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

Big data analytics for fraud detection: techniques and tools

[1]Amal Unnikrishnan Nair

[1]Department of Information Technology
[1]KSD'S Model College, Dombivli, India

ABSTRACT

Technological improvements have enabled a massive and constant influx of real-time data that can be utilised to optimise every part of a financial institution, from client targeting to cybersecurity. On the other hand, the shift to digital banking has brought with it a slew of cybersecurity problems and fraud dangers. According to Boston Consulting Group, financial services firms are 300 times more likely than other businesses to be victims of a cyberattack.

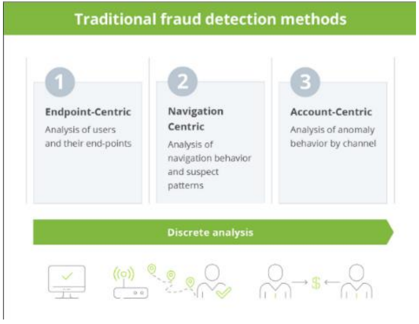
It has prompted businesses that deal with massive data sets to incorporate fraud detection technologies into their risk management strategies. 75% of responders who have integrated AI and machine learning in their risk management strategies employ Big Data analytics for fraud detection.

For decades, fraud has been a persistent issue for businesses and organisations. With the growth of digital technologies and online transactions, fraud has grown even more frequent and sophisticated. In recent years, Big Data Analytics has emerged as a significant tool for fraud detection and prevention

IndexTerms: Big Data, Security, Threat, Privacy, Fraud.

INTRODUCTION

Banking is one of the modern economy's most data-intensive industries. With a growing number of bank offices closing around the world, evolving digitization gave rise to online banking. Cashless transactions and web services have substantially reduced face-to-face contact with clients, paving the path for more important customer data to be collected through digital channels. According to a KPMG Banking survey, 42% of respondents claimed that digital channels are used to offer half of their banking services. Historically, the banking industry has been at the forefront of the fight against financial crime. However, due to a global digital transformation and an increase in e-commerce and online banking, financial institutions have made cybersecurity a major concern. Fraud detection is the most effective technique to avoid data breaches that jeopardise sensitive customer information and an organisation's reputation. Identity theft, mortgage fraud, and laundering of money are the most common types of financial fraud affecting banks. According to the Federal Trade Commission, credit card fraud is the most common sort of identity theft, accounting for 41.8% of incidents in 2019. (Twice as much compared to previous years). Because of the expansion in online transactions, wire transfers have become the most used method for cybercrime. Such data breaches are time-sensitive, which means that detecting the fraud on the first day will cost a consumer only $3, but it will cost $1,061 if it goes undiscovered for five months. Structured data, and discrete analysis were used in traditional fraud detection methods. Suspicious behaviour was spotted by a rule-based algorithm and then manually inspected by investigators. The main disadvantages of this strategy are time, human errors, and the inability to detect irregular and odd patterns of behaviour that could lead to fraud.


Traditional fraud detection methods
1
Endpoint-Centric
Analysis of users and their end-points
2
Navigation Centric
Analysis of navigation behavior and suspect patterns
3
Account-Centric
Analysis of anomaly behavior by channel
Discrete analysis
$

WHAT IS BIG DATA?

Big data analytics is the use of modern analytics to massive quantities of data of both structured and unstructured data in order to generate important insights for enterprises. It is widely utilised in a variety of industries, including health care, education, insurance, artificial intelligence, retail, and manufacturing, in order to understand what works and what doesn't in order to enhance processes, systems, and profitability. Big data analytics is the act of gathering, reviewing, and analysing massive amounts of data in order to find market trends, insights, and patterns that can assist businesses in making better business decisions. This information is promptly and efficiently available, allowing businesses to be agile when crafting plans that preserve their competitive advantage. Business intelligence (BI) tools and systems assist organisations in integrating unstructured and structured data from numerous sources. Users (usually workers) enter queries into these tools in order to gain a better understanding of business operations and performance. To find relevant insights and come up with solutions, big data analytics utilizes the four data analysis approaches.

