

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data

Mingchen Feng¹, Jiangbin Zheng¹, Jinchang Ren^{2,3}, Senior Member, IEEE, Amir Hussain⁴, Senior Member, IEEE, Xiuxiu Li⁵, Yue Xi¹, and Qiaoyuan Liu⁶

¹School of Computer Science, Northwestern Polytechnical University, Xi'an, 710072, China

²Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, G11XW, UK

³School of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, 030024, China

⁴Cognitive Big Data and Cybersecurity Research Lab, Edinburgh Napier University, Edinburgh, EH11 4DY, UK

⁵School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China

⁶Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, 130000, China

Corresponding author: Jiangbin Zheng (zhengjb@nwpu.edu.cn); Jinchang Ren (jinchang.ren@strath.ac.uk).

This work has been supported by HGJ, HJSW and Research & Development plan of Shaanxi Province (Program No. 2017ZDXM-GY-094, 2015KTZDGY04-01) and the National Natural Science Foundation of China under grant Nos.61502382.

ABSTRACT Big Data Analytics (BDA) is a systematic approach for analyzing and identifying different patterns, relations and trends within a large volume of data. In this paper we apply BDA to criminal data where exploratory data analysis is conducted for visualization and trends prediction. Several state-of-the-art data mining and deep learning techniques are used. Following statistical analysis and visualization, some interesting facts and patterns are discovered from criminal data in San Francisco, Chicago and Philadelphia. The predictive results show that the Prophet model and Keras stateful LSTM perform better than neural network models, where the optimal size of the training data is found to be three years. These promising outcomes will benefit for police departments and law enforcement organisations to better understand crime issues and provide insights that will enable them to track activities, predict the likelihood of incidents, effectively deploy resources and optimize the decision making process.

INDEX TERMS Big Data Analytics (BDA), Data Mining, Data Visualization, Neural Network, Time Series Forecasting.

I. INTRODUCTION

In recent years, Big Data Analytics (BDA) has become an emerging approach for analyzing data and extracting information and their relations in a wide range of application areas [1]. Due to continuous urbanization and growing populations, cities play important central roles in our society. However, such developments have also been accompanied by an increase in violent crimes and accidents. To tackle such problems, sociologists, analysts, and safety institutions have devoted much effort towards mining potential patterns and factors [34]. In relation to public policy however, there are many challenges in dealing with large amounts of available data. As a result, new methods and technologies need to be devised in order to analyze this heterogeneous and multi-sourced data [35]. Analysis of such big data enables us to effectively keep track of occurred events, identify similarities from incidents, deploy resources and make quick decisions accordingly [36]. This can also help further our understanding of both historical

issues and current situations, ultimately ensuring improved safety/security and quality of life, as well as increased cultural and economic growth.

The rapid growth of cloud computing and data acquisition and storage technologies, from business and research institutions to governments and various organizations, have led to a huge number of unprecedented scopes/complexities from data that has been collected and made publicly available [37]. It has become increasingly important to extract meaningful information and achieve new insights for understanding patterns from such data resources. BDA can effectively address the challenges of data that are too vast, too unstructured, and too fast moving to be managed by traditional methods [2]. As a fast-growing and influential practice, DBA can aid organizations to utilize their data and facilitate new opportunities. Furthermore, BDA can be deployed to help intelligent businesses move ahead with more effective operations, high profits and satisfied customers.

Consequently, BDA becomes increasingly crucial to organizations to address their developmental issues [3].

As one of the fundamental techniques of BDA, data mining is an innovative, interdisciplinary, and growing research area, which can build paradigms and techniques across various fields for deducing useful information and hidden patterns from data [4]. Data mining is useful in not only the discovery of new knowledge or phenomena but also for enhancing our understanding of known ones. With the support of such techniques, BDA can help us easily identify crime patterns which occur in a particular area and how they are related with time. The implications of machine learning and statistical techniques on crime or other big data applications such as traffic accidents or time series data, will enable the analysis, extraction and understanding of associated patterns and trends, ultimately assisting in crime prevention and management.

In this paper, state-of-the-art machine learning and big data analytics algorithms are utilized for the mining of crime data from three US cities, i.e. San-Francisco, Chicago and Philadelphia. After preprocessing, including data filtering and normalization, Google maps based geo-mapping of the features are implemented for visualization of the statistical results. Various approaches in machine learning, deep learning, and time series modelling are utilized for future trends analysis. The major contribution of this paper can be summarized as follows:

- 1) A series of investigative explorations are conducted to explore and explain the crime data in three US cities;
- 2) We propose a novel visual representation which is capable of handling large datasets and enables users to explore, compare, and analyze evolutionary trends and patterns of crime incidents;
- 3) A combination and comparison of different machine learning, deep learning and time series modeling algorithms to predict trends with the optimal parameters, time periods and models.

The rest of this paper is organized as follows. Following a literature review of related work in Section 2, Section 3 details the techniques used for data processing and visualization. Section 4 describes the deep learning and machine learning techniques for trends analysis. Experimental results are presented in Section 5, followed by some concluding remarks summarized in Section 6.

II. RELATED WORK

A. BIG DATA ANALYTICS

Big data analytics (BDA) has been extensively applied and studied in the fields of data science and computer science for quite some time. While machine learning provide both opportunities and challenges when meets large amount of big data [52]. Raghupathi et al. [5] described the promise and potential of BDA in healthcare and summarized current

challenges. Archenaa et al. [6] conducted a survey on the applications of BDA in healthcare and the government. Londhe et al. [7] presented various software frameworks available for BDA and discussed some widely used data mining algorithms. Grady et al. [8] illustrated the implications of an agile process for the cleansing, transformation, and analytics of data in BDA. Vatrappu et al. [9] demonstrated the suitability and effectiveness of BDA for conceptualizing, formalizing and analyzing of big social data from content-driven social media platforms e.g. Facebook. The so-called Social Set Analysis was used for studying events such as unexpected crises and coordinated marketing campaigns. Zhang et al. [10] proposed an overall BDA architecture for the lifecycle of products, where BDA and service-driven patterns were integrated to assist in the decision making and overcome barriers. Ngai et al. [11] reviewed BDA and its applications on electronic markets. Liu et al. [12] exemplified the utilization of BDA in the tourism domain in terms of the datasets, data capture techniques, analytical tools, and analysis results to provide insights for Destination Management and Marketing. Fisher et al. [13] explained the conception of big data in BDA, its analytics and the associated challenges when interacting among them.

B. CRIME DATA MINING, VISUALIZATION & TRENDS FORECASTING

In criminology literature the relationship between crime and various factors has been intensively analyzed, where typical examples include historical crime records [14], unemployment rate [15], and spatial similarity [16]. Using data mining and statistical techniques, new algorithms and systems have been developed along with new types of data. For instance, classification and statistical models are applied for mining of crime patterns and crime prediction [17, 18], where transfer learning has been employed to exploit spatio-temporal patterns in New York city [19]. Wu et al. [20] developed a system to automatically collect crime-logged data for mining of crime patterns, in order to achieve more effective crime prevention in/around a university campus. In Vineeth et al. [21], a random forest was applied on the obtained correlation between crime types to classify the state based on their crime intensity point. Unsupervised learning based methods have also been used for mining of crime patterns and crime hotspots, such as memetic differential fuzzy cluster [22] for forecasting of criminal patterns, and fuzzy C-means algorithm [23] to cluster criminal events in space. Noor et al. [24] derived association mining rules to determine relationships between different crimes. MohammadNoor et al. [53] conducted a survey to summary data mining techniques on social media.

