iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57384

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

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

Nagaravya Marasu[1], Dr. M. Arathi[2]
JNTU College OF Engineering, India

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 the 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 organizations 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.***

Keywords: Visualization, BDA-Big Data Analytics, LSTM-Long Short Term Memory, Optimize, Predictive

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. In relation to public policy however, there are many challenges in dealing with large amounts of available data. As a result, new methodsand technologies need 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 ensuringimproved 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 ofunprecedented 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 resource

II. LITERATURE SURVEY

1) Beyond the Hype: Big Data Concepts, Methods, and Analytics

Size is the first, and at times, the only dimension that leaps out at the mention of big data. This paper attempts to offer a broader definition of big data that captures its other unique and defining characteristics. The rapid evolution and adoption of big data by industry has leapfrogged the discourse to popular outlets, forcing the academic press to catch up. Academic journals in numerous disciplines, which will benefit from a relevant discussion of big data, have yet to coverthe topic. This paper presents a consolidated description of big data by integrating definitions from practitioners and academics. The paper's primary focus is on the analytic methods used for big data. A particular distinguishing feature of this paper is its focus on analytics related to unstructured data, which constitute 95% of big data. This paper highlights the need to develop appropriate and efficient analytical methods to leverage massive volumes of heterogeneous data in unstructured text, audio, and video formats. This paper also reinforces the need to devise new tools for predictive analytics for structured big data. The statistical methods in practice were devised to infer from sample data. The heterogeneity, noise, and the massive size of structured bigdata calls for developing computationally efficient algorithms that may avoid big data pitfalls, suchas spurious correlation.

©IJRASET: All Rights are Reserved

2) An Integrated Big Data Analytics-Enabled Transformation Model Application To Health Care

Big data analytics enabled transformation model based on practice-based view is developed which reveals the causal relationships among big data analytics capabilities, IT-enabled transformation practices, benefit dimensions and business value. This model was then tested in healthcare setting.

Through analyzing big data implementation cases, we sought to understand how big data analytics capabilities transform organizational practices, thereby generating potential benefits. In addition to conceptually defining four big data analytics capabilities, the model offers a strategic view of big data analytics. Three significant path-to-value chains were identified for healthcare organizations by applying the model, which provides practical insights for managers.

3) A Survey of Data Mining Techniques For Analyzing Crime Patterns

In recent years 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 reviews the literatures on various data mining applications, especially applications that applied to solve the crimes. Survey also throws light on research gaps and challenges of crime datamining. In additional to that, it 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 datamining.

4) A Survey of Big Data Analytics In Healthcare And Government

It is an insight of how it can uncover additional value from the data generated by healthcare and government. Large amount of heterogeneous data is generated by these agencies. But without proper data analytics methods these data became useless. Big Data Analytics using Hadoop plays an effective role in performing meaningful real-time analysis on the huge volume of data and able to predict the emergency situations before it happens. It describes about the big data use cases in healthcare and government.

5) Platforms For Big Data Analytics: Trend Towards Hybrid Era

The primary objective of this paper is to present detailed analysis of various platforms suitable for Big Data processing. In this paper, various software frameworks available for Big Data analytics are surveyed and in-detail assessment of their strengths and weaknesses is discussed. In addition to this, widely used data mining algorithm are discussed for their adaptation for Big Data analysis

w.r.t their suitability for handling real-world application problems. Future trends of Big Data processing and analytics can be predicted with effective implementation of these well established and widely used data mining algorithms by considering the strengths of software frameworks and platforms available. Hybrid approaches (integration of two or more platforms) may be more appropriate for a specific data mining algorithm and can be highly adaptable as well as perform real-time processing.

III. PROPOSED METHODOLOGY

BDA can effectively address the challenges of data that are too vast, too unstructured, and too fast moving to be managed by traditional method. As a fast-growing and influential practice, DBA can aid organizations to utilize their data and facilitate new opportunities.