WHY IS BIG DATA ANALYTICS IMPORTANT?

Big data analytics is significant because it allows businesses to analyse their information to uncover areas for improvement and optimisation. Increasing efficiency leads to more intelligent operations, bigger earnings, and satisfied consumers across all business segments. Big data analytics assists businesses in lowering costs and developing more customer-centric products and services. Data analytics can assist generate insights that improve how our society works. Big data analytics in the healthcare sector not only keeps track of and analyses individual records, but it also plays an important role in assessing COVID-19 outcomes on a worldwide scale. It advises health ministries in every nation on how to proceed with vaccinations and develops solutions for future outbreaks of pandemics.

BENEFITS OF BIG DATA ANALYTICS

Incorporating big data analytics into a business or organisation has numerous advantages. These are some examples:

Cost savings: Storing all business data in one location can save money. Tracking analytics also assists businesses in finding ways to work more efficiently in order to reduce costs wherever possible.

Product development: Developing and selling new products, services, or brands is considerably easier when data from client needs and desires is used. Big data analytics also assists organisations in understanding product viability and staying current with trends.

Strategic business decisions: The capacity to continuously analyse data assists firms in making better and faster decisions, such as cost and supply chain optimisation.

Customer experience: Data-driven algorithms aid marketing efforts (for example, tailored adverts) and raise customer happiness by providing a better customer experience. Risk management: Companies can discover hazards by analysing data patterns and designing risk management solutions.

Entertainment: Personalized movie and music recommendations based on a customer's particular preferences have transformed the entertainment sector. (Think Spotify and Netflix).

Education: Big data assists schools and educational technology businesses in developing new curriculums as well as enhancing existing programmes in response to needs and desires.

Government: To help manage the public sector, big data can be used to collect data from CCTV and traffic cameras, satellites, body cameras and sensors, emails, telephones, and other sources.

HealthCare: Medical history monitoring assists doctors in detecting and preventing diseases.

Marketing: Using customer information and preferences, personalised advertising campaigns with a good return on investment may be created. (ROI).

Banking: Data analytics can aid in the detection and monitoring of unlawful money laundering.

TYPES OF BIG DATA ANALYTICS

1. Descriptive analytics: Data that is easily read and analysed is referred to as descriptive analytics. This data is used to generate reports and visualise information about a company's profitability and sales.

2. Diagnostics analytics: Diagnostics analytics assists businesses in determining why an issue happened. Big data technologies and tools enable users to mine and recover data that aids in the investigation of a problem and its prevention in the future.

3. Predictive analytics: To develop forecasts, predictive analytics examines both past and present data. Users can predict market trends using artificial intelligence (AI), machine learning, and data mining.

4.Prescriptive analytics: Prescriptive analytics solves an issue by using AI and machine learning to collect data and use it for risk management.

TOOLS USED IN BIG DATA ANALYTICS

Hadoop: It is a free and open-source system for storing and analysing big data sets. Hadoop can manage and analyse both structured and unstructured data.

Spark: Spark is an open-source cluster computing framework for real-time data processing and analysis.

Data integration software: Applications that enable massive data to be streamlined across multiple platforms, such as MongoDB, Apache Hadoop, and Amazon EMR.

Stream analytics tools: Systems that filter, aggregate, and analyse data from many platforms and formats, such as Kafka.

Distributed storage: Databases, such as Cassandra, which can extend data over various servers and identify lost or faulty data.

Predictive analytics hardware and software: Systems that use machine learning and algorithms to anticipate future outcomes, such as fraud detection, marketing, and risk assessments.

Data mining tools: Applications that enable users to search structured and unstructured huge data.

NoSQL databases: They are non-relational data management solutions that are suited for dealing with unstructured and raw data.

