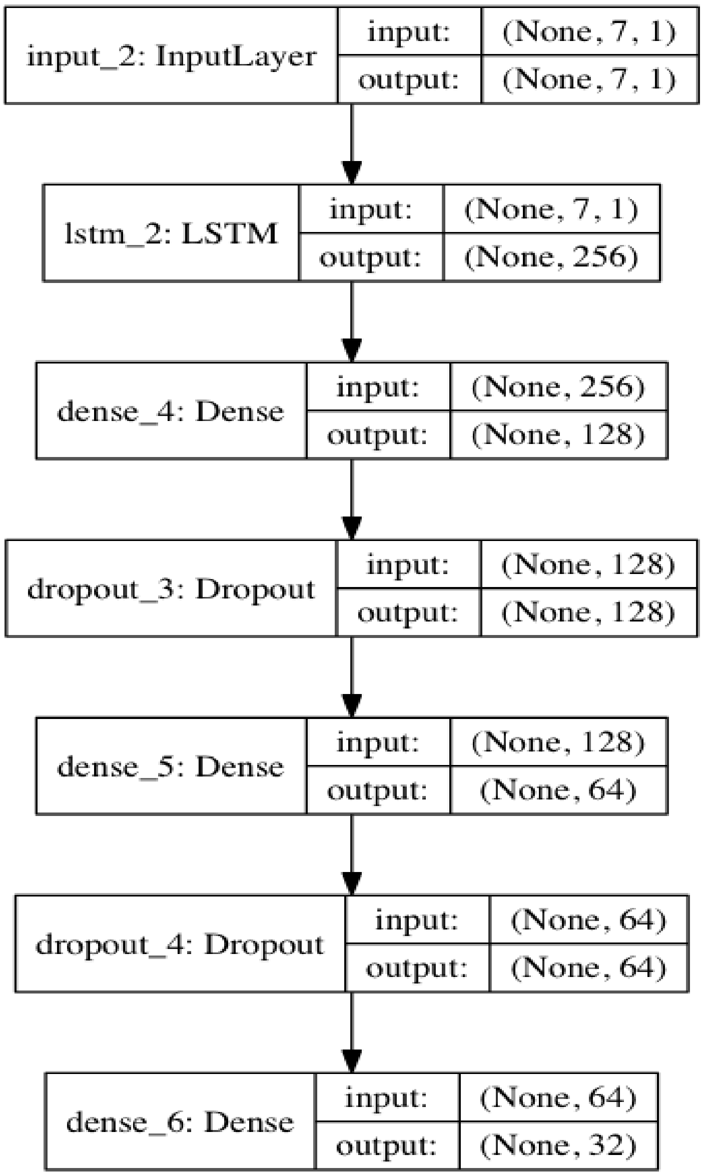
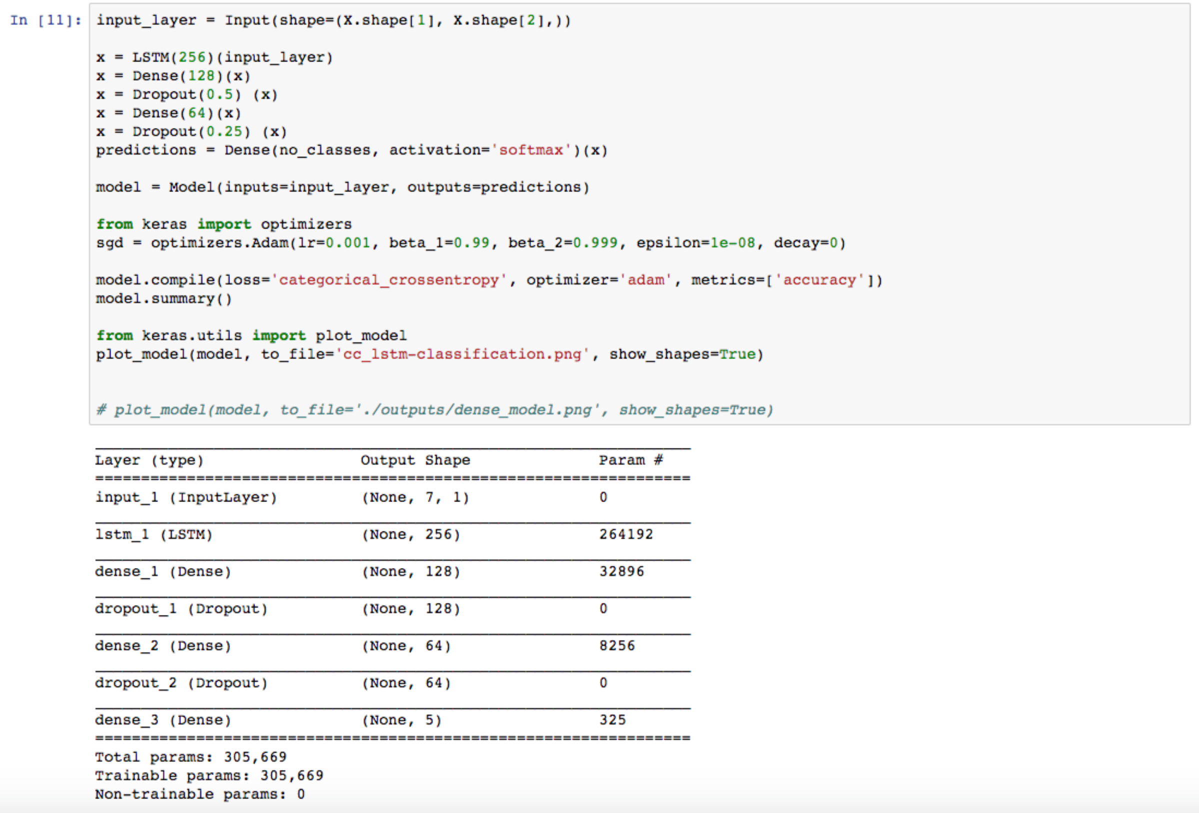


Fig. 5.5.1: Long Short Term Memory model

LSTM (Long short-term memory) is used as one of our classification model. The model takes input in the (7, 1) shape and returns the same size output. In this model it could extract up to 256 features. We have used 100 epoch for every model. Then we used dropout 2 times as our data was over fitted while running the algorithm. Then we flattened the output of last convolution layer. In the last layer this model got 32 classes which is the number of the unique crime types in the dataset.

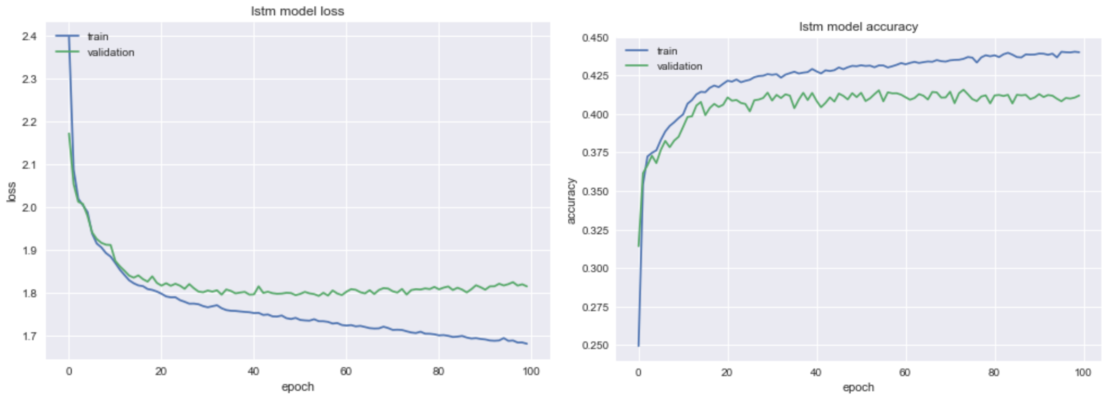


```
In [11]: input_layer = Input(shape=(X.shape[1], X.shape[2],))

         x = LSTM(256)(input_layer)
         x = Dense(128)(x)
         x = Dropout(0.5) (x)
         x = Dense(64)(x)
         x = Dropout(0.25) (x)
         predictions = Dense(no_classes, activation='softmax')(x)

         model = Model(inputs=input_layer, outputs=predictions)

         from keras import optimizers
         sgd = optimizers.Adam(lr=0.001, beta_1=0.99, beta_2=0.999, epsilon=1e-08, decay=0)

         model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
         model.summary()

         from keras.utils import plot_model
         plot_model(model, to_file='cc_lstm-classification.png', show_shapes=True)


