Analysing Crime Patterns using Machine Learning: A case study in Chicago

MSc Research Project
Data Analytics

Himanshi Himanshi
Student ID: x20218001

School of Computing
National College of Ireland

Supervisor: Dr.Paul Stynes, Musfira Jilani, Dr.Pramod Pathak

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Himanshi Himanshi
Student ID:	x20218001
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Dr.Paul Stynes, Musfira Jilani, Dr.Pramod Pathak
Submission Due Date:	15/08/2022
Project Title:	Analysing Crime Patterns using Machine Learning: A case study in Chicago
Word Count:	4684
Page Count:	14

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	19th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Analysing Crime Patterns using Machine Learning: A case study in Chicago

Himanshi Himanshi
x20218001

Abstract

Crime patterns analysis is systematic examination of high-volume crimes. Analysis of crime patterns is useful for predicting a crime before it occurs. Crime rates,hot spots, and types of crimes can all be reliably estimated from past patterns.Developing a machine learning model for predicting crime pattern with great accuracy is a challenge. By being able to forecast crimes based on time, location and so forth, law enforcement can provide useful information from a strategic standpoint. This research proposes the framework of Crime pattern analysis to predict the crime rate with an acceptable amount of accuracy using deep learning and supervised machine learning methods. The proposed framework uses deep learning methodology with traditional supervised machine learning approach. Chicago crime dataset for crime prediction starting from 2001 to December 2021 consisting of 62,59,111 crime records is used to train the deep learning for time series forecasting of crime rate using 24 classes of crime types. Feature engineering and data modelling are done to train six different time series model namely Weighted Moving Average, Exponential Moving Average, Simple Moving Average, Bidirectional LSTM, CNN-LSTM and Random Forest Regression. Result of six models is presented in this paper based on Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Bidirectional LSTM model outperforms all other models in terms of RMSE value. This paper forms a method to analyse crime patterns by doing the case study on Chicago Crime data and consequently creating deductive deductions with the aid of solid and trustworthy Machine Learning algorithms.
Keywords – Crimes, Crime Pattern Analysis and Forecast, Deep Learning

1 Introduction

Crime is the biggest issue every nation is facing Elyta et al. (2022), and huge amount of money has spent over the years to handle this issue Laycock (2013). There are numerous factors like poverty, large population, lack of resources etc. that drives criminal activities and analysing crime based on all these factors is a challenge. Deep learning models for time series forecasting can improve the prediction of future crime rate. Accuracy can be improved Devi and Kavitha (2021) by 20% using RNNLSTM as compared to traditional regression model. Researchers Safat et al. (2021) aimed to test machine learning techniques to predict crime but got the accuracy of 42% which is quite less, and this accuracy can be improved further if data tuning can be applied.

The aim of this research is to predict the crime pattern with great accuracy in different districts of Chicago city to reduce future criminal activities. The major contribution of this research is a novel hybrid deep learning LSTM based approach for predicting crime rate. A minor contribution of this research is to analyse the crime pattern in Covid-19 pandemic period. In order to find the optimal deep learning model this research compares Weighted Moving Average, Exponential Moving Average, Simple Moving Average, Bidirectional LSTM, CNN-LSTM and Random Forest Regression. A model was used for the crime analysis framework based on Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

Research Question: How deep learning methods are better than machine learning algorithms to predict Chicago city crime pattern?

The use of deep learning methods Huang et al. (2015) along with traditional Machine Learning methods for time series forecasting offers several promising features, such as automatic detection of temporal relationships and the automatic handling of time series structures like trends and seasonality. Through this an additional focus can be provided on a particular component with deep learning techniques. As far as we are aware, this hybrid model has not been used to crime prediction.

This paper discusses deep learning methods used for crime rate analysis and forecasting in section 2 related work. The research methodology is discussed in section 3. Section 4 discusses the design components for the LSTM deep learning framework. The implementation of this research is discussed in section 5. Section 6 presents and discusses the evaluation results. Section 7 concludes the research and discusses future work.

2 Related Work

A wider analytical approach to crime has been formed as a result of technological advancements in every aspect of human life. Borowik et al. (2018) developed a forecasting model to analyse time series of criminal events using prophet forecasting model. Researchers aimed to perform case study on Polland crime events, and their main objective was to create models that contained information about criminal behaviour and its propagation. Results show that Mean Absolute Percentage Error (MAPE) varies between 10% to 40% depending upon the type of crime. According to Borowik et al. (2018), adding additional features to time series history should be part of future research to improve the accuracy of overall predictive model.

The threat of crime and violation is intended to be controlled in order to protect justice. Safat et al. (2021) proposed crime prediction and crime forecasting model and applied eight different models like Logistic Regression, SVM, Naïve Bayes, k-nearest neighbors (KNN), decision tree, multilayer perceptron (MLP), random forest, and eXtreme Gradient Boosting (XGBoost), LSTM and ARIMA model of machine learning to get accurate prediction result. According to the prediction results, crimes of all types and crime rates of the past few years are identified along with areas with high crime density. Foresight of future crime trends and analytics were also undertaken using the ARIMA model for time series analyses. Safat et al. (2021) checked the stationarity of data before applying LSTM model to get more accurate result and achieved RMSE and MAE value of 12.66 and 11.70 respectively. Research can be better if data visualization can be added. Kane et al. (2014) compared the performance of LSTM and Random Forest model for time series forecasting and found out that random forest outperforms in terms of predictive ability. The limitation of their research is that they did not mention about layers used in LSTM model.

