IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

A CASE STUDY ON FINANCIAL FRAUD DETECTION WITH BIG DATA ANALYTICS

A.Suraj Kumar[1] ,D Govardhan[2] ,W Sai Kaushik[3] ,Ch Harshith[4] ,U Madhu[5] ,

[1] Assistant Professor, Department of CSE (Data Science), Nadimpally Satyanaraya RajuInstitute of Technology, Sontyam, Visakhapatnam

[2*] Student of Department of CSE (Data Science), Nadimpally Satyanarayana Raju Institute OfTechnology, Sontyam, Visakhapatnam [3*] Student of Department of CSE (Data Science) Nadimpally Satyanarayana Raju Institute OfTechnology, Sontyam, Visakhapatnam [4*]Student of Department of CSE (Data Science) Nadimpally Satyanarayana Raju Institute OfTechnology, Sontyam Visakhapatnam [5*] Student of Department of CSE (Data Science) Nadimpally Satyanarayana Raju Institute OfTechnology, Sontyam, Visakhapatnam

Abstract : The financial sector is currentlyundergoing digital transformation across products, services, and business models. This digitization aims to automate most of the manualfinancial transactions and otherrelated services. Therefore, detectingfraud 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 paymenttransactions. The dataset is obtained from Kaggle and consists of 6 millionevent records and 10 features with anevent 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 highcomputing power to process and trainmachine learning algorithms.Additionally, another challenge is thevery uneven distribution between thefraudulent (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. Computationalefficiency is achieved by implementing Apache Spark, which provides distributed processing forlarge volumes of data. The best performance is achieved using the random forest algorithm on the oversampled dataset with a precisionof 99.95, an F1 score of 0.999, arecall value of 0.999, a geometricmean of 99.9%, and a model trainingtime of 13.9 minutes. This article provides valuable insights into usinglarge-scale, highly imbalanced big data sets to predict and generate financial fraud alerts.

Keywords*: Fraud Detection, Big dataanalytics, Apache Spark*

INTRODUCTION

Big Data is a large amount ofinformation collected from varioussocial 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 drawconclusions 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 thecloud, the Internet of Things,machine learning, and open banking, big data is one of the financial industry's favorites. When acustomer enters a bank for the first time, he brings with him manypossibilities, such as the possibility of becoming a regular customer, the possibility of making appropriateinvestments, a short-termrelationship, or even the possibility of fraud. Banks need to focus on their customers from a 360-degree angle tovisualize their behavior patterns, repayment habits, and financial needs.

Banks deal with millions of potentialpeople 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 inthe banking sector. Big Data has emerged as the savior of the bankingindustry.

With the help of big data, companiesproviding financial services havechanged their operating methods. Bigdata 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 companiesdealing with public markets to gobeyond 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 todetect and prevent evolving andsophisticated fraud schemes.

Private companies and governmentsrecognize the enormous potential of this information to create real value for consumers and increaseproductivity over time. Government agencies use big data to assess the systemic risk of major financialmarkets in order to implement safeguards against threats such as bubbles and recessions. At the same time, companies are taking proactivemeasures to avoid sanctions that could threaten their viability and corebusiness. 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 focusingon digitizing India by connecting all people to ministries through broadband services, regardless of urban or rural services. Demonetization and covid ushered ina new era that accelerated digitizationand paved the way for online bankingand 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 ina number of ways, such as analyzingcustomer 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 theirfraud 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 FraudCard 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 fraudulentpurchases) and authorized pushpayment 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 debitcard to make fraudulent purchases oraccesses their credit card informationand makes fraudulent purchases without the card.

Merchant acquires fraud

The Merchant Acquisition scam describes the types of scams thattarget 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 identifypatterns of behavior or suspicious activity that may indicate fraud has occurred or is occurring. Ideally, a bank's fraud detection system detectsfraud before the money leaves thecustomer's account. This fraud detection system monitors transactions for an unauthorized activityor access to sensitive information.Transaction monitoring relies on specific tools, techniques, andstrategies to detect fraud, including:

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

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

1.2 The Importance of Big DataAnalytics in Terms of Fraud Prevention

As the number of online purchases, payments, and money transfertransactions increases, so do the risksof potential fraud from these transactions. It has been very difficultfor companies to process and analyzethe huge amount of data generated by these transactions and use it to detectfraud. . Here we find an important facilitating tool: big data analysis forfraud detection. The use of

big data analysis in some areas of fraud detection offers many advantages. One of the most important points in detecting fraud is to act quickly. It can take a 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 thisassessment 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 andprecise 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 havebecome experts at hijacking Internet sessions, stealing customer credentials, and using malware to defraud unsuspecting accounts. MarcGoodman 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 astaggering £768.8 million in 2016, a 2% increase in 2015. Meanwhile, thevalue of fraud prevented in 2016 was £1.38 billion. Anti-fraud efforts by banks and card companies havehelped save up to £6.0 for every £10 of attempted fraud. [1]

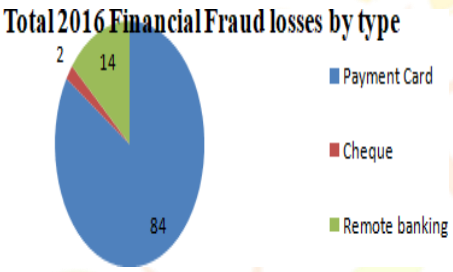

Fig.1

2 Data analysis software:

One of these options is the use of dataanalysis software, which in most cases ensures flawless fraud detection. Modern systems allowfraud investigators to analyzebusiness data and check how well theinternal control system is working. As a result, they may show events that indicate fraudulent activity or anincreased risk of fraud. Variousanalytical measures can be applied tofight fraud. This varies from the context of individual fraud investigations to the repeatableanalysis of financial processes susceptible to criminal activity.

If the risk of fraud is really high, financial and banking institutions canuse 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 thatrequires operator intervention.However, traditional securitysystems can perform well in real-timefraud 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 areenhanced with the performancemeasurement 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 onthe high and low values of the statistical perimeters, which are the indicators of fraud. This method willhelp 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 whichfrauds 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 fraudanalytics helps you attain that. [5,11]

4AI TECHNOLOGY AND FRAUD PREVENTION:

Artificial intelligence automates dataprocessing, taking away the pressureof 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 theircustomers. Efficient AI-based fraud prevention systems also uphold the financial firm's image. The clientswho have suffered from the repercussions of the fraudulent operations may migrate to othercompanies.[12]

With the machine learning approach,banks and fintech can actively predictfraud instead of always staying one step back. The advance of deep learning techniques helps financialinstitutions create a safer environment for their customers.[7]

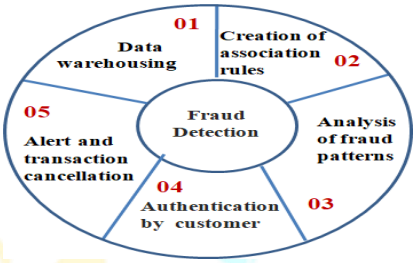

FIG.2 FRAMEWORK OF AI IN FRAUDDETECTION

The company maintains a working model and a challenger model, whichit 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 empowersfinancial institutions to detect threatssuch 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 byeliminating complex fraud cases and protecting their brand. At elinext, weoffer a variety ofprograms to developservices to thrive in this market environment. With nearly 300 professionals, we tackle challengesfrom every angle to quickly understand the data, workflows, compliance requirements, and mathbehind securities, trading, andinvesting. [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 toidentify 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 fraudprevention by notifying themwhenever there is unusual activity ontheir accounts. You can ask them to participate or simply notify them automatically if, for example, a creditcard payment originates far from their home. Since electronic fraud is so common these days, you may wantto offer customers an alert every timetheir 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 youdecide 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 are an open invitation to hackers. Anyone with knowledge can easily exploit information shared over anunencrypted network. They can stealinformation such as account numbersor login credentials and inject malware into a device connected to apublic 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 anorganization. In addition, timely detection of fraud directly impactsthe business in a positive way byreducing future potential losses. Effective detection techniques such as AI and statistical data analysis serve as a deterrent to potential fraudsters. As regulatoryrequirements and compliancedemands have grown, it has become extremely important to implement a robust fraud detection and prevention program.[14,15]

References:

https://www.elinext.com/blog/fraud-management-detection-and-prevention- in-banking-industry/ https://www.formica.ai/blo g/big-data-analytics- problems-in-fraud-detection
https://www.researchgate.net/publication/334768195_ Big_Data_for_Fraud_Detection 17 Actionable Fraud Prevention Tips for

Institution(fpsgold.com).
https://www.indiumsoftware.com/blog/fraud-analytics-for-banking/ https://www.hyperverge.co/blog/types-of-fraud- detection-techniques- systems
https://nexocode.com/blog/posts/ai-based-fraud- detection-in-banking-and- fintech-use-cases-and- benefits
https://www.analyticsinsight.net/the-importance-of- data-and-analytics-in-fraud-prevention/
http://psychologyandeducation.net/pae/index.php/pae/article/view/2518 https://medium.com/@fenjiro/analytics-techniques-for-fraud-detection-86b5e3816b92
https://www.forbes.com/sites/louiscolumbus/2019/07/09/top-9-ways-artificial-intelligence-prevents- fraud/?sh=2012e53e14b4
https://www.enformion.com/fraud-prevention-techniques/
https://www.researchgate.net/publication/35316322_A_Case_Study_in_Financial_Fraud_Detection_usin g_Big_Data_Analytics
https://www.researchgate.net/publication/328625161_Financial_fraud_detection_and_big_data_analytics