         # plot_model(model, to_file='./outputs/dense_model.png', show_shapes=True)
```

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
input_1 (InputLayer)         (None, 7, 1)              0         
_________________________________________________________________
lstm_1 (LSTM)                (None, 256)               264192    
_________________________________________________________________
dense_1 (Dense)              (None, 128)               32896     
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_2 (Dense)              (None, 64)                8256      
_________________________________________________________________
dropout_2 (Dropout)          (None, 64)                0         
_________________________________________________________________
dense_3 (Dense)              (None, 5)                 325       
=================================================================
Total params: 305,669
Trainable params: 305,669
Non-trainable params: 0
```

Fig. 5.5.2: Code of Long Short Term Memory model

5.5.1 Loss and accuracy for LSTM model



Loss **Accuracy**

Fig. 5.5.1.1: Loss and accuracy for Long Short Term Memory Model

LSTM (Long short-term memory) model did generate a very good result. This model is far better than Convolutional Neural Network (CNN) model and Shallow Dense model for our dataset. It has got an accuracy of 44.4% and loss 1.67. In this graphs green line indicates test result and green line indicates train result. The loss got lower and accuracy got higher with the higher number of epoch in terms of both train and test data. From the graphs it can be said that LSTM is the best model in terms of unclassified dataset.

5.6 Imbalanced Dataset

A dataset set is imbalanced when the classification categories are not represented approximately equally. Two different strategies are used to deal with imbalanced dataset problem. The first one is a brute force method to reduce number of classes by only picking the classes with highest frequencies. The second strategy is to synthesize more samples or reduce samples to make classes more balanced.

5.7 Reducing Classes

Our dataset is a very complex and it has too many classes. We tried to reduce the classes to simplify the imbalanced dataset so that the classifier can compensate for minority classes. In our dataset most of the primary crime types are very few in number but some of the types are very large in number. So, we have selected the most occurred classes for our next series of prediction. The classes are given below.

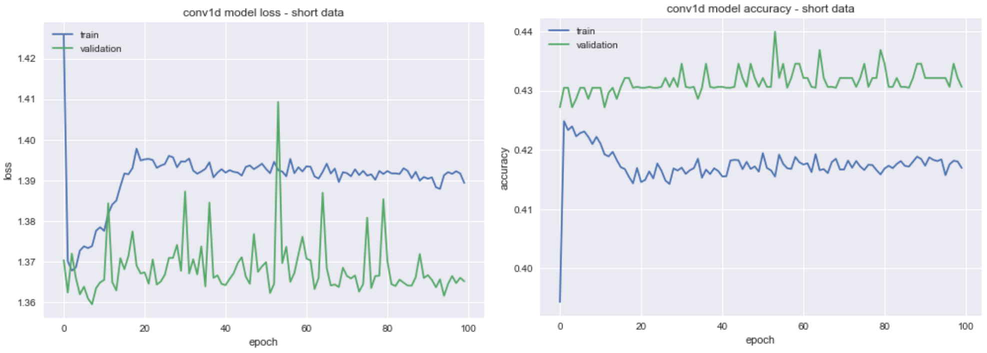
Table 5.7.1: Selected Primary Crime Types

THEFT
BATTERY
CRIMINAL_DAMAGE
NARCOTICS
ASSAULT

5.8 Results

We reduce the classes to simplify the imbalanced dataset and after that some results have changed. Results after reducing classes are given below.

