**ABSTRACT**

Big data analytics (BDA) is a method for analyzing and discovering patterns, relationships, and trends in massive amounts of data. We use BDA to criminal data for this research, with exploratory data analysis for visualization and trend prediction. A variety of splitting data mining and deep learning techniques are employed. Some intriguing facts and patterns are revealed from criminal data in the United States after statistical analysis and visualization. These promising results will help police departments and law enforcement agencies better comprehend crime issues and provide insights that will allow them to analyze activities, estimate the likelihood of incidents, deploy resources effectively, and optimize decision-making processes.

Table of Contents

[**1.** **Introduction to Big Data** 1](#_Toc101164418)

[**2.** **Introduction to the Dataset** 2](#_Toc101164419)

[**2.1 Problem Formulation:** 2](#_Toc101164420)

[**2.2** **Aims and Objectives** 3](#_Toc101164421)

[**3.** **Related Works** 4](#_Toc101164422)

[**A. Big Data Analytics** 4](#_Toc101164423)

[**4. Methodologies** 5](#_Toc101164424)

[**5. Result and Discussion** 6](#_Toc101164425)

[**6. Conclusion** 7](#_Toc101164426)

[References 9](#_Toc101164427)

1. **Introduction to Big Data**

We all have smartphones, but have you ever considered how much data they generate in the form of messages, phone calls, emails, images, videos, searches, music, and other types of data? According to (Department, 2022), in the next several years, global mobile data usage is expected to skyrocket, with an average of 34 gigabytes per smartphone per month in 2026. Consider multiplying this number by 7 billion smartphone users. Isn't that a lot for our minds to comprehend? This vast volume of data is what we refer to as "Big Data," as it is far too much for typical computing systems to handle.

Let's have a look at how much data is generated on the internet every minute. According to (Lewis, 2021) research, 3.4 million snaps are posted on Snapchat, 3.8 million Google searches are made, 1.4 million individuals log on to Facebook, 500 hours of content are uploaded to YouTube, and 197.6 million emails are received, among other things. That's a lot of information, so how do you categorize it as Big Data? Volume, Velocity, Variety, Veracity, and Value are the five V's that make this feasible. (Shuwei, et al., 2020)

To better understand this, consider an example from the healthcare industry article published by (Sasubilli & Kumar, 2020). Hospitals and clinics all across the planet generate massive volumes of knowledge. Petabytes of knowledge are collected a day within the sort of patient records and test results, with the medical domain accounting for 75% of the entire. All of this data is generated at a rapid speed, adding to the data's velocity. Variety refers to the varied data types, like structured, semi-structured, and unstructured data. Examples include Excel spreadsheets, log files, and x-ray images, to call a couple of. The accuracy and trustworthiness of massive data could also be described as its value, and reviewing all of this data will enhance the medical sector by allowing earlier disease detection, better treatment, and reduced expenses.

Now comes the question of how to store and process all of this data. Various frameworks, such as Cassandra, Hadoop, and Spark, are used to do this task.

1. **Introduction to the Dataset**

Criminality is one of the country's most serious social issues, affecting public safety, children's development, and adult socioeconomic standing. Understanding what variables contribute to increased crime rates is crucial for policymakers seeking to reduce crime and improve citizens' quality of life. Understanding how to control crime is critical since, in the United States, exposure to violence and crime has been abnormally high for several decades and, while reducing, remains high. According to **Invalid source specified.**, in the United States in 2020, there were 21,570 reported cases of murder or non-negligent manslaughter.

The dataset is extracted from [data.world](https://data.world/kentaris/1985-2014violentcrimerateuspolicedepartments/workspace/file?filename=1985-2014_Violent_Crime_Rate_US_Police_Departments/Murder_rate.csv). The raw dataset contains 1056 rows of data and 32 columns that reflect the names of the police agencies, the state, and the crime rates for each year between 1985 to 2014. This information allows researchers to look at the overall number of crimes, total crime rates, and average crime rates in each police department overtime. The raw dataset contains 1160 rows of data and 16 columns that reflect the names of the police agencies, the state, and the number of crime rates for each year between 1985 and 2014. This information allows researchers to look at the overall number of crimes, total crime rates, and average crime rates in each police department overtime. The main reason for using this dataset is to see how much crime is prevalent in various US police departments over the past years. Certain trends would be detected, studied, and discussed based on the data analysis and visualization in order to make well-informed judgments so that law and order can be properly maintained and individuals feel comfortable and secure.