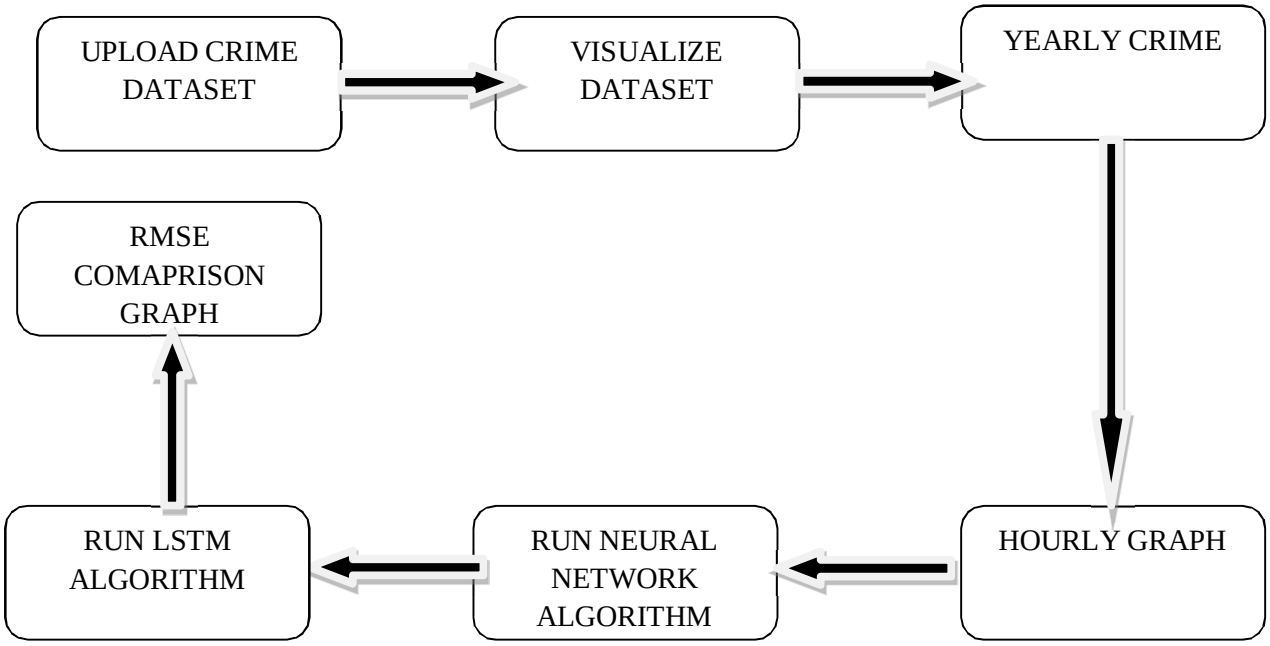
Furthermore, BDA can be deployed to help intelligent businesses move ahead with more effective operations, high profits and satisfied customers.

Consequently, BDA becomes increasingly crucial to organizations to address their developmental issues. In this project I am implementing LSTM and Neural Network and then comparing RSME (root mean square error) between them.

A. Advantages Of Proposed System

1) High accuracy
2) More effective
3) High profits and satisfied customers.

B. Architecture



IV. MODULES OF WORK

To carry out the aforementioned project, we created the modules listed below.

1) Data Exploration: we will put data into the system using this module.
2) Processing: we will read data for processing using this module.
3) Using this module, data will be separated into train and test groups.
4) Model generation: Neural network model and long short-term memory (LSTM).
5) Output screens: Click on the output data to ensure the require

V. IMPLEMENTATION

Using big data, visualizing the useful information and then using Deep Machine LearningAlgorithms such as LSTM (Long short-term memory) and neural network, can forecast the crime for next year. Using this prediction police people can take necessary decision on time. Here I am using Neural Network, LSTM to predict crimes but LSTM is giving better performance.
In this project I am implementing LSTM and Neural Network and then comparing RSME (root mean square error) between them.