Han et al. (2020) combined LSTM and ST-GCN methods to present a crime prediction model on daily basis. Doing this, they can identify the factors that affects theft crime and areas of Chicago city where theft crime is maximum. Their idea of adding spatial- temporal features in between different communities present in Chicago City help to get the exact occurrence of crime incidents and predict the crime pattern.Han et al. (2020) mentioned that their future work should include the effect of external social factors on crime prediction.Safat et al. (2021) andHan et al. (2020) applied LSTM model but consider different external factors on crime prediction model. Stec and Klabjan (2018) merge crime dataset with weather dataset, public transportation and census data and achieve the accuracy of 75.6%.

As far as types of crimes are concerned, the COVID-19 pandemic Tasnim et al. (2022) has a varied relationship with crimes. Yang et al. (2021) analysed the impact of Covid-19 on criminal activities in Chicago. They have done the case study on Chicago city taking different types of crime and found out that crime rate varies drastically in pandemic period like crimes like robbery, battery, theft, and burglary have reduced but cybercrime, domestic violence and homicide have not been affected much. They have used spatial point pattern test to check the difference between pandemic i.e., 2020 dataset and other years datasets. It was noticed that crime distribution changed significantly in 2020 as compared to others.Tasnim et al. (2022) identified that fusion model of one layer LSTM based on Spatial factor and bidirectional LSTM based on Spatial and temporal factor performed better in predicting crime. Their study done the experimental analysis of crime prediction on San Francisco and Chicago cities. This study proposed a novel method for integrating two separate models to improve the performance of the model and reap the benefits of both. However, with fusion, the model yielded a MAE value 0.008. Their model has one weakness which is it cannot perform well for data more than three years. Instead, on considering spatial and temporal features separately, Yi et al. (2018) proposed an integrated model which is clustered continuous conditional random field that considers spatial and temporal factor in an integrated way to predict crime.

Dash et al. (2018) suggested that social parameters affected crime so they instead of predicting crime of whole Chicago city, they predicted crime based on communities and identified that the communities which have similar social factors like police stations, libraries etc. have same rate of crime. Researchers built a network of 77 communities of Chicago city and evaluated the similarity matrix within communities. They carried out their research in step wise manner like first, network of different communities was built then found out the features between that network and after all this they applied polynomial regression, Support vector regression and auto regressive model to predict crime rate in different Chicago communities. RMSE Mansour and Lundy (2019) was used to evaluate the model and found out that SVR outperforms all the models. This research used district to analyse crime in Chicago city. Their research will not give accurate result on large dataset that can cover more than 10 years of data.

According to Vijayarani et al. (2021), analysis of longitudinal data, including multivariate time series, can be done using the ARIMA method. Their study predicted only five types of crime which includes assault, battery, narcotics, battery criminal damage and theft. Researcher used ARIMA model to predict crime rate and got the accuracy of 91.4%. They have analysed the data for three years and seasonal patterns were identified for each month, week, and day. Their model is very simple but effective in terms of prediction. The use of ARIMA Butt et al. (2020) shows promise in predicting crime. Research can be better if some modifications can be made in model.

In conclusion, the state of art indicates that several methods such as logistic regression, support vector machine (SVM), k-nearest neighbours (KNN), decision tree, XGBoost, time series analysis through LSTM and ARIMA were used to analyse and predict crime data of Chicago and Los angles cities. This research analysed and forecast crime rates in 23 districts of Chicago city using ARIMA, one layer LSTM, bidirectional LSTM and moving average models. State of art used Chicago and Los angles crime data from 2001 to November 2019. Current research performed time series analysis on Chicago crime dataset from 2001 to December 2021. This research proposes the framework of Crime pattern analysis to predict the crime rate with an acceptable amount of accuracy using deep learning methods. Current research focused on data pre-processing to handle outliers in order to get more accuracy in crime rate prediction.

3 Methodology

The research methodology consists of five steps namely data gathering, data preprocessing, data transformation, data modelling and conversion, evaluation and results as shown in Figure 1. This research is influenced by CRISPDM methodology.