5.8.1 Loss & Accuracy for CNN model

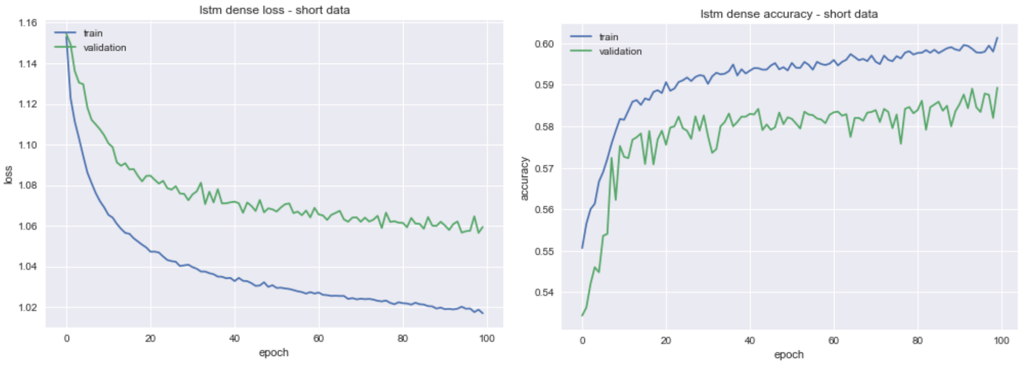


Loss **Accuracy**

Fig. 5.8.1.1: Loss & Accuracy for CNN model

Even after reducing classes Convolutional Neural Network (CNN) did not generate a very good result but we can say it is decent. It has got an accuracy of 43% and loss 1.39 which is far better than the previous result. In the loss graph the training and the validation loss are far away from each other. Even with the reduced class Convolutional Neural Network (CNN) model has an unstable training accuracy and loss found from the graphs.

5.8.2 Loss & Accuracy for Shallow Dense Classification

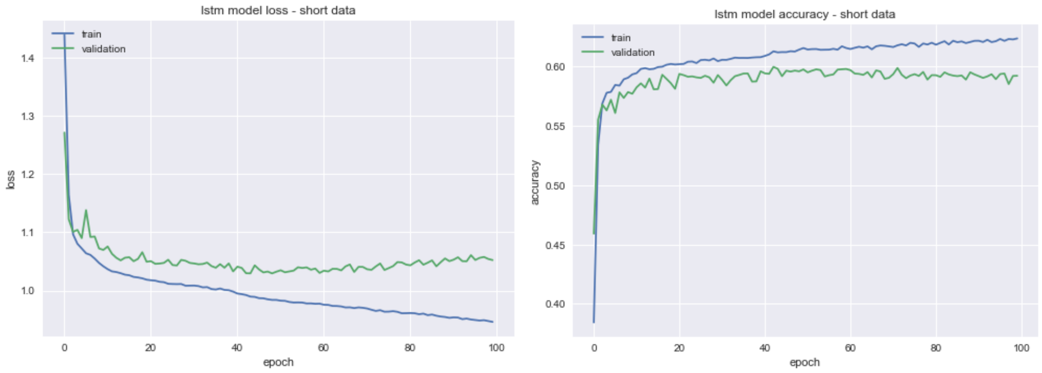


Loss **Accuracy**

Fig. 5.8.2.1: Loss & Accuracy for Shallow Dense Classification

Shallow dense classification model did better than Convolutional Neural Network (CNN) model for our dataset. It has got an accuracy of 60.5% and loss 1.06. The loss got lower and accuracy got higher with the higher number of epoch in terms of both train and test data. So it got a huge improvement from the previous time which was 41.8% accuracy rate. As it got a high accuracy rate with only 100 epoch and a limited number of dataset, it can get higher accuracy rate if we train it with more dataset.