A. Neural Network Model

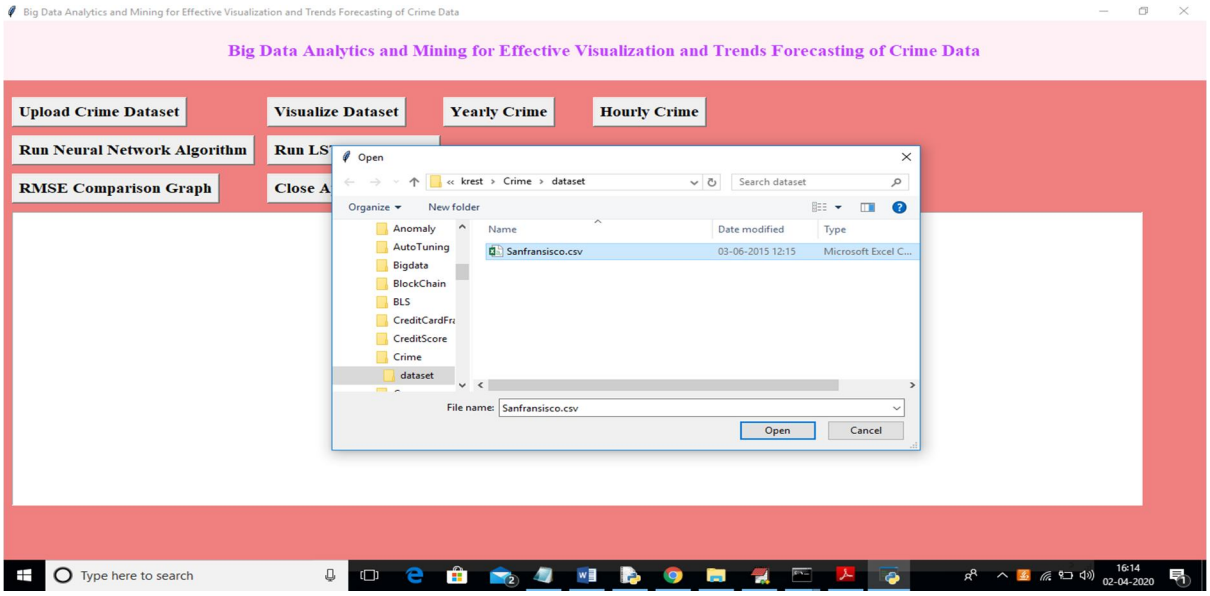
A neural network is composed of a certain number of neurons, namely nodes in the network, whichare organized in several layers and connected to each other cross different layers. There are at least three layers in a neural network wide, the input layer of the observations, a non-observable hiddenlayer in the middle, and an output layer as the predicted results. In this project I 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.

B. LSTM Model

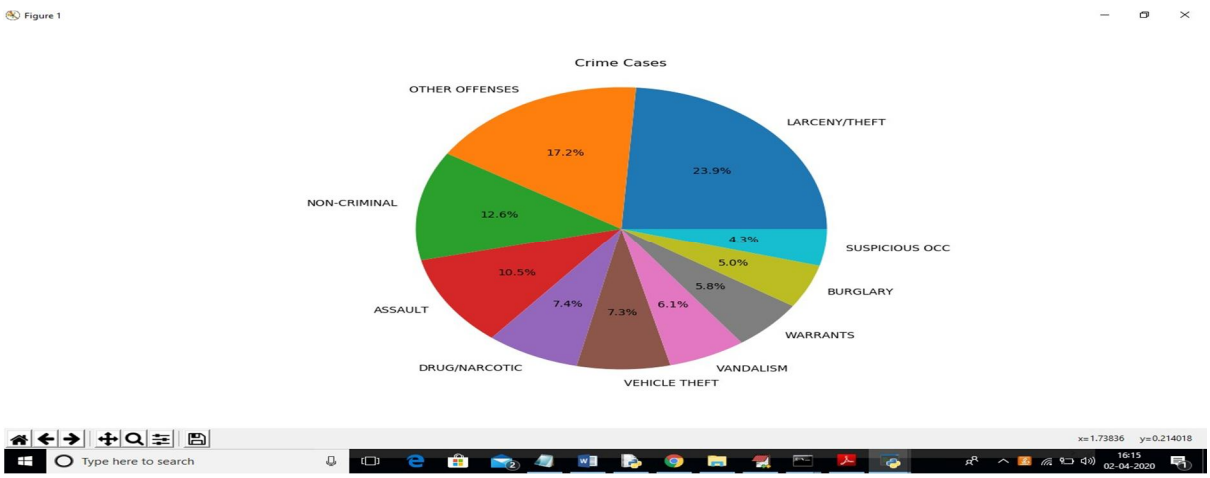
LSTM model is a powerful type of recurrent neural network (RNN), capable of learning long-term dependencies. For time series involves autocorrelation. the presence of correlation between the time series and lagged versions of itself, LSTMs are particular useful in prediction due to their capability of maintaining the state while recognizing patterns over the time series. The recurrent architecture enables the states to be 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.