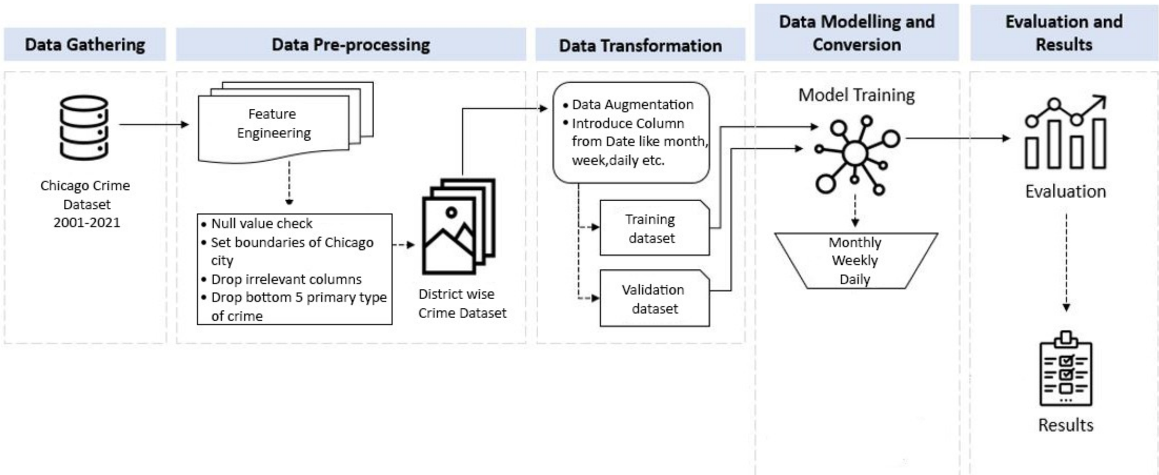

Figure 1: Research Methodology

The first Step, *Data Gathering* The first Step, Data Gathering involves downloading the Chicago crime incidents dataset from 2001 to December 2021 [1] from Chicago government data portal.

The second Step, *Data Pre-processing* involves removal of irrelevant attributes and crime instances. In the Chicago dataset, 91,59,279 crimes were initially detected, and 3033173 have been removed for incorrect formatting like missing values, data, fates etc. Dataset contains 22 attributes like id, case number, primary type etc. This study considers only relevant attributes and removes unnecessary columns like ID, updated on, location etc. drop_duplicates method is used to check and remove the duplicate rows and 1.6 million rows were removed after this step [2]. After data pre-processing, this study has 7360231 incidents, and 16 attributes to be studied.

[1] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2

[2] https://www.kaggle.com/code/vivekhn/fb-prophet-auto-arima-time-series-on-crime-data/notebook

The third step, *Data Transformation* involves creation of new columns like month, dayOfWeek, dayOfMonth and weekOfMonth from Date of crime incident column. This helps in analysing the past crime patterns. Initially Latitude column in dataset present as object but to check the hotspots area, this needs to be changed into float. Using value counts method, this study checked the top categories of crime type and there are 33 different types of primary types of crime category and from the count we can see that few categories like OBSCENITY and below have very less crime numbers and can be removed from the data (treated as outliers). Finally, this study analysed 24 primary types of crime. Only top 50 locations where crime count is maximum considers in this study. Similarly, few districts have almost 0 crimes, so this study considers 23 districts out of 31 districts. Finally, cleaned and transformed dataset contains 62,59,111 rows and 24 columns is ready for time series forecasting model. Dataset was then split into training and validation datasets.

The fourth step, *Data Modelling and Conversion* involves training of model, conversion of model, and model deployment. The models were trained with training dataset that contains crime incidents from 2001 to December 2020 and validation dataset contains crime incidents from January 2021 to December 2021. For district wise prediction for a month, week and day, list of unique districts is needed.

The fifth step, *Evaluation and Results* involves evaluating the performance of each of deep learning forecasting models using RMSE and MAE value. The six models were compared and evaluated. The optimal deep learning time series forecasting model was selected from the experimentation. The trend that the anticipated value is following in comparison to the actual values can be seen by plotting the actual values and the predicted value on a graph.

4 Design Specification

The hybrid deep learning framework architecture combines a deep learning time series forecasting model and traditional machine learning methods i.e., random forest regression based on district wise prediction for a month, week, and day as shown in Figure 2. The components of deep learning forecasting model includes Hybrid Prediction model inference as mentioned in 4.1 and Further analysis of crime pattern is mentioned in section 4.2.

4.1 Hybrid Crime Prediction Model

Hybrid model of predicting crime rate in Chicago city includes deep learning models and traditional machine learning models. Deep learning models includes Simple/Weighted/ Exponential Moving average, Single layer LSTM/bi-directional LSTM and CNN-LSTM models. In Simple moving average calculation, average data are calculated over a given period, while in weighted moving average, current data are given higher weightage. However, an exponential moving average also emphasizes recent crime incident, but it decreases at an exponential rate between one incident and its preceding incident. The main reason of using LSTM model in crime prediction is that a sequence of inputs is made up of components with both present and past information. This is beneficial to predict future crime pattern based on crime happened in past. On the other hand, traditional machine learning model includes random forest regression and it felt easier and more robust in comparison of other models. An inference is generated by this hybrid framework