5.8.3 Loss & Accuracy for LSTM



Loss **Accuracy**

Fig.5.8.3.1: Loss & Accuracy for LSTM

Again, LSTM (Long Short Term Memory) model did generate a very good result for our dataset. This model is better than Convolutional Neural Network (CNN) model and Shallow Dense model as before for our dataset. It has got an accuracy of 64.2% and a very little number of loss which is 0.03. From the graphs it can be said that LSTM is the best model for our crime type prediction and classification.

5.9 Comparative analysis of different strategies

Three deep learning models are taken to predict crime category from Chicago Crime dataset. The original dataset with 32 classes gives both low accuracy and loss. Loss and accuracy improves if only 5 classes are considered.

Table 5.9.1: Classification on Original Classes

Algorithm	**Accuracy**	**Loss**
LSTM	44.4%	1.67
Shallow dense	41.8%	1.78
CNN	29%	2.23

The highest accuracy achieved by LSTM Model. However, CNN gives better loss value than any other model.

Table 5.9.2: Classification on Balanced Classes

Algorithm	**Accuracy**	**Loss**
LSTM	64.2%	0.03
Shallow dense	60.5%	1.06
CNN	43%	1.39

By comparing two results we can see that, the LSTM is the best model for our data set. It gave the best accuracy and the best loss value among them.

CHAPTER 6

Conclusion and Future Work

6.1 Conclusion

Throughout the research it has been evident that basic details of a criminal activities in an area contains indicators that can be used by Deep learning agents to classify a criminal activity given a location. Even though the learning agent suffers from imbalanced categories of the dataset, it was able to overcome the difficulty. LSTM model successfully classified criminal activities based on the location. With an accuracy of 64.2%, it was able to outpace other deep learning models. Imbalanced classes are one of the main hurdles to achieve a better result. Though the deep learning agent was able to create a predictive model out of a big crime data, a demographic dataset would probably help to further improve the result and solidify it.

6.2 Future Work

For this research, only crime data has been used, but as many researched have showed that a particular area's socio-economic standard is also a key indicator of possible criminal activity. This Deep learning agent could incorporate those data and might perform better.

This model can be also used for other geographic locations. This would also help to analyze crimes occurring in different locations and build a better understanding of different crimes and its relation with particular demography.

This model can be used for time series prediction and for Geo-spatial prediction.

Also, there are many advanced Deep learning approaches that can be explored. Deep Learning & Neural Networks can provide a more balanced understanding of criminal activities. As it has been seen on this research, imbalanced classes has been a major issue in dealing with the particular database. Advanced techniques to deal with imbalanced classes are also something that remains to be explored.

REFERENCES

[1] Kaggle.com. (2017). Crimes in Chicago | Kaggle. [online] Available at: https://www.kaggle.com/currie32/crimes-in-chicago [Accessed 13 Dec. 2017].

[2] W. David. “Place-based policing”, in Ideas in American Policing, 2008, pp. 1–16.

[3] Freeman R. B. The economics of crime. Handbook of labor economics, 3:3529–3571, 1999.

[4] Patterson E. B. Poverty, income inequality, and community crime rates. Criminology, 29(4):755–776, 1991.

[5] Braithwaite J. Crime, Shame and Reintegration. Ambridge: Cambridge University Press, 1989.

[6] Ehrlich I. On the relation between education and crime. 1975.

[7] Hyeon-Woo Kang and Hang-Bong Kang. Multi-modal data using deep learning to predict the crime occurrence.

[8] Wang X., Gerber M.S and BrownD. E. Auto-matic crime prediction using events extracted from twitter posts. In Social Computing, Behavioral-Cultural Modeling and Prediction, pages 231–238. Springer, 2012.

[9] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014, November). Once upon a crime: towards crime prediction from demographics and mobile data. In Proceedings of the 16th international conference on multimodal interaction(pp. 427-434). ACM.

[10] Iqbal, R., Murad, M. A. A., Mustapha, A., Panahy, P. H. S., & Khanahmadliravi, N. (2013). An experimental study of classification algorithms for crime prediction. Indian Journal of Science and Technology,6(3), 4219-4225.

