

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

Durairam R 1st, Dineshkumar M 2nd, Gokul S 3rd, Arjun T 4th,
Dhamu Sri S 5th

*1st Assistant Professor, 2nd, 3rd, 4th, 5th UG Scholar (B.E), Department of Computer Science and Engineering,
Mahendra Institute of Technology, Mahendhirapuri.*

Abstract

Enormous Data Analytics (BDA) is an efficient methodology for examining and recognizing various examples, relations and patterns inside a huge volume of information. In this paper we apply BDA to criminal information where exploratory information examination is directed for representation and patterns expectation. A few best-in-class information mining and profound learning methods are utilized. Following measurable investigation and representation, some intriguing realities and examples are found from criminal information in San Francisco, Chicago and Philadelphia. The prescient outcomes show that the Prophet model and Kera stateful LSTM perform in a way that is better than neural organization models, where the ideal size of the preparation information is discovered to be three years. These promising results will profit for police divisions and law requirement associations to more readily comprehend wrongdoing issues and give bits of knowledge that will empower them to follow exercises, foresee the probability of occurrences, adequately send assets and improve the dynamic cycle. With the help of such strategies, BDA can help us without any problem recognize wrongdoing designs which happen in a specific territory and how they are connected with time. The ramifications of AI and measurable methods on wrongdoing or other enormous information applications, for example, car crashes or time arrangement information, will empower the examination, extraction and comprehension of related examples and patterns, at last aiding wrongdoing anticipation and the executives. separating and standardization, Google maps-based Geo mapping of the highlights are actualized for perception of the measurable outcomes. Different methodologies in machine learning, profound learning, and time arrangement demonstrating are used for future patterns examination. 1) A progression of insightful investigations are directed to investigate and clarify the wrongdoing information in three US urban communities; 2) We propose a novel visual portrayal which is equipped for taking care of huge datasets and empowers clients to investigate, think about, and examine developmental patterns and examples of wrongdoing occurrences; 3) A mix and correlation of various AI, profound learning and time arrangement displaying calculations to foresee patterns with the ideal boundaries, time spans and models.

Keywords: *Crime data Forecast , Visualization of Crime data, Big data analytics for crime data analytics*

Date of Submission: 01-06-2022

Date of acceptance: 13-06-2022

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. 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. 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. 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 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.

II. LITERATURE SURVEY

A Novel Spatio-Temporal Model for City-Scale Traffic Speed Prediction KUN NIU et al. has been proposed. In this paper City-scale traffic speed prediction provides significant data foundation for Intelligent Transportation System (ITS), which enriches commuters with up-to-date information about traffic condition. However, predicting on-road vehicle speed accurately is challenging, as the speed of vehicles on urban roads is affected by various types of factors. These factors can be categorized into three main aspects, which are temporal, spatial, and other latent information. In this paper, we propose a novel spatio-temporal model named L-U-Net based on U-Net as well as Long Short-Term Memory (LSTM) architecture, and develop an effective speed prediction model, which is capable of forecasting city-scale traffic conditions. It is worth noting that our model can avoid the high complexity and uncertainty of subjective features extraction, and can be easily extended to solve other spatio-temporal prediction problems such as flow prediction. The experimental results demonstrate that the prediction model we proposed can forecast urban traffic speed effectively. We propose a novel spatio-temporal prediction model named L-U-Net by utilizing LSTM neural network combined with U-Net architecture. The model can not only capture features both in temporal and spatial dimension for traffic speed prediction, but also extract features without extensive features engineering. Our method can reduce the workload of feature engineering effectively, and we have demonstrated that it can predict traffic conditions in future well across the real dataset.

