IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Online Banking Fraud Detection Using Machine Learning

**Anisha Santosh Lanke**
Department Of Computer Science And Application, JSPM University Pune.

**Ajay Nagne**
Assistant Professor, Faculty of Science And Technology, JSPM University Pune.

**Abstract**: As digital banking becomes increasingly central to financial activity, the risk of fraud has grown significantly. Traditional fraud detection methods, which often rely on static rules, are proving insufficient against the sophisticated strategies of modern cybercriminals. This research focuses on utilizing machine learning (ML) approaches to enhance the detection of fraudulent behaviour in online banking systems. By examining extensive transaction datasets, ML models can learn to recognize unusual patterns that may signal fraud, allowing for timely and accurate detection. The system developed in this study applies supervised learning techniques such as logistic regression, decision trees, and random forest classifiers trained on labelled data to distinguish between legitimate and suspicious transactions. Through careful feature extraction and data pre-processing, the models achieve improved precision and recall. Findings indicate that machine learning offers a dynamic and efficient framework for combating fraud, ultimately strengthening the safety of online banking and fostering greater user confidence.

Index Terms: Online banking, fraud detection, machine learning, financial fraud, Decision Tree, Random Forest, SVM, Neural Networks, XG Boost, Transactions, Risk Scoring, User Behaviour Analysis, Transaction Data, Time-Series Analysis, Python, Tensor Flow.

## I)Introduction:

The rise of online banking has significantly changed how people and businesses manage their financial activities, delivering unmatched convenience, speed, and around-the-clock access. However, this rapid digital evolution has also made banking systems more attractive targets for cybercriminals. As fraud techniques become more advanced, attackers exploit system weaknesses, leading to serious financial damage and a decline in user trust [5], [15]. Traditional fraud detection methods—often rule-based and manually updated—have difficulty keeping up with the fast-changing tactics used by fraudsters[11] .Machine learning (ML) provides a dynamic and data-driven approach to addressing this issue. By analysing large volumes of past transaction data, ML algorithms can detect patterns and outliers that might indicate fraudulent behaviour [2], [4]. Unlike rigid rule-based systems, ML models continuously learn and improve, allowing them to better detect subtle and emerging fraud scenarios while reducing false alarms [18]. Commonly used techniques such as logistic regression, decision trees, support vector machines, and ensemble models have proven effective in identifying fraud in real-time environments[10], [12]. It reviews key ML algorithms, data pre-processing methods, and performance evaluation metrics used to build smart and adaptive detection systems. Ultimately, the objective is to show how machine learning can strengthen the security of digital banking platforms while offering a more scalable and proactive defence against financial fraud [17]

## II) System Architecture & Methodology:

System Architecture :The system developed for detecting fraud in online banking using machine learning is designed as a structured, multi-layered pipeline. Each layer performs a critical function, helping to ensure that fraudulent activity is identified accurately and efficiently. Below is a simplified breakdown of the system's core components:

1. Data Collection :This stage gathers all relevant data needed for analysis. It includes transaction records, customer profiles, device and browser details, IP addresses, and any previous instances of fraud. The system works with both well-structured data (like amounts, timestamps, and locations) and less structured behavioural data (such as usage patterns and login frequency).
2. Data Pre-processing :Before feeding data into a machine learning model, it must be cleaned and organized. This step involves :Cleaning: Removing duplicate entries and filling in or discarding missing data. Normalization : Making sure all numeric data is on a consistent scale. Encoding : Converting non-numeric (categorical) data into numeric form using techniques like label or one-hot encoding. Feature Engineering: Creating new features that help the model better understand user behaviour for example, calculating how often a user transacts or the time between their last few transactions.
3. Fraud Detection Engine: The Machine learning is a heart of the system. It has two main parts: Training Module: The model is trained using historical data that's labelled to show which transactions were fraudulent and which weren't.