VI. EXPERIMENTAL RESULTS

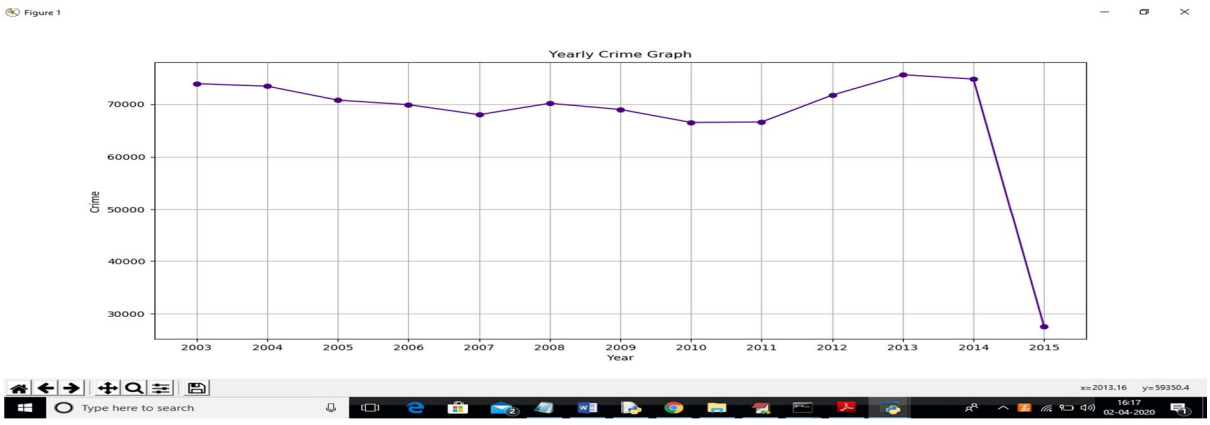
'Upload Crime Dataset' button and upload dataset