With deep learning and neural networks, new models have been developed to predict crime occurrence [25]. As deep learning [26,27,28] and artificial intelligence [29,30] have achieved great success in computer vision, they are also

applied in BDA for predicting trends and classification. In Zhao et al. [31] and Dai et al. [32], Long Short-Term Memory networks were successfully applied to predict stock price and gas dissolved in power transformation. In Kashef et al. [33], a neural network was applied on a smart grid system to estimate the trends of power loss. Zheng et al. [55] proposed a big data processing architecture for radio signals analysis. Zhao et al. [56] proposed a neural network model to predict travel time and gained high accuracy. Niu et al. [57] utilized LSTM and developed an effective speed prediction model to solve spatio-temporal prediction problems. Peral et al. [58] summarized analytics techniques and proposed an architecture for online forum data mining.

In our proposed system, a similar but more comprehensive workflow has been adopted, which include statistical analysis, data visualization, and trends prediction. We explored crime data in three US cities, Chicago, San Francisco and Philadelphia. Data visualization and mining techniques are used to show the extracted statistical relationships among different attributes within the huge volume of data. State-of-the-art machine learning and deep learning algorithms are deployed to forecast trends and obtain optimal models with the highest accuracy.

III. DATA ANALYSIS AND VISUALIZATION

The three crime datasets we used for analysis are publicly available, which cover 3 cities in US, i.e. San-Francisco, Chicago, and Philadelphia. The San-Francisco crime data contains 2,142,685 crime incidents from 01/01/2003 to 11/08/2017 [38]. Data from Chicago has a total number of 5,541,398 records, dating back from 2017 to 2003 [39]. In the Philadelphia dataset, there are 2,371,416 crime incidents which were captured from 01/01/2006 to 12/31/2017 [40]. Detailed analysis of these dataset is presented as follows.

A. FEATURED ATTRIBUTES

For each entry of crime incidents in the datasets, the following 13 featured attributes are included:

- 1) IncidentNum - Case number of each incident;
- 2) Dates - Date and timestamp of the crime incident;
- 3) Category - Type of the crime. This is the

- 4) Descript - A brief note describing any pertinent details of the crime;
- 5) DayOfWeek - Day of the week that crime occurred;
- 6) PdDistrict - Police Department District ID where the crime is assigned;
- 7) Resolution - How the crime incident was resolved (with the perpetrator being, say, arrest or booked);
- 8) Address - The approximate street address of the crime incident;
- 9) X - Longitude of the location of a crime;
- 10) Y - Latitude of the location of a crime;
- 11) Coordinate - Pairs of Longitude and Latitude;
- 12) Dome - whether crime id domestic or not;
- 13) Arrest - Arrested or not;

B. DATA PREPROCESSING

Before implementing any algorithms on our datasets, a series of preprocessing steps are performed for data conditioning as presented below:

1) Time is discretized into a couple of columns to allow for time series forecasting for the overall trend within the data.

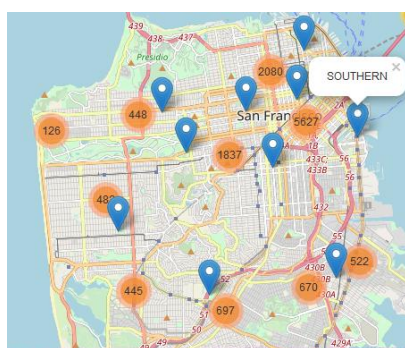
2) For some missing coordinate attributes in Chicago and Philadelphia datasets, we imputed random values sampled from the non-missing values, computed their mean, and then replaced the missing ones [41].

3) The timestamp indicates the date and time of occurrence of each crime, we deduced these attributes into five features: Year (2003-2017), Month (1-12), Day (1-31), Hour (0-23), and Minute (0-59).

4) We also omit some features that unneeded like incidentNum, coordinate.

C. NARRATIVE VISUALIZATION

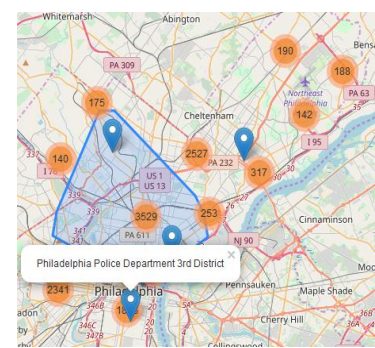
Considering the geographic nature of the crime incidents, an interactive map based on Google map was used for data visualization, where crime incidents are clustered according to their latitude/longitude information. As shown in Fig. 1, the blue label stands for the distribution of police stations in each city, where the round label with numbers are for crime hot-spots and the associated number of incidents.



(a)



(b)



(c)

target/label that we need to predict in the classification stage;

FIGURE 1. Visualization of crime activities on openstreet maps for cities of (a) San-Francisco, (b) Chicago, and (c) Philadelphia

The marker cluster algorithm [42] can help us manage multiple markers at different zooming levels, corresponding to various spatial scales or resolutions. When zooming out, the markers will gather together into clusters to view a broader geographic range on the map yet with a much coarser scale. When viewing the map at a high zooming level, the individual markers can be shown on the map to indicate the exact location of the incident. By using this interactive map, users can look up crimes on specific dates and streets. This can provide some preliminary guidance of the distributions of the associated incidents. Actually, it will show more crimes at the downtown and less crimes around the edge of a city.

Fig. 2 summarized crime incidents in each year for the three cities. In San-Francisco, the number of crime incidents seems to soar since 2012 and reaches its peak in 2013, whilst the numbers for the other two cities tend to decreasing yet following certain patterns. Actually, crime incidents in these two cities tend to increase from January and reach the peaks in the middle of the year and then begin to decline until the end of year. It seems that at the beginning and the end of each year, crime incidents always reach the lowest point. This may due to it is the time when everyone is engaged in celebrating the new year. We also learned from US economic growth statistics [50] that San-Francisco is one of the 9 cities with the worst income inequality, which will lead to wealth gap and cause more crimes [51].

Among 30+ categories of reported crime incidents available in the datasets, the distribution of these categories is heavily skewed. As such, we focus mainly on the Top-10 frequently occurred crimes and plot their distributions as percentages in Fig. 3. As seen, theft is a major problem that threatens each city, along with some violent crimes such as assaults and burglary.

The hourly trends unravel some interesting crime facts as shown in Fig. 4. As seen, crimes by hour in all three cities share similar patterns, where 5-6 am is the safest part of the day whilst 0 am is the most dangerous hour with the most

crime incidents reported. Also, 12 pm is also very dangerous during a day and in fact the hour where the maximum incidents reported under some crime categories. This actually is reasonable as it is the time when more people go out for lunch and also the time when tired criminals become aggressive under limited police available [43]. As a result, more police resources should be allocated in the shift from noon to midnight if the police force is limited. We also noticed that Chicago has more crimes than the other two cities, and San Francisco is only one-third of the size of Philadelphia but has the same number of criminal cases. From American census data we know that Chicago has more population and larger area, which explains somehow more crimes it has. When calculating the population density as person/square mile, however, it becomes 17179 for San-Francisco, 11841 for Chicago, and 11379 for Philadelphia, this actually is another main factor that boosts crimes [44].

In Fig. 5, we show the keywords of crimes for the three cities, where wordcloud plots are used to illustrate the significance of different categories of crimes. As seen, San-Francisco is now facing the problem of theft and property related issues. Chicago, however, is plagued by domestic and battery crime as well as vehicle-related crimes. For Philadelphia, the crime description is relatively simple, and the major problem of the city we discovered is related to theft, vehicle and assaults.