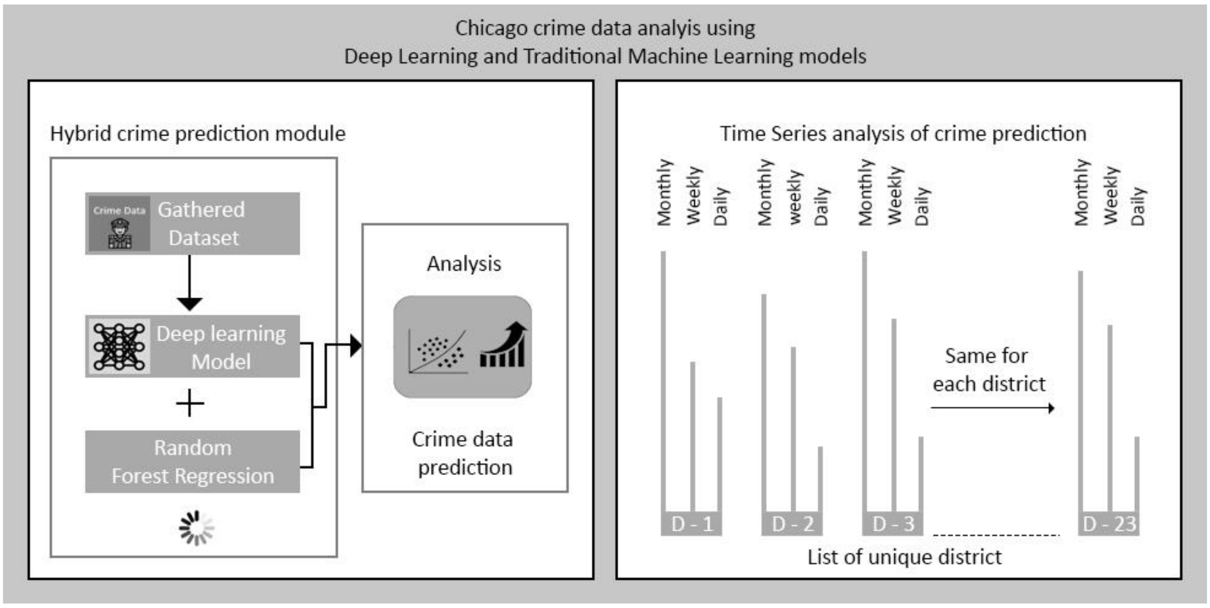

Figure 2: Hybrid Framework for Crime prediction Architecture

for predicting crime by providing a corresponding RMSE and MAE values. To ensure the future crime prediction is accurate, 75% of confidence interval is considered in this study.

4.2 Time Series analysis of crime prediction

Time series analysis of all the models applied on crime datasets was done based on districts present in Chicago city. Instead of doing time series analysis on Chicago, this study presents a novel approach of district wise prediction. In this research, district wise analysis was done on the basis of month, week and day because it is necessary to see the trend followed by past crime so that precaution can be taken for future. Entire world faced a pandemic period i.e., Covid-19 starting from 2020 January till 2022 May, most of the things has changed in that period like it is mandatory to stay at home, people are doing work from home etc. It is interesting to check how criminal activities were affected in that pandemic period. Time series analysis will help us to analyse future crime data based on previous crime pattern.

5 Implementation

The hybrid deep learning framework was implemented as time series forecasting model for Chicago City. Throughout this section, we will explain the multiple steps we took in order to obtain the research objectives by forecasting crime rate in 23 districts of Chicago City. Python was used to code all the models in a Jupyter notebook since it is flexible and capable of handling large datasets. This study made use of the Anaconda Python distribution that is open source. This study used Chicago crime dataset from 2001 to December 2021, downloaded from Chicago government data portal in csv format with a size of 6 GB. 9159279 rows and 22 columns of data were imported into Jupyter using the "Panda" Python library. Missing values were detected by using the isnull()

function. After handling the missing values, 2033173 rows were removed from 9159279 rows originally loaded into Jupyter. Current research considers all the location of Chicago City that is bounded[3] by latitude and longitude as (41.6439, -87.9401; 41.9437, -87.5878) and removed all the rows of location outside bounded box. After analysing crime type using value counts function, it has noticed that there are 33 types of crime but few crimes like Obscenity, Human Trafficking, public indecency etc. are very less in number this study focussed on 24 types of crime and dropped all the rows that are related to minor categories. Similarly, this study considers only top 50 locations and 23 districts of Chicago City. After all this analysis, 6259111 rows are left which are enough to perform experiments. To improve performance of model, the Pandas.to datetime() function was used since the date in the crime data frame was originally in string form and therefore, some more columns like month, day of week, day of month, week of month added in dataset and used in exploratory data analysis. Irrelevant columns like CaseNumber, ID, updatedOn, Location were dropped from dataset. Finally, this study performed experiments on 24 columns.

Time series forecasting model was then performed using LSTM model present in keras package in python because it performs better if longer-term trend is present in dataset. It is important to normalize or scale data when we are using neural network [13], so this study used MinMaxScaler class present in sklearn.preprocessing library. The LSTM models applied in this study has multiple layers. As a first parameter, this study needs to determine how many neurons or nodes are required in LSTM layer. Current research has different models with different node which is covered under section 5. Since the models applied has more layers, so second parameter i.e. return sequence always sets to true. Then Training of models was done on training dataset and then models were evaluated on validation dataset. This study evaluates the models on the basis of error measures which are scale-dependent in nature namely, MAE and RMSE values.

