Chapter 10
Leveraging Machine Learning in Financial Fraud Forensics in the Age of Cybersecurity

Md Ariful Haque
Old Dominion University, USA

Sachin Shetty
Old Dominion University, USA

ABSTRACT

Financial sectors are lucrative cyber-attack targets because of their immediate financial gain. As a result, financial institutions face challenges in developing systems that can automatically identify security breaches and separate fraudulent transactions from legitimate transactions. Today, organizations widely use machine learning techniques to identify any fraudulent behavior in customers' transactions. However, machine learning techniques are often challenging because of financial institutions' confidentiality policy, leading to not sharing the customer transaction data. This chapter discusses some crucial challenges of handling cybersecurity and fraud in the financial industry and building machine learning-based models to address those challenges. The authors utilize an open-source e-commerce transaction dataset to illustrate the forensic processes by creating a machine learning model to classify fraudulent transactions. Overall, the chapter focuses on how the machine learning models can help detect and prevent fraudulent activities in the financial sector in the age of cybersecurity.

DOI: 10.4018/978-1-7998-8386-9.ch010

1. INTRODUCTION

Cybersecurity is one of the biggest concerns because of the growing number of fraudulent activities both online and offline. Online fraud activities are disrupting and manipulating real-time data transmission and thus stealing the credentials which the hackers could use legitimate identity to make online transactions. For example, in e-commerce transactions, the exchange is happening in real-time. Therefore, any attempt to steal and use legitimate credentials to make purchases fraudulently is an online fraud activity. On the other hand, offline fraud activities can be planting malware or spyware in the target system and then collecting the credentials by monitoring the system's message communications. One such example is the recent fraud activity of the Bangladesh Central Bank's SWIFT payment systems (Zetter, 2016). The hackers had stolen the credentials by planting Spywares in the SWIFT system and then making transaction instructions during non-business hours. Thus, there are no questions on the need for developing cyber forensic capabilities in almost all sectors, specifically in the financial sectors, because of massive loss to the institutions and customers.

According to PwC's global economic crime and fraud survey, cybercrime increases 34 percent year over year (Global, 2020). Identity theft accounted for approximately 20.33 percent of the nearly 3.2 million fraud cases reported to the Federal Trade Commission (FTC) in 2019, according to the Ascent (Ascent, 2020). Credit card fraud is the most prevalent form of identity theft, accounting for 41.8 percent of all identity theft reports, according to Ascent's report, "Identity theft and credit card fraud statistics for 2020." In February 2016, the Bangladesh Central Bank was the target of a cyberattack. Hackers used the SWIFT network to send 35 fraudulent instructions to transfer nearly US$1 billion from a Bangladesh Bank account at the Federal Reserve Bank of New York. Five out of thirty-five fraudulent instructions were successful in stealing US$101 million. The investigation into this case is ongoing, and efforts to reclaim the stolen funds keep continuing.

The statistics above reflects the underlying reasons why the institutions are overly concerned to handle cybercrime issues and therefore, investing in developing preventive mechanisms for cybercrime. According to Mckinsey & Company (Hasham, Joshi, & Mikkelsen, 2019), private companies spent approximately $8.2 billion on anti–money laundering controls alone in 2017. Thus, there are always ongoing efforts to develop preventive measures and forensic capabilities to deal with cybercrimes in the financial sectors. In this age of cybersecurity and advanced analytics, machine-learning-based models and preventive mechanisms gain popularity because of the model's accuracy to predict and classify the potential anomalies in cyberspace. In this chapter, we would start with some background discussions on the cybercrime prevention techniques currently in place and then how we can leverage machine