Fig. 6 shows the top streets where the maximum crime incidents were reported in the three cities. In San-Francisco the top three dangerous streets were Bryant Street, Market Street, and Mission Street. In Chicago these become Ohare Street, State Street, and Cicero Avenue. In Philadelphia, Roosevelt BLVD, Frankford Avenue, and Market Street were the three streets with the most frequent crimes. These streets are all close to the downtown, where many financial districts, fortune 500 companies, and luxury hotels located nearby. This may be one reason that crime incidents boosts.

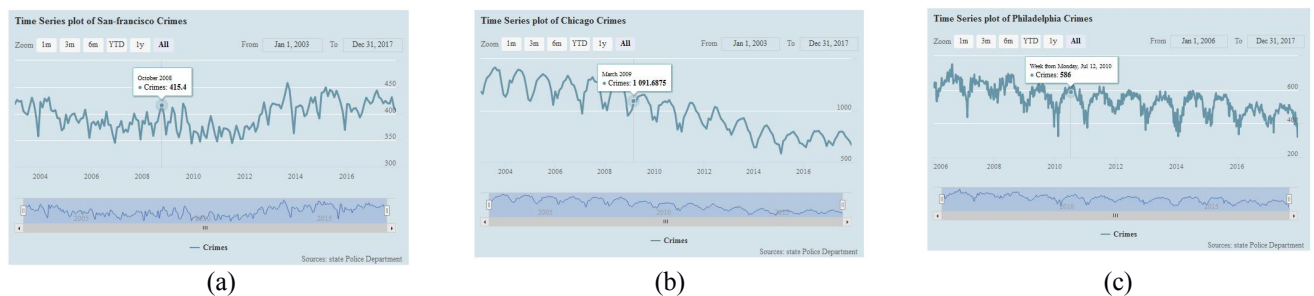


FIGURE 2. Time Series plot of crime incidents for cities of (a) San-Francisco, (b) Chicago, and (c) Philadelphia.

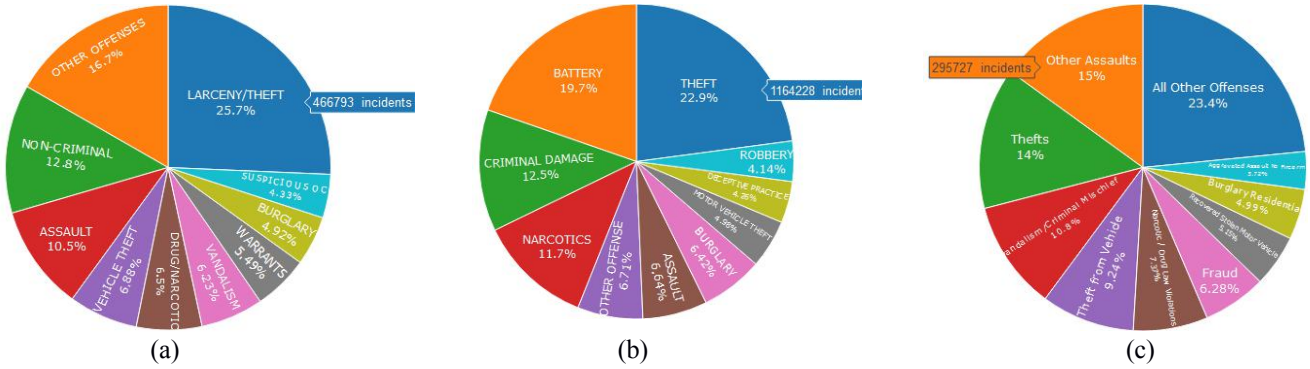


FIGURE 3. Top-10 Crime cases for cities of (a) San-Francisco, (b) Chicago, and (c) Philadelphia.

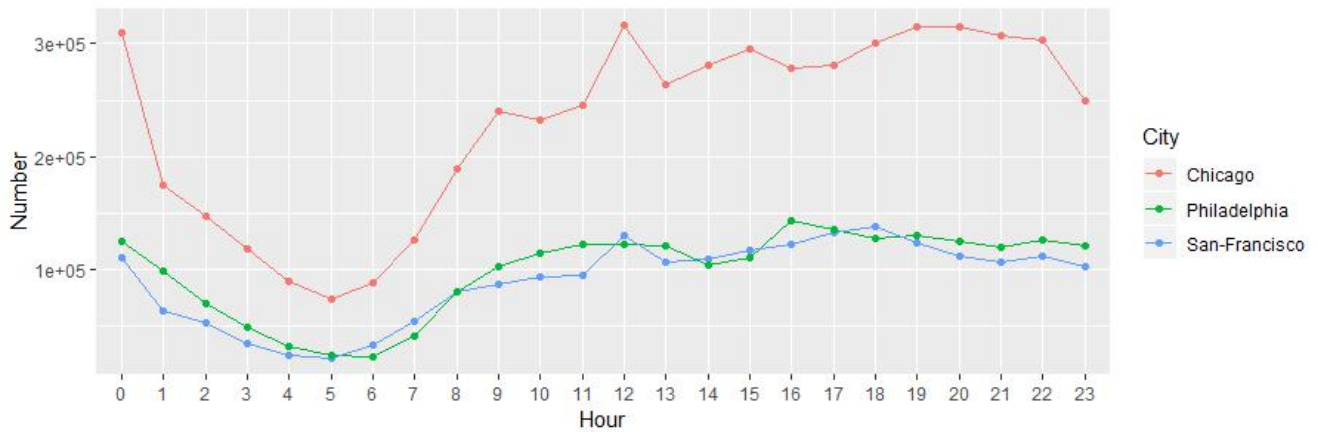


FIGURE 4. Hourly trend of crime in each city.

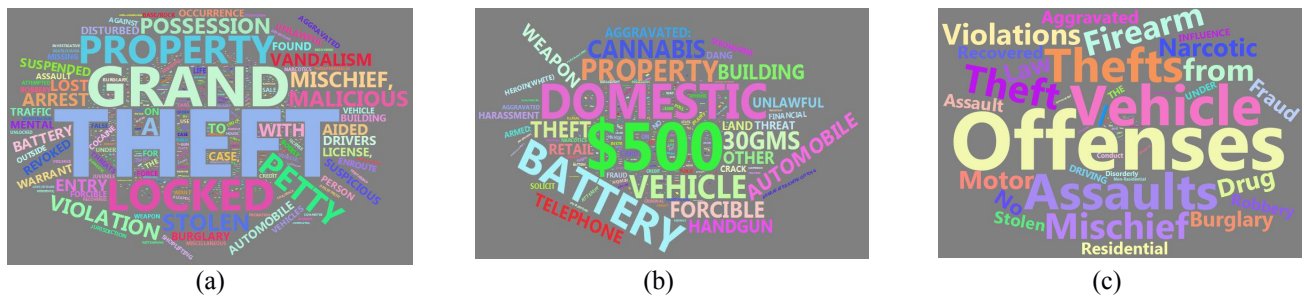


FIGURE 5. Wordcloud of crime description in each city: (a) San-Francisco, (b) Chicago, (c) Philadelphia.

