A CASE STUDY ON FINANCIAL FRAUD DETECTION WITH BIG DATA ANALYTICS

A.Suraj Kumar¹, D Govardhan², W Sai Kaushik³, Ch Harshith⁴, U Madhu⁵,

¹Assistant Professor, Department of CSE (Data Science), Nadimpally Satyanaraya Raju Institute of Technology, Sontyam, Visakhapatnam
²Student of Department of CSE (Data Science), Nadimpally Satyanaraya Raju Institute OfTechnology, Sontyam, Visakhapatnam
³Student of Department of CSE (Data Science) Nadimpally Satyanaraya Raju Institute OfTechnology, Sontyam, Visakhapatnam
⁴Student of Department of CSE (Data Science) Nadimpally Satyanaraya Raju Institute OfTechnology, Sontyam, Visakhapatnam
⁵Student of Department of CSE (Data Science) Nadimpally Satyanaraya Raju Institute OfTechnology, Sontyam, Visakhapatnam

Abstract: The financial sector is currently undergoing digital transformation across products, services, and business models. This digitization aims to automate most of the manual financial transactions and other related services. Therefore, detecting fraud in financial transactions has become an important priority for all financial institutions. With modern technology and global communication, fraud has greatly increased and caused great damage. The focus of this paper is to test different approaches to detect fraud on a real data set of financial payment transactions. The dataset is obtained from Kaggle and consists of 6 million event records and 10 features with an event label of “fraudulent” or “non-fraudulent”. These functions are investigated through exploratory data analysis and only 6 are kept for testing, such as payment type, account balance, transaction amount, etc. Two supervised machine learning algorithms, a random forest, and a support vector classifier are used to detect fraudsters transactions. The dataset is large and requires high computing power to process and train machine learning algorithms. Additionally, another challenge is the very uneven distribution between the fraudulent (0.1%) and non-fraudulent (99.9%) classes. This study aims to address both of these issues. To address the class imbalance, oversampling of minority class data using the Synthetic Minority Oversampling Technique (SMOTE) and undersampling of the majority class using random sub-sampling are investigated. Computational efficiency is achieved by implementing Apache Spark, which provides distributed processing for large volumes of data. The best performance is achieved using the random forest algorithm on the oversampled dataset with a precision of 99.95, an F1 score of 0.999, an recall value of 0.999, a geometric mean of 99.9%, and a model training time of 13.9 minutes. This article provides valuable insights into using large-scale, highly imbalanced big data sets to predict and generate financial fraud alerts.

Keywords: Fraud Detection, Big data analytics, Apache Spark

INTRODUCTION

Big Data is a large amount of information collected from various social media, questionnaires, and voluntarily given product purchases. This information is stored in computer databases and analyzed using software designed to process large and complex data sets and draw conclusions faster and faster.

Big data has significantly impacted many sectors of the global economy, such as healthcare, manufacturing, and retail. It is transforming the world and no industry has been left untouched by its immense benefits, and banking is no exception. Like the cloud, the Internet of Things, machine learning, and open banking, big data is one of the financial industry’s favorites. When a customer enters a bank for the first time, he brings with him many possibilities, such as the possibility of becoming a regular customer, the possibility of making appropriate investments, a short-term relationship, or even the possibility of fraud. Banks need to focus on their customers from a 360-degree angle to visualize their behavior patterns, repayment habits, and financial needs.

Banks deal with millions of potential people every day, and for all that, they need data, lots of it. When potential customers come in, banks have to process a lot of potential data. There is no shortage of information in the banking sector. Big Data has emerged as the savior of the banking industry.

With the help of big data, companies providing financial services have changed their operating methods. Big data reduces the risk of fraud detection, enforcement, and portfolio management. This risk reduction, combined with the optimization of a winning strategy, can give
financial services companies a significant competitive advantage. Big data has enabled new strategies for companies dealing with public markets to go beyond simple improvements. As financial systems and products become more complex and sophisticated, they can provide opportunities for fraudsters to commit fraud. To protect against fraud and risk, financial companies must move quickly to big data to detect and prevent evolving and sophisticated fraud schemes.

Private companies and governments recognize the enormous potential of this information to create real value for consumers and increase productivity over time. Government agencies use big data to assess the systemic risk of major financial markets in order to implement safeguards against threats such as bubbles and recessions. At the same time, companies are taking proactive measures to avoid sanctions that could threaten their viability and core business. This systemic change has forced financial firms to evolve or perish. Big data may flow through businesses and economies, but data science is the real game changer.