Now click on 'Visualize Dataset' button to view top 10 crimes



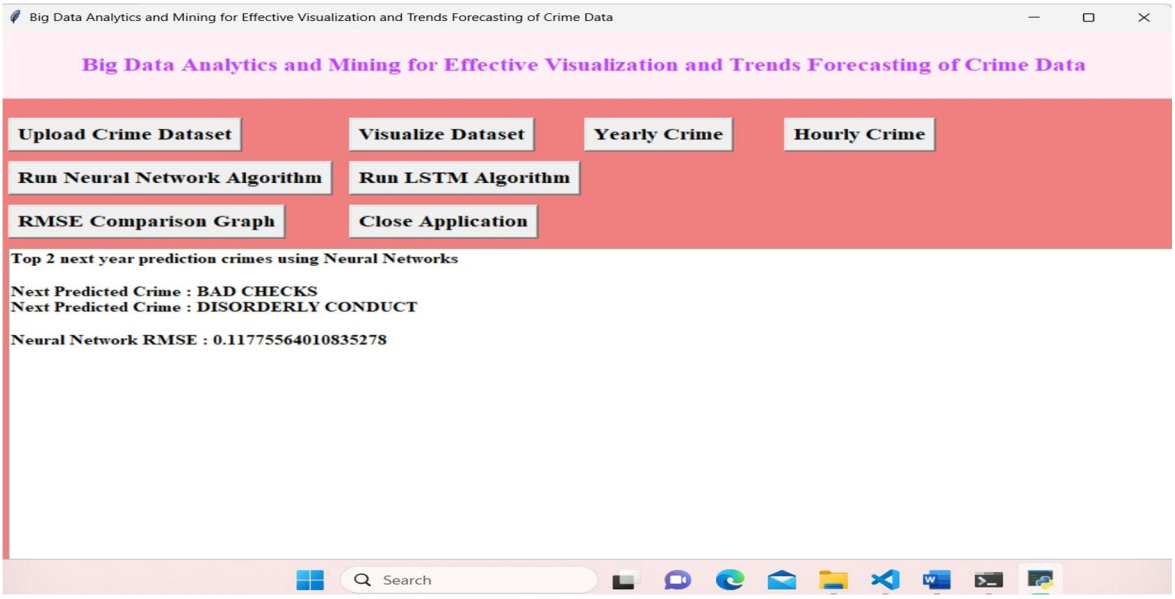
'Yearly Crime' to visualize below graph



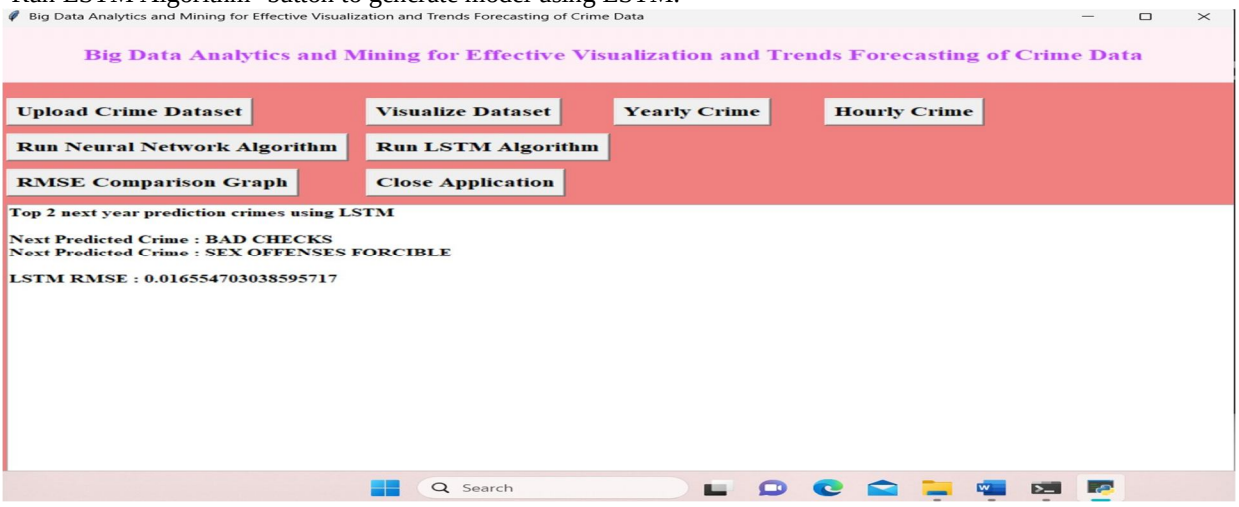
Hourly Graph' button to visualize below graph