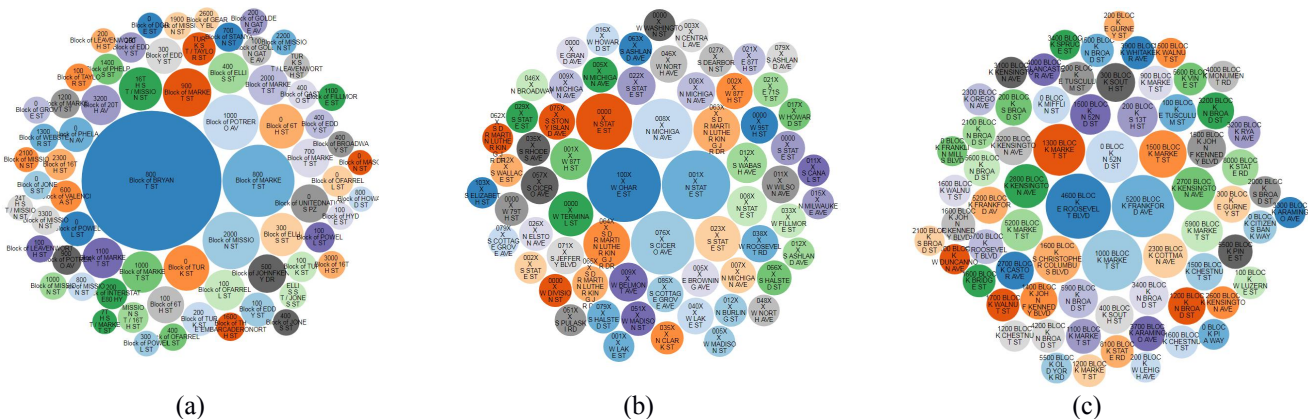


FIGURE 6. Bubble plot for street crime analysis: (a) San-Francisco, (b) Chicago, (c) Philadelphia.

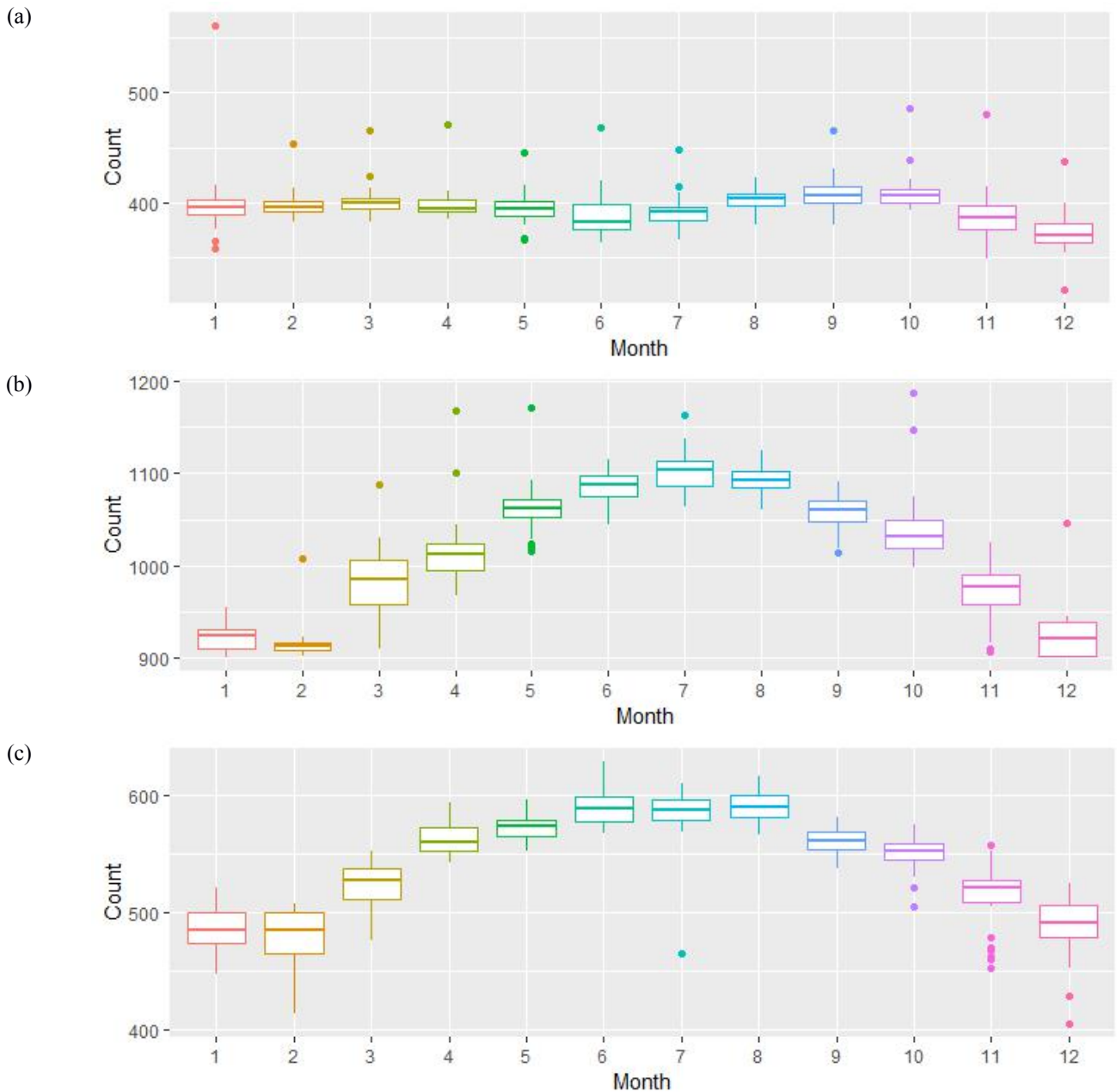


FIGURE 7. Box plot of Monthly Statistics of crime in each city: (a) San-Francisco, (b) Chicago, (c) Philadelphia.

Fig. 7 demonstrates the monthly based crime statistics, in which we can clearly see that the number of crime incidents in San-Francisco is quite stable with an average of 400 cases per day. The month of December seems the safest and is observed to have the least crime incidents whilst October is least safe month with the maximum number of reported crime incidents. For Philadelphia, the mean value of reported crimes in each day is 400-600 incidents with February the safest and June-August the least safe months. While Chicago has the most crimes reported among the three cities with a mean value of 1000 incidents per day, where February is the safest month followed by December and January. However, June-August are the most notorious months with the highest

numbers of crime incidents. **We found that monthly crime rate is very likely linked to local climates:** With a Mediterranean climate, San-Francisco is warm hence people may share similar working and life patterns throughout the year. Whilst Chicago and Philadelphia are temperate continental climate with a cold winter and hot summer in June- August, significant more outdoor living/activities can be found in summer than winter days thus the associated higher or lower crime incidents in different seasons.

IV. PREDICTION MODELS

In order to tackle the problem of crime trends forecasting we explored several state-of-the-art machine learning and deep

learning algorithms and time series models. A time series is a sequence of numerical data points successively indexed or listed/graphed in the time order. Usually, the successive data points within a time series are equally spaced in time, hence these data are discrete in time. Fig. 8 demonstrates how the amount of crime incidents changed over time, which clearly show the potential trend and seasonality in the data as analysed and discussed in the following sections.

A. PROPHET MODEL

The Prophet model is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and/or daily seasonality, plus holiday effects [36]. It works best with time series that have strong seasonal effects and cover several seasons of historical data. The Prophet model is robust to missing data and shifts in the trend, and typically it handles outliers well. The Prophet model is designed to handle complex features in time series, it also designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model.

The Prophet model decomposes time series into three main components, i.e. the trend, seasonality, and holidays. They are combined in the following equation:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

where $g(t)$ is the trend of any non-periodic changes in the time series, $s(t)$ represents periodic changes (e.g., weekly and yearly seasonality), and $h(t)$ represents the holiday effects of any potentially irregular schedules over one or more days. The error term ε_t represents any random effects which are not accommodated by the model.

For the trend function $g(t)$, we utilized a linear trend with limited change points, where a piece-wise constant rate of growth provides a parsimonious and useful model.

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (2)$$

$$a_j(t) = \begin{cases} 1, & \text{if } t \geq s_j, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

where k is the growth rate, δ is the rate adjustment, m is the offset parameter, and γ is set to $-s_j \delta_j$ to make the function continuous.