In India, the government is focusing on digitizing India by connecting all people to ministries through broadband services, regardless of urban or rural services. Demonetization and COVID ushered in a new era that accelerated digitization and paved the way for online banking and e-commerce businesses in India. Since the whole process involves a lot of data, Big Data can be a game changer.

1.1 Fraud Detection in Banking

Fraud detection is the process of determining whether a transaction is fraudulent or not. This can be done in a number of ways, such as analyzing customer behavior or looking for patterns in data that may indicate fraudulent incidents. Bank fraud detection describes the tools and processes banks use to monitor transactions and payments for suspicious activity. If an event or pattern of behavior raises a red flag, the bank's fraud team can intervene.

Banks have been developing their fraud detection capabilities for years. Today, most banks rely on previous-generation machine learning tools programmed to detect certain types of activity. As we will soon see, this fraud detection has its limitations. Frauds in the banking sector:

Payment Fraud
Card Fraud
Merchant acquires fraud

Payment fraud describes a variety of scams, including account hijacking (when someone gains access to a victim's passwords and payment information to make fraudulent purchases) and authorized push payment fraud (when someone uses social engineering to convince a victim to send them money).

Card fraud
Card fraud occurs when a fraudster either steals a person's credit or debit card to make fraudulent purchases or accesses their credit card information and makes fraudulent purchases without the card.

Merchant acquires fraud

The Merchant Acquisition scam describes the types of scams that target merchants, the intermediaries that bridge the banking and business sectors. Such fraud includes transaction laundering and chargeback fraud.

Different ways to detect bank fraud:

Banks monitor transactions and review transaction data to identify patterns of behavior or suspicious activity that may indicate fraud has occurred or is occurring. Ideally, a bank's fraud detection system detects fraud before the money leaves the customer's account. This fraud detection system monitors transactions for an unauthorized activity or access to sensitive information. Transaction monitoring relies on specific tools, techniques, and strategies to detect fraud, including:

Data mining using artificial intelligence to identify patterns in transaction data. Modeling customer behavior using machine learning. Estimating the level of risk using statistical methods such as regression analyzes and probability distributions.

The system then flags suspicious events for manual review. [7]

1.2 The Importance of Big Data Analytics in Terms of Fraud Prevention

As the number of online purchases, payments, and money transfer transactions increases, so do the risks of potential fraud from these transactions. It has been very difficult for companies to process and analyze the huge amount of data generated by these transactions and use it to detect fraud. Here we find an important facilitating tool: big data analysis for fraud detection. The use of big data analytics in some areas of fraud detection offers many advantages. One of the most important points in detecting fraud is that it is a quick, long time to identify suspects from the irregular data generated by these large transactions.

As a result of these long analyses, some events may seem suspicious and misinterpreted. During this assessment process, people are still required, namely manual workloads, to analyze the data and check for suspicious events or misinterpretations.

Using data analyzed with techniques in big data analytics can provide low costs, more accurate and precise detections, optimized workflows, and efficiency of systems and better services to customers.
1.3 DETECTION AND PREVENTION IN THE BANKING INDUSTRY:

Today, the banking sector is facing an acute problem of fraud. The problem is global and no country is completely immune.Fraudsters have become experts at hijacking Internet sessions, stealing customer credentials, and using malware to defraud unsuspecting accounts. Marc Goodman explains in his book "Future Crimes" that "criminals are often the first to take advantage of new technologies and turn their sophistication against their users." According to a report by Financial Fraud UK, financial fraud losses from payment cards, remote banking, and checks accounted for a staggering £768.8 million in 2016, a 2% increase in 2015. Meanwhile, the value of fraud prevented in 2016 was £1.38 billion. Anti-fraud efforts by banks and card companies have helped save up to £6.0 for every £10 of attempted fraud. [1]

![Total 2016 Financial Fraud losses by type](image)

Fig. 1

2 Data analysis software:

One of these options is the use of data analysis software, which in most cases ensures flawless fraud detection. Modern systems allow fraud investigators to analyze business data and check how well the internal control system is working. As a result, they may show events that indicate fraudulent activity or an increased risk of fraud. Various analytical measures can be applied to fight fraud. This varies from the context of individual fraud investigations to the repeatable analysis of financial processes susceptible to criminal activity.

If the risk of fraud is really high, financial and banking institutions can use a continuous or continuous approach to fraud detection. This works particularly well in situations where preventive control is not practical or effective. Most modern financial services companies have increased information management requirements as audit engagement shifts from traditional cyclical behavior to a risk-based, long-term model. Many banks use special transaction monitoring systems to detect fraud. They generally represent native software that requires operator intervention. However, traditional security systems can perform well in real-time fraud detection at individual stores. But this is only the tip of the iceberg.

2.1 ANALYTICAL TECHNIQUES USED TO DETECT FRAUD:

With the advent of advanced fraud analytics, a whole new dimension to fraud detection techniques can be seen. Fraud detection capabilities are enhanced with the performance measurement which helps standardize and maintain control for constant improvement is possible with the use of these tools.

STATISTICAL ANALYTICS TECHNIQUES:

Anomalies must be detected based on the high and low values of the statistical perimeters, which are the indicators of fraud. This method will help you go beyond finding regular frauds and will help detect abnormal ones as well as transactional anomalies.[10]

3 ADVANTAGES:

Fraud analytics helps in the identification of scenarios, new trends, and patterns under which frauds take place. It is not a replacement of the existing rule-based methods, but a build-up of the traditional methods. Deriving value from unstructured data is an unexplored goldmine and fraud analytics helps you attain that.[5,11]

4 AI TECHNOLOGY AND FRAUD PREVENTION:

Artificial intelligence automates data processing, taking away the pressure of manual evaluation which is less likely to be efficient with the information in large quantities. Many kinds of fraud affect directly not only businesses but also their customers. Efficient AI-based fraud prevention systems also uphold the financial firm's image. The clients who have suffered from the repercussions of the fraudulent operations may migrate to other companies.[12]

With the machine learning approach, banks and fintech can actively predict fraud instead of always staying one step back. The advance of deep learning techniques helps financial institutions create a safer environment for their customers.[7]
The company maintains a working model and a challenger model, which continuously evolves as threats change. As the challenger model becomes more effective, it replaces the first model and a new contender appears. Another company, Theta Ray, offers a platform that empowers financial institutions to detect threats such as credit fraud, ATM hacking, money laundering, and cyber attacks. Strategic use of artificial intelligence and machine learning will become an integral part of security policies of banking organizations within a few years. Artificial intelligence can save banks significant money by eliminating complex fraud cases and protecting their brand. At efinext, we offer a variety of programs to develop services to thrive in this market environment. With nearly 300 professionals, we tackle challenges from every angle to quickly understand the data, workflows, compliance requirements, and the math behind securities, trading, and investing.[1]

PREVENTIONS TECHNIQUES OF FRAUD:

4.1 Data mining:

Data analysis is an integral part of transaction fraud detection and has become a major area of research in the last few years. This technique enables classification and data segmentation to find correlations between data sets that help in scammers’ attempts to steal money from banks and other financial institutions.[8]

NEURAL NETWORKS:

TECHNOLOGY IS USED TO PREDICT FRAUD-RELATED DATA THAT CAN BE MAPPED AGAINST FINANCIAL DOCUMENTS AND AUDITS.

MACHINE LEARNING:

ML algorithms are employed to identify previous fraud patterns and also identify the same in future activities and transactions. [6]

4.2 ALERT CUSTOMERS WHEN THERE IS SUSPICIOUS ACTIVITY

Engage your customers in fraud prevention by notifying them whenever there is unusual activity on their accounts. You can ask them to participate or simply notify them automatically if, for example, a credit card payment originates far from their home. Since electronic fraud is so common these days, you may want to offer customers an alert every time their card is charged.[9]

REQUIRE TWO-FACTOR AUTHENTICATION AND CHALLENGE QUESTIONS

Two-factor authentication or challenge questions make fraud very tough, even with stolen credentials. You can claim them when you log in and make high-risk transactions. There are several two-factor authentication methods. If you decide to offer, consider ease of use for legitimate customers.[13]

4.3 DO NOT USE PUBLIC WI-FI

Don’t use public Wi-Fi connections as an open invitation to hackers. Anyone with knowledge can easily exploit information shared over an unencrypted network. They can steal information such as account numbers or login credentials and inject malware into a device connected to a public Wi-Fi network.[4]

CONCLUSION:

Fraud detection and prevention need to be a top priority for any business. A well-designed and implemented fraud detection system can significantly reduce the chances of fraud occurring within an organization. In addition, timely detection of fraud directly impacts the business in a positive way by reducing future potential losses. Effective detection techniques such as AI and statistical data analysis serve as a deterrent to potential fraudsters. As regulatory requirements and compliance demands have grown, it has become extremely important to implement a robust fraud detection and prevention program.[14,15]

References:


https://www.researchgate.net/publication/334768195_Big_Data_for_Fraud_Detection 17 Actionable Fraud Prevention Tips for