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

Madhuri C A[1], Nalini J[2], Meghala S[3], Navya S[4]

[1]*Assistant Professor, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India*
[2-4]*Students, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India*

Abstract-*Big Data Analytics (BDA) is a structured approach for examining and recognizing different patterns relations, and trends within a huge amount of information. The BDA is applied to criminal data where preliminary data analysis is conducted for visualization and trends prediction. Various data mining and deep learning techniques are used. The predictive results show the Keras stateful LSTM perform better than neural network models, where the optimal size of training data are found to be three years. The system shows the graphs as end result using the LSTM model which shows the time series graph. These end result will be helpful for police departments and law enforcement organization to keep track of activities, predict the chance of incidents, effectively locate resources and enhance the decision-making process.*

Key Words: LSTM Model, Deep Learning, Neural Network Model

1. INTRODUCTION

Big data analytics refers to the process of analyzing huge amount of data. This big data is gathered from a wide variety of machines learning repository. Due to uninterrupted civilization and increasing populations, cities play a major role in our society. However, such developments have also been accompanied by 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 relevancy public policy, however there are many challenges in coping with large amounts of accessible data. As a result, new methods and technologies need to be devised to analyze this heterogeneous and multi-sourced data. Analysis of huge amount data helps us to keep track of the crimes. 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 rising of cloud computing and data acquisition, and storage technologies, from business, and research institutions to governments and various organizations, have led to many unprecedented scopes/complexities from data that have been collected and made publicly available It has became increasingly important to extract meaningful information and achieve new insights for understanding patterns from such data resources. BDA can effectively address the challenges that are too vast, too unstructured, and too fast- paced to be managed by traditional methods as a quick growing and influential practice, BDA can aid organizations to utilize their data and facilitate new opportunities. Furthermore, BDA may be deployed to assist intelligent businesses move ahead with more practical operations, high profits and satisfied customers.

Fig:- 1: Two example of time series plot of crime incident in two cities

2. METHODOLOGY

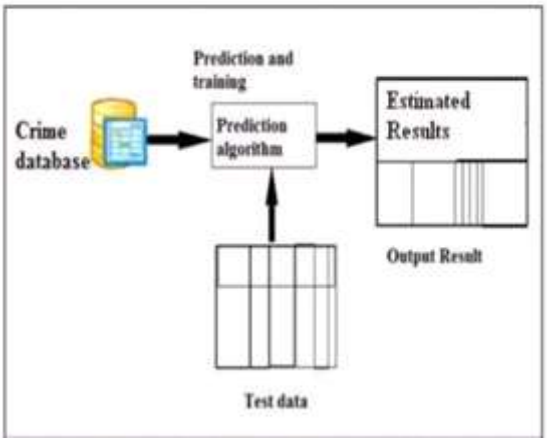

Fig:- 2.1: System Architecture

The figure above shows the construction of the architecture of the system. The prediction algorithm rn allerxlule, which includes preprocessing, extraction and classification of the data. The report consists of a variety of report on violence. System can break into research and instruction, and identify patterns in crime.

2.1 Data Preprocessing

In Machine Learning data preprocessing may be crucial step that helps enhance the standard of knowledge to push the extraction of meaningful insights from the info. It's the process of transforming of raw data into understandable format. Typically, real-world data is incomplete, inconsistent, inaccurate, and often lacks specific attribute values/trends.

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

2) For some missing coordinate attribute, we assign random values sampled from the non-missing values, computed their mean, and then replaced the missing one.

2.2 Feature Extraction and classification

Feature extraction is initial step of measured data and constructs the derived data which is meaningful and non- redundant. When the input data to an algorithm is too vast to be processed and it may contain redundant data, then such data must be transformed into a reduced set of feature. The given data is classified into two types of classes i.e., structured and unstructured data. After the classification process the mapping function is approximated from the given input variables to discrete output variables. The main objective is to recognize which class the new data will fall into.

2.3 LSTM

LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies. 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 while recognizing patterns over the time series. The recurrent architecture enables the states to be continued, or interfaced between updated weights. Moreover, the LSTM cell architecture and enhance the RNN by enabling long-term persistence in addition to short term.

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

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

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

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

Where, ft is a sigmoid function to indicate whether to keep the previous state, iiCt-1 is the old cell state, Ct is the updated cell state, wf, wi, and wc are the previous values in each layer, ht and xt, is the input value, bf, bi,and bc are constant values, it decides which value will be used to update state, ct stands for the new candidate values

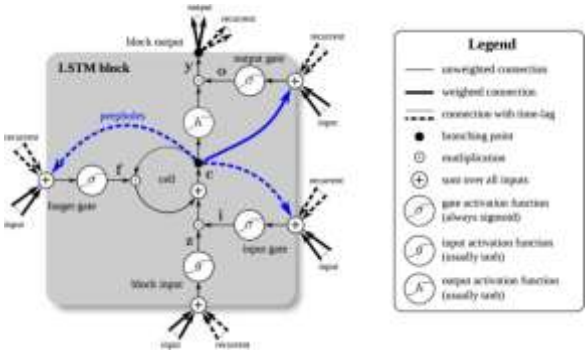

Fig: - 2.3: Schematic unit of LSTM

3. CONCLUSIONS AND FUTURE WORKS

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 cities, which allowed us to identify patterns and obtain trends. By exploring the deep learning algorithm LSTM. The LSTM algorithm act well than conventional neural network models. 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 the future, it is capable of processing different types of data for broad range of application in different platforms. We also plan to incorporate multivariate visualization, graph mining techniques and fine-grained spatial analysis to uncover more potential patterns and trends within these datasets. In addition, our goal is to conduct more realistic case studies to further evaluate the

effectiveness and scalability of the different models in our system.

REFERENCES

[1] Gandomi A, Haider M., Vol. 35, n. 2, pages 137-144, April 2015.

[2] Zakir J, Seymour T, et al., LARGE DATA ANALYTICS. Problemsin Information Systems, Vol. 16, n. 2, pages 81 to 90.2015.

[3] Wang Y, Kung L A et al. An Integrated transformation model enabled for Big Data analysis: Application to Healthcare. Information and management, Vol. 55, n, 1, pages 64-79, January 2018.

[4] Thongsatapornwatana U, A survey on data mining techniques to analyze criminal models. In: 2nd Asian Conf. On Defense Technology, ChiangMai, Thailand, 2016, pp. 123-128.

[5] Raghupathi W, Raghupathi V, Analysis of big data in healthcare: promise and potential. IT and health systems, vol. 2, n. 1, pages 1-10, February 2014.

[6] Archenaa J, Anitha E A M, An investigation of big data analysis in the area of health and government, Procedia Computer Science, vol. 50, pp. 408-413,April 2015.