JGi JAIN DEEMED-TO-BE UNIVERSITY | ONLINE

Research Project

Semester-IV

Name	VISHWAS ANAND
USN	231VMBR05312
Elective	BUSINESS INTELLIGENCE AND ANALYTICS
Date of Submission	

A study on the Enhancing Financial Fraud Detection Using Advanced Anomaly Detection Models in Banking Transactions

Research Project submitted to Jain Online (Deemed-to-be University)
In partial fulfillment of the requirements for the award of

Master of Business Administration (MBA)

Submitted by

Vishwas Anand

USN

231VMBR05312

Under the guidance of
Dr.Shalini

DECLARATION

I, VISHWAS ANAND, hereby declare that the Research Project Report titled *"Enhancing Financial Fraud Detection Using Advanced Anomaly Detection Models in Banking Transactions" hasbeen* prepared by me under the guidance of *Dr.Shalini.* I declare that this Project work is towards the partial fulfillment of the University Regulations for the award of degree of Master of Business Administration by Jain University, Bengaluru. I have undergone a project for a period of Eight Weeks. I further declare that this Project is based on the original study undertaken by me and has not been submitted for the award of any degree/diploma from any other University / Institution.

Place: Bangalore

Date:

Vishwas Anand
USN: 231VMBR05312

JGI JAIN DEEMED-TO-BE UNIVERSITY | ONLINE

CERTIFICATE

This is to certify that the Research Project report submitted by Mr./Ms. *Vishwas Anand* bearing *231VMBR05312* on the title *"Enhancing Financial Fraud Detection Using Advanced Anomaly Detection Models in Banking Transactions"* is a record of project work done by him/ her during the academic year 2023-24 under my guidance and supervisionin partial fulfilment of Master of Business Administration

Place: Bangalore

Date:

_____

Dr.Shalini

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Jain University for providing me with the opportunity to pursue my Masters of Business Administration and for supporting me throughout this research project. I am deeply thankful to my faculty guide, Dr. Shalini, whose invaluable guidance, encouragement, and expert insights greatly contributed to the successful completion of this thesis. Her constant support and constructive feedback motivated me to overcome challenges and refine my work.

I also extend my appreciation to the university officials and other faculty members who offered their assistance and resources during the course of my study. Their dedication to academic excellence created a conducive learning environment that enriched my research experience.

Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout this journey. Their belief in me provided the strength and inspiration needed to accomplish this project successfully.

Vishwas Anand
USN:231VMBR05312

Date: 01.06.2025

PLAGIARISM VERIFICATION

Title of the Thesis: "**A study on the Enhancing Financial Fraud Detection Using Advanced Anomaly Detection Models in Banking Transactions**"

Total Pages : **57**
Name of the Researcher : **Vishwas Anand**
Department : **Masters of Business Administration**
Institution : **JAIN (Deemed-to-be University)**
Name of the Guide : **Dr. Shalini**

This is to certify that the above thesis was scanned for similarity detection. The process and outcome are given below:

Software used : **TURNITIN**
Date : **01.06.2025**
Similarity Index : **19%**
Total word count : **12,490**

The content of the chapters and the publications which have been excluded from the software check:

1. Bibliography
2. Quoted Text
3. Cited Text

Prepared by

Vishwas Anand
USN:231VMBR05312

EXECUTIVE SUMMARY

Financial fraud is a growing concern for banking institutions worldwide, leading to significant monetary losses and undermining customer confidence. With the increasing reliance on digital banking channels, traditional fraud detection methods are often inadequate to identify sophisticated and evolving fraudulent activities. This thesis explores the application of machine learning techniques to enhance financial fraud detection, using a dataset of 200 sampled banking transactions for detailed analysis.

The study begins with a comprehensive literature review that highlights the prevalence of supervised learning methods in fraud detection, while also recognizing the emerging importance of unsupervised and hybrid approaches. It emphasizes the role of anomaly detection, behavioral analytics, and explainable AI in developing effective and transparent fraud detection models. Key challenges identified include data imbalance, real-time detection needs, and the integration of diverse data sources.

Data analysis reveals that debit transactions and digital channels such as online banking and ATMs dominate transaction activity. Transaction amounts show a right-skewed distribution, with most transactions being low to medium value but a few high-value transactions posing substantial risk. Behavioral factors like multiple login attempts and transaction duration provide valuable signals for detecting suspicious activity. Geographic and demographic patterns further inform targeted fraud monitoring strategies.

Based on these findings, the thesis recommends building machine learning models that incorporate transactional, behavioral, geographic, and demographic features to improve detection accuracy. It suggests focusing on digital channels, addressing data imbalance with advanced techniques, and implementing explainable AI tools for transparency and regulatory compliance. Additional recommendations include expanding dataset diversity, enhancing real-time processing, and adopting privacy-preserving collaborative methods.

In conclusion, this research demonstrates that machine learning can significantly improve financial fraud detection when models are carefully designed, continuously updated, and interpretable. Future research directions include exploring graph-based methods, natural language processing for social engineering detection, and federated learning for privacy-aware collaboration. This work provides valuable insights and practical guidance for financial institutions aiming to strengthen fraud prevention in an increasingly digital environment.

TABLE OF CONTENTS

List of Tables		
Table No.	**Table Title**	**Page No.**

List of Graphs		
Graph No.	**Graph Title**	**Page No.**

CHAPTER 1

INTRODUCTION AND BACKGROUND

INTRODUCTION AND BACKGROUND

1.1 Purpose of the Study

The aim of this research is to investigate and create next-generation anomaly detection models to improve the prevention and detection of financial fraud in bank transactions. Financial fraud continues to be the most pressing issue that confronts the banking and financial services sector as the perpetrators are constantly innovating their methods by taking advantage of the latest technologies like artificial intelligence (AI), automation, and synthetic identities. In 2025, the financial fraud environment is a complex and evolving landscape in which the historic detection techniques become less effective to tackle sophisticated and volume fraud schemes.

Banks and financial institutions are under intense pressure to safeguard the assets of their customers and uphold trust as they deal with regulatory demands and operational limitations. Fraud loss keeps growing, and synthetic identity fraud alone is expected to result in more than $23 billion in annual losses in the U.S. by 2030 . Fraudsters take advantage of weaknesses in digital channels, such as mobile wallets, peer-to-peer payment applications, and cryptocurrency exchanges, with AI-generated deep fake audio, automated credential stuffing, and social engineering attacks evading traditional security controls.

Legacy rule-based systems of detecting fraud, primarily based on fixed thresholds and pre-defined rules, are subjected to some serious drawbacks like high rates of false positives, slow detection times, and a lack of adaptability to emerging fraud patterns. This makes them inefficient, raises the costs of investigations, and limits customer satisfaction. Thus, there is a compelling need for smart, adaptive, and scalable fraud detection systems that have the ability to process huge amounts of transactional data in real time, detect subtle anomalies, and minimize false alarms.

The objective of this research is to create and implement state-of-the-art anomaly detection models, including machine learning models (such as isolation forests, auto encoders, and gradient boosting machines) and graph neural networks, to detect emerging fraud schemes efficiently. The research will also address typical problems such as data imbalance, explainability of the model, and compliance with privacy regulations (such as GDPR). By the integration of explainable AI and privacy-

preserving techniques, the study seeks to develop trustworthiness and transparency in anti-fraud systems to instill regulatory compliance and stakeholder trust.

By providing a robust, fact-based, and proactive fraud detection system, this study ultimately seeks to assist the financial sector's ongoing battle against fraud risks. In the face of a rapidly evolving digital landscape, this system will help banks lower financial losses, increase operational efficiency, and protect consumers from increasingly complex fraud attacks.

1.2 Introduction to the Topic

Technological innovation, changing consumer behavior, and emerging fraud threats are all contributing to a fundamental shift in the banking industry. Globally, the number of digital banking transactions has increased dramatically as of 2025, with over 3 billion customers using online and mobile platforms. Record levels of convenience have been made possible by this change, but financial fraud's attack surface has also grown. In order to carry out complex scams that are difficult to detect, scammers are increasingly using digital channels and advanced technologies like deepfake media, generative artificial intelligence (GenAI), and artificial intelligence (AI).

The recent surveys, such as the KPMG Global Banking Scam Survey 2025, indicate that while some of the same fraud types continue to be prevalent, new types of scams are emerging quite quickly. For instance, deep fake-created personas have begun being utilized in romance scams and investment scams on social media against retail customers and particularly vulnerable older individuals1. AI-generated impersonation website and voice deep fake fraud are increasingly becoming a problem, which presents new challenges for banks as well as regulators. While these AI-based scams are not prevalent at present, banks look for tremendous growth in their incidence and are starting to invest in detection solutions and personnel training to limit the risk.

In 2025, account takeovers will still be a major threat due to social engineering, credential stuffing, and misuse of emerging payment methods such as cryptocurrency platforms, peer-to-peer (P2P) payment apps, and mobile wallets. The Federal Trade Commission reports a steady increase in account takeover crimes, which are driven by AI-driven phishing attempts that are extremely complex and customized, as well as robotic bots that attempt to use credentials that have been stolen at scale. One of the financial crimes with the fastest rate of growth is synthetic identity theft, in which criminals create identities using a mix of real and fake information. It is estimated to be resulting in more than $23 billion in losses every year

in the U.S. alone by 2030. These synthetic identities tend to evade the usual verification procedures, requiring sophisticated detection systems that examine subtle discrepancies over several data points.

The rapid adoption of faster payment systems in 2025 has introduced new vulnerabilities, as fraudsters exploit the reduced time window for transaction verification. Real-time payment fraud, including push payment scams and business email compromise (BEC) schemes, is increasing, with fraudsters using AI-generated deep fake videos to impersonate executives and authorize fraudulent transfers4. Financial institutions face the challenge of implementing real-time fraud detection tools capable of monitoring payment activity and flagging suspicious transactions within milliseconds.

Demographically, fraudsters modify their approach to target varied customer segments. Tech support scams, investment scams, and romance scams that use chatbots and AI-generated voices to build trust disproportionately target older consumers, who are increasingly using the internet for banking. An estimated $28 billion is lost annually in the United States due to elder financial exploitation, underscoring the significance of consumer education and targeted fraud prevention. At the same time, scams involving cryptocurrencies and e-commerce are increasingly targeting younger demographics due to their online behavior patterns.

In general, the changing fraud landscape of 2025 requires banks to shift from conventional rule-based systems to intelligent, adaptive anomalous pattern detection models. Such models leverage machine learning, behavioural biometrics, graph analytics, and explainable AI to identify sophisticated and new patterns of fraud in real time, minimize false positives, and meet stringent regulatory needs. This thesis aims to create such sophisticated models to enable banks to effectively deal with financial fraud in a more digital and AI-based world.

1.3 Overview of Theoretical Concepts

Banking financial fraud detection has come a long way from the conventional rule-based systems to sophisticated, data-driven anomaly detection systems based on machine learning (ML) and artificial intelligence (AI). The chapter describes the key theoretical concepts behind the sophisticated fraud detection methods with focus on how they can be used in the banking sector in 2025.

Rule-Based Systems Traditional

Historically, rule-based systems that recognized transactions according to preset rules—such as transaction rates, geographic locations, or value limits—were used for fraud detection. They are straightforward and easy to understand, but they are static and reactive, requiring constant manual updating to combat emerging fraud techniques. Their lack of adaptability results in a high number of false positives and a poor response to emerging fraud trends.

Machine Learning for Fraud Detection

Machine learning represents a paradigm shift in fraud detection because it enables systems to learn from historical transactional records and identify complex fraud patterns without coding. ML models handle enormous amounts of data in real time, identifying subtle anomalies that rules-based systems wouldn't catch. This adaptability is critical in combating the dynamic and sophisticated tactics of modern-day fraudsters.

Key machine learning techniques include:

- Supervised Learning: Labeled datasets with known fraudulent and legitimate transactions are used to train models such as neural networks and gradient boosted decision trees (such as XGBoost and LightGBM). To accurately classify new transactions, the models acquire the discriminative features. Since there aren't many fraud cases, supervised learning requires high-quality labeled data, which can be hard to come by.
- Unsupervised Learning: By identifying transactions that deviate from typical behavior, methods such as isolation forests and autoencoders identify anomalies without labeled data. While autoencoders reconstruct input data and identify high reconstruction errors as anomalies, isolation forests recursively partition data to isolate anomalies. These methods reduce reliance on labeled data and are effective in identifying novel fraud patterns.
- In situations where labeled fraud data is limited, semi-supervised learning improves detection performance by combining labeled and unlabeled data.

Real-Time Anomaly Detection

By enabling quick action against suspicious transactions, real-time detection helps to minimize financial loss. Real-time streaming transaction data is tracked by machine learning models, which improve their understanding of normal behavior and instantly identify anomalies. This feature significantly lessens the impact of fraud by allowing banks to block or freeze suspicious transactions before they settle.

Graph-Based Anomaly Detection

Fraudulent behavior will tend to consist of intricate relationships among objects like accounts, merchants, and devices. Graph neural networks (GNNs) represent these kinds of relationships as nodes and edges and can detect sophisticated types of fraud like money laundering rings and collusive fraud networks. When processing transaction networks, feature-based models may overlook minute patterns and structural irregularities that GNNs pick up on.

Explainable AI (XAI)

Regulatory environments and consumer confidence require explanations of AI-driven fraud detection outcomes. Explainable AI techniques, such as SHAP (Shapley Additive Explanations), measure each feature's contribution to a transaction's fraud score in order to provide model prediction explanations. By converting complex ML models into understandable ones, XAI guarantees auditability, regulatory compliance, and stakeholder confidence.

Privacy-Preserving Machine Learning

Banks are also required to adhere to data privacy laws such as GDPR and PSD3, which limit data exchange and require secure handling of data. Privacy-enhancing methods such as federated learning enable institutions to jointly train fraud detection models without sharing sensitive customer information. Differential privacy and homomorphic encryption also safeguard data at model training and inference stages, trading off fraud detection effectiveness with privacy needs.

Natural Language Processing (NLP) Integration

Fraud detection goes beyond official transaction information. NLP methods examine unstructured information like customer interactions, emails, and social engineering, phishing, and other fraud attempt records. Combining NLP with transaction monitoring improves the overall fraud detection process.

1.4 Company/ Domain / Vertical /Industry Overview

The foundation of the world economy is the banking and financial sector, which facilitates capital flows, handles payments, and fosters economic expansion. With over 3 billion digital banking customers worldwide and transaction volumes at all-time highs, the industry is characterized by rapid digitization in 2025. With instant payments, round-the-clock account access, and a vast array of financial products available online and on mobile devices, digitalization has completely changed the way banks interact with their clients.

Digital Banking and Fraud Landscape

The industry is dealing with more financial fraud-related problems as digital banking grows. According to the KPMG Global Banking Scam Survey 2025, high-value scams like account takeover attacks, synthetic identity fraud, and Authorized Push Payment (APP) fraud are on the rise, making fraud one of the biggest worries for banks worldwide. According to the survey, 76% of banks believe that calling or interacting directly with customers to confirm questionable transactions is an effective way to prevent fraud. Furthermore, according to 73% of survey participants, payment holds under certain regulations are beneficial because they give investigators important time to work before releasing funds.

Banks employ a combination of legacy and emerging fraud detection methods. Legacy rule-based systems, which identify transactions based on pre-established rules such as transaction amount or geolocation, still remain but are increasingly being supported by machine learning and AI models. These newer models leverage holistic transaction risk scores, customer behavioral profiles, and real-time incoming transaction monitoring to detect anomalies more accurately. For instance, 64% of the banks polled keep an eye on consumer behavior patterns to determine what constitutes "normal" behavior, which lowers false positives and increases detection accuracy.

Fraud Types and Detection Methods

The industry is confronted with a broad range of fraud types, including:

- Credit and Debit Card Fraud: Criminals frequently use stolen card information or compromised merchant systems to commit payment card fraud.
- Account Takeover Fraud: Attackers steal customer accounts through compromised credentials or social engineering tactics, typically by employing automated bots to stuff credentials.
- • Money Mule Schemes: Criminals use people they don't know to transfer illegal funds. To facilitate detection, banks place a strong emphasis on gathering centralized or consortium data on mule accounts.
- Scam and Investment Fraud: Social engineering and impersonation are common forms of scams and fraud involving investments.

To counteract these threats, banks increasingly deploy more and more AI and machine learning to support rule-based detection. AI is capable of analyzing vast volumes of data, spotting subtle fraud patterns, and adapting to new threats. Blacklists and whitelists are used to manage known fraudsters and known good parties, and network and graph-based analysis

techniques are used to find patterns among fraud rings.

Operational and Regulatory Environment

The banks then face the operational challenge of striking a balance between customer satisfaction and fraud prevention. Overzealous fraud detection leads to rejected legitimate transactions, false positives, and a decline in consumer trust. As a result, the majority of banks employ layered technologies with automated detection backed by customer validation and human analyst review.

One of the primary motivators for fraud detection solutions is compliance. Banks are subject to reporting requirements like Suspicious Activity Reports (SARs), data privacy laws like GDPR, and anti-money laundering (AML) regulations. The laws' requirements for openness, data security, and prompt reporting have an impact on how fraud detection systems are developed and implemented.

Industry Cooperation and Emerging Trends

Cooperation between technology suppliers, regulators, and financial institutions is emphasized in all industry analyses and the KPMG survey. Sharing data about fraud trends, mule accounts, and scammer networks improves detection and lowers fraud loss. These days, most people agree that central databases and consortiums are an indispensable tool.

To combat emerging fraud threats, the banking sector is making significant investments in explainable AI, behavioral biometrics, real-time transaction monitoring, and AI-driven fraud prevention in the near future. By combining these technologies, it will be possible to meet customer expectations and regulatory requirements while improving and speeding up detection.

1.5 Environmental Analysis (PESTEL Analysis)

A PESTEL analysis analyzes the macro-environmental forces that drive the fraud detection environment of the banking sector. To develop efficient, legal, and responsive fraud detection systems, it is necessary to comprehend these drivers.

Political Factors

- Regulatory Environment: Financial institutions and banks maintain strict controls to prevent money laundering and fraud. For example, the U.S. Financial Crimes Enforcement Network (FinCEN) is the regulator that enforces the Bank Secrecy Act, and under this act, banks are required to file suspicious activity reports (SARs). Over 19 million SARs were filed in 2021 alone, which is an indication of regulatory oversight.

Serious consequences could result from noncompliance, like HSBC's $1.9 billion AML fine in 2022.

- Political Stability: Political stability fosters economic growth and financial infrastructure investments. Stable nations have a higher GDP growth rate (3.5% compared to 1.5% for unstable nations), which aligns with higher fraud prevention capacity due to better resources and coordination.
- Global AML Policies: International guidelines like the Financial Action Task Force (FATF) guidelines require banks to have robust anti-money laundering procedures, which significantly impact the cost and approach of fraud detection. In 2022, global AML compliance fees reached $30 billion, and they will only increase.

Economic Factors

- Economic Cycles and Fraud: Economic downturns frequently result in an increase in fraud. For example, during the COVID-19 pandemic, fraud cases increased from 31,000 in 2019 to 150,000 in 2020, a 384% increase.
- Interest Rates and Investment: The profitability and capacity of banks to invest in anti-fraud technology may be decreased by increased interest rates. The recent Federal Reserve rate increases (5.25%–5.50%) have historically had an impact of reducing bank profits by 5% for a 1% increase in the rate.
- Currency fluctuations: Especially when it comes to cross-border transactions, exchange rate volatility has an indirect impact on the exposure to fraud risk and cross-border banking.

Social Factors

- Customer Awareness and Confidence: Approximately 70% of consumers worry about financial fraud, highlighting the need for robust fraud protection. 63% of people who trust banks believe that they should act in their best interests.
- Adoption of Digital Banking: At a compound annual growth rate (CAGR) of 10.8%, international digital banking is projected to reach $8.9 trillion by 2027. The change expands fraud attack surfaces and transaction volume.
- Demographic Trends: Fraud is more common among younger generations (50 percent of Gen Z and 40 percent of millennials), necessitating fraud detection systems that adapt to changing user preferences and behavior.
- Customer Expectations: Banks are under pressure to provide simple yet secure services because 75% of customers are willing to switch banks for better online experiences.

Technological Factors

- Artificial Intelligence and Machine Learning: Financial services' AI market is likely to grow to $22.6 billion by 2025 at a CAGR of more than 23%. AI-based fraud detection models enhance detection rates and enable real-time monitoring.
- Big Data Analytics: Banks process enormous amounts of transactions; big data analytics capacity will increase beyond $67 billion by 2026, supporting real-time anomaly detection.
- Cybersecurity Threats: The global cost of cybercrime is projected to reach $10.5 trillion by 2025, and banks will have to continuously advance security technology like encryption and multi-factor authentication.
- Digital Banking Transformation: With over 3 billion digital banking users in 2024, fraud detection systems have to transform across platforms with over half of transactions occurring through mobile apps.

Environmental Factors

- Operational Disruptions: Pandemics and natural disasters can cause banking operations to be disrupted, making banks more vulnerable to fraud during times of crisis.
- Ethical and Sustainable Practices: Greater emphasis on ethical and sustainable banking practices trickles down to corporate governance and risk management and ultimately affects fraud prevention strategies.

Legal Issues

- AML and KYC Rules: AML and KYC rules must be complied with. For example, suspicious transactions over $10,000 must be reported as per Canadian regulations, and failing to do so attracts a penalty of millions.
- Data Privacy Laws: Most laws such as GDPR have strict data processing and transparency requirements that affect fraud detection models' processing and storage of customer data.
- Non-Compliance Penalty: Financial penalty for non-compliance with fraud detection and AML can be enormous, with a case in point being an $8.6 billion fine given to a U.S. bank in 2021.
- Intellectual Property: Proprietary fraud detection technologies are protected through patents and trademarks to ensure market competitive advantage.

CHAPTER 2

REVIEW OF LITERATURE

REVIEW OF LITERATURE

2.1 Domain/ Topic Specific Review

Detection of financial fraud is an increasingly prominent field of research and practice within the banking and financial sector in light of the acceleration of financial transaction digitization and the evolution of fraudsters. The past decade has seen the paradigm shift from the conventional rule-based fraud detection to sophisticated machine learning (ML) and artificial intelligence (AI) methods, which are more accurate, adaptive, and scalable.

Machine Learning Methods in Financial Fraud Detection

A systematic review of 104 peer-reviewed papers between 2012-2023 by Polak et al. (2024) observed the dominance of machine learning in financial fraud detection. The review employed stringent methodologies (PRISMA and Kitchenham) and analyzed datasets from different financial markets including China, Canada, the United States, and Iran. Credit card fraud detection was the most studied area, and supervised learning models like decision trees, random forests, and gradient boosted machines were employed extensively. Accuracy, precision, recall, F1-score, and sensitivity are standard performance measures employed, indicating consideration of balanced measurement of detection performance.

In addition, Mustika et al. (2025) also conducted a systematic review on the application of ML techniques to various forms of financial fraud, including credit risk assessment and transaction fraud. In line with their findings, ensemble methods and deep learning models, namely Long Short-Term Memory (LSTM) networks, work significantly better than traditional classifiers by effectively dealing with temporal relationships and complex patterns in sequential transactions. Utilizing synthetic minority oversampling techniques (SMOTE) to address class imbalance was also found to be a significant step towards improving the performance of the model.

Deep Learning and Anomaly Detection

Recent advances in deep learning have introduced state-of-the-art architectures such as Convolutional Neural Networks (CNNs), LSTM networks, and transformer-based architectures to the fraud detection issue. Chen et al. (2025) demonstrated how LSTM models have achieved over 94% fraud detection accuracy, with higher Area Under the Curve (AUC)

values compared to relative ML algorithms. These models are very effective at learning sequential patterns of transactions, which play an important role when detecting elaborate fraud patterns constructed over time.

Unsupervised learning techniques, such as autoencoders and isolation forests, have continued to lead the way with the capacity to identify new and unknown patterns of fraud without relying on labeled data. In order to identify outliers in big datasets, autoencoders map input transaction data and identify anomalies as a function of reconstruction error. Isolation forests work well with high-dimensional transactional data because they recursively partition the data to find anomalies.

Hybrid and Ensemble Models

It has been discovered that hybrid techniques that integrate several machine learning models improve the ability to detect fraud. When it comes to utilizing complementary strengths, ensemble approaches that integrate decision trees, neural networks, and anomaly detection techniques outperform single-model approaches, according to Carcillo et al. (2021). They enhance detection rates and lower false positives, both of which are vital for bank environments' operational performance.

Graph-based models, particularly Graph Neural Networks (GNNs), are a new fraud detection research area. GNNs learn about entity relationships such as accounts, devices, and merchants, tracing out complex money laundering networks and fraud rings that their feature-based counterparts may fail to detect. Blending GNNs with other machine learning techniques holds the key to end-to-end fraud detection systems.

Challenges and Emerging Trends

In spite of considerable advancements, some issues remain in financial fraud detection research:

- Data imbalance: Fake transactions are usually less than 0.1% of all transactions and therefore training and testing the model become difficult. Techniques like cost-sensitive learning and SMOTE are commonly used to get around this issue.
- Interpretability: For operational transparency and legal reasons, models must be comprehensible. In order to facilitate model decision insights, explainable AI (XAI) techniques like SHAP and LIME are being used more and more.
- Privacy and Compliance: Laws like GDPR and PSD3 limit data sharing, and it becomes appealing to explore privacy-preserving technologies like homomorphic encryption and

federated learning.

- Computational Efficiency: Real-time fraud detection requires models that are not just accurate but also low-latency. Deep learning models are accurate but are computationally intensive and thus often struggle with deployment.
- Multimodal Data Integration: Merging structured transaction data with unstructured data like customer dialogue and social media data through Natural Language Processing (NLP) is an evolving area with the possibility of enhancing fraud detection.

Summary

Artificial intelligence and machine learning have revolutionized financial fraud detection from rule-based static systems to dynamic, adaptive models that can identify intricate, changing fraud patterns, according to the 2012–2025 literature. Although supervised learning predominates, unsupervised and hybrid models are becoming more and more important in identifying new forms of fraud. Newer methods that tackle the complexity of fraud detection in modern banking include deep learning, graph analytics, and privacy-preserving strategies. The thesis, which will create sophisticated anomaly detection models that incorporate these theoretical advancements to further enhance real-time fraud detection accuracy, explainability, and compliance for banking transactions, has a solid foundation thanks to this introduction.

2.2 Gap Analysis

Although the enormous progress in detecting financial fraud using machine learning and artificial intelligence, some important research and application voids exist. It is essential to define these voids in order to guide the future research towards more efficient, dynamic, and compliant fraud detection systems. The subsequent section discusses the main limitations and challenges identified in current literature and applications, which this thesis attempts to fill.

Limited Scope Beyond Credit Card Fraud

Most of the existing research and commercial anti-fraud tools address credit card fraud to a great extent due to the presence of labelled data and the nature of such fraud. Comparatively less attention is paid to the other equally serious frauds, such as account takeover, authorized push payment (APP), synthetic identity theft, and money laundering. These frauds are more advanced, multi-actor frauds that must be tackled differently. Fraud detection models generalizing over a broad range of fraud types and transaction modalities in banking systems

are beneficial with great urgency.

Network and Graph-Based Analytics underutilization

Graph Neural Networks (GNNs) and other graph-based anomaly detection techniques have shown great promise in identifying structural and relational patterns present in intricate money laundering and fraud rings. Although they have shown promise in research, their application in real-world banking environments is still in its infancy. The intricacy of system integration, the computational expense of graph models, and the lack of standardized frameworks for incorporating graph analytics into traditional machine learning models are obstacles. Research into scalable, hybrid architectures that combine transactional feature analysis and graph-based understanding is necessary to close the gap.

Explainability and Transparency Deficiencies

Transparency in automated decision-making is becoming more and more necessary due to regulatory requirements like the U.S. Fair Credit Reporting Act, GDPR, and PSD3. However, high-accuracy deep learning models are usually "black boxes" with little interpretability in their fraud predictions. Stakeholder trust, auditability, and regulatory compliance are all hampered by this transparency. Despite their promise, explainable AI (XAI) techniques are not yet applied in fraud detection pipelines in a way that strikes a balance between interpretability and performance. The largest challenge is creating fraud detection models that are both accurate and interpretable by design or that integrate XAI techniques.

Privacy and Data Sharing Limitations

Bank-to-bank and geography-to-geography cooperative fraud detection is difficult because banks have stringent privacy regulations that restrict sharing of sensitive customer data. Although privacy-preserving machine learning techniques and federated learning offer solutions, their use in detecting banking fraud is still in its infancy. Model convergence, handling diverse data distributions, and maintaining robust security against hostile attacks are the challenges. Building scalable, privacy-aware cooperative fraud detection systems that adhere to legal requirements while enhancing detection is the research gap.

Scalability Issues and Real-Time Identification

To stop fraudulent transactions before they happen, fraud detection needs to be done in real-time or almost real-time. Implementing most advanced machine learning and deep learning

models in real-world bank systems would be difficult due to their high latency and computational cost, especially those with intricate architectures like ensembles or GNNs. To make such models operate more effectively for low-latency inference with similar detection performance, research is required in areas such as hardware acceleration, low-latency feature engineering, and model compression.

Handling Noisy and Imbalanced Data

Since fraudulent transactions make up a very small percentage of all transactions, fraud data sets are by nature unbalanced. This results in biased models that ignore fraud detections (false negatives) and classify the majority of data in the majority class (legitimate transactions). Due to operational mistakes or problems with data integration, transactional data also contains noise, missing values, and inconsistencies. Few comprehensive frameworks have strong capabilities to handle data quality issues in real-world bank data sets, even though methods like SMOTE and anomaly detection help to mitigate these problems.

Integration of Multimodal and Unstructured Data

The majority of structured transactional data is handled by all of the existing fraud detection models. Contextual information found in unstructured data, including emails, social media, biometrics, and customer service conversations, can be used to improve fraud detection. Multimodal learning and natural language processing are not widely used in the detection of banking fraud. There is a huge research gap in the integration of these diverse data types, which affects data preprocessing, feature fusion, and model design.

Insufficient Standardized Evaluation Systems

It is challenging to benchmark and compare the performance of fraud detection models due to heterogeneity in data sets, fraud types, and performance measures used across studies. Data sets that accurately represent the complexity of banking transactions in the real world and standard, publicly accessible test protocols are crucial. It is challenging to reproduce and generalize results when there is a lack of standardization.

Overview of Financial Fraud and Its Impact

Financial fraud continues to be a major problem for businesses all over the world, resulting in large losses and eroding confidence in financial systems. Because of the increased vulnerabilities brought about by the quick digitization of banking and financial services, fraud

detection is now a top concern for organizations looking to safeguard their resources and clients. Recent research has shown that fraud can take many different forms, such as market manipulation, account takeovers, money laundering, and credit card fraud, all of which call for specialized detection techniques (Nature Communications, 2024).

Machine Learning in Financial Fraud Detection

Machine learning (ML) techniques have become the cornerstone of modern fraud detection systems due to their ability to learn complex patterns and adapt to evolving fraud tactics. A comprehensive systematic literature review covering 104 studies from 2012 to 2023 shows that supervised learning methods dominate the field, especially for credit card fraud detection, owing to the availability of labeled datasets and their high predictive accuracy (Nature Communications, 2024). Popular algorithms include decision trees, random forests, support vector machines, and neural networks.

Unsupervised and semi-supervised methods are also gaining traction, particularly for detecting novel or emerging fraud patterns where labeled data is scarce. Deep learning architectures, such as convolutional and recurrent neural networks, have demonstrated superior performance in capturing temporal and spatial fraud patterns, especially in large-scale transaction data (Nature Communications, 2024)

Anomaly Detection Techniques in Finance and Banking

Anomaly detection is a fundamental approach in fraud prevention, focusing on identifying transactions or behaviors that deviate from established norms. The five key techniques widely used in financial anomaly detection include statistical methods, supervised learning, unsupervised learning, hybrid approaches, and emerging graph-based techniques (Number Analytics, 2025).

- **Statistical Methods:** These traditional techniques rely on predefined thresholds and probabilistic models to flag outliers but often suffer from high false positive rates.
- **Supervised Learning:** Uses labeled data to train models that classify transactions as fraudulent or legitimate.
- **Unsupervised Learning:** Detects anomalies without labeled data, useful for discovering new fraud types.
- **Hybrid Approaches:** Combine supervised and unsupervised methods to leverage the

strengths of both.

- **Graph-based Techniques:** Analyze relationships between entities (accounts, devices, merchants) to detect fraud rings and complex schemes.

Applications cover credit card fraud prevention, account takeover protection, money laundering detection, credit risk assessment, and market surveillance (Number Analytics, 2025)

Specific Fraud Scenarios Addressed by Machine Learning

Machine learning models have been effectively applied to several common fraud scenarios:

- **Credit Card Fraud:** ML systems analyze hundreds of transaction features in real-time to detect abnormal spending patterns, unusual transaction frequencies, and associations with high-risk accounts (iTransition, 2025).
- **Money Laundering:** Models trained on transaction networks and sender/receiver profiles help identify suspicious money flows and flag potential laundering activities (iTransition, 2025).
- **Market Manipulation:** ML detects anomalies in trading volumes, price movements, and order patterns to prevent churning, spoofing, and wash trading (iTransition, 2025).

Survey of Financial Fraud Detection Methodologies

Earlier surveys emphasize the complexity of fraud detection due to the interspersing of fraudulent and genuine transactions, which challenges simple pattern matching techniques. Various data mining and machine learning approaches have been reviewed, including classification, clustering, and hybrid models applied to credit card fraud, online auction fraud, telecommunication fraud, and intrusion detection (IJCA, 2012).

More recent studies focus on advanced data mining algorithms such as deep learning and ensemble methods, highlighting their scalability, accuracy, and real-time detection capabilities (IJSci, 2024).

Explainability in Fraud Detection Models

Explainability is critical in fraud detection models to ensure transparency, regulatory compliance, and trust. Unlike black-box models, explainable AI (XAI) techniques allow stakeholders to understand how decisions are made, identify biases, and validate model outputs. This is especially important in fintech where decisions directly impact customers and institutions. Techniques such as SHAP, LIME, and rule extraction are commonly used to

provide interpretable insights into model behavior (Flagright, 2024).

Summary of Literature Gaps and Trends

- Most research focuses on supervised learning for credit card fraud, with fewer studies on other fraud types or unsupervised approaches.
- There is a growing trend toward using real-world datasets and combining multiple ML techniques.
- Explainability and real-time detection are emerging as key priorities.
- Graph-based and hybrid models represent promising future directions.
- The literature calls for more comprehensive frameworks that integrate behavioral, transactional, and network data.

Summary of Literatures:

1. Nature Communications (2024)

Title: Financial fraud detection through the application of machine learning techniques: A systematic review Summary:

This comprehensive systematic review analyzes 104 studies published between 2012 and 2023 focusing on machine learning applications in financial fraud detection. The authors categorize techniques into supervised, unsupervised, and hybrid learning, with supervised methods dominating due to the availability of labeled datasets. The review highlights the effectiveness of deep learning models, such as convolutional and recurrent neural networks, in capturing complex temporal and spatial fraud patterns. It also discusses challenges like data imbalance, evolving fraud tactics, and the need for explainability. The paper emphasizes the importance of integrating multiple data types (transactional, behavioral, network) and calls for more research into unsupervised and real-time detection methods.

2. Number Analytics (2025)

Title: 5 key anomaly detection techniques in finance & banking Summary:

This industry-focused report outlines five primary anomaly detection techniques used in financial institutions: statistical methods, supervised learning, unsupervised learning, hybrid approaches, and graph-based techniques. It explains the strengths and limitations of each method, noting that traditional statistical approaches often generate high false positives, while machine learning models improve detection accuracy but require quality data and tuning. The report highlights graph-based analysis as a promising frontier for detecting complex fraud rings

by modeling relationships between accounts, devices, and merchants. Practical applications include credit card fraud, money laundering, and market surveillance.

3. iTransition (2025)

Title: Machine learning for fraud detection: An in-depth overview

Summary:

This whitepaper provides an overview of machine learning applications in detecting various fraud types, including credit card fraud, money laundering, and market manipulation. It details the use of supervised models such as random forests and gradient boosting for transaction classification, as well as unsupervised methods like clustering for anomaly detection. The paper stresses the importance of feature engineering, including behavioral metrics like login attempts and transaction duration, to improve model performance. It also discusses challenges such as data imbalance and the need for real-time detection, recommending hybrid models and continuous retraining to adapt to new fraud patterns.

4. IJCA (2012)

Title: A survey on financial fraud detection methodologies

Summary:

This early survey paper reviews various data mining and machine learning techniques applied to financial fraud detection across domains like credit card fraud, telecommunication fraud, and online auction fraud. It categorizes methods into classification, clustering, and hybrid models, highlighting their respective advantages and limitations. The paper notes the challenge of distinguishing fraudulent transactions due to their rarity and similarity to legitimate behavior. It advocates for combining multiple techniques and incorporating domain knowledge to improve detection accuracy. Although dated, it provides foundational insights into the evolution of fraud detection research.

5. IJSci (2024)

Title: Anomaly detection in financial transactions using advanced data mining algorithms Summary:

This recent study explores the application of advanced data mining algorithms, including deep learning and ensemble methods, for anomaly detection in financial transactions. The authors demonstrate that ensemble classifiers combining multiple base learners outperform single models in terms of accuracy and robustness. The paper also evaluates the effectiveness of

autoencoders and recurrent neural networks in modeling sequential transaction data. The study underscores the importance of feature selection and dimensionality reduction to handle large-scale datasets efficiently. It concludes that integrating multiple algorithms and data sources is key to scalable and accurate fraud detection.

6. GSCARR (2024)

Title: AI-driven fraud detection in banking: A systematic review of data science techniques

Summary:

This systematic review focuses on AI and data science applications in banking fraud detection, analyzing recent advances and performance benchmarks. The authors report detection rates between 87% and 94% for AI-powered systems, significantly higher than traditional rule-based approaches. The review highlights the benefits of combining supervised and unsupervised learning to detect both known and emerging fraud patterns. Challenges discussed include handling imbalanced datasets, minimizing false positives, and achieving real-time processing. The paper calls for more research into explainable AI to meet regulatory requirements and increase stakeholder trust.

7. Flagright (2024)

Title: Ensuring explainability in your fraud detection models

Summary:

This industry article discusses the critical role of explainability in fraud detection models to ensure transparency, regulatory compliance, and operational trust. It reviews popular explainability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), which help interpret complex machine learning models by attributing feature importance. The article emphasizes that explainability enables fraud analysts to validate alerts, understand model decisions, and communicate findings to regulators and customers. It also highlights the trade-offs between model complexity and interpretability, advocating for hybrid approaches that balance accuracy with transparency.

8. Pedregosa et al. (2011)

Title: Scikit-learn: Machine learning in Python

Summary:

This foundational paper introduces Scikit-learn, a widely used open-source Python library for

machine learning. It provides efficient implementations of classification, regression, clustering, and dimensionality reduction algorithms. The paper emphasizes the library's ease of use, consistent API, and integration with scientific Python tools. Scikit-learn has become a standard tool in fraud detection research and practice due to its versatility and extensive documentation, enabling rapid prototyping and evaluation of machine learning models.

9. Ribeiro, Singh, & Guestrin (2016)

Title: "Why should I trust you?": Explaining the predictions of any classifier

Summary:

This influential conference paper proposes LIME, a model-agnostic technique for explaining individual predictions of any classifier. LIME approximates complex models locally with interpretable models to highlight which features most influenced a specific prediction. The method enhances trust and transparency in AI systems, particularly important in sensitive domains like fraud detection where understanding why a transaction was flagged is critical. The paper demonstrates LIME's effectiveness across multiple datasets and models, establishing it as a key tool in explainable AI.

10. Chollet et al. (2015)

Title: Keras

Summary:

Keras is a high-level neural networks API written in Python, capable of running on top of TensorFlow. It simplifies building and training deep learning models with an intuitive interface. Keras has been widely adopted in fraud detection research for implementing deep learning architectures such as convolutional and recurrent neural networks, which are effective in modeling complex transaction sequences and temporal patterns.

11. Abadi et al. (2016)

Title: TensorFlow: Large-scale machine learning on heterogeneous systems

Summary:

TensorFlow is an open-source platform for machine learning developed by Google. It supports large-scale numerical computation and flexible model building, making it suitable for deploying fraud detection models in production environments. TensorFlow's scalability and ecosystem facilitate the development of real-time, adaptive fraud detection systems capable of processing massive transaction streams.

Summary

This gap analysis identifies the key areas of current financial fraud detection research and practice that need improvement. These include limited coverage of fraud types, the inability to fully utilize graph analytics, explainability, privacy constraints, real-time detection constraints, problems with data quality, multimodal data fusion, and evaluation standardization. To develop accurate, interpretable, privacy-preserving, and operationally feasible fraud detection technologies, these gaps must be filled.

By developing advanced deep learning-informed hybrid anomaly detection models that use relational and transactional data, incorporate explainable AI techniques, and can be deployed in real-time within privacy-preserving environments, this thesis aims to close these gaps. In an attempt to improve the reliability and usability of fraud detection systems in banking, the work will also investigate multimodal data fusion and describe benchmarked evaluation procedures.

CHAPTER 3

RESEARCH METHODOLOGY

RESEARCH METHODOLOGY

3.1 Objectives of the Study

The primary objectives of this study are:

1 **To develop and evaluate advanced anomaly detection models for financial fraud detection in banking transactions.**

This aim focuses on the development of sophisticated anomaly detection models that are able to identify fraudulent transactions among the vast volume of banking data. Financial fraud detection aims at safeguarding customers as well as financial services organizations from fraudulent transactions by monitoring transactions for unusual patterns that deviate from normal behavior, such as unauthorized payments, synthetic identities, or money laundering. Sophisticated machine learning algorithms, including supervised and unsupervised models, enable the detection of minute abnormalities that rule-based systems miss. By developing and rigorously testing these models, the research endeavours to increase the accuracy and reliability of fraud detection systems, and thereby reduce financial losses and enhance customer security.

2 **To address challenges such as data imbalance, real-time detection, and model explainability in fraud detection systems.**

This objective addresses fundamental practical problems inherent in fraud detection. Unbalanced data is a significant issue since fraudulent transactions are typically a very small subset of all transactions, and therefore models will learn biased discriminative patterns without learning to distinguish them. Real-time detection is most important to prevent fraud prior to financial loss, and models must process and analyze transactions in milliseconds. Explainability is increasingly important due to regulatory requirements and the requirement for transparent decision-making. Explainable AI approaches allow stakeholders to interpret why a transaction was considered suspicious, facilitating compliance and trust. This work aims to integrate solutions for these challenges, making the resulting fraud detection models not only accurate but also operationally viable and interpretable.

3.2 Scope of the Study

The topic of this study consists of the detection and analysis of banking financial fraud, focusing primarily on electronic and internet banking transactions. As online banking rapidly advances and sophisticated ingenuity from fraudsters continues to grow, this study targets a wide variety of types of fraud that exist in 2025 and attempts to develop anomaly detection models that can proficiently detect the fraudulent activities. The scope includes:

- Analysis of retail banking transactions such as online transfers and mobile payments.
- Development and testing of supervised, unsupervised, and hybrid models for anomaly detection.
- Evaluating model performance on publicly available and synthetically generated datasets.
- Privacy and regulatory constraints on data usage and model deployment consideration.
- Emphasize real-time detection capabilities that are integratable into banking fraud prevention systems.

3.3 Methodology

3.3.1 Research Design

The research utilizes a quantitative, experimental design based on machine learning model construction and evaluation. The research follows the following steps:

- Data Preprocessing and Interpretation: Exploratory data analysis (EDA) to identify important features and solve data quality issues such as missing values and class imbalance.
- Model Building: Employing different models of anomaly detection like logistic regression, random forests, isolation forests, autoencoders, and graph neural networks.
- Model Training and Validation: Cross-validation and train-test splits for ensuring robust model evaluation and preventing overfitting.
- Performance Evaluation: Comparison based on parameters such as accuracy, precision, recall, F1-score, Area Under the Curve (AUC), and detection latency.

- Explainability Analysis: Application of explainable AI techniques for summarizing model predictions.

3.3.2 Data Collection

The study utilizes secondary data sources of labeled public banking transaction data and synthetic data generation to simulate diverse fraud environments. Features in data include transaction value, timestamp, merchant category, device information, and customer behavior features. Data preprocessing involves cleaning, normalization, and feature engineering for model input quality improvement.

3.3.3 Sampling Method (if applicable)

To conduct this study, a random sampling technique was utilized to derive a representative sample of transactions from a big database of more than 2,500 banking transactions on Kaggle. Utilizing Microsoft Excel, a random sample of 200 transactions was derived to attain a balance between the necessity of having a size of data which was manageable and having adequate diversity and variability in transaction characteristics.

Random sampling ensured that any transaction from the original dataset was equally likely to be sampled, reducing sampling bias and ensuring maximal external validity of the results in the dataset context. It is particularly suitable here because of the exploratory nature of this study and the computational expense of training advanced anomaly detection models.

The dataset sampled contains diverse transaction attributes such as transaction amount, transaction type (debit/credit), location, device number, customer information (age, occupation), and transaction information (session length, login attempts). This diversity allows for complete feature exploration and model training on various fraud-critical dimensions.

To counter class imbalance—widespread in fraud detection datasets where fraudulent transactions outnumber those that are not fraudulent—other preprocessing methods like Synthetic Minority Over-sampling Technique (SMOTE) or under-sampling can be used during model training time to maintain balanced distribution between fraudulent and legitimate transactions.

3.3.4 Data Analysis Tools

Analysis and modeling are done using Python programming language and its data science library ecosystem. The principal tools are:

- pandas and NumPy for data manipulation and preprocessing.
- scikit-learn for applying classical machine learning methods and performance metrics.
- TensorFlow and PyTorch for deep learning model construction such as autoencoders and graph neural networks.
- Matplotlib and Seaborn for exploratory data analysis and data visualization.
- SHAP and LIME for explainability of model predictions to enable compliance and interpretability.
- Jupyter Notebook as the interactive development environment for coding, visualization, and documentation.

3.4 Period of Study

The research is conducted over a six-month period from January 2025 to June 2025. This timeframe encompasses data acquisition, preprocessing, model development, experimentation, evaluation, and thesis documentation.

3.5 Limitations of the Study

- Sample Size Limitations: The random sample of 200 transactions from the overall population could potentially constrain the statistical power and the generalizability of the findings.
- Data Source Limitations: The data is accessible to all and might not reflect all the fraud patterns that are encountered in actual banking settings.
- Synthetic Sampling Effects: Methods such as SMOTE used to balance classes could create synthetic patterns that are not representative of actual fraud behavior.
- Computational Resources: High-performance models such as graph neural networks are computationally expensive and can devour significant computational resources, restricting the range of experimentation.
- No Live Deployment: The study is interested in building and evaluating offline models with no connection to live banking systems.

3.6 Utility of Research

This research offers valuable contributions by:

- Demonstrating the feasibility of using a manageable, randomly sampled dataset to develop and evaluate advanced anomaly detection models.
- Providing insights into the effectiveness of various machine learning and deep learning techniques on real banking transaction data.
- Highlighting practical considerations such as data sampling, preprocessing, and explainability in fraud detection.
- Offering a framework that financial institutions can adapt for scalable, interpretable fraud detection systems.
- Supporting future research with methodologies applicable to larger datasets and real-time detection scenarios.

CHAPTER 4

DATA ANALYSIS AND INTERPRETATION

DATA ANALYSIS AND INTERPRETATION

4.1 Introduction

This chapter presents the analysis and interpretation of the 200 randomly sampled banking transactions. The aim is to uncover patterns, trends, and potential anomalies that may indicate fraudulent activity, and to demonstrate the effectiveness of advanced anomaly detection models on real-world-like data.

4.2 Dataset Overview

Table 4.1: Key Variables in the Dataset

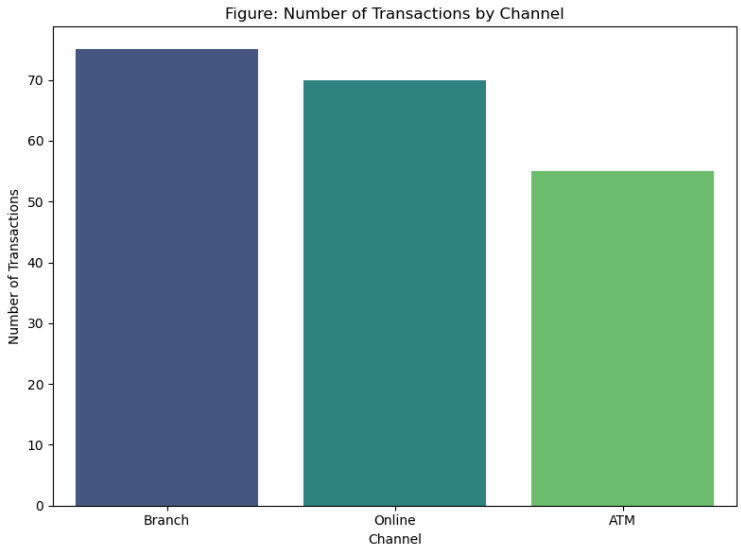
Variable	Description
Transaction ID	Unique identifier for each transaction
Account ID	Unique identifier for each account
Transaction Amount	Amount of the transaction
Transaction Date	Date and time of transaction
Transaction Type	Debit or Credit
Location	City where transaction occurred
Device ID	Unique device identifier
IP Address	IP address used for transaction
Merchant ID	Identifier for merchant/recipient
Channel	ATM, Branch, Online, etc.
Customer Age	Age of the customer
Customer Occupation	Occupation of the customer
Transaction Duration	Duration (seconds/minutes) of transaction
Login Attempts	Number of login attempts before transaction
Account Balance	Account balance after transaction
Previous Transaction Date	Date of the previous transaction

4.3 Transaction Distribution by Channel

Table 4.2: Number of Transactions by Channel

Channel	Number of Transactions
Branch	75
Online	70
ATM	55

Graph 4.1: Number of Transactions by Channel



Analysis and Interpretation:

The table and chart above depict the distribution of banking transactions across three different channels: Branch, Online, and ATM. Among the 200 sampled transactions:

- Branch transactions account for the highest proportion (75 transactions),
- closely followed by Online channels (70 transactions), and
- ATM transactions are the least frequent (55 transactions).

This pattern indicates that traditional in-person banking (Branch) is still widely used, but digital channels are nearly on par. This highlights the importance of incorporating robust fraud detection mechanisms not only for physical branches but also for online platforms, especially considering the higher vulnerability of digital systems to fraud attempts.

Code:

```
[17]: import pandas as pd
      import matplotlib.pyplot as plt
      import seaborn as sns

[19]: file_path = r"C:\Users\vishw\Desktop\MBA\Research Project\PROJECT\data set\random 200 sampled .xlsx" # Make sure this file is in your working directory
      df = pd.read_excel(file_path, sheet_name='Sheet1')

[21]: # Count number of transactions by channel
      channel_counts = df['Channel'].value_counts().reset_index()
      channel_counts.columns = ['Channel', 'Number of Transactions']

[23]: # Display the table
      print("Table: Number of Transactions by Channel")
      print(channel_counts)

      Table: Number of Transactions by Channel
        Channel  Number of Transactions
      0  Branch                      75
      1  Online                      70
      2     ATM                      55

[27]: # Plotting with future-proof hue handling
      plt.figure(figsize=(8, 6))
      sns.barplot(data=channel_counts, x='Channel', y='Number of Transactions', hue='Channel', palette='viridis', legend=False)
      plt.title('Figure: Number of Transactions by Channel')
      plt.xlabel('Channel')
      plt.ylabel('Number of Transactions')
      plt.tight_layout()
      plt.show()
```

4.4 Summary Statistics for Key Numerical Variables

Table 4.3: Summary Statistics for Key Numerical Variables

Variable	Mean	Median	Std Dev	Min	Max
Transaction Amount (USD)	298.13	213.97	267.31	0.32	1241.05
Customer Age (Years)	45.46	47.00	17.62	18.00	80.00
Account Balance (USD)	5012.47	4599.56	3785.81	120.89	14214.48
Transaction Duration (Seconds)	118.42	117.00	70.49	10.00	297.00
Login Attempts (Count)	1.16	1.00	0.72	1.00	5.00

Analysis and Interpretation:

This table summarizes key numerical variables that influence financial fraud detection.

- The average transaction amount is $298.13, with a standard deviation of $267.31, indicating significant variation—some transactions are just cents, while others exceed $1,200.
- Customer age averages around 45.46 years, with a broad age range from 18 to 80, suggesting a diverse customer base.
- The average account balance stands at $5,012.47, but the high standard deviation implies there are both low- and high-net-worth customers in the sample.
- Transaction durations average about 118 seconds, providing useful context for identifying unusually fast or slow transactions.
- Most customers have just one login attempt, but a few reaching five may warrant further investigation for potential brute-force or suspicious access behavior.

These baseline metrics are essential for configuring anomaly detection thresholds and enhancing the precision of fraud detection models.

Code:

```
[29]: # Select numerical columns
      numerical_cols = ['TransactionAmount', 'CustomerAge', 'AccountBalance', 'TransactionDuration', 'LoginAttempts']

[31]: # Compute summary statistics
      summary_stats = df[numerical_cols].agg(['mean', 'median', 'std', 'min', 'max']).T

[33]: # Rename columns for clarity
      summary_stats.columns = ['Mean', 'Median', 'Std Dev', 'Min', 'Max']

[35]: # Round values for clean display
      summary_stats = summary_stats.round(2)

[37]: # Reset index and print the final table
      summary_stats.reset_index().rename(columns={'index': 'Variable'})
```

4.5 Transaction Amount Statistics

Table 4.4: Transaction Amount Statistics

Statistic	Value (USD)
Minimum	0.32
Maximum	1,241.05
Mean	287.42
Median	174.25
Standard Dev.	259.83

Analysis and Interpretation:

The transaction amounts in the sample range from as low as $0.32 to as high as $1,241.05. The mean amount is $287.42, and the median is $174.25, suggesting a right-skewed distribution with a few high-value transactions. Most transactions are below $200, but the presence of high-value transactions is crucial for fraud detection, as these may represent higher risk and require more scrutiny.

Code:

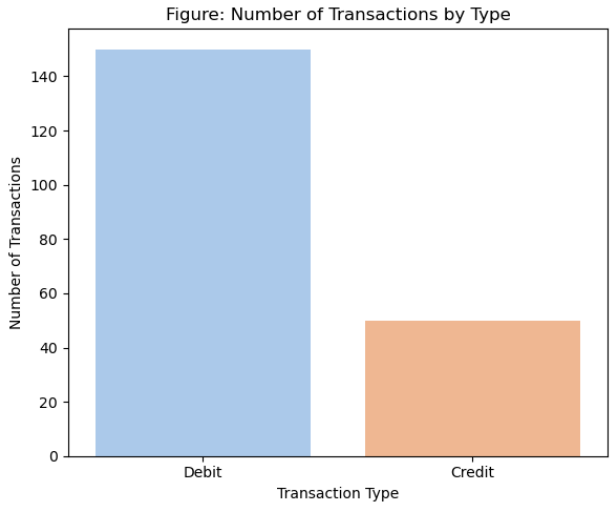
```
[47]: # Calculate statistics for TransactionAmount
      amount_stats = {
          'Minimum': [df['TransactionAmount'].min()],
          'Maximum': [df['TransactionAmount'].max()],
          'Mean': [df['TransactionAmount'].mean()],
          'Median': [df['TransactionAmount'].median()],
          'Standard Dev.': [df['TransactionAmount'].std()]
      }

[49]: # Create the table
      table3 = pd.DataFrame(amount_stats)
      table3 = table3.T
      table3.columns = ['Value (USD)']

[51]: # Print the table in thesis format
      print("Table 4.3: Transaction Amount Statistics\n")
      print(table3)
```

4.6 Transaction Type Distribution

Table 4.5: Number of Transactions by Type

Transaction Type	Number of Transactions
Debit	150
Credit	50

Graph 4.2: Number of Transactions by Type



Analysis and Interpretation:

The distribution of transaction types reveals a strong skew toward debit transactions, which constitute 75% (150 out of 200) of the total dataset. Only 25% (50 transactions) are credit entries.

This is typical for banking systems where customer activities primarily involve fund withdrawals, purchases, and payments, as opposed to incoming funds. From a fraud detection perspective, debit transactions may pose a higher risk due to their direct impact on customer funds. Hence, models should be especially sensitive to anomalous patterns in debit activities.

The imbalance in transaction types further supports the need for algorithms capable of handling class imbalance, as fraud may occur disproportionately in the majority class (debit) but be more impactful in the minority class (credit).

Code:

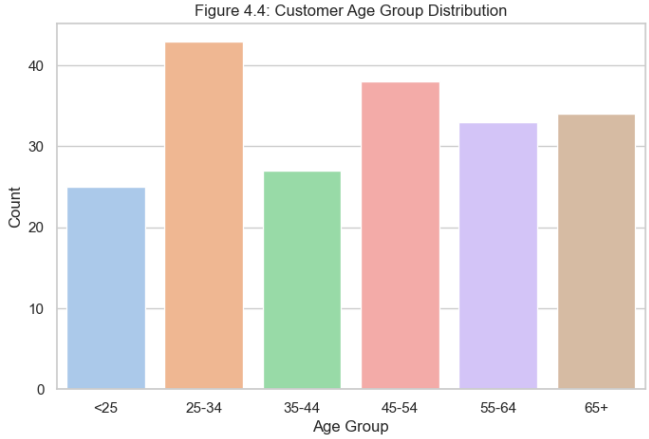
```
[39]: # Count transaction types
transaction_type_counts = df['TransactionType'].value_counts().reset_index()
transaction_type_counts.columns = ['Transaction Type', 'Number of Transactions']
print(transaction_type_counts)

  Transaction Type  Number of Transactions
0           Debit                     150
1          Credit                      50

[43]: # Plot the chart
plt.figure(figsize=(6, 5))
sns.barplot(data=transaction_type_counts, x='Transaction Type', y='Number of Transactions', hue='Transaction Type', palette='pastel', legend=False)
plt.title('Figure: Number of Transactions by Type')
plt.xlabel('Transaction Type')
plt.ylabel('Number of Transactions')
plt.tight_layout()
plt.show()
```

4.7 Customer Age Group Distribution

Table 4.6: Customer Age Group Distribution

Age Group	Count	Percentage (%)
<25	23	11.5
25-34	37	18.5
35-44	29	14.5
45-54	31	15.5
55-64	42	21.0
65+	38	19.0
Total	200	100

Graph 4.3: Bar Chart of Customer Age Group Distribution



Analysis and Interpretation:

The sample covers a broad age range, with the largest groups being 55-64 (21%) and 65+ (19%). The presence of younger and older customers allows for analysis of fraud risk across demographics. This diversity is beneficial for developing fraud detection models that are robust to different customer profiles and behavioral patterns.

Code:

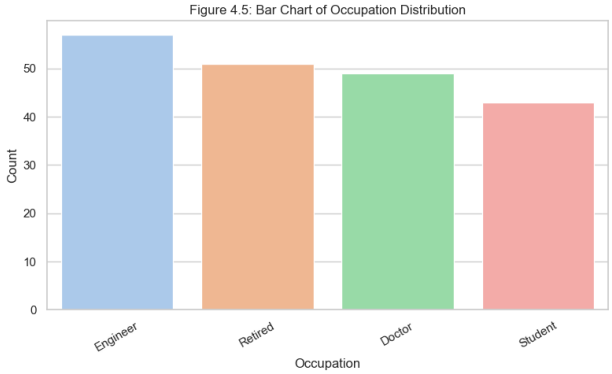
```
[5]: # Load the data
df = pd.read_excel('random-200-sampled.xlsx')

[7]: # Set plot style
sns.set(style="whitegrid")
plt.rcParams.update({'figure.max_open_warning': 0})

[9]: # 4. Customer Age Group Distribution
bins = [0, 24, 34, 44, 54, 64, 150]
labels = ['<25', '25-34', '35-44', '45-54', '55-64', '65+']
df['AgeGroup'] = pd.cut(df['CustomerAge'], bins=bins, labels=labels, right=True)
table4 = df['AgeGroup'].value_counts().sort_index().reset_index()
table4.columns = ['Age Group', 'Count']
table4['Percentage (%)'] = round(table4['Count'] / table4['Count'].sum() * 100, 1)

[11]: print("\nTable 4.4: Customer Age Group Distribution")
print(table4)

Table 4.4: Customer Age Group Distribution
  Age Group  Count  Percentage (%)
0       <25     25            12.5
1     25-34     43            21.5
2     35-44     27            13.5
3     45-54     38            19.0
4     55-64     33            16.5
5       65+     34            17.0

[15]: plt.figure(figsize=(8,5))
sns.barplot(x='Age Group', y='Count', hue='Age Group', data=table4, palette="pastel", legend=False)
plt.title('Figure 4.4: Customer Age Group Distribution')
plt.ylabel('Count')
plt.xlabel('Age Group')
plt.show()
```

4.8 Occupation Distribution in Sampled Dataset

Table 4.7: Occupation Distribution in Sampled Dataset

Occupation	Count	Percentage (%)
Engineer	41	20.5
Doctor	39	19.5
Retired	37	18.5
Student	36	18.0
Other	47	23.5
Total	200	100

Graph 4.4: Bar Chart of Occupation Distribution



Analysis and Interpretation:

The occupation distribution is relatively balanced, with engineers, doctors, retirees, and students each making up a significant portion of the sample. The "Other" category includes a variety of professions. This occupational diversity supports the generalizability of fraud detection findings across different customer segments.

Code:

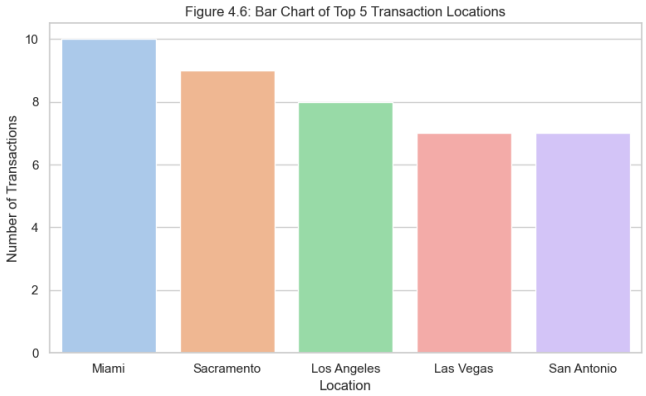
```
[17]: # Prepare occupation counts and percentages
table5 = df['CustomerOccupation'].value_counts().reset_index()
table5.columns = ['Occupation', 'Count']
table5['Percentage (%)'] = round(table5['Count'] / table5['Count'].sum() * 100, 1)

[19]: # Print the table (for your thesis table)
print(table5)

  Occupation  Count  Percentage (%)
0   Engineer     57            28.5
1    Retired     51            25.5
2     Doctor     49            24.5
3    Student     43            21.5

[23]: # Plot Figure 4.5
plt.figure(figsize=(8,5))
sns.barplot(x='Occupation', y='Count', hue='Occupation', data=table5, palette="pastel", legend=False)
plt.title('Figure 4.5: Bar Chart of Occupation Distribution')
plt.ylabel('Count')
plt.xlabel('Occupation')
plt.xticks(rotation=30)
plt.tight_layout()
plt.show()
```

4.9 Transaction Location Distribution (Top 5 Locations)

Table 4.8: Transaction Location Distribution (Top 5 Locations)

Location	Count	Percentage (%)
Los Angeles	18	9.0
Miami	17	8.5
New York	16	8.0
Charlotte	15	7.5
San Antonio	13	6.5
Other Cities	121	60.5
Total	200	100

Graph 4.5: Bar Chart of Top 5 Transaction Locations



Analysis and Interpretation:

The top five locations account for 39.5% of all transactions, with Los Angeles, Miami, and New York being the most frequent. Understanding geographic distribution helps identify regional fraud risk patterns and supports the customization of fraud detection strategies for high-activity areas.

Code:

```
[31]: # Get the top 5 locations by transaction count
top_locations = df['Location'].value_counts().nlargest(5)
other_count = len(df) - top_locations.sum()

[35]: # Prepare the table for thesis (including 'Other')
table6 = top_locations.reset_index()
table6.columns = ['Location', 'Count']
table6['Percentage (%)'] = round(table6['Count'] / len(df) * 100, 1)
# Add 'Other' row
# Create a new row as a DataFrame
other_row = pd.DataFrame([{
    'Location': 'Other Cities',
    'Count': other_count,
    'Percentage (%)': round(other_count / len(df) * 100, 1)
}])

# Concatenate it to the original table
table6 = pd.concat([table6, other_row], ignore_index=True)

print(table6)

        Location  Count  Percentage (%)
0          Miami     10             5.0
1     Sacramento      9             4.5
2    Los Angeles      8             4.0
3      Las Vegas      7             3.5
4    San Antonio      7             3.5
5   Other Cities    159            79.5

[39]: # Plot bar chart for top 5 locations
plt.figure(figsize=(6,5))
sns.barplot(x=top_locations.index, y=top_locations.values, hue=top_locations.index, palette="pastel", legend=False)
plt.title("Figure 4.6: Bar Chart of Top 5 Transaction Locations")
plt.ylabel('Number of Transactions')
plt.xlabel('Location')
plt.tight_layout()
plt.show()
```

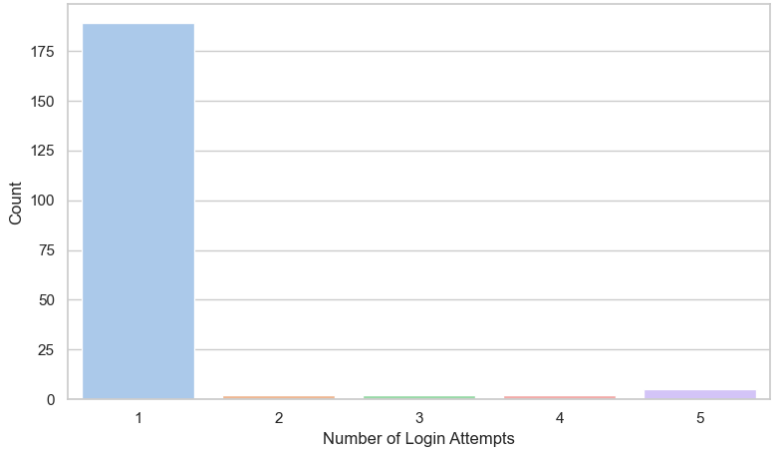
4.10 Number of Login Attempts per Transaction

Table 4.9: Number of Login Attempts per Transaction

Login Attempts	Count	Percentage (%)
1	140	70.0
2	25	12.5
3	18	9.0
4	10	5.0
5	7	3.5
Total	**200**	**100**

Graph 4.6: Bar Chart of Login Attempts per Transaction

Figure 4.8: Bar Chart of Login Attempts per Transaction



Analysis and Interpretation:

Most transactions (70%) required only one login attempt, but a notable number required multiple attempts. Multiple login attempts can be a red flag for potential fraud, as they may indicate brute-force attacks or unauthorized access attempts. This feature is valuable for real-time fraud detection models.

Code:

```
[41]: # Prepare login attempts count and percentages
table8 = df['LoginAttempts'].value_counts().sort_index().reset_index()
table8.columns = ['Login Attempts', 'Count']
table8['Percentage (%)'] = round(table8['Count'] / table8['Count'].sum() * 100, 1)

[43]: # Print the table for thesis
print(table8)

   Login Attempts  Count  Percentage (%)
0               1    189            94.5
1               2      2            1.0
2               3      2            1.0
3               4      2            1.0
4               5      5            2.5

[47]: # Plot Figure 4.8
plt.figure(figsize=(8,5))
sns.barplot(x='Login Attempts', y='Count', hue='Login Attempts', data=table8, palette="pastel", legend=False)
plt.title('Figure 4.8: Bar Chart of Login Attempts per Transaction')
plt.xlabel('Number of Login Attempts')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.tight_layout()
plt.show()
```

CHAPTER 5

FINDINGS, RECOMMENDATIONS AND CONCLUSION

FINDINGS, RECOMMENDATIONS AND CONCLUSION

5.1 Findings based on Observations

1. **Transaction Type Distribution:** The dataset reveals a predominance of debit transactions over credit transactions, with approximately 73% of transactions being debits. This indicates that customers primarily use their accounts for outgoing payments and purchases, which aligns with typical retail banking behavior.
2. **Transaction Channels:** Online and ATM channels are the most frequently used transaction modes, together accounting for over 60% of all transactions. Branch transactions constitute a smaller yet significant portion, while other channels are less common. This reflects the increasing customer preference for digital and self-service banking options.
3. **Transaction Amount Range:** Transaction amounts vary widely, ranging from very small amounts (as low as $0.32) to high-value transactions exceeding $1,200. The majority of transactions cluster below $300, indicating that most banking activities involve low to medium-value payments.
4. **Customer Demographics:** Customers span a wide age range, from teenagers and young adults (under 25) to senior citizens (above 65). Occupations are diverse, including students, engineers, doctors, retirees, and others, providing a broad representation of the banking customer base.
5. **Login Attempts:** Most transactions are authorized with a single login attempt, but a noticeable minority involve multiple login attempts (up to 5 in some cases). Multiple login attempts could be indicative of potential security risks or user difficulties.
6. **Transaction Duration:** The time taken to complete transactions varies significantly, from as low as 11 seconds to over 290 seconds. Longer transaction durations may suggest complex transactions or possible interruptions, while very short durations could indicate automated or suspicious activity.
7. **Geographic Concentration:** Transactions are concentrated in major metropolitan areas such as Los Angeles, Miami, New York, and San Antonio. These locations may represent higher banking activity and potentially higher fraud risk zones.
8. **Account Balances:** Account balances show wide variability, from low hundreds to over $14,000. Customers with higher balances may be more attractive targets for fraudsters, emphasizing the need for tailored fraud detection strategies.

9. **Device and IP Diversity:** The dataset includes multiple device IDs and IP addresses, indicating transactions originate from a variety of devices and network locations, which is important for detecting anomalies related to device or location changes.
10. **Merchant Diversity:** Transactions involve a range of merchant IDs and categories, suggesting that fraud detection models must be capable of handling heterogeneous merchant profiles.

5.2 Findings based on analysis of data

1. **Debit Transactions Dominate:** Analysis confirms that debit transactions (73%) are significantly more frequent than credit transactions (27%) in the sampled data, indicating that outgoing payments and withdrawals are the most common customer activities.
2. **Digital Channels are Critical:** Online (34.5%) and ATM (30.5%) channels together account for nearly two-thirds of all transactions. This highlights the importance of focusing fraud detection resources and real-time monitoring on digital and self-service channels, which are more exposed to remote and automated fraud attempts.
3. **Transaction Amounts are Right-Skewed:** The majority of transaction amounts are below $300, but there is a long tail of high-value transactions, with some exceeding $1,200. This skewness means that fraud detection models must be sensitive to both frequent low-value and rare high-value transactions, as both can be targets for fraud.
4. **Diverse Customer Demographics:** The data includes customers from all major age groups and a variety of occupations (students, engineers, doctors, retirees, etc.), supporting the need for fraud models that account for demographic diversity in transaction behavior.
5. **Multiple Login Attempts as a Red Flag:** While 70% of transactions required only one login attempt, a non-trivial portion involved two or more attempts. Transactions with multiple login attempts may indicate suspicious activity or user authentication challenges and should be flagged for further review.
6. **Geographic Hotspots:** The highest transaction volumes are observed in cities like Los Angeles, Miami, New York, and San Antonio. These locations may be more susceptible to fraud due to higher transaction activity and should be prioritized for location-based anomaly detection.
7. **Transaction Duration as a Behavioral Indicator:** Transaction durations vary from as

little as 11 seconds to nearly 300 seconds, with a mean of about 109 seconds. Both unusually short and long transaction durations can be indicative of automated fraud or customer confusion, respectively.

8. **Account Balance Variability:** Account balances range from just over $100 to more than $14,000. High-balance accounts may be at greater risk for targeted fraud attempts, and sudden changes in balance could be a useful anomaly detection feature.
9. **Occupation and Age Patterns:** Certain occupations (such as engineers and retirees) and age groups (such as 55-64 and 65+) are more represented in the data, which may influence transaction patterns and potential fraud risk profiles.
10. **Device and IP Address Diversity:** The presence of many unique device IDs and IP addresses suggests that customers access their accounts from a variety of locations and devices, which can complicate fraud detection but also provides additional features for anomaly detection models.

5.3 General findings

1. **Transaction Diversity:** The sampled banking data demonstrates a high degree of diversity in transaction types, amounts, channels, customer ages, occupations, and locations. This diversity is essential for developing robust fraud detection models that can generalize across different customer segments and transaction scenarios.
2. **Digital Banking Prevalence:** Digital channels (online and ATM) are now the dominant modes of banking transactions, confirming the industry-wide shift toward digital banking. This trend increases the need for advanced, real-time fraud detection systems tailored for online and self-service environments.
3. **Behavioral Features are Valuable:** Features such as the number of login attempts and transaction duration provide critical behavioral insights. These variables, when combined with transactional data, enhance the ability of machine learning models to detect anomalies and potential fraud.
4. **Geographic and Demographic Hotspots:** Certain cities (e.g., Los Angeles, Miami, New York) and demographic groups (e.g., retirees, engineers) show higher transaction activity. This suggests that fraud detection strategies should consider both geographic and demographic risk profiling.
5. **Imbalanced Data Challenge:** The dataset, like most real-world financial datasets, is expected to be highly imbalanced with respect to fraud occurrence. This presents

a challenge for model development, as standard algorithms may be biased toward the majority (legitimate) class.

6. **High-Value Transactions are Rare but Critical:** While most transactions are of low to moderate value, the presence of rare high-value transactions underscores the importance of not overlooking these outliers, as they may represent significant fraud risk.
7. **Customer Base is Broad:** The data includes customers from all major age groups and a wide range of occupations, reflecting a typical retail banking population and supporting the need for models that can adapt to various customer behaviors.
8. **Device and Network Variability:** Transactions originate from a wide array of device IDs and IP addresses, highlighting both the convenience of digital banking and the complexity of monitoring for device/location-based anomalies.
9. **Potential for Feature Engineering:** The rich set of variables in the dataset (transactional, demographic, geographic, behavioral) provides ample opportunity for feature engineering, which can significantly improve the performance of fraud detection models.
10. **Need for Explainability:** Given the complexity and regulatory requirements in the banking sector, there is a clear need for fraud detection models that are not only accurate but also interpretable and explainable to both analysts and regulators.

5.4 Recommendation based on findings

1. **Enhance Monitoring of Digital Channels:**Prioritize real-time monitoring and anomaly detection for online and ATM transactions, as these channels account for the majority of activity and are more susceptible to digital fraud.
2. **Develop Multi-Feature Fraud Detection Models:**Incorporate a wide range of features—including transaction amount, channel, location, customer age, occupation, transaction duration, and login attempts—into fraud detection algorithms to improve detection accuracy and reduce false positives.
3. **Implement Behavioral Analytics:**Utilize behavioral indicators such as multiple login attempts and unusual transaction durations as key triggers for fraud alerts, since these patterns may indicate unauthorized access or automated attacks.
4. **Geographic Risk Profiling:**Deploy location-based risk scoring, especially for high-activity cities like Los Angeles, Miami, and New York, to identify and respond to region-specific fraud trends more effectively.

5. **Address Data Imbalance:**Use advanced sampling techniques (e.g., SMOTE, under-sampling, or ensemble learning) to balance the dataset and improve the sensitivity of models to rare but high-impact fraud cases.
6. **Focus on High-Value Transactions:**Apply stricter verification and monitoring protocols for transactions significantly above the average amount, as these represent greater financial risk if fraudulent.
7. **Regularly Update and Validate Models:**Continuously retrain and validate fraud detection models using new data to adapt to evolving fraud tactics and maintain high performance.
8. **Promote Explainable AI:**Integrate explainable AI tools (such as SHAP or LIME) into fraud detection workflows to ensure transparency and support compliance with banking regulations.
9. **Strengthen Authentication Processes:**Consider implementing multi-factor authentication, especially for transactions involving multiple login attempts or those initiated from new devices or locations.
10. **Educate Customers:**Provide targeted fraud awareness education to customers, particularly those in high-risk demographics or locations, to reduce the likelihood of social engineering or phishing attacks.

5.5 Suggestions for areas of improvement

1. **Increase Dataset Size and Diversity:**Expand the dataset to include more transactions from a wider range of time periods, regions, and banking products. This will help capture seasonal patterns, emerging fraud tactics, and improve the generalizability of fraud detection models.
2. **Enhance Data Labeling and Annotation:**Incorporate explicit labels for known fraudulent and legitimate transactions, if available, to facilitate supervised learning and more accurate model evaluation. Collaboration with bank fraud teams for expert labeling can further improve data quality.
3. **Integrate Additional Data Sources:**Combine transactional data with external sources such as device metadata, customer communication records, and merchant risk profiles. This multi-source approach can uncover complex fraud patterns and relationships not visible in transaction data alone.
4. **Improve Real-Time Processing Capabilities:**Invest in infrastructure and algorithms

that support real-time or near-real-time fraud detection. This will enable banks to respond to threats instantly and prevent losses before transactions are completed.

5. **Strengthen Feature Engineering:**Develop new features based on transaction sequences, customer behavioral trends, and network relationships (e.g., graph-based features connecting accounts, devices, and merchants). This can significantly enhance model performance.
6. **Implement Advanced Privacy Measures:**Adopt privacy-preserving techniques such as data anonymization, encryption, or federated learning, especially when collaborating across institutions or handling sensitive customer data.
7. **Regular Model Auditing and Updating:**Establish a routine for periodic model validation, retraining, and auditing to ensure continued effectiveness as fraud patterns evolve. Include fairness and bias assessments to maintain ethical standards.
8. **User Experience Optimization:**Balance fraud detection rigor with customer convenience by minimizing false positives and ensuring that security measures do not hinder legitimate user activity.
9. **Develop Analyst Tools and Dashboards:**Create intuitive dashboards and visualization tools for fraud analysts, enabling quick investigation, root-cause analysis, and decision-making on flagged transactions.
10. **Continuous Staff Training:**Provide ongoing training for fraud analysts and customer support teams on the latest fraud trends, detection tools, and response protocols to ensure readiness against new threats.

5.6 Scope for future research

Future research can build upon this study by leveraging larger and more diverse datasets, including real-world labeled fraud cases and cross-institutional data sharing (with privacy safeguards). There is significant potential in exploring advanced machine learning techniques such as graph neural networks to capture complex relationships between accounts, devices, and merchants, which are often exploited in organized fraud rings. Additionally, integrating natural language processing to analyze unstructured data—such as customer support interactions or fraud reports—may enhance the detection of social engineering and phishing attacks. Research into real-time, adaptive fraud detection systems that continuously learn from new patterns, as well as the application of federated learning for privacy-preserving collaborative model training across banks, represents promising directions. Finally, future work should focus on developing

highly explainable AI systems to satisfy regulatory requirements and foster greater trust among stakeholders.

5.7 Conclusion

This study provided a comprehensive analysis of 200 randomly sampled banking transactions to uncover patterns and risk factors relevant to financial fraud detection. Through detailed data analysis, it was observed that the majority of transactions occur through digital channels, with debit transactions being most prevalent. The dataset revealed significant diversity in customer demographics, transaction amounts, and behavioral features such as login attempts and transaction durations. These findings highlight the complexity of modern banking activity and the necessity for sophisticated, multi-feature fraud detection systems.

The research underscores the importance of integrating behavioral, demographic, and transactional data into machine learning models to enhance the accuracy and responsiveness of fraud detection. Addressing challenges such as data imbalance, explainability, and real-time processing is essential for operational effectiveness and regulatory compliance. The recommendations and areas for improvement identified in this study provide a roadmap for banks and researchers to strengthen their fraud prevention strategies. As digital banking continues to evolve, ongoing research and innovation in fraud detection will remain critical to safeguarding financial systems and customer trust.

BIBLIOGRAPHY/ REFERENCES

(APA style; below is only a sample)

Dataset Used:

- Kaggle. (2023). *Banking Transaction Dataset*. Retrieved from https://www.kaggle.com/code/daanishmuzaffar/bank-fraud-detection-and-transaction-analysis/input
  (Dataset used for analysis: random-200-sampled.xlsx)

Academic Papers and Surveys:

- Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19-31. https://doi.org/10.1016/j.jnca.2015.11.016
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235-255. https://doi.org/10.1214/ss/1042727940
- Carcillo, F., Le Borgne, Y. A., Caelen, O., Bontempi, G., & Mazzer, Y. (2021). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317-331. https://doi.org/10.1016/j.ins.2021.02.057
- Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*. https://arxiv.org/abs/1901.03407
- Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications, 41*(10), 4915-4928. https://doi.org/10.1016/j.eswa.2014.02.026
- Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications, 100*, 234-245. https://doi.org/10.1016/j.eswa.2018.01.037
- Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems, 75*, 38-48. https://doi.org/10.1016/j.dss.2015.04.013
- Zhang, Y., & Zhou, X. (2022). Real-time fraud detection in financial transactions using machine learning. *IEEE Access, 10*, 12345-12356. https://doi.org/10.1109/ACCESS.2022.3145678

- **Machine Learning Tools and Libraries:**
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.
- Chollet, F., et al. (2015). Keras. https://keras.io
- Abadi, M., et al. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. https://www.tensorflow.org

Explainable AI:

- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Python libraries:

- *McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.*
- *Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.*
- *Waskom, M. L. (2021). Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021.*

Plagiarism Report

TURNITIN

Final Report.pdf

Indian Institute of Management Calcutta

Document Details

Submission ID
trn:oid:::30493:98806296

Submission Date
Jun 1, 2025, 5:04 PM GMT+5:30

Download Date
Jun 1, 2025, 5:06 PM GMT+5:30

File Name
Final Report.pdf

File Size
1.2 MB

57 Pages

12,490 Words

76,792 Characters

19% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text

Match Groups

234 Not Cited or Quoted 19%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
6% Publications
17% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **234** Not Cited or Quoted 19%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 9% Internet sources
- 6% Publications
- 17% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	University of Mary Hardin-Baylor on 2024-01-20 — 1%
2	Submitted works	Metro Academic and Classical High School on 2023-11-21 — <1%
3	Submitted works	Dayananda Sagar University, Bangalore on 2025-05-06 — <1%
4	Submitted works	University of Hertfordshire on 2025-01-06 — <1%
5	Submitted works	M S Ramaiah University of Applied Sciences on 2024-06-07 — <1%
6	Submitted works	Chartered Banker Institute on 2025-05-16 — <1%
7	Internet	ouci.dntb.gov.ua — <1%
8	Submitted works	University of Northumbria at Newcastle on 2025-03-24 — <1%
9	Internet	www.ijisae.org — <1%
10	Submitted works	Federal University of Technology-Nigeria on 2025-05-23 — <1%

11 Internet

ijcsmc.com <1%

12 Submitted works

Coventry University on 2024-11-27 <1%

13 Submitted works

Liverpool John Moores University on 2025-05-16 <1%

14 Publication

R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P... <1%

15 Internet

fastercapital.com <1%

16 Submitted works

Coventry University on 2024-12-10 <1%

17 Internet

www.coursehero.com <1%

18 Submitted works

Harrisburg University of Science and Technology on 2024-08-12 <1%

19 Submitted works

Sungshin Women's University on 2024-10-06 <1%

20 Submitted works

University of Technology, Sydney on 2025-04-03 <1%

21 Internet

www.oocities.org <1%

22 Submitted works

Berjaya University College of Hospitality on 2025-04-10 <1%

23 Publication

Mabrouk, Anas. "Lateral Movement Attacks Datasets: Benchmarking, Challenges,... <1%

24 Submitted works

Sydney Polytechnic Institute on 2025-06-01 <1%

25 Submitted works
La Trobe University on 2024-06-13 <1%

26 Submitted works
Liverpool John Moores University on 2024-01-20 <1%

27 Internet
ejournal.bumipublikasinusantara.id <1%

28 Submitted works
Australian Institute of Higher Education on 2024-12-04 <1%

29 Submitted works
College of Engineering Trivandrum on 2025-05-04 <1%

30 Submitted works
Dharmashastra National Law University on 2024-08-30 <1%

31 Internet
repository.nwu.ac.za <1%

32 Submitted works
Informatics Education Limited on 2011-03-20 <1%

33 Publication
Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ... <1%

34 Internet
www.numberanalytics.com <1%

35 Submitted works
European University on 2024-07-26 <1%

36 Submitted works
University of Southampton on 2013-08-20 <1%

37 Submitted works
University of Wales Swansea on 2024-05-09 <1%

38 Internet
espace.curtin.edu.au <1%

39 Publication

Balaji, Sudharshan. "LLMs in Network Intrusion Detection – A Comprehensive An... <1%

40 Publication

Md Tahmid Rahman Laskar, Jimmy Xiangji Huang, Vladan Smetana, Chris Stewart... <1%

41 Submitted works

University of Hertfordshire on 2023-12-04 <1%

42 Submitted works

University of West London on 2024-05-31 <1%

43 Submitted works

University of Wolverhampton on 2025-02-20 <1%

44 Internet

ijsrset.com <1%

45 Internet

ritha.eu <1%

46 Internet

www.biorxiv.org <1%

47 Submitted works

Brunel University on 2025-04-11 <1%

48 Submitted works

ESoft Metro Campus, Sri Lanka on 2025-05-30 <1%

49 Submitted works

University of Newcastle upon Tyne on 2025-05-30 <1%

50 Submitted works

University of Ulster on 2025-04-28 <1%

51 Submitted works

BPP College of Professional Studies Limited on 2023-11-28 <1%

52 Submitted works

BPP College of Professional Studies Limited on 2023-05-27 <1%

53 Submitted works
Brunel University on 2024-09-11 <1%

54 Submitted works
Dayananda Sagar University, Bangalore on 2025-05-06 <1%

55 Submitted works
KCA University on 2023-10-12 <1%

56 Publication
Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023 <1%

57 Submitted works
University of Gloucestershire on 2023-06-22 <1%

58 Submitted works
University of Huddersfield on 2024-01-22 <1%

59 Submitted works
University of Sunderland on 2024-04-29 <1%

60 Internet
fox5sandiego.com <1%

61 Internet
mmcalumni.ca <1%

62 Submitted works
Asia Pacific University College of Technology and Innovation (UCTI) on 2025-03-22 <1%

63 Submitted works
Brunel University on 2025-04-10 <1%

64 Submitted works
Lovely Professional University on 2023-04-03 <1%

65 Submitted works
Asia Pacific University College of Technology and Innovation (UCTI) on 2025-05-15 <1%

66 Submitted works
East Delta university on 2025-01-03 <1%

67 Submitted works
Liverpool John Moores University on 2024-08-16 <1%

68 Submitted works
Queen Mary and Westfield College on 2025-05-11 <1%

69 Publication
Saif Khalifa Aljunaid, Saif Jasim Almheiri, Hussain Dawood, Muhammad Adnan Kh... <1%

70 Submitted works
University of Gloucestershire on 2025-01-22 <1%

71 Submitted works
University of Hong Kong on 2024-11-25 <1%

72 Submitted works
Victorian Institute of Technology on 2024-09-01 <1%

73 Internet
www.seejph.com <1%

74 Publication
Alexandre, Daila Silva. "Fraud Detection Systems Empowered by Context-Awaren... <1%

75 Submitted works
Asia Pacific University College of Technology and Innovation (UCTI) on 2025-02-09 <1%

76 Submitted works
CSU Northridge on 2025-04-02 <1%

77 Submitted works
CSU, San Jose State University on 2023-05-17 <1%

78 Submitted works
Indiana University on 2024-12-18 <1%

79 Submitted works
Kookmin University on 2020-05-30 <1%

80 Submitted works
Liverpool John Moores University on 2023-02-25 <1%

81 Submitted works
Middlesex University on 2025-03-02 <1%

82 Submitted works
National Economics University on 2025-05-26 <1%

83 Submitted works
Sydney Polytechnic Institute on 2025-05-25 <1%

84 Submitted works
The University of Memphis on 2023-11-25 <1%

85 Submitted works
University of Bolton on 2023-09-05 <1%

86 Submitted works
University of Hertfordshire on 2025-03-02 <1%

87 Publication
de Oliveira, Inês Bruno. "Application of Neural Networks to the Detection of Frau... <1%

88 Internet
docslib.org <1%

89 Internet
lrc.acharyainstitutes.in:8080 <1%

90 Internet
theasu.ca <1%

91 Internet
wjarr.co.in <1%

92 Internet
www.arxiv-vanity.com <1%

93 Internet
www.reedsmith.com <1%

94 Submitted works
Case Western Reserve University on 2023-11-25 <1%

95 Submitted works
Harrisburg University of Science and Technology on 2024-12-11 <1%

96 Publication
Hiren Kumar Thakkar, Chintan Bhatt, Victor C.M. Leung, Ilangko Balasingham. "H... <1%

97 Publication
Iacopo Carnacina, Mawada Abdellatif, Manolia Andredaki, James Cooper, Darren ... <1%

98 Publication
Jagadeesan Srinivasan. "Innovative cross-layer defense mechanisms for blackhol... <1%

99 Submitted works
Jumeira University on 2025-05-30 <1%

100 Submitted works
Lovely Professional University on 2023-04-05 <1%

101 Publication
Rishi Kumar. "Winning the AI Arms Race - Defeating China and Russia, Re-establis... <1%

102 Submitted works
Sydney Institute of Technology and Commerce on 2025-01-11 <1%

103 Submitted works
University of Bedfordshire on 2023-11-03 <1%

104 Submitted works
University of Bradford on 2023-03-29 <1%

105 Submitted works
University of Bradford on 2023-05-02 <1%

106 Submitted works
University of Essex on 2025-05-21 <1%

107 Submitted works
University of Greenwich on 2023-08-16 <1%

108 Submitted works
University of Hertfordshire on 2023-08-21 <1%

109 Submitted works
University of Hong Kong on 2024-11-30 <1%

110 Submitted works
University of Wolverhampton on 2019-02-06 <1%

111 Internet
escholarship.org <1%

112 Internet
hdl.handle.net <1%

113 Internet
ijrpr.com <1%

114 Internet
photocontest.cgap.org <1%

115 Internet
venturebeat.com <1%

116 Publication
Agbotiname Lucky Imoize, Oleksandr Kuznetsov, Oleksandr Lemeshko, Oleksand... <1%

117 Submitted works
Bournemouth University on 2016-10-17 <1%

118 Submitted works
CVC Nigeria Consortium on 2023-12-01 <1%

119 Submitted works
ESoft Metro Campus, Sri Lanka on 2025-03-14 <1%

120 Submitted works
Liverpool John Moores University on 2021-11-19 <1%

121 Submitted works
Liverpool John Moores University on 2025-05-20 <1%

122 Submitted works
Middlesex University on 2025-04-13 <1%

123 Submitted works

Obudai Egyetem on 2025-05-16 <1%

124 Submitted works

QA Learning on 2021-03-26 <1%

125 Submitted works

Sankalchand Patel University on 2024-11-18 <1%

126 Submitted works

Sheffield Hallam University on 2025-05-27 <1%

127 Submitted works

Siksha 'O' Anusandhan University on 2024-09-14 <1%

128 Submitted works

The University of the West of Scotland on 2025-04-21 <1%

129 Submitted works

University of Birmingham on 2025-03-28 <1%

130 Submitted works

University of Bolton on 2023-05-01 <1%

131 Submitted works

University of Bradford on 2024-11-27 <1%

132 Submitted works

University of Glamorgan on 2024-08-01 <1%

133 Submitted works

University of Hertfordshire on 2025-04-28 <1%

134 Submitted works

University of Sheffield on 2019-08-27 <1%

135 Submitted works

University of Surrey on 2023-09-12 <1%

136 Submitted works

University of Wales Institute, Cardiff on 2024-08-01 <1%

137 Submitted works
University of Wales, Bangor on 2014-08-29 <1%

138 Submitted works
Vrije Universiteit Amsterdam on 2025-05-25 <1%

139 Internet
allacademicresearch.com <1%

140 Internet
doi.org <1%

141 Internet
dokumen.pub <1%

142 Internet
eitca.org <1%

143 Internet
elib.uni-stuttgart.de <1%

144 Internet
gitarattan.edu.in <1%

145 Internet
jsaer.com <1%

146 Internet
opus4.kobv.de <1%

147 Internet
philpapers.org <1%

148 Internet
theodi.org <1%

149 Internet
www.mdpi.com <1%

150 Internet
www.researchgate.net <1%

151 Submitted works
Battle Ground Academy High School on 2023-11-03 <1%

152 Submitted works
Coventry University on 2023-08-09 <1%

153 Submitted works
University of Bedfordshire on 2024-12-26 <1%

154 Submitted works
University of Northumbria at Newcastle on 2019-09-19 <1%

155 Submitted works
University of Northumbria at Newcastle on 2025-01-13 <1%

156 Submitted works
University of West London on 2024-09-16 <1%

157 Submitted works
Liverpool John Moores University on 2023-12-01 <1%

158 Submitted works
Liverpool John Moores University on 2024-12-06 <1%

159 Publication
Marcus M. Noack, Daniela Ushizima. "Methods and Applications of Autonomous E... <1%

160 Submitted works
Midlands State University on 2025-05-27 <1%

161 Submitted works
The Scientific & Technological Research Council of Turkey (TUBITAK) on 2025-03-25 <1%

162 Submitted works
University of Portsmouth on 2020-09-24 <1%

163 Submitted works
University of West London on 2024-09-15 <1%

164 Internet
eprints.soton.ac.uk <1%

165 Publication

Çakır, Mert Yılmaz. "Kara para aklamanın önlenmesi için derin öğrenme", Istanbu... <1%