Chapter 3
Analysis of Crime Report by Data Analytics Using Python

G. Maria Jones
Saveetha Engineering College, India

S. Godfrey Winster
SRM Institute of Science and Technology, India

ABSTRACT

The ever-rapid development of technology in today's world tends to provide us with a dramatic explosion of data, leading to its accumulation and thus data computation has amplified in comparison to the recent past. To manage such complex data, emerging new technologies are enabled specially to identify crime patterns, as crime-related data is escalating. These digital technologies have the potential to manipulate and also alter the pattern. To combat this, machine learning techniques are introduced which have the ability to analyse such voluminous data. In this work, the authors intend to understand and implement machine learning techniques in real time data analysis by means of Python. The detailed explanation in preparing the dataset, understanding, visualizing the data using pandas, and performance measure of algorithm is evaluated.

INTRODUCTION

According to FBI (Federal Bureau of Investigation), crimes are defined as an offensive which involves threat of forces. The violent crimes are comprised with main four forms: murder, rape, robbery and aggravated followed by a property crimes ("Violent Crime - Crime in the United States 2009," n.d.). From traditional days to modern days, crimes have been evolved in various forms like computer crime, computer- related crimes, cyber-crimes and digital crimes. Computer crimes are defined as any crime act committed via computers and when computer involved in criminal act it is known as computer-related crimes. Cybercrime encompassed of criminal act including misuse of computers which are connected to internet. Finally, a relatively new and advanced crime called digital crime where attacks includes un-authorized network access, manipulation and dissemination of sensitive information, theft of data, child

DOI: 10.4018/978-1-7998-2566-1.ch003

pornography, narcotics traffickers and many more. As more number of digital devices evolved when compared to olden days, the data stored on digital gadgets are massive. Initially, hacking was coined by Massachusetts Institute of Technology (USA) in 1960s.

Computer Criminal activities have existed for decades and came to forefront in mid-80's which exposed the vulnerability of system data. In 1986, a German hacker used personal computer and modem for exploiting sensitive information by tapping into a military database (Britz, n.d.). In 1988, a student crippled over 6000 computers and damaged millions of computers by Morris Worm which infected 10% of computers connected in a network by exploiting UNIX operating system security holes. In mid-80's, many Phreakers were involved in manipulation of telecommunication system (Britz, n.d.). The following table.1 showcases the evolution of cyber-attacks starting from 1960. Industrial Big data, Mobile Devices, Internet of Things, Network connected devices and data-driven techniques are enabled and accessed through networks (LAN,WAN, etc.) to accumulate the enormous amount of information from connected machines and turn the big machinery data into actionable information (Zhao et al., 2019). The automation of systems has immense impact as it is proficient in detecting fraudulent activities from huge volumes of data, to design and develop machine learning algorithms to detect fraud. The growing addiction to technology in the present setup is an invitation to attempt cyber-crime and so counter measures are essential to tackle such criminal activities. With the accessibility of latest smartphones with 2G, 3G, 4G and upcoming 5G technologies, the user has the opportunity to communicate and exchange every piece of information using e-services which includes social networks, e-mails, blogs, etc (Méndez, Cotos-yañez, & Ruano-ordás, 2019).

The availability of data sets is enormous and since there is huge volume of data, storage and sophisticated software are at hand, the potential threat incidents are very likely to occur and so data analysis aids in preventing and detecting the crime. Further, collaborative efforts among investigation departments, researchers, and businesses has led to the development of data analytical techniques which have effective accompanying tools, such as variety of programming languages, software programs, applications, etc. To extract useful information from raw data and make appropriate decisions from this data, machine learning techniques have been recognized as a powerful solution. As a branch of Artificial Intelligence (AI), Machine Learning (ML) and deep learning models provide a platform to represent the data, classify and predict data patterns of information processing. About 93,000 fraud cases had been registered in China regarding mobile phones in the form of mails, messages, live chats, calls, social networking posts etc.(Wei, Sunny, & Liu, 2019).

The machine learning algorithms are classified into three types. They are supervised learning, unsupervised learning and reinforcement learning. Since machine learning is an advanced technique, it has the capacity to detect the crime with highest accuracy. In this chapter, the author offers techniques of data analytics methodologies to detect the criminal pattern which can be advantageous for criminal investigation to be effective and also to prevent crime. There are many facts that can allow law enforcement departments to provide and use their sources in crime scene to prevent from manipulating of original data and also helps to monitor the crime. These measures can efficiently prevent and respond quickly to criminal activities (Catlett, Cesario, Talia, & Vinci, 2019). The chapter considered a case study of Chicago, San Francisco and India crime report for analysing and visualizing. All the crime report has been gathered from online repository with more than 70 thousand crime event details including X and Y axis co-ordinates. The experimental results show the effectiveness of achieving great accuracy in ML algorithms.