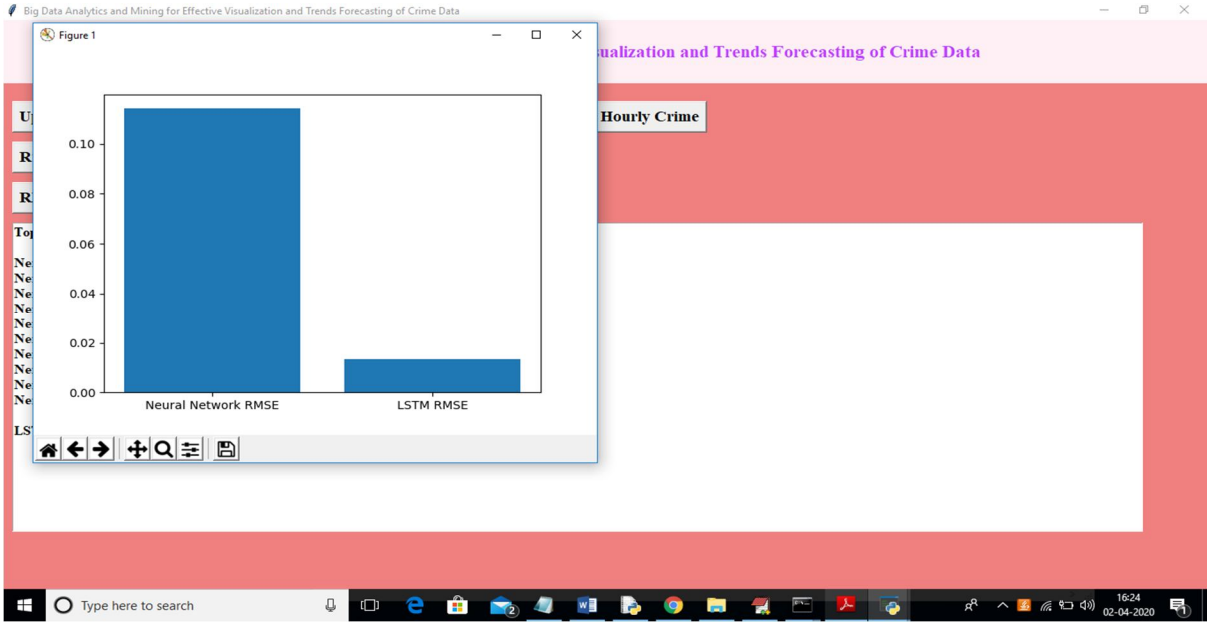
After training model we can get next year top 2 crime prediction in below screen



click on 'Run LSTM Algorithm' button to generate model using LSTM.



Now click on 'RMSE Comparison Graph' button to get RMSE comparison graph between LSTMand neural network.In Below graph x-axis represents algorithm name and y-axis represents RMSE value. From abovegraph we can say LSTM got less RMSE so it performance is better.