Common algorithms used include Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), XG Boost, and Neural Networks [4], [7], [10], [19].
Prediction Module: Once trained, the model can be used to check new transactions in real time (or in batches), flagging any that look suspicious.
4. Evaluation: To measure how well the model performs, it is tested using several key metrics: Accuracy (overall correctness), Precision (correct fraud predictions), Recall (how many actual frauds are caught), F1-score (balance between precision and recall), and ROC-AUC (model discrimination ability).Techniques like K-Fold Cross-Validation and Confusion Matrices are used to confirm[12], [13]. that the model isn't just memorizing the data but can generalize to new cases.
5. Alert and Response: If the system detects a suspicious transaction, it automatically sends an alert. These alerts can be reviewed by the bank's fraud team or used to trigger automatic responses like account locks or transaction blocks.
6. Monitoring and Reporting: The system includes a user-friendly dashboard where analysts and administrators can track model performance, monitor detected fraud, and generate regular reports [15], [17]. for further investigation or policy adjustments.

- Methodology

Steps: Each step ensured that the final model was accurate, efficient, and adaptable to real-world banking environments.
Step 1: Collecting the Data: The process starts by gathering real or publicly available datasets that include labelled transaction data [1]. (i.e., marking which transactions are fraudulent). An example of such a dataset is the credit card fraud dataset available on Kaggle.
Step 2: Preparing the Data :Next, the data is cleaned and prepared for modelling :Unnecessary or duplicate data entries are removed .Outliers and missing values are handled carefully.
Feature selection or reduction techniques, such as PCA (Principal Component Analysis), may be applied to simplify [3]. the dataset while keeping useful patterns intact.
Step 3: Exploring the Data :Exploratory Data Analysis (EDA) is performed to understand how the data behaves. This involves visualizing distributions, identifying correlations between features, and discovering trends that might help the model distinguish between fraud and normal activity.
Step 4: Choosing the Right Model: Several machine learning algorithms are selected and compared to find the best fit. Simpler models like logistic regression are used as baselines, while more complex ones like random forests or XG Boost are tested for better performance. Hyper parameter tuning methods such as grid search [7], [12] are used to improve model settings.
Step 5: Training and Testing :The model is trained on the training set and evaluated on the test set using appropriate performance metrics to make sure it works well and isn't over fitting[12], [18].
Step 6: Deploying the Model :Once the best-performing model is selected, it is integrated into the banking environment. It can be used to analyse transactions in real time or in scheduled batches [10]. Thresholds are set to determine when a transaction should be flagged as potentially fraudulent.
Step 7: Feedback and Continuous Learning :The system is updated regularly using new data and feedback from flagged transactions—whether they were actually fraud or false alarms. This helps keep the model current and improves accuracy over time as fraud tactics evolve.

## III) Standard Databases For Model Training & Evaluation:

1. IEEE-CIS Fraud Detection Dataset
Where to find it: Available on Kaggle.
What it is: This dataset contains information about online payment transactions, including both legitimate and fraudulent transactions. Size: Around 500,000 transactions .What's in it: The data includes features like transaction ID, amount, device info, product codes, email addresses, and more. Why use it: It's one of the most comprehensive datasets for fraud detection [5], [8],[15]. Things to know: The dataset's feature names are anonym zed, so you'll need to interpret them during preprocessing.
2. Credit Card Fraud Detection Dataset
Where to find it: Available on Kaggle.
What it is: This dataset contains transactions from European credit card holders from 2013.
What's in it: Features include time, amount, and a class label (0 = non-fraud, 1 = fraud), along with anonym zed features.
Why use it: It's widely used for benchmarking machine learning models due to its accessibility. Things to know: It's an imbalanced [1], [12]. dataset, so the models need to handle this properly.
3. PaySim (Synthetic Dataset)
Where to find it: Check it out on GitHub.
What it is: This dataset is synthetic, created using a simulation of real mobile money transactions.
Size: It has around 6 million transactions. What's in it: You get transaction types, amounts, sender/receiver balances, and a fraud flag .Why use it: It's great for situations where you don't have access to real-world data, and it mimics real transaction patterns. Things to know: While the data is synthetic, it's realistic enough for fraud detection experiments.
4. BankSim Dataset: Where to find it: It's available on Kaggle.
What it is: Another synthetic dataset designed to simulate bank transaction behaviors.
What's in it: The dataset includes transaction type, time, location, and transaction amount.
Why use it: Perfect for testing fraud detection algorithms when real data isn't available[6].
Things to know: This one is also synthetic, so while it's not real data, it's designed to reflect real-world patterns.
5. Financial Datasets from Open ML & UCI: There are also smaller financial datasets available on platforms like Open ML and the UCI Machine Learning Repository. They can be useful for general fraud detection tasks, but they may not be as specific or large as some of the others listed here. Examples include datasets like "Credit Approval" or "German Credit", which can be used to practice classification algorithms.
Important Evaluation Metrics:
When you start building models for fraud detection, you'll want to keep an eye on a few key evaluation metrics:
- Precision, Recall, and F1-Score are essential to understand model performance in fraud detection.
- The ROC-AUC and PR-AUC (Precision-Recall AUC) are commonly used to assess how well your model handles the imbalanced classes.

