**CHAPTER 1**

**INTRODUCTION**

1. **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. 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. 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. 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. 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. 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 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.

1. **About the Seminar Topic**

This topic focus on the BDA technique that has been applied to process the crime reports of the data collected for three years in San Francisco, Chicago, and Philadelphia. This data after analysis is helpful to obtain patterns following statistical analysis and visualization, some interesting facts and patterns are discovered. Several the state-of-the art data mining and deep learning techniques are used.

1. **Outline of Seminar**

This seminar throws light on how series of state-of-the-art big data analytics and visualization techniques are the effective ways to find statistical visualization and identify patterns of criminal data of three US cities. Results obtained will provide new insights into crime trends and will assist both police departments and law enforcement agencies in their decision making. The Prophet model, a neural network model, and the deep learning algorithm Long Short Term Memory (LSTM) are used and show better perform than neural network models.

The outline of this topic 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. 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.

**CHAPTER 2**

**LITERATURE SURVEY**

The related work presents the literature review about different data mining techniques and analysis methodologies of some of the research papers. It contains other attempts to solve same problem. In this there are Big Data Analytics technique to obtain trends in crime reports and related attributes of criminal data. This helps as a basis and reference for continuing further work.

1. **Description of models**

**A Survey of Data Mining Techniques for Analyzing Crime Patterns**

The data mining is data analyzing techniques that used to analyze crime data previously stored from various sources to ﬁnd patterns and trends in crimes. In additional, it can be applied to increase efﬁciency in solving the crimes faster and also can be applied to automatically notify the crimes. However, there are many data mining techniques. In order to increase efﬁciency of crime detection, it is necessary to select the data mining techniques suitably. This paper throws light on research gaps and challenges of crime data mining and provides insight about the data mining for ﬁnding the patterns and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining. It presents some of the analyzing techniques that are used in data mining. They are Association Rule Mining, Clustering, and Classiﬁcation. Crime are characterized which change over time and increase continuously. The changing and increasing of crime lead to the issues of understanding the crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Research interests have tried to solve these issues. However, these researches still have gaps in the crime detection accuracy. This leads to the challenges in the ﬁeld of crime detection. The challenges include modeling of crimes for ﬁnding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time [1].

**Crime Pattern Detection, Analysis & Prediction**

This paper aims to provide a comprehensive review of theory and research with respect to the prevention of the crime in the society and to implement different data analysis algorithms which address the connections between crime and its pattern. The data for the project are collected from the legitimate government sources.The crime patterns are detected by feeding previous year’s record of crimes taken from legitimate online portal of India listing various crimes such as murder, kidnapping and abduction, dacoits, robbery, burglary, rape and other such crimes. A regression model is created and the crime rate for the following years in various states can be predicted. It uses supervised, semi-supervised and unsupervised learning technique on the crime records for knowledge discovery and to help in increasing the predictive accuracy of the crime. This work will be helpful to the local police stations in crime suppression.

Technologies such as Weka Tool and R Tool are used for mining various crime pattern and analysis. The developed model will reduce crimes and will help the crime detection field in many ways that is from arresting the criminals to reducing the crimes by carrying out various necessary measures [2].

**An Visual Analytics Approach to Explore Criminal Patterns Based On Multidimensional Data**

Comprehensive crime data sets, which contain multidimensional numeric attributes, have been collected over time. To understand the pattern and trend of such attributes, this paper presents an interactive system using visual analytics approach to explore the crime data of Fuzhou, China in 2011. Specifically, users could analyze multidimensional data simultaneously in various views and be able to discover high-relevant and valuable information through a variety of combination. This approach is able to analyze and detect the spatiotemporal distributions, changes, patterns and potential impacts of crime events. And it is also designed for different views and supports the analysis by efficient filtering of time and space. It provides a case study using criminal data sets in Fuzhou, China over 2011 to demonstrate the effectiveness of this approach and to provide insights from multidimensional crime data through users’ involvement [3].

**LSTM Networks for the Trend Prediction of Gases Dissolved in Power Transformer Insulation Oil**

Forecasting the accurate dissolved gases content in power transformer oil can provide important basis for transformer condition assessment. In this paper, a new forecasting model based on long short-term memory networks is proposed. The sequence data of gas dissolved in oil are trained by long short-term memory network to explore the forecast characteristic. The characteristics contain the correlations between historical status monitoring information and the gases content in the forecasting time. The development trend of gases content is automatically extracted. The case studies show that the proposed method can effectively forecast gases content dissolved in transformer oil. The model achieves greater forecasting accuracy than grey model, back propagation network model and support vector machine model. It overcomes drawbacks of low stability in traditional methods and shortcoming of considering only one characteristic gas. In view of this, the prediction method of dissolved gas concentration in oil based on correlation is firstly analyzed. Then a prediction model of dissolved gas content in transformer oil based on long and short memory (LSTM) network is proposed [4].

**Systematic Review of Crime Data Analytics**

Big data is the voluminous and complicated assortment of information that comes from completely different sources like sensors, content denote on social media web sites, sale purchase dealings etc. Such voluminous knowledge becomes robust to method victimization ancient process application. There are distinct tools and techniques within the marketplace for big data analytics. With regularly increasing population, crimes and rate analyzing connected knowledge may be a large issue for governments to form strategic elections thus to maintain law and order. This can be extremely necessary to stay the voters of the country safe from crimes. the most effective place to seem up to seek out area for improvement is that the voluminous data that's generated on a daily basis from varied sources by applying Big Data Analytics (BDA) that helps to investigate sure trends that has to be discovered, in order that law and order are often maintained properly and there's a way of safety and well-being among the voters of the country. It provides retrospective read of many methodologies and technologies, significantly big data analytics methodology for enhancing security and criminal investigations [5].

**CHAPTER 3**

**CONCEPTS**

The concepts of State of-the-art machine learning and deep learning algorithms are deployed to forecast trends and obtain optimal models with the highest accuracy. In order to tackle the problem of crime trends forecasting the prediction models are used.

**Prediction Models**

In order to tackle the problem of crime trends forecasting it has been 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. It demonstrates how the amount of crime incidents changes over time, so as to clearly show the potential trend and seasonality in the data as analyzed and discussed in the following.

**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. 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.

**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. 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 system it 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.

**C. LSTM MODEL**

LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies. For time it series involves auto-correlation, i.e. the presence of correlation between the time series and lagged versions of itself, LSTMs are particularly 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

**CHAPTER 4**

**ANALYSIS**

The analysis of the researches in predicting the trends and forecasting it by training huge unstructured data using different techniques so as to obtain insights into it for decision making awakens to specifically distinguish each technique uniquely. The Big Data Analytics therefore aid to train and explore new trends from voluminous data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| References | Techniques | Tasks | Gaps | Research Challenges |
| A Survey of Data Mining Techniques for Analyzing Crime Patterns[1] | Apriori algorithm | Extracting elements from data sources and analyzing the crime patterns | False detection | Improve precise detection |
| Association analysis concept | Analyzing the crime patterns | No concerning with solving processing time and visualization | Improve performance of processing time and visualization |
| Neural network-based entity extractor | Collecting and extracting the data obtained from police reports | No crime model and visualization | Model the crime future attacks and improve visualization |
| An enhanced Decision Tree Algorithm | Classifying e-mails in relation to crime activities | No data collection, crime prediction and crime model | Collect data and model the crime future attacks |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Crime Pattern Detection, Analysis &  Prediction [2]. | Association Mining (Aprori) along with Clustering(k-Means) Naive Bayes classification is used. | The data sets containing crime reports and other attributes related to crime are collected. | Data acquisition and staging of huge data | Generating the crime hot-spots that will help in deployment of police at most likely places of crime for any given window of time |
| An Visual Analytics Approach to Explore Criminal Patterns based on Multidimensional  Data [3]. | visual analytics method is used, which helps users to attain useful insights from multidimensional data sets | The criminal data set that includes criminal events with multidimensional attributes are collected | There exists information complexity and display complexity. | To make interactive visual analytics to explore the criminal information on multidimensional data sets. |
| LSTM Networks for the Trend Prediction of Gases Dissolved in Power Transformer Insulation Oil [4] | Long short-term memory network is used to explore the forecast characteristic. | The sequence data of gas dissolved in oil are trained by long short-term memory network | Prediction accuracy is moderate. | To improve precession detection |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Systematic Review of Crime Data Analytics [5]. | Data analytics techniques such as text mining, neural network, Intelligent agent are used | It involves collecting crime reports , classification and presentation | No when data for analysis is less | Data handling and pre processing |

Table 4.1: Comparative study of different Big Data Analytical techniques

The analysis of Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data, in this the three crime datasets are 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 Data from Chicago has a total number of 5,541,398records, dating back from 2017 to 2003.In the Philadelphia dataset, there are 2,371,416 crime incidents which were captured from 01/01/2006 to 12/31/2017. Detailed analysis of these dataset is presented by Featured attributes, Data preprocessing and Narrative visualization.

**CHAPTER 5**

**Applications**

The big data analytics in forensic science has many types of applications in criminal justice system from profiling criminals to locating data and presenting evidenced based arguments to jurors at trial. They are listed as follows:

* It helps in studying repeated offenders and repeated victims.
* It helps in analyzing social media data.
* Looking into criminal history.
* Analyzing communications.
* Observing hot spots.
* Prediction of crime and measures can be taken to prevent it.
* The data mining techniques can be applied to increase efﬁciency in solving the crimes faster and also can be applied to automatically notify the crimes.
* For the detection and prevention of cybercrimes.
* It helps to generate the crime hot-spots that will help in deployment of police at most likely places of crime.
* The use of big data analytics will be helpful in identifying criminal activities and to

Predict future terrorism incidents.

* Many application domains such as healthcare, transportation systems environmental monitoring for real time decision making and control.

**CHAPTER 6**

**Conclusion**

Big data analytics and the different data mining techniques used to analyze data, predict trends help to deal with velocity, veracity, voluminous and variety of data. In this the BDA is used to analyze the criminal reports this analysis of crime data provides insights into the crimes and helps in prediction of suspects. The statistical analysis and visualization technique helps to find some interesting facts and patterns to be discovered from criminal data. These promising outcomes will beneﬁt for police departments and law enforcement organizations to better understand crime issues. By analysis the Prophet model, a neural network model, and the deep learning algorithm LSTM, it is found that both the Prophet model and the LSTM algorithm perform better than conventional neural network models. The BDA has been the significant technique to be used to unearth the hidden patterns and helps in analysis and decision making and this technique used in forensic science helps in crime prediction, prevention and punishment of criminals. The future work in this helps to become Geospatial in Crime Pattern Detection that is the system could be enriched to support crime mapping over India through which terrestrial model can be created declaring various crimes and the degree of such crimes performed. This terrestrial model would help us to compare the various crime rates in the diverse states of India and to cultivate new strategies to abate the crime rate in that particular area.

**REFERENCES**

[1] U. Thongsatapornwatana, ‘‘A survey of data mining techniques for analyzing crime patterns,’’ in Proc. 2nd Asian Conf. Defence Technol., Chiang Mai, Thailand, 2016, pp. 123–128.

[2] S. Yadav, A. Yadav, R. Vishwakarma, N. Yadav, and M. Timbadia, ‘‘Crime pattern detection, analysis & prediction,’’ in Proc. IEEE Int.Conf.Electron.,Commun.Aerosp.Technol.,Coimbatore,India,Apr.2017, pp. 225–230.

[3] Daichao Lia, Yingjie Wanga, Sheng Wu, Junhui Qia, Tingting Wang, “An visual analytics approach to explore criminal patterns based on multidimensional Data,” Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, A11 Datun Rd,Chaoyang District, Beijing China

[4] J. Dai, G. Sheng, X. Jiang, and H. Song, ‘‘LSTM networks for the trend prediction of gases dissolved in power transformer insulation oil, ’’in Proc. 12th  Int. Conf. Properties Appl. Dielectr. Mater, Xi’an, China, 2018, pp. 666–669.

[5] Sheena Rewari, Dr. Williamjeet Singh, “Systematic Review of Crime Data Analytics,” in

IEEE, International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017).

We also explored crime

trend using time series model. In the future, we plan to add more features that related to

crime incidents and using more state-of-the-art algorithms to get higher accuracy for

multi-class classiﬁcation problem. Besides, as Deep Learning [14–16] and Artiﬁcial

Intelligence [17,18] has achieved big success in computer vision, we also want to

perform more deep learning models and other learning algorithms [19,20] on our

datasets.

We also explored crime

trend using time series model. In the future, we plan to add more features that related to

crime incidents and using more state-of-the-art algorithms to get higher accuracy for

multi-class classiﬁcation problem. Besides, as Deep Learning [14–16] and Artiﬁcial

Intelligence [17,18] has achieved big success in computer vision, we also want to

perform more deep learning models and other learning algorithms [19,20]