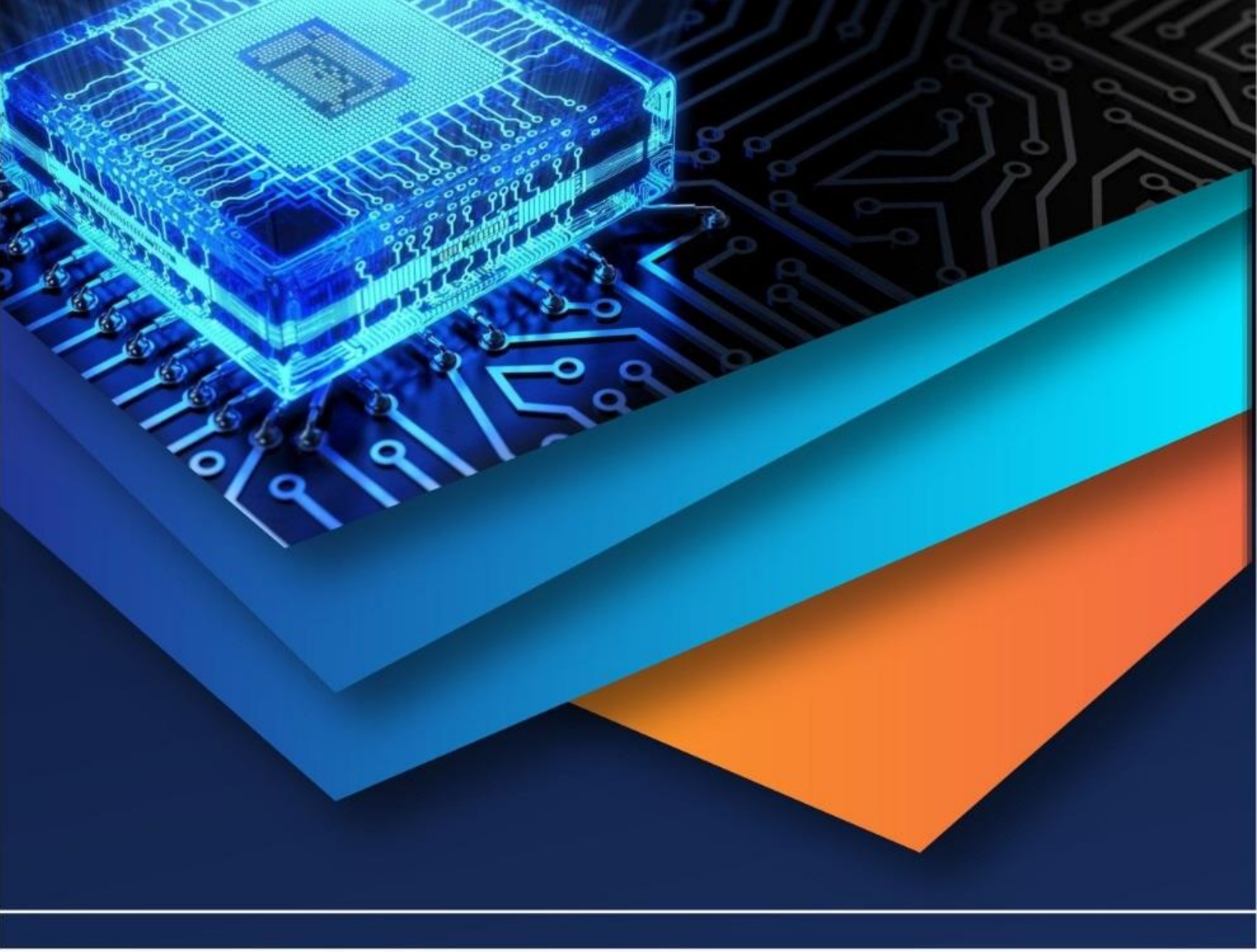


VII. CONCLUSION

In this project a series of state-of-the-art big data analytics and visualization techniques were utilizedto analyze crime big data from three US cities, which allowed us to identify patterns and obtain trends. By exploring the neural network model, and the deep learning algorithmLSTM, I found that the LSTM algorithm perform better than conventional neural network models and also the optimal time period for the training sample to be 3 years, in order to achieve the best prediction oftrends in terms of RMSE and spearman correlation. Optimal parameters for the LSTM model 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 makin

REFERENCES

[1] A. Gandomi and M. Haider, ''Beyond the hype: Big data concepts, methods, and analytics,'' Int. J. Inf. Manage., vol. 35, no. 2, pp. 137–144, Apr. 2015.
[2] J. Zakir and T. Seymour, ''Big data analytics,'' Issues Inf. Syst., vol. 16, no. 2, pp. 81–90, 2015.
[3] Y. Wang, L. Kung, W. Y. C. Wang, and C. G. Cegielski, ''An integrated big data analytics- enabled transformation model: Application to health care,'' Inf. Manage., vol. 55, no. 1, pp. 64– 79, Jan. 2018.
[4] 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.
[5] W. Raghupathi and V. Raghupathi, ''Big data analytics in healthcare: Promise and `potential,'' Health Inf. Sci. Syst., vol. 2, no. 1, pp. 1–10, Feb. 2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)