## IV)Applications & Uses Case:

In today's digital age, online banking has become the norm. But with the convenience also comes risk—fraudulent activities are on the rise, and financial institutions are under pressure to stay one step ahead. This is where machine learning (ML). steps in as a game-changer. Machine learning gives banks and payment platforms the ability to detect fraud faster, more accurately, and in real-time, unlike traditional rule-based systems that struggle to adapt. Let's explore some of the practical ways ML [1]. is used to catch online fraud before it causes harm:

1. Real-Time Fraud Detection: One of the biggest strengths of machine learning is spotting fraud as it happens. By continuously analyzing transaction data, ML algorithms can flag anything that seems unusual the moment it occurs.

Example: Imagine someone trying to transfer a large amount of money from your account in a different country than you usually transact from. The system sees this as out of the ordinary and instantly blocks the payment or asks for extra verification.

2. Understanding Customer Behavior: Machine learning doesn't just look at what's happening—it understands how you usually behave. It learns things like your spending habits, login times, devices, and even location.

Example: If you typically log in from Mumbai around 9 AM using your phone, but suddenly there's a login attempt at 2 AM from another country, the system picks up on this unusual pattern and flags it as suspicious.

3. Scoring Transaction Risk: Every time a transaction happens, machine learning can assign a risk score to it. This score helps decide whether the transaction should go through or be reviewed further.

Example: A large money transfer to a first-time payee in a high-risk location may be given a high fraud score and paused until verified.

4.Spotting Outliers with Anomaly Detection: Some fraudulent transactions don't look obviously fake—they just seem a little "off". ML models can notice these rare, unexpected patterns and bring them to attention.

Example: A normally quiet account suddenly starts receiving small deposits from multiple foreign sources. That's unusual and might indicate money laundering.

5. Keeping Up with New Fraud Tactics :Fraudsters keep changing their methods, and that's why machine learning's ability to learn from new data is so valuable. Instead of relying on fixed rules, ML updates itself as new fraud patterns appear.

Example: If a new trick for phishing or account takeover starts circulating, the model can learn to catch it early based on fresh data.

6. Strengthening Identity Verification: ML also helps verify who is actually using the bank account—not just by passwords, but by tracking unique behavioral data like typing speed or how a person moves their mouse.

Example: If a hacker gets your login details but uses the keyboard or app differently than you normally do, the system might ask them to verify their identity another way.

7. Helping Analysts Investigate Faster; Machine learning also helps human fraud analysts by grouping suspicious transactions and presenting them in clear dashboards. This helps teams focus their energy where it's most needed.

Example: A visual tool might show that five different accounts, all created in the same hour, are behaving suspiciously—possibly a fraud ring.

8.Automating Compliance and Reporting: Banks are required to report suspicious activity to regulators. Machine learning simplifies this by automatically generating alerts and summaries when fraud is detected.

Example: If a flagged transaction meets the criteria for legal reporting, the system can instantly create a report for authorities like RBI or other regulatory bodies.

9. Smarter Chat bots for Customer Alerts

: Modern banks use AI-powered chat bots to keep customers informed. When machine learning detects unusual activity, a bot can send a message or even interact to confirm if the transaction was genuine.

Example: You get a text asking, "Did you make this ₹50,000 transaction?" If you reply "No," the bot triggers an investigation or freezes your card.