Data warehouses: Large volumes of data collected from many sources that are typically stored using pre-set schemas.

INNOVATIVE FRAUD DETECTION METHODS BASED ON BIG DATA

Customer contact and interaction are migrating online as digital transformation accelerates, translating valuable insights into vast volumes of unstructured real-time data. While it is a tremendous benefit for businesses, it may also be their largest security flaw. AI and machine learning algorithms have facilitated the creation of fraud detection solutions that make use of Big Data to analyse massive data sets and avoid security breaches. Ad-hoc and predictive analysis are examples of discrete analysis approaches used to evaluate specific actions.

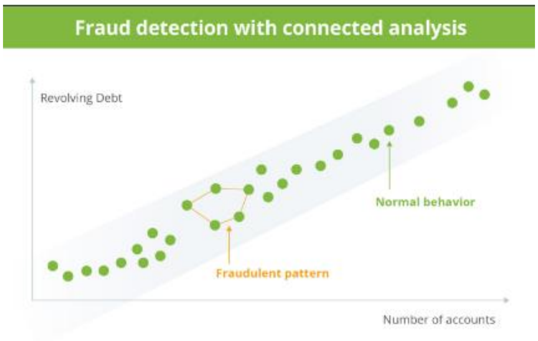
AD-HOC ANALYSIS AND SAMPLING

Ad-hoc testing is intended to uncover specific data about its application area by examining transactions for potentially hazardous activities. This technique employs a hypothesis as a starting point for detecting probable fraud in transactions. Ad-hoc testing is based on formulas and queries, which necessitate manual labour and take time. The sampling technique is frequently used to supplement ad hoc research by providing samples of transactions with fraud risks that can highlight some differences. These strategies produce good results on small data sets but are ineffective on large data sets.

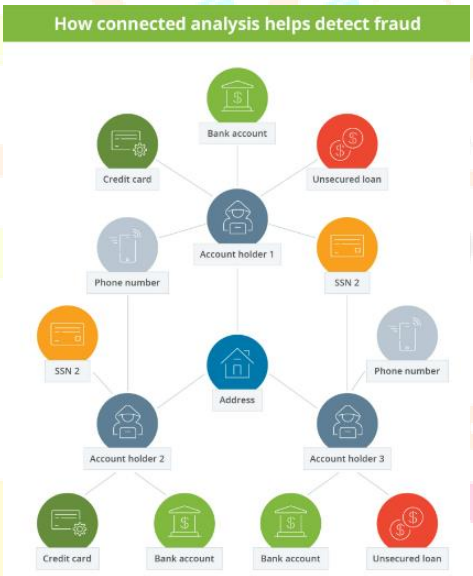
PREDICTIVE ANALYSIS

The primary purpose of predictive analytics is to develop a model that can forecast the occurrence of a given target. (Suspicious or fraudulent activities). If the training set is appropriate for the context, predictive analysis is correct. One key disadvantage of this technology is that it cannot detect fraud that did not exist in the historical data set from which it was trained.

CONNECTED ANALYSIS



With the growth and diversification of data, a lot of important information is now concealed in the constant streams of real-time unstructured data, numbers, text, speech, and pictures, including mouse movements and gyro sensor readings, which represent customer behaviour patterns. When only Google processes 20 petabytes of data each day, existing methods' analysis capability is insufficient. This is where connected data analysis may help. This method provides a more comprehensive picture of connected behaviours and relationships between people, which aids in the detection of potentially fraudulent activities.



CONTINUOUS ANALYSIS

This technology enables users to continuously monitor transactions and user activities. This method's algorithm can process data and revise patterns in real time. When compared to the more traditional rule-based method, the main advantage of continuous analytics is its ability to generate fresh insights depending on its results. Most fraud instances go unnoticed for roughly 18 months; therefore consistency is essential for fraud detection.

SOCIAL NETWORK ANALYSIS (SNA)

Based on the correlation of the analysed subjects, SNA delivers useful insights into vast datasets along the network. By extracting additional value from the subject's relationships, social network analysis broadens the scope of valuable data. SNA is a powerful way for connecting many and diverse data sources for not only fraud detection but also prediction. While traditional data-driven methods focus on statistical procedures, connected analytical techniques rely on subject relationships.

ADVANCED BEHAVIORAL AND COGNITIVE ANALYTICS

With Operational Data Lakes (ODL), which can store both organised and unstructured raw data, Big Data technology has altered data storage. Popular Big Data processing platforms such as Apache Hadoop, Apache Storm, Google Big Query, and others offer open-source frameworks that enable parallel processing by distributing massive data sets across multiple computers.

It has paved the way for a new approach to data processing. Deep analytics techniques represent a transition from discrete structured data analysis to connected unstructured and real-time data analysis. The deep analytics system may detect abnormalities and red flag possibly fraudulent activities by analysing each customer's behavioural patterns

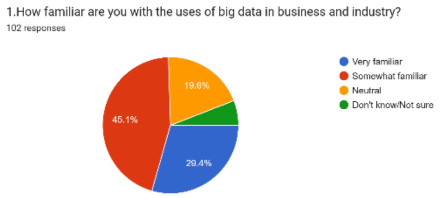
(Spending, transaction patterns, average balance, geolocation, etc.). It searches for patterns and connections between similar attacks and develops real-time algorithms for detecting suspicious activity. For instance, Danske Bank struggled with cybercrime over a long time, with a 40% fraud detection ratio and over 1,200 incorrectly identified alarms every single day. The bank decreased false-positive cases by 60% after using deep learning technologies combined with sophisticated analytics, contributing to an improvement in operating profit.

These adaptive behavioural and cognitive analytics are effective at discriminating between real atypical behaviours and fraud. These algorithms create profiles based on the nature and frequency of a customer's behaviour. When an interaction happens, data on the type of device, location, and number of purchases is analysed. These behaviours create a client portrait that can be utilised to detect credit card fraud accurately. Biometric data, such as mouse movements and clicks, enable machine learning algorithms to study the basic nature of user interactions with banking websites and create a scenario indicating the risk of fraudulent conduct.

AVOIDING FINANCIAL LOSSES WITH BIG DATA AND MACHINE LEARNING

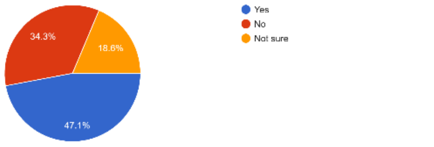
Risk measurement is a primary priority for any financial institution. When dealing with millions of sensitive personal files, identifying a potential threat as soon as feasible becomes a main goal. According to Javelin Strategy & Research, it takes a bank a total of forty days primarily to discover a possible fraudulent activity, which might result in significant financial losses. According to IBM, the typical breach of data will cost a company $3.92 million (12% more than 2018 forecasts). It highlights the significance of using predictive analytics to prevent fraud. Big Data-powered fraud detection systems are the most effective strategy to combat global cybersecurity threats. When combined with machine learning and cloud computing, such systems make real-time processing and analysis of huge amounts of streaming data possible. For example, American Express used a machine-learning model to detect suspicious and fraudulent activities by matching customer-related data with algorithms. The corporation has saved $2 billion in possible yearly incremental fraud incidences because to the data-driven approach. Furthermore, fraud detection systems contribute significantly to the company's information security and protection from external and internal threats. Internal fraud has been more prevalent among employees in recent years. Internal fraud detection mechanisms are like credit card types. It is an action-based method that collects data from phone calls, online visits, employee transactions, and other work-related activities for specific employee roles. All these data allow for the development of behaviour patterns for each function, which aids in the prevention of internal fraudulent actions.