## **2.1 Problem Formulation:**

Every organization in today's world has ever-increasing issues that must be met quickly and efficiently. With an ever-increasing population, assessing crime data and making strategic decisions to maintain law and order has become a major concern for governments. This is absolutely required in order to protect the residents of the country from criminal activity. With the increasing in crime rate, it is difficult for the resident to determine safe and danger zones. The enormous raw data that is generated on a regular basis from numerous sources by applying Big Data Analytics is the finest area to look for room for improvement. With an ever-increasing population and crime rates, certain trends must be discovered, examined, and discussed in order to make well-informed judgments so that law and order may be properly maintained and individuals feel safe and secure.

## **Aims and Objectives**

The primary goal of this report is to determine how much crime has increased in various US police departments throughout the years. On the basis of data analysis and visualization, certain trends would be recognized, examined, and discussed in order to make well-informed decisions so that law and order may be properly maintained and individuals feel safe and secure.

The remainder of this report is organized in the following manner. Following a review of related literature in [Section 3](#_3.__Related), The methodologies for trend analysis described in [Section 4](#_4.__Methodology) are deep learning and machine learning. [Section 5](#_5._Result_and) explores into the data processing and visualization approaches used and presents the experimental results, which are followed by some concluding notes in [Section 6](#_6._Conclusion).

Section III

# **Related Works**

## **A. Big Data Analytics**

For a long time, big data analytics (BDA) has been widely used and studied in the domains of data science and computer science. When dealing with massive amounts of big data, machine learning presents both potential and challenges. (Lin, et al., 2018) analyzes contemporary research efforts and provides a complete review of medical big data work, specifically as it relates to chronic diseases and health monitoring. The writers of a book released by (Khanna, et al., 2021) discuss the issues that the healthcare business has, such as gathering, storing, finding, sharing, and evaluating data. Furthermore, the book highlights the issues that data scientists, engineers, and physicians face while using Big Data and provides solutions, with a focus on data repositories, challenges, and concepts for data scientists, engineers, and clinicians. (Monika & Bhat, 2019) are familiar with the societal problem of crime utilizing Apache Pig and Hadoop, which entails finding similar vicious crime episodes selectively using incident-level crime data provided by previous identical crime incidents. (Londhe & Rao, 2017) addressed some of the most extensively used data mining algorithms as well as the many software frameworks available for BDA.

Section IV

# **4. Methodologies**

Section V

# **5. Result and Discussion**

Section VI

# **6. Conclusion**

# References

Department, S. R., 2022. *Statista.com.* [Online]   
Available at: https://www.statista.com/statistics/489169/canada-united-states-average-data-usage-user-per-month/  
[Accessed 10 March 2022].

Khanna, A., Gupta, D. & Dey, N., 2021. *Applications of Big Data in Healthcare.* 1st ed. Delhi, India: Elsevier Wordmark.

Lewis, L., 2021. *allaccess.com.* [Online]   
Available at: https://www.allaccess.com/merge/archive/32972/infographic-what-happens-in-an-internet-minute  
[Accessed 10 March 2022].

Lin, R., Ye, Z., Wang, H. & Wu, B., 2018. Chronic Diseases and Health Monitoring Big Data: A Survey. *IEEE Reviews in Biomedical Engineering,* 11(DOI: 10.1109/RBME.2018.2829704), pp. 275 - 288.

Londhe, A. & Rao, P. P., 2017. Platforms for big data analytics: Trend towards hybrid era. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS),* Issue DOI: 10.1109/ICECDS.2017.8390056, pp. 3235-3238.

Monika & Bhat, A., 2019. An analysis of Crime data under Apache Pig on Big Data. *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* Issue DOI: 10.1109/I-SMAC47947.2019.9032565, pp. 330-335.

Sasubilli, G. & Kumar, A., 2020. *Machine Learning and Big Data Implementation on Health Care data.* Madurai, India, IEEE.

Shuwei, S. et al., 2020. Application and research of Big data analysis in commercial Banks. *2020 International Conference on Big Data and Social Sciences,* Issue DOI: 10.1109/ICBDSS51270.2020.00025, pp. 76-79.