A Review of the Analytics Techniques for an Efficient Management of Online Forums: an Architecture Proposal J. Pera, A. Ferrández et al. has proposed. In this paper E-Learning is a response to the new educational needs of society and an important development in Information and Communication Technologies (ICT) because it represents the future of the teaching and learning processes. However, this trend presents many challenges, such as the processing of online forums which generate a huge number of messages with an unordered structure and a great variety of topics. These forums provide an excellent platform for learning and connecting students of a subject but the difficulty of following and searching the vast volume of information that they generate may be counterproductive. The main goal of this paper is to review the approaches and techniques related to online courses in order to present a set of learning analytics techniques and a general architecture that solve the main challenges found in the state-of-the-art by managing them in a more efficient way: (1) efficient tracking and monitoring of forums generated; (2) design of effective search mechanisms for questions and answers in the forums; (3) extraction of relevant Key Performance Indicators with the objective of carrying out an efficient management of online forums. In our proposal, Natural Language Processing, Clustering, Information Retrieval, Question Answering, and Data Mining techniques will be used. The ODL platforms have become very popular in recent years. Forums are a central communication tool in many courses included in online educational platforms. These courses rely mainly on discussion forums for interaction among students. However, the learning advantages that these tools should provide are very often not exploited. Forums do not support learning if many messages are produced, especially when they are posted in a disordered and unstructured way which makes it difficult and time consuming for the user to analyze the information.

Stream Story: Exploring Multivariate Time Series on Multiple Scales Luka Stopar, Primož Skraba et al. has proposed. In this paper This paper presents an approach for the interactive visualization, exploration and interpretation of large multivariate time series. Interesting patterns in such datasets usually appear as periodic or recurrent behavior often caused by the interaction between variables. To identify such patterns, we summarize the data as conceptual states, modeling temporal dynamics as transitions between the states. This representation can visualize large datasets with potentially billions of examples. We extend the representation to multiple spatial granularities allowing the user to find patterns on multiple scales. The result is an interactive web-based tool called Stream Story. Stream Story couples' abstraction with several tools that map the abstractions back to domain-specific concepts using techniques from statistics and machine learning. It is aimed at users who are not experts in data analytics, minimizing the number of parameters to configure out-of-the-box. We use three real-world datasets to demonstrate how Stream Story can be used to perform three main visual analytics tasks: identify the main states of a complex system and map them back to data-specific concepts, find high-level and long-term periodic behavior and traverse the scales to identify which scales exhibit interesting phenomena. We find and interpret several known, as well as previously unknown patterns in these datasets

III. EXISTING METHOD

Manual explanation issue for object discovery in optical RSIs by proposing a feebly directed learning (WSL) system. As perhaps the most cost-effective learning draws near, WSL just requires a frail name for the preparation pictures to indicate whether the picture contains the object of interest or not. To this end, in contrast to customary directed learning draws near, which depend on physically marked bouncing boxes for preparing object identifier, exact areas and sizes of the objective items are not required in the WSL system. Article identification utilizing WSL will in general settle restriction of the objects of interest in every certain preparation picture (programmed comment) and item indicator preparing utilizing programmed explanations (identifier learning) at the same time. In this work, which cannot effectively distinguish between normal and anomalous instances on NVCR Dataset. Distance measures between instances can be challenging when the data are complex. Cannot guarantee critical Crime record information that belongs to a data set may not be at a great “distance” from the other “normal” points, and may end up being classified as “normal. It tends to lie in a less dense area with respect to a intellectual local neighborhood. Identifying critical areas such as the pervious hit points which do not lie at a great “distance” from the rest of the points.

IV. PROPOSED SYSTEM

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 .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 multisource data .Big data analytics (BDA) has been extensively applied and studied in the fields of data science and computer science for quite some time. The conception of big data in BDA, its analytics and the associated challenges when interacting among them. On research gaps and challenges of crime data mining. In additional to that, this project insight about the data mining for finding the patterns 13 and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining.As a consequence, the management and analysis with huge data are very difficult and complex. In order to increase efficiency of crime detection, it is necessary to select data mining techniques suitably. various data mining applications, especially applications that applied to solve the crimesApriori algorithm to find the effective association rule and to reduce the amount of processing time. Additionally, there are several techniques that have been developed in order to analyze associations between two itemset more effectively such as mutual information concept but the algorithm was increased the more amount of time. 3.4

Improves classification accuracy. Filtering imbalance is overcome by fine turning. It can provide to very close to the class boundary and are sensitive to small changes in attribute values. Best accuracy to classify nugget data information’s. High performance. Low time consumption. It can support multidimensional real-world data sets. also helped to reduce the number of false positives and It support false negatives data values and, thus, significantly improving the overall accuracy of classification tasks.

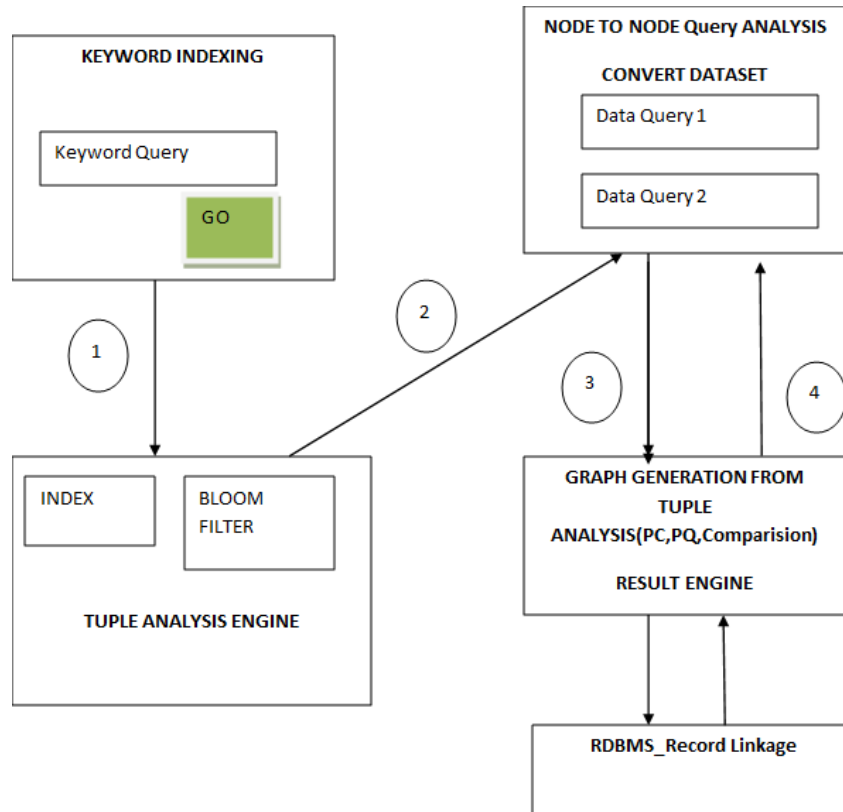


Fig 4.1 Architecture Diagram

V. KEY RESULTS

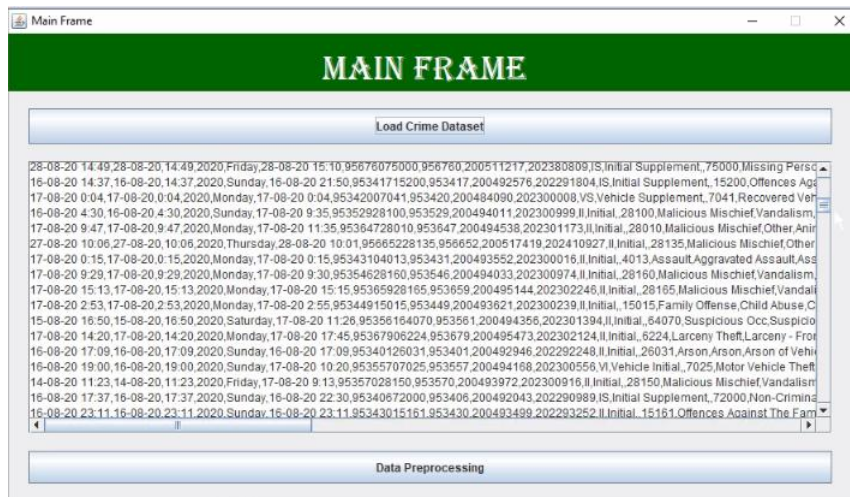


Fig 5.1 Data Preprocessing

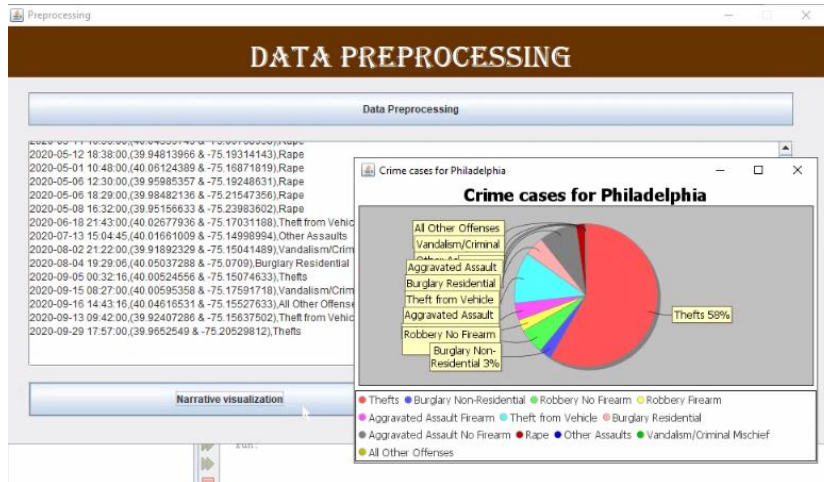


Fig 5.2 Crime data Classification

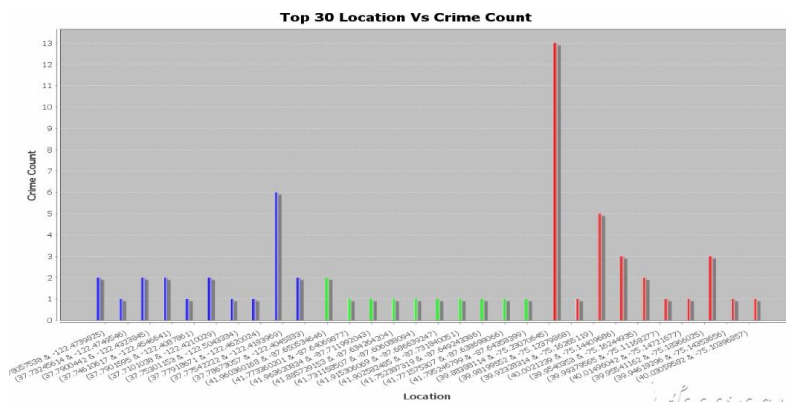


Fig 5.3 Location vs Crime Count



Fig 5.4 Data Loading

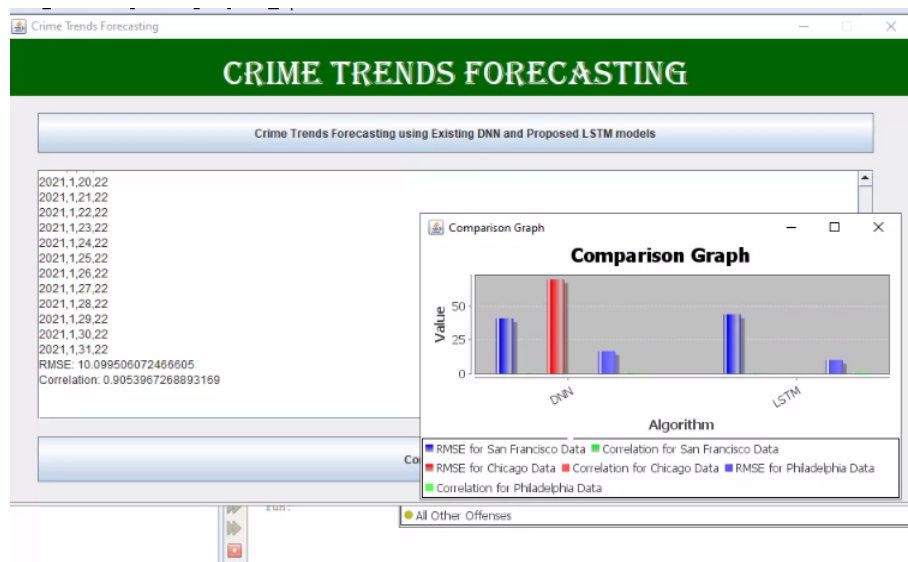


Fig 5.4 Crime Analysis

VI. CONCLUSION AND FUTURE ENHANCEMENTS

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 visualizationgraph mining techniques and fine grained spatial analysis 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.

REFERENCES

- [1]. 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.
- [2]. 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, Feb. 2019.
- [3]. Stopar L, Skraba P, Grobelsnik 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.
- [4]. 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.
- [5]. 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.
- [6]. Musa S., Smart Cities-A Road Map for Development. IEEE Potentials, vol. 37, no. 2, pp. 19-23, Mar. 2018.
- [7]. 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.
- [8]. 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.
- [9]. 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.
- [10]. 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.