6 Results and Discussion

This study has performed multiple experiments to get the accurate model for time series forecasting of Chicago crime rate. Explanation of experiments that involves deep learning and Supervised Machine learning models performed are mentioned below.

6.1 Experiment 1: Exploratory Data Analysis

The aim of this experiment is to analyse the pattern followed by crime data like which year, month, week, day has maximum or minimum crime. Which area of Chicago has highest crime rate. This experiment is needed to know the impact of pandemic period on crime rate. After analysing the crime dataset from 2001 to December 2021, this study made few decisions like Friday Pradhan (2018) is the most common day for crimes, but no day is crime-prone than another as shown in Figure 3.

Similarly, from Figure 4 it is apparent that there is no significant difference between crimes occurring in specific months, but February has the lowest number of crimes occurring. One reason for this may be the number of days in that month.There is pattern in overall crime rate shown in Figure 6 that there is always an increase in crime rates

[3]https://boundingbox.klokantech.com/

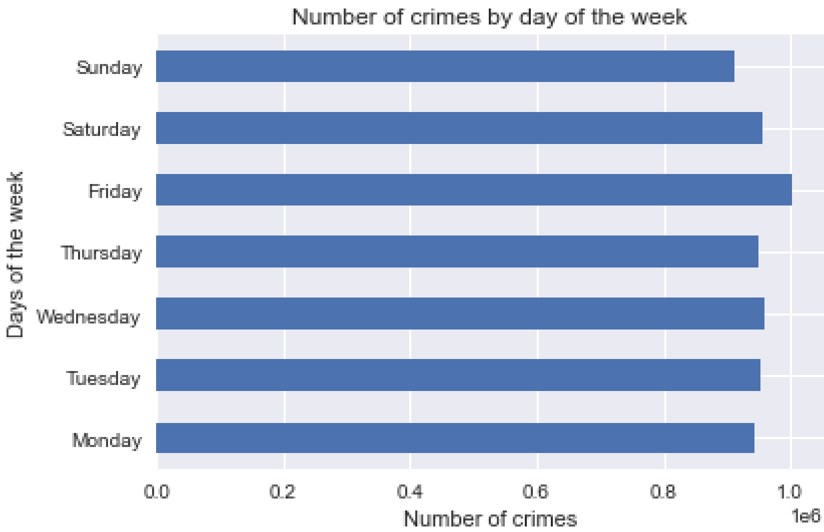

Figure 3: Number of crimes by day of week

month by month, followed by a decline later in the year but overall crime rate is in downward direction. It is evident that January 2020 to 2021 December has lowest number of crimes, and this is because of Covid-19 pandemic period. From Exploratory data analysis, this study noticed that THEFT, BATTERY, CRIMINAL DAMAGE, NARCOTICS, ASSAULT are top five crime category. Out of all districts, district 8, 11, 6, 7 has maximum crime rate but according to ARREST attribute, people don't get arrested in maximum crime.Figure 5 indicates that Year 2020-2021 experienced lesser number of crimes and this is the same period when pandemic was on its peak.

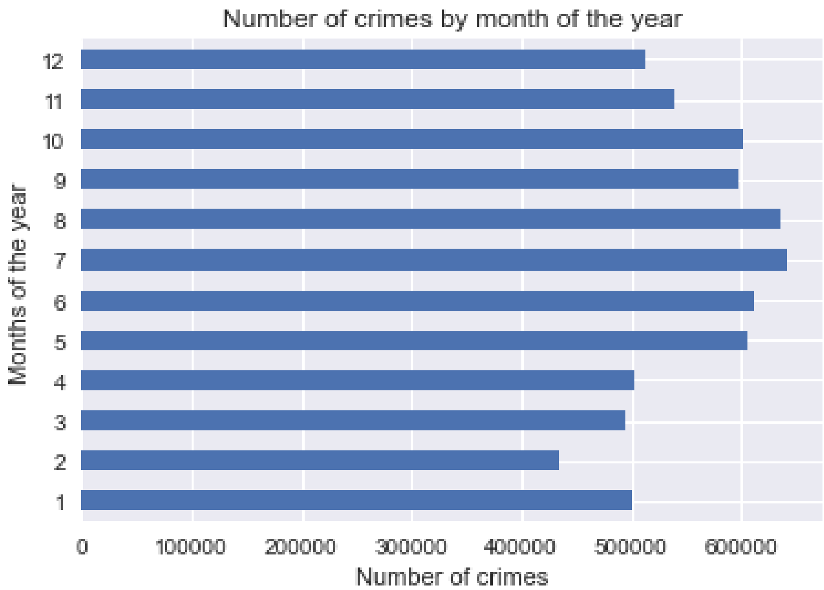

Figure 4: Number of crimes by month of year

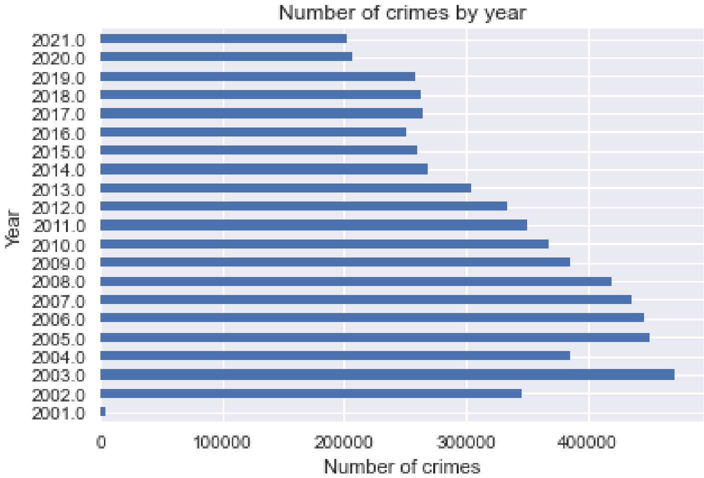

Figure 5: Number of crimes by Year

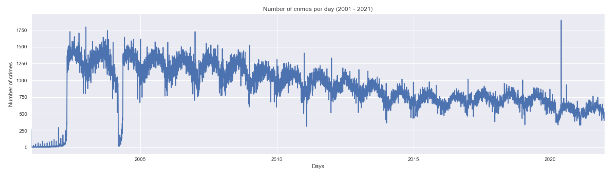

Figure 6: Overall Crime Pattern

6.2 Experiment 2: Replication of State of Art

The aim of this experiment is to replicate the state of art Safat et al. (2021) that includes analysis of time series by long-short term memory (LSTM) and autoregressive integrated moving average (ARIMA) model on Chicago. First, stationarity of data has checked using Dickey-Fuller test[4] and it has observed that ADF statistic exceeds any critical value and p-value is greater than the significance level of 0.05 therefore, data is stationary in nature. After that, LSTM model has applied with learning rate as 0.01, number of layers as 1, hidden size as 4 and input size as 1. All these parameters are not mentioned in state of art, so this study made assumptions regarding these parameters. Also, the number of epochs is mentioned as 40 in state of art, but LSTM model is acting abruptly by keeping this as 40 so, for this experiment 2500 is used as number of epochs. For ARIMA model implementation, p (Auto Regressive), d (Integrated) and q (Moving Average) are necessary to provide, and this study kept these as 0,1 and 3 respectively. A scale-dependent error and a percentage error are usually measured for time-series data. LSTM and ARIMA performance metrics which are RMSE, MAE and R2 values are listed in Table 1, which proposes the model performance in validation dataset. As a result of the 13th iteration, the outcomes of the epochs were the same in case of LSTM.

This result indicates that LSTM provides adequate performance for time series analysis, especially for R2 value, where it can categorize the data based on their variation. ARIMA model is not following the trend of the test data that's why R2 value is negative. The result of LSTM approach to predict crime patterns shows less difference between

[4]https://machinelearningmastery.com/time-series-data-stationary-python/

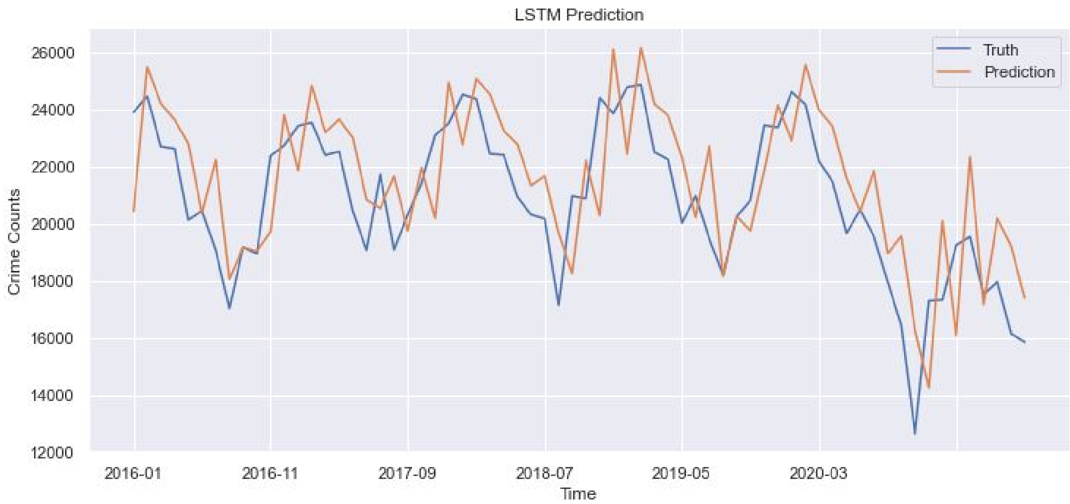

Figure 7: Crime Prediction through LSTM

Table 1: Model Comparison by evaluation metrics

Models	**RMSE Value**	**MAE Value**	**R2 Score**
LSTM	3553	2278	0.962
ARIMA	3309	2867	-0.287

predicted and actual chart as shown in Figure 7 The actual and prediction pattern is showing more or less same trend.

6.3 Experiment 3: Application of Deep learning models to forecast Chicago crime rate based on districts

The aim of this experiment is to forecast the crime rate based on different districts of Chicago. This research considered one district at one time while doing this experiment and performed modelling on three time periods which are day, week and month using Exponential Moving Average, Simple Moving Average, Weighted Moving Average, 1-layer LSTM, bidirectional LSTM and CNN-LSTM models. Compared to previous experiment, this experiment made some changes on training and validation datasets, splitting of datasets is done like data from year 2001 to 2020 December is training data and from January 2021 to 2021 December is considered as validation dataset. To start modelling on district wise prediction, list of unique districts found out using unique function and there are 23 districts total. After that training and testing dataset having district wise count of crimes grouping by year and month was calculated. Simple Moving Average with window size five is the first performed model in which actual value and predicting value of each district along with graph representation was made. Then weighted moving average with weights (1, 2, 3, 4, 5) and window as five was performed. After that Exponential moving average with alpha value as 0.6 was performed. Now, single layer LSTM model was performed with window value as 3, epochs as 200 and reshaping of data has done using reshape function with feature value as 1. This study used adam optimizer and mse loss function.

After performing single model training in which model learns from the sequence of input provided to the model, this experiment performed bidirectional LSTM model in which first model trained using provided sequential input and second model get learnt by reverse of first model sequence. Keeping all the parameters like previous model, just added one more additional layer and performed the modelling. Another LSTM model performed well for time series analysis Huang et al. (2015) for Chicago crime prediction and keeping window as 4, number of steps as 2 and epochs as 500, CNN LSTM has performed over district wise crime dataset. All the performed models were evaluated on three metrics which are RMSE, MAE and R2 values and performed over year, month, and week. Comparison of models are presented in Figure 8. This result indicates that in view of week-wise predictions, RMSE scores are good, but this study cannot model the year-end dip in crime numbers, but Day-by-day predictions smooth out this problem and RMSEs have improved by a fantastic amount from week to day-by-day LSTM approaches have shown improvements over simple naive approaches. The result show promise for CNN LSTM with RMSE scores less for large dataset and it is necessary to predict on recent crime dataset as crime pattern varies over time.

Models	**RMSE**	**MAE**
Monthly		
Simple Moving Average	96	74
Weighted Moving Average	94	71
Exponential Moving Average	101	70
1 Layer LSTM	104	58
Bi-directional LSTM	72	54
CNN LSTM	82	61
Weekly		
Simple Moving Average	26.4	17.7
Weighted Moving Average	26	17.7
Exponential Moving Average	30	19
1 Layer LSTM	22.5	15
Bi-directional LSTM	21	14
CNN LSTM		
Day		
Simple Moving Average	5.8	4
Weighted Moving Average	5.8	3.2
Exponential Moving Average	6	5
1 Layer LSTM	10	4
Bi-directional LSTM	9.5	3.5
CNN LSTM	8	2.8

Figure 8: Crime Prediction through LSTM

6.4 Experiment 4: Application of supervised machine learning model to predict Chicago crime rate based on different districts

The aim of this experiment is to check how supervised machine learning model helps to predict the crime rate in Chicago city. So, Randomforestregressor present in sklearn package of python used to perform modelling. Keeping window as 3, random state as 20 and number of jobs performed as 1, Random Forest regression has performed Kane et al. (2014). The result indicates that during volatile times, models have trouble predicting sudden spikes and drops in outbreaks, such as 2019–2021 in random forest model. In terms of root mean square error, the Random Forest model is not better predictor as compared to LSTM models for future crime prediction on large dataset, according to the simulated prospective analyses. RMSE and MAE values are used to evaluate random forest model and through this model RMSE and MAE are 1943 and 3216 respectively. The plot for actual values vs predicted values is mentioned in Figure 9.

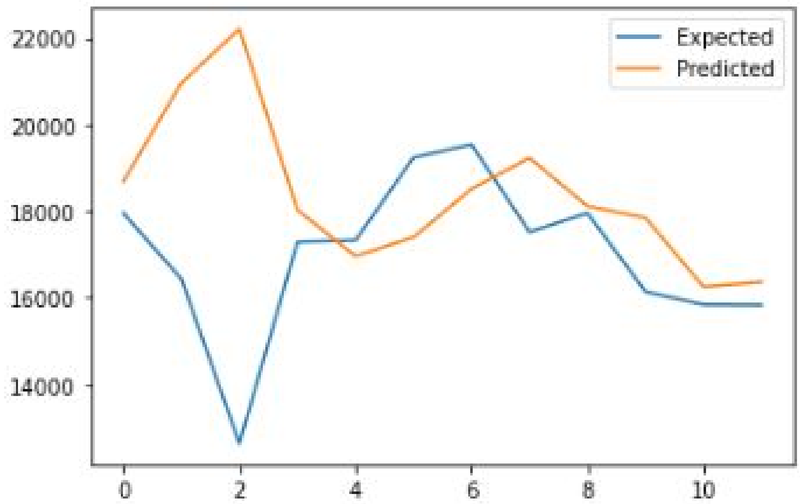

Figure 9: Crime prediction through random forest regression

7 Conclusion and Future Work

The aim of this research was to enhance the safety of metropolitan computationally because crimes pose serious threats to human safety, society, and sustainable development. This research proposes the framework of Crime pattern analysis to predict the crime rate with an acceptable amount of accuracy using Deep learning and supervised machine learning methods. Results demonstrate that RMSEs of LSTM approaches that are based on day-wise data, have dramatically decreased as compared to weekly and monthly prediction. As compared to simple naive approaches, LSTM-based approaches certainly improved. It is evident from the annual crime trend in Chicago that crime rates are declining significantly.A limitation of this study is not able to consider spatial temporal factors like weather, roads condition, proximity to police station etc. that may affect crime rate. When measured week-by-week, this study cannot predict the year-end dip in crime numbers. In addition to models applied, exploratory data analysis provides useful information about criminal activities of Chicago city.

This research can potentially enhance the predictive technologies which detect harmful patterns of police behaviour in advance. In order to improve this work, the models can be optimized, and a statistical methodology set up to determine which final model should be used for the framework based on factors like RMSE, MAE, and size. For different crime datasets, different learning techniques will be applied with corresponding visual data and satellite imagery for the purposes of expanding this study. A future work plan will develop visual images and location maps that will provide effective anticipation of the criminal events to be expected. This will provide police with a chance to improve patrolling regulation. This provides proactive information about crime trends to law enforcement officers and develops assessments of threats, vulnerabilities, and risks. More research has to be carried out to develop crime detection model that will consider spatial and temporal factors of that area.

8 Acknowledgement

To begin with, I want to thank my research supervisors, Paul Stynes, Pramod Pathak and Musfira Jilani for constantly motivating and supporting me. I received continuous assistance and guidance from my supervisors over a period of 13 weeks. It is my pleasure to express my sincere gratitude to my family for their continuous and extraordinary support, understanding, and love. Lastly, I would like to thank my friends for being such a source of inspiration for me.

References

Borowik, G., Wawrzyniak, Z. M. and Cichosz, P. (2018). Time series analysis for crime forecasting, *2018 26th International Conference on Systems Engineering (ICSEng)*, IEEE, pp. 1–10.

Butt, U. M., Letchmunan, S., Hassan, F. H., Ali, M., Baqir, A. and Sherazi, H. H. R. (2020). Spatio-temporal crime hotspot detection and prediction: a systematic literature review, *IEEE Access* **8**: 166553–166574.

Dash, S. K., Safro, I. and Srinivasamurthy, R. S. (2018). Spatio-temporal prediction of crimes using network analytic approach, *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 1912–1917.

Devi, J. V. and Kavitha, K. (2021). Automating time series forecasting on crime data using rnn-lstm, *International Journal of Advanced Computer Science and Applications* **12**(10).

Elyta, E., Mujiono, D. I. and Sagena, U. W. (2022). Facing the dangers in indonesia's waters: Government's efforts in proposing illegal, unreported and unregulated fishing as transnational organized crime, *Intermestic: Journal of International Studies* **6**(2): 336–352.

Han, X., Hu, X., Wu, H., Shen, B. and Wu, J. (2020). Risk prediction of theft crimes in urban communities: an integrated model of lstm and st-gcn, *IEEE Access* **8**: 217222–217230.

Huang, Z., Xu, W. and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging, *arXiv preprint arXiv:1508.01991* .

Kane, M. J., Price, N., Scotch, M. and Rabinowitz, P. (2014). Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks, *BMC bioinformatics* **15**(1): 1–9.

Laycock, G. (2013). Defining crime science, pp. 3–24.

Mansour, H. A. and Lundy, M. (2019). Crime types prediction, *International Conference on Computing*, Springer, pp. 260–274.

Pradhan, I. (2018). Exploratory data analysis and crime prediction in san francisco.

Safat, W., Asghar, S. and Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques, *IEEE Access* **9**: 70080–70094.

Stec, A. and Klabjan, D. (2018). Forecasting crime with deep learning, *arXiv preprint arXiv:1806.01486* .

Tasnim, N., Imam, I. T. and Hashem, M. (2022). A novel multi-module approach to predict crime based on multivariate spatio-temporal data using attention and sequential fusion model, *IEEE Access* **10**: 48009–48030.

Vijayarani, S., Suganya, E. and Navya, C. (2021). Crime analysis and prediction using enhanced arima model, *Journal homepage: www. ijrpr. com ISSN* **2582**: 7421.

Yang, M., Chen, Z., Zhou, M., Liang, X. and Bai, Z. (2021). The impact of covid-19 on crime: A spatial temporal analysis in chicago, *ISPRS International Journal of Geo-Information* **10**(3): 152.

Yi, F., Yu, Z., Zhuang, F., Zhang, X. and Xiong, H. (2018). An integrated model for crime prediction using temporal and spatial factors, *2018 IEEE International Conference on Data Mining (ICDM)*, IEEE, pp. 1386–1391.