Crime data is time series with multi-period seasonality, e.g. daily, weekly and annually. To this end, the seasonal function relies on Fourier series defined below, where P denotes the regular period:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (4)$$

As for holidays related events, they are defined by generating a matrix of regressors:

$$Z(t) = [l(t \in D_1), \dots, l(t \in D_L)] \quad (5)$$

$$h(t) = Z(t)k \quad (6)$$

where D_i is the set of past and future dates for holidays, $k \sim \text{Normal}(0, v^2)$, l is an indicator function representing

whether time t is during holiday i , $Z(t)$ is the regressor matrix.

B. NEURAL NETWORK MODEL

A neural network is composed of a certain numbers of neurons, namely nodes in the network, which are organized in several layers and connected to each other cross different layers [45]. There are at least three layers in a neural network, i.e. the input layer of the observations, a non-observable hidden layer in the middle, and an output layer as the predicted results. In this paper we explored the multilayer feed-forward network, where each layer of nodes receives inputs from the previous layer. The outputs of the nodes in one layer will become the inputs to the next layer. The inputs to each node are combined using a weighted linear combination below [46]:

$$z_j = b_j + \sum_{i=1}^n \omega_{i,j} x_i \quad (7)$$

The hidden layer will modify the input above using a nonlinear function by

$$s(z) = \frac{1}{1 + e^{-z}} \quad (8)$$

C. LSTM MODEL

LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies [47]. For time series involves auto-correlation, i.e. the presence of correlation between the time series and lagged versions of itself, LSTMs are particular useful in prediction due to their capability of maintaining the state whilst recognizing patterns over the time series. The recurrent architecture enables the states to be persisted, or communicate between updated weights as each epoch progresses. Moreover, the LSTM cell architecture can enhance the RNN by enabling long term persistence in addition to short term [48].

$$f(t) = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (11)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (12)$$

Where, f_t is a sigmoid function to indicate whether to keep the previous state, C_{t-1} is the old cell state, C_t is the updated cell state, W_f , W_i , and W_C are the previous value in each layer, h_{t-1} and x_t is the input value, b_f , b_i , and b_C are constant values, i_t decides which value will be used to update the state, C_t stands for the new candidate values.

V. EXPERIMENTAL RESULTS

A. DECOMPOSE OF TIME SERIES

Time series can exhibit a variety of patterns, and it is always helpful to decompose a time series into several components, each representing an underlying pattern category. Fig. 8 illustrates a decomposed crime time series, where for each original time series on the top, the three decomposed parts can respectively show the estimated trend component, seasonal component, and irregular component, respectively. The estimated trend component has shown that the overall crimes in San-Francisco slightly decreased from 2003 to 2013, followed by a steady increase from then on to 2017. However, crimes in Chicago seemed to decrease quickly from 2003 to 2015 and then became quite stable, whilst in

Philadelphia the number of crimes had a downward trend yet with some undulations until it became stable after 2016. Regarding the seasonal component, it changes slowly over the time, where a quite strong annual periodic pattern can be observed for Chicago and Philadelphia than San-Francisco. This may be due to the more apparent annual climate changes in the two cities, where the crimes reached the peak in the middle of the year when the temperature becomes the hottest. As such, time series models will perform well on these datasets to forecast crimes in the future.

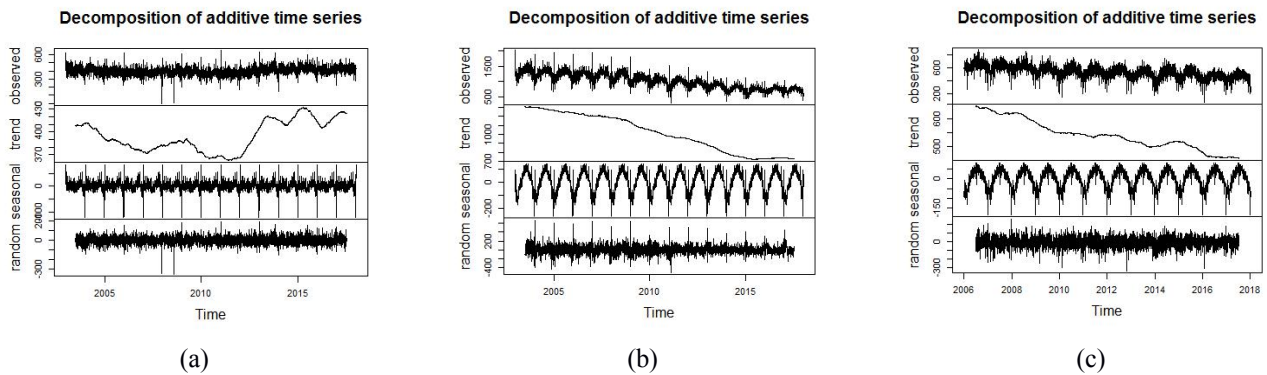


FIGURE 8. Decomposed time series to show how crime evolved over time in three cities of San Francisco (a), Chicago (b), and Philadelphia (c). For each original time series (top), we show the estimated trend component (2nd top), the estimated seasonal component (3rd top), and the estimated irregular component (bottom).

B. RESULTS ANALYSIS

We have explored deep learning algorithms and time series forecast models to predict crime trends. For performance evaluation, the Root Mean Square Error (RMSE) and spearman correlation [49] are used in terms of different parameters and different sizes of training samples. The RMSE and spearman correlation are defined as follows.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{Y}_i)^2} \quad (13)$$

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (14)$$

where Y_i and Y are the true values; \tilde{Y}_i and X are predicted values; cov is the covariance; σ_X is the standard deviation of X , and σ_Y is the standard deviation of Y .

To train our models for predicting trends, we first summarized the number of crime incidents per day, and then transformed these data into a “tibbletime” format, and then we divided the data into training and testing sets, where the training set contains data from 2003 to 2016 and the testing set has data from 2017, for training

process we set 1 year’s data as validation set. We evaluated the performance of the prediction models whilst changing the number of training years from 1 to 10 and the results are summarized in Table I. As seen in Table I, more training data do not necessarily lead to better results although too little training data also fails to generate good results. The optimal time period for crime trends forecasting is 3 years where the RMSE is the minimum and the spearman correlation is the highest. The results also showed that Prophet model and LSTM model performed better than traditional neural network models as demonstrated in table. I that neural network seems has lower RMSE but the correlation between predicted values and the real ones is low. The visualization of the trends in Fig. 10, 11, and 12 also conforms this conclusion.

Besides, we also evaluated the effects of some key parameters in the best two approaches, the Prophet and LSTM models. For Prophet model, after training we can obtain trends and seasonality of the dataset, but for holiday components we have to manually input the value. As shown in Fig. 9, we summarize top 10 dates with the most and least crime incidents respectively, thus we set these 20 dates as holidays. Moreover, we analyzed different changepoint ranges, referring to the proportion of history in which trend changepoints will be estimated. According to the results shown in Table II, the best

change point range is determined as 0.8, which means that the first 80 points are specified as change points. For the LSTM model, we first investigated the effect of different echos, i.e. the total number of forward/backward propagation iterations. Generally, the more iterations, the better the model performs [46]. However, in our experiments, we found the optimal number of epoch for the best results is 300 for the three

datasets. In addition, we analyzed in detail the impact of different layers and neurons as shown in table. III and table. IV, the best number of layers for LSTM is computed as 50. The number of neurons in our model is affected largely by a parameter called cell state, we found that when the cell state is set as 60, we get the optimal results.

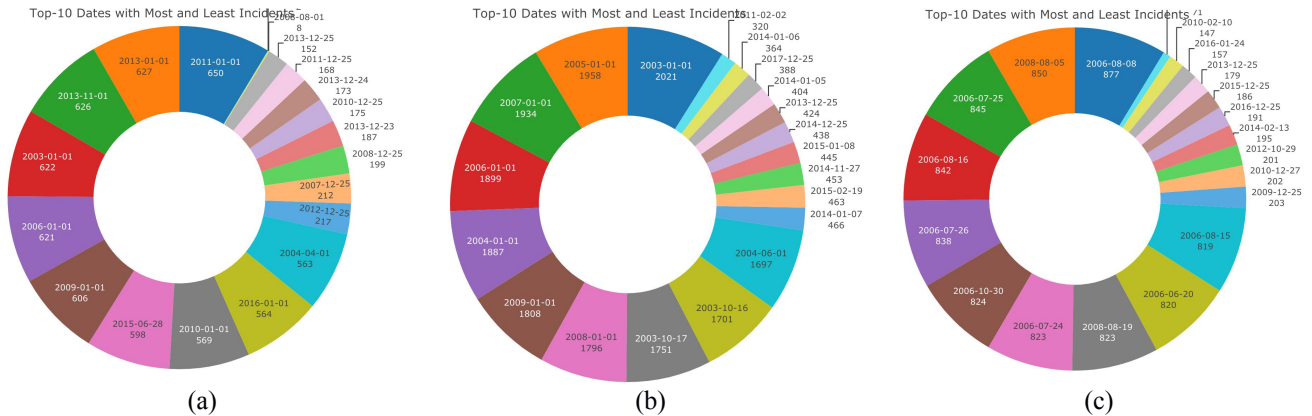


FIGURE 9. Top-10 dates with the most and least crimes: (a) San-Francisco, (b)Chicago, (c)Philadelphia.

TABLE I Comparison of different algorithms/models in terms of RMSE and spearman correlation under different sizes of training samples

city	Years for training	RMSE-Prophet	Correlation-Prophet	RMSE-LSTM	Correlation-LSTM	RMSE-Neural Network	Correlation-Neural Network
San-Francisco	10	38.21	0.384	55.21	0.354	54.79	0.097
San-Francisco	5	35.70	0.402	54.18	0.365	48.04	0.232
San-Francisco	4	36.18	0.415	53.96	0.411	48.04	0.236
San-Francisco	3	35.65	0.398	45.65	0.423	41.62	0.291
San-Francisco	2	91.93	0.087	160.2	0.098	41.17	0.128
San-Francisco	1	100.56	0.182	95.96	0.122	-	-
Chicago	10	76.89	0.560	77.01	0.532	77.19	0.367
Chicago	5	68.21	0.652	69.12	0.549	92.74	0.551
Chicago	4	66.75	0.654	67.45	0.612	88.04	0.492
Chicago	3	66.68	0.658	67.15	0.625	75.06	0.505
Chicago	2	67.42	0.632	68.14	0.576	75.90	0.552
Chicago	1	100.51	0.02	78.98	0.459	-	-
Philadelphia	10	51.83	0.716	50.65	0.709	82.21	0.422
Philadelphia	5	56.07	0.728	51.23	0.698	71.19	0.486
Philadelphia	4	55.37	0.728	49.22	0.714	67.74	0.588
Philadelphia	3	48.73	0.729	48.15	0.725	63.68	0.537
Philadelphia	2	50.35	0.718	57.16	0.705	170.12	0.128
Philadelphia	1	100.71	0.098	140.63	0.562	-	-

TABLE II Comparison of different change point ranges in the Prophet model in terms of RMSE and spearman correlation

Range of changepoint	RMSE-San-Francisco	Correlation-San-Francisco	RMSE-Chicago	Correlation-Chicago	RMSE-Philadelphia	Correlation-Philadelphia
0.9	36.17	0.405	68.12	0.649	48.79	0.721
0.8	35.65	0.398	66.68	0.658	48.73	0.729
0.7	36.42	0.395	67.14	0.652	49.25	0.715
0.6	36.15	0.401	67.22	0.649	49.85	0.702
0.5	37.89	0.389	68.23	0.632	50.12	0.698
0.4	41.26	0.362	68.52	0.612	50.23	0.692
0.3	45.18	0.285	68.28	0.645	50.98	0.691
0.2	49.68	0.273	69.36	0.611	51.11	0.688
0.1	51.12	0.259	69.89	0.605	51.15	0.689

TABLE III Comparison of different layers in the LSTM model in terms of RMSE and spearman correlation

Layers	RMSE-San-Francisco	Correlation-San-Francisco	RMSE-Chicago	Correlation-Chicago	RMSE-Philadelphia	Correlation-Philadelphia
80	42.52	0.393	68.42	0.596	50.86	0.734
70	42.23	0.413	67.56	0.622	49.23	0.721
60	42.58	0.402	67.15	0.614	48.19	0.725
50	41.79	0.419	67.15	0.625	48.11	0.743
40	43.53	0.376	69.16	0.618	49.87	0.692
30	44.74	0.325	71.23	0.607	52.54	0.685
20	43.32	0.208	72.18	0.598	55.19	0.699

TABLE IV Comparison of different number of neuros (cell state) in the LSTM model in terms of RMSE and spearman correlation

Number of neurons (cell state)	RMSE-San-Francisco	Correlation-San-Francisco	RMSE-Chicago	Correlation-Chicago	RMSE-Philadelphia	Correlation-Philadelphia
44351 (70)	87.91	0.097	350.20	-0.11	362.01	-0.19
37101 (60)	41.79	0.419	67.15	0.625	48.11	0.743
30651(50)	44.63	0.395	68.67	0.621	49.51	0.704
25001 (40)	48.69	0.408	77.97	0.618	58.93	0.689
20151 (30)	69.89	0.325	85.15	0.607	67.59	0.598
16101 (20)	70.12	0.208	83.67	0.379	84.33	0.463

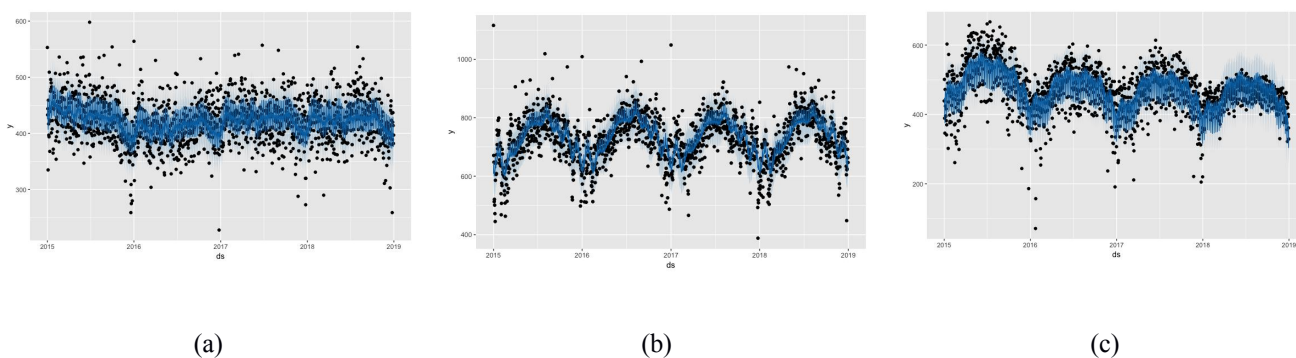
With the derived best training period as three years and the optimal parameters for the Prophet and the LSTM models, we predict the crimes for the three cities in 2018 as shown in Fig. 9, Fig. 10 and Fig. 11 respectively. All predictive results indicate that crime in 2018 may decrease but still boost in some months. When compared with part of the data available in 2018, our results show relatively low RMSE and high correlation with the new data. As such we suggest that police department deploy more forces in summer and citizens be more cautious when going out.

VI. CONCLUSION & FUTURE WORK

In this paper a series of state-of-the-art big data analytics and visualization techniques were utilized to analyze crime big data from three US cities, which allowed us to identify patterns and obtain trends. By exploring the Prophet model, a neural network model, and the deep learning algorithm LSTM, we found that both the Prophet model and the LSTM algorithm perform better than conventional neural network

models. We also found the optimal time period for the training sample to be 3 years, in order to achieve the best prediction of trends in terms of RMSE and spearman correlation. Optimal parameters for the Prophet and the LSTM models are also determined. Additional results explained earlier will provide new insights into crime trends and will assist both police departments and law enforcement agencies in their decision making.

In future, we plan to complete our on-going platform for generic big data analytics which will be capable of processing various types of data for a wide range of applications. We also plan to incorporate multivariate visualization[59], graph mining techniques [54] and fine-grained spatial analysis[60] to uncover more potential patterns and trends within these datasets. Moreover, we aim to conduct more realistic case studies to further evaluate the effectiveness and scalability of the different models in our system.

**FIGURE 10.** Crime trends in 2018 using Prophet Model with 3 years training data in (a)San-Francisco, (b)Chicago, (c)Philadelphia

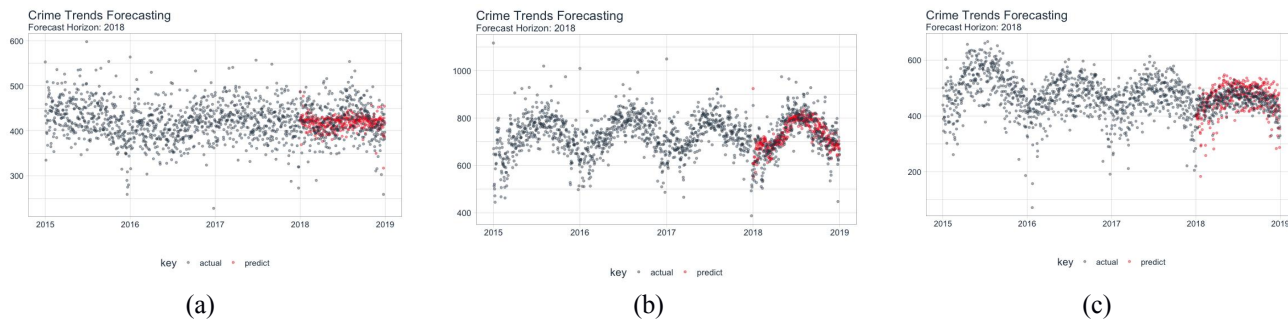


FIGURE 11. Crime trends in 2018 using State LSTM Model with 3 years training data in (a)San-Francisco, (b)Chicago, (c)Philadelphia

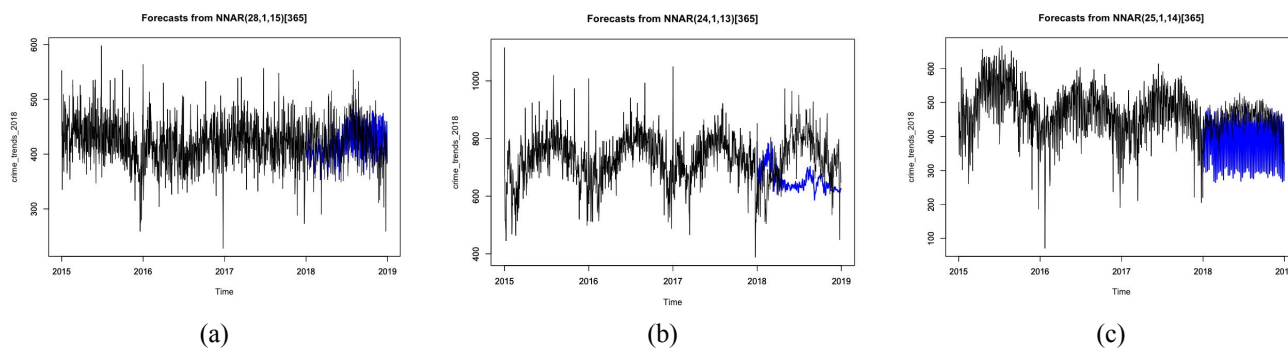


FIGURE 12. Crime trends in 2018 using Neural Network Model with 3 years training data in (a)San-Francisco, (b)Chicago, (c)Philadelphia

REFERENCES

[1] Gandomi A, Haider M. , vol.35, no. 2, pp. 137-144, Apr. 2015.

[2] Zakir J, Seymour T, et al., BIG DATA ANALYTICS. Issues in Information Systems, vol. 16, no. 2, pp. 81-90, 2015.

[3] Wang Y, Kung L A, et al. An integrated big data analytics-enabled transformation model: Application to health care. Information & Management, vol. 55, no. 1, pp. 64-79, Jan. 2018.

[4] Thongsatopornwatana U. A survey of data mining techniques for analyzing crime patterns. In: 2nd Asian Conf. On Defence Technology, Chiang Mai, Thailand, 2016, pp. 123-128.

[5] Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems, vol.2, no. 1, pp. 1-10, Feb. 2014.

[6] Archenaa J, Anita E A M. A survey of big data analytics in healthcare and government. Procedia Computer Science, vol. 50, pp. 408-413, Apr. 2015.

[7] Londhe A., Rao P., Platforms for big data analytics: Trend towards hybrid era, 2017 Int. Conf. on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 3235-3238.

[8] Grady W., Payne A. et al., Agile big data analytics: AnalyticsOps for data science, In 2017 IEEE Int. Conf. On Big Data, Boston, MA, USA, 2017, pp. 2331-2339.

[9] Vatrapu R., R. Mukkamala R., Hussain A. et al., Social Set Analysis: A Set Theoretical Approach to Big Data Analytics, IEEE Access, vol. 4, pp. 2542-2571, Apr. 2016.

[10] Zhang Y, Ren S, Liu Y, et al. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of Cleaner Production, vol. 142, no. 2, pp. 626-641, Jan. 2017.

[11] Ngai E W T, Gunasekaran A, et al. Big data analytics in electronic markets. Electronic Markets, vol. 27, no. 3, pp. 243-245, Aug. 2017.

[12] Liu Y Y, Tseng F M, Tseng Y H. Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. Technological Forecasting and Social Change, vol. 130, pp. 123-134, May 2018.

[13] Fisher D, DeLine R, et al. Interactions with big data analytics. interactions, vol. 19, no. 3, pp. 50-59, June 2012.

[14] Yu C., Ding W.: et al., Hierarchical Spatio-Temporal Pattern Discovery and Predictive Modeling. IEEE Trans. on Knowledge and Data Engineering, vol. 28, no. 4, pp. 979-993, Apr. 2016.

[15] Musa S., Smart Cities-A Road Map for Development. IEEE Potentials, vol. 37, no. 2, pp. 19-23, Mar. 2018.

[16] Wang S., Wang X., et al.: Parallel Crime Scene Analysis Based on ACP Approach. IEEE Transactions on Computational Social Systems, vol. 5, no. 1, pp.244-255, Jan. 2018.

[17] Yadav S, Timbadia M., et al: Crime pattern detection, analysis & prediction. In: IEEE Int. Conf. on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2017, pp. 225-230.

[18] Baloian N. et al.: Crime prediction using patterns and context. In: 21st IEEE Int. Conf. on Computer Supported Cooperative Work in Design, Wellington, New Zealand, 2017, pp. 2-9.

[19] Zhao X., Tang J.: Exploring Transfer Learning for Crime Prediction. In Proc. IEEE Int. Conf. on Data Mining Workshops, New Orleans, LA, 2017, pp. 1158-1159.

[20] Wu S. et al.: Spatial-temporal campus crime pattern mining from historical alert messages. In: Int. Conf. on Computing, Networking and Communications, Santa Clara, CA, USA, 2017, pp. 778-782.

[21] Vineeth K., Pandey A., et al.: A novel approach for intelligent crime pattern discovery and prediction. In: Int. Conf. on Advanced Communication Control and Computing Technologies, Ramanathapuram, Ramanathapuram, India, 2016, pp. 531-538.

[22] Rodriguez C., Gomez D., et al.: Forecasting time series from clustering by a memetic differential fuzzy approach: An application to crime prediction. IEEE Symposium Series on Computational Intelligence, Honolulu, HI, 2017, pp. 1-8.

[23] Joshi A., Sabitha A. S., et al.: Crime Analysis Using K-Means Clustering. In: 2017 3rd Int. Conf. on Computational Intelligence and Networks, Odisha, 2017, pp. 33-39.

[24] Noor N., Ghazali A., et al.: Supporting decision making in situational crime prevention using fuzzy association rule. In: Int. Conf. on Computer, Control, Informatics and Its Applications (IC3INA), Jakarta, 2013, pp. 225-229.

- [25] Wang M., Zhang F., et al.: Hybrid neural network mixed with random forests and Perlin noise. In: 2nd IEEE Int. Conf. On Computer and Communications (ICCC), Chengdu, China, 2016, pp. 1937-1941.
- [26] Wang Z, Ren J, et al.: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. *Neurocomputing*, vol. 287, pp. 68-83, Apr. 2018.
- [27] Ren J, Jiang J.: Hierarchical modeling and adaptive clustering for real-time summarization of rush videos. *IEEE Transactions on Multimedia*, vol.11, no. 5, pp. 906-917, Aug. 2009.
- [28] Han J, Zhang D, et al.: Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325-3337, Dec. 2014.
- [29] Yan Y, Ren J, et al.: Cognitive Fusion of Thermal and Visible Imagery for Effective Detection and Tracking of Pedestrians in Videos. *Cognitive Computation*, vol. 10, no. 1, pp. 94-104, Feb. 2018.
- [30] Yan Y, Ren J, et al.: Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. *Pattern Recognition*, vol. 79, pp. 65-78, July 2018..
- [31] Zhao Z., Rao R. et al., Time-Weighted LSTM Model with Redefined Labeling for Stock Trend Prediction, 2017 IEEE 29th Int. Conf. on Tools with Artificial Intelligence (ICTAI), Boston, MA, USA, 2017, pp. 1210-1217.
- [32] Dai J., Song H. et al., LSTM networks for the trend prediction of gases dissolved in power transformer insulation oil, 2018 12th Int. Conf. on the Properties and Applications of Dielectric Materials , Xi'an, China, 2018, pp. 666-669.
- [33] Kashef H., Mahmoud K. et al., Power loss estimation in smart grids using a neural network model, 2018 Int. Conf. on Innovative Trends in Computer Engineering (ITCE), Aswan, 2018, pp. 258-263.
- [34] Hassani H, Huang X, Silva E S, et al. A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 9, no. 3, pp. 139-154, Apr. 2016.
- [35] Jia Z, Shen C, Yi X, et al. Big-data analysis of multi-source logs for anomaly detection on network-based system, 2017 13th IEEE Conf. on Automation Science and Engineering (CASE). Xi'an, China, 2017, pp. 1136-1141.
- [36] Agresti A. An introduction to categorical data analysis. Wiley, 3rd ed., 2018.
- [37] Huda M, Maseleno A, Atmotiyoso P, et al. Big data emerging technology: insights into innovative environment for online learning resources. *International Journal of Emerging Technologies in Learning (IJET)*, vol. 13, no. 1, pp. 23-36, Jan. 2018.
- [38] San Francisco crime data : <https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmfn-yvry>.
- [39] Chicago crime data : <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>.
- [40] Philadelphia crime data : <https://www.opendataphilly.org/dataset/crime-incidents>.
- [41] Viswanathan V, Viswanathan S. R Data Analysis Cookbook. Packt Publishing Ltd, 2nd ed. USA, 2015, pp.30-39.
- [42] Boix R, Hervás - Oliver J L, et al., Micro - geographies of creative industries clusters in Europe: From hot spots to assemblages. *Papers in Regional Science*, vol. 94, no. 4, pp. 753-772, Jan. 2015.
- [43] Krizan Z, Herlache A D. Sleep disruption and aggression: Implications for violence and its prevention. *Psychology of Violence*, vol. 6, no. 4, pp. 542-552, Oct. 2016.
- [44] US population density by city: <http://www.governing.com/blogs/by-the-numbers/most-densely-populated-cities-data-map.html>.
- [45] da Silva I.N., Hernane Spatti D., et al.. Introduction. In: *Artificial Neural Networks*. Springer, Cham, pp. 3-19, Aug. 2017.
- [46] Hyndman R J, Athanasopoulos G. *Forecasting: principles and practice*. OTexts, 2nd ed., 2018, pp. 333-339.
- [47] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222-2232, July 2017.
- [48] Lai J, Chen B, Tan T, et al. Phone-aware LSTM-RNN for voice conversion, 2016 IEEE 13th Int. Conf. on Signal Processing (ICSP). Chengdu, China, 2016, pp. 177-182.
- [49] Schober P, Boer C, Schwarte L A. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763-1768, May 2018.
- [50] American cities with the worst income inequality: <https://www.cbsnews.com/media/9-american-cities-with-the-worst-income-inequality/>.
- [51] Lofstrom M, Raphael S. Crime, the criminal justice system, and socioeconomic inequality. *Journal of Economic Perspectives*, vol. 30, no. 2, pp. 103-26, Mar. 2016.
- [52] Zhou L , Pan S , et al. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, vol. 237, pp. 350-261, May 2017.
- [53] Injadat M N , Salo F , Nassif A B . *Data Mining Techniques in Social Media: A Survey*. *Neurocomputing*, vol. 214, pp. 654-670, Nov. 2016..
- [54] Gao Y, Xia Y, Wu S, et al. Solution to gang crime based on Graph Theory and Analytical Hierarchy Process. *Neurocomputing*, vol. 140, pp. 121-127, Sep. 2014.
- [55] Zheng S, Chen S, Yang L, et al. Big Data Processing Architecture for Radio Signals Empowered by Deep Learning: Concept, Experiment, Applications and Challenges. *IEEE Access*, vol. 6, pp. 55907-55922, Sep. 2018.
- [56] Zhao J, Gao Y, Qu Y, et al. Travel time prediction: Based on gated recurrent unit method and data fusion. *IEEE Access*, vol. 6, pp. 70463-70472, Oct. 2018.
- [57] Niu K, Zhang H, Zhou T, et al. A Novel Spatio-Temporal Model for City-Scale Traffic Speed Prediction. *IEEE Access*, vol. 7, pp. 30050 - 30057, Feb. 2019.
- [58] Peral J, Ferrández A, Mora H, et al. A Review of the Analytics Techniques for an Efficient Management of Online Forums: an Architecture Proposal. *IEEE Access*, vol. 7, pp. 12220 - 12240, Fewb. 2019.
- [59] Stopar L, Skraba P, Grobelnik M, et al. Streamstory: exploring multivariate time series on multiple scales. *IEEE transactions on visualization and computer graphics*, vol. 25, no. 4, pp. 1788-1802, Apr. 2019.
- [60] Guo L, Cai X, Hao F, et al. Exploiting fine-grained co-authorship for personalized citation recommendation. *IEEE Access*, vol. 5, pp.12714-12725, June 2017.