[11] Shojaee, S., Mustapha, A., Sidi, F., & Jabar, M. A. (2013). A study on classification learning algorithms to predict crime status. International Journal of Digital Content Technology and its Applications,7(9), 361.

[12] Redmond M, Baveja A., “A Data-driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments”, European Journal of Operational Research, Science Direct, vol. 141, no. 3, pp. 660–678, 2002.

[13] Sadhana, C. S. (2015). Survey on Predicting Crime Using Twitter Sentiment and Weather Data israce .2015

[14] Beckmann, M., Ebecken, N. F., & de Lima, B. S. P. (2015). A KNN under sampling approach for data balancing. Journal of Intelligent Learning Systems and Applications,7(4), 104.

[15] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research,16, 321-357.

[16] Maloof, M. A. (2003, August). Learning when data sets are imbalanced and when costs are unequal and unknown. In ICML-2003 workshop on learning from imbalanced data sets II(Vol. 2, pp. 2-1).

[17] Sas.com. (2017). Machine Learning: What it is and why it matters. [online]Available:https://www.sas.com/en_us/insights/analytics/machine-learning.html [Accessed 13 Dec. 2017].

[18] WhatIs.com. (2017). What is machine learning? – Definition from WhatIs.com.[online]Available:http://whatis.techtarget.com/definition/machine-learning [Accessed 14 Nov. 2017].

[19] Digitalocean.com. (2017). An Introduction to Machine Learning, DigitalOcean.[online]Available:https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning [Accessed 14 Dec. 2017].

[20] Brownlee, J. (2017). Supervised and Unsupervised Machine Learning Algorithms - Machine Learning Mastery. [online] Machine Learning Mastery Available at:https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/ [Accessed 14 Dec. 2017].

[21] Mathworks.com. (2017). What Is Deep Learning? | How It Works, Techniques&Applications.[online]Availableat:https://www.mathworks.com/discovery/deep-learning.html [Accessed 14 Dec. 2017].

[22] Baldi, P. and Sadowski, P. (2017). The dropout learning algorithm.

[23] Analytics Vidhya. (2017). Fundamentals of Deep Learning - Activation Functions and their use. [online] Available at: https://www.analyticsvidhya.com/blog/2017/10/fundamentals-deep-learning-activation-functions-when-to-use-them/ [Accessed 17 Jun. 2017].

[24] Brownlee, J. (2017). Gentle Introduction to the Adam Optimization Algorithm for Deep Learning - Machine Learning Mastery. [online] MachineLearningMastery.Availableat:https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/ [Accessed 9 Dec. 2017].

[25] Brownlee, J. (2017). A Gentle Introduction to Long Short-Term Memory Networks by the Experts - Machine Learning Mastery. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/ [Accessed 14 Dec. 2017].

[26] Nielsen, M. (2017). Neural Networks and Deep Learning. [Online] Neuralnetworksanddeeplearning.com.Available:http://neuralnetworksanddeeplearning.com/chap1.html [Accessed 14 Dec. 2017].

[27] Medium. (2017). Machine Learning for Humans, Part 2.1: Supervised Learning. [online] Available at: https://medium.com/machine-learning-for-humans/supervised-learning-740383a2feab [Accessed 14 Dec. 2017].

[28] Exsilio Blog. (2017). Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures - Exsilio Blog. [online] Available at: http://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/ [Accessed 1 Dec. 2017].

[29] Lts2.epfl.ch. (2017). Categorical Cross Entropy – Semester project progresslog.[online]Available:https://lts2.epfl.ch/blog/dennis/2015/12/08/categorical-cross-entropy/ [Accessed 14 Dec. 2017].

[30] Medium. (2017). Machine Learning for Humans, Part 2.2: Supervised Learning. [online] Available at: https://medium.com/machine-learning-for-humans/supervised-learning-740383a2feab [Accessed 14 Dec. 2017]