SURVEY RESULTS



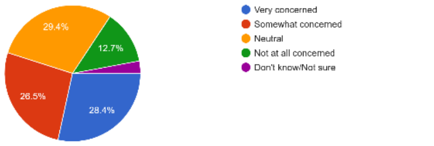
2.Have you ever knowingly provided your personal data (such as name, age, address, or email) to a company for their big data analysis?

102 responses



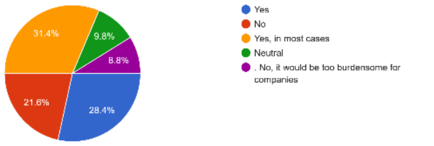
3.Are you concerned about the privacy and security of your personal data when it is used in big data analysis?

102 responses



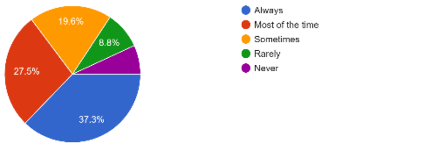
4.Do you think companies should be required to obtain explicit consent from individuals before using their personal data in big data analysis?

102 responses



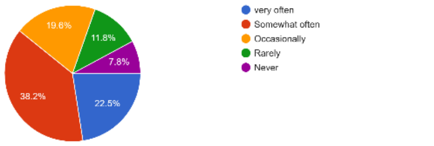
5.How often do you read the privacy policy of a company before sharing your personal data with them?

102 responses



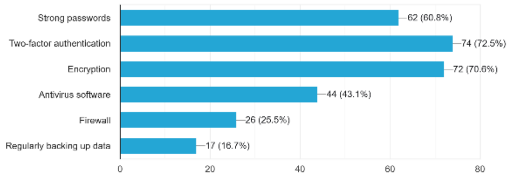
6.How often do you hear news about data breaches or data misuse by companies?

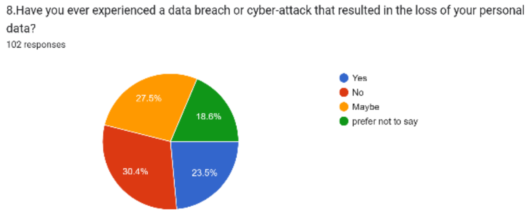
102 responses



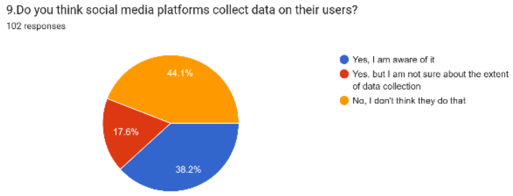
7.Which of the following methods do you use to secure your personal data? (Select all that apply)

102 responses

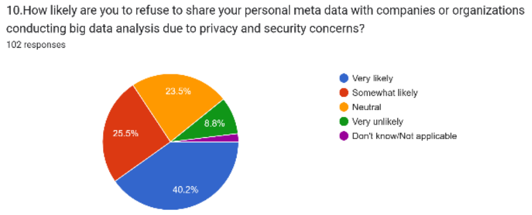



8.Have you ever experienced a data breach or cyber-attack that resulted in the loss of your personal data?

102 responses


9.Do you think social media platforms collect data on their users?

102 responses


10.How likely are you to refuse to share your personal meta data with companies or organizations conducting big data analysis due to privacy and security concerns?

102 responses

CONCLUSION

Big Data analytics is a strong and sophisticated tool for not only detecting security issues and fraudulent activities, but also preventing them from occurring in the first place. After all, data is a company's most precious asset. Fraud detection is critical not only for financial reasons, but also for client retention. When it comes to their security, customers are quite concerned. The majority of responders said they would rather have their personal images leaked than their financial information jeopardised.In today's environment, data is more than simply an IT asset; it is a critical component of the banking and finance industry's digital transformation. The latter necessitates a high level of cybersecurity, which organisations may assure by utilising current fraud detection tools.

REFERENCES

www.infopulse.com

www.researchgate.com

www.analyticsvidhya.com