10. Preventing Account Takeovers: Fraudsters sometimes steal login credentials to take over someone's bank account. Machine learning can pick up on the early signs of such activity and act before serious damage is done.

Example: If someone logs in and immediately tries to change your email, phone number, and PIN, the system sees that as a red flag and blocks access.

- In Summary

Machine learning is transforming how we defend against online banking fraud. Its ability to learn from patterns, adapt to new threats, and act in real-time makes it an essential tool for financial institutions.,

## V) Flowchart:

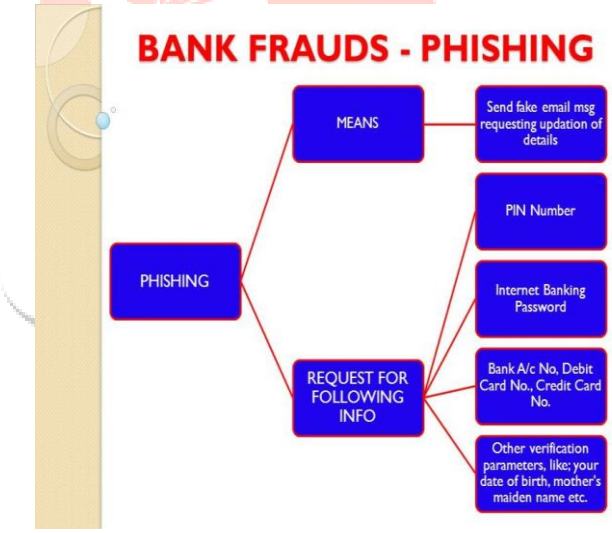


**Fig.Bank fraud phishing**

## VI) Comparative Analysis With Existing technologies:

Online banking fraud is constantly evolving, with scammers coming up with new tricks every day. For a long time, banks have relied on traditional fraud detection methods—mainly based on fixed rules and human judgment. While those approaches worked well in the past, they're now struggling to keep up with today's fast, complex fraud patterns .Let's explore how machine learning (ML) stands apart, and how it improves fraud detection compared to older, rule-based systems.

- Traditional Methods: Rule-Based Systems

These systems follow predefined rules that were set by experts. For example:

"If a user tries to transfer more than ₹1,00,000 to a foreign account, block it."

While this sounds effective, it only works for known fraud scenarios. Anything new or unexpected might slip through unnoticed.

- Benefits:

Easy to understand and explain.

Reliable for spotting fraud that follows a clear pattern.

- Limitations:

Cannot detect new or unusual fraud methods .Often blocks genuine transactions by mistake (false alarms).Needs regular manual updates and supervision.

- Modern Approach: Machine Learning

Machine learning doesn't rely on fixed rules. Instead, it learns from past data—how people usually spend, where they log in from, what devices they use, and more. It uses this knowledge to spot even slight changes in behavior that may signal fraud.

- Benefits:

Learns and improves automatically as it sees more data. Capable of catching unknown fraud patterns. Reduces the number of false alerts, saving time and frustration. Works in real-time, analyzing thousands of transactions instantly.

- Limitations:

Needs high-quality, well-labeled data to be accurate. Can be harder to explain how a decision was made (especially in complex models).Requires technical knowledge to set up and maintain properly. Fast; handles huge data volumes :Setup & Maintenance Easier to set up, but manual Needs data and expertise. Explain ability Very transparent May require added explanation tools

- Real-World Example:

-Let's say someone suddenly transfers ₹2,00,000 from an account that usually only pays small bills.

- A rule-based system might block the transaction, even if it's for something genuine—like paying for a wedding venue.

- A machine learning system would analyze the user's history, device, login location, and past spending. If everything checks out, it might approve the payment or just ask for a quick OTP confirmation.

- This shows how machine learning adds more intelligence and flexibility compared to older systems.

- The Best Approach? Combine Both:

Many banks today use a hybrid system—they keep basic rules in place for common fraud patterns and layer machine learning on top to catch what the rules miss. This combination provides better safety without annoying the customer.

- Conclusion

Traditional fraud detection has served us well, but today's digital threats are more complex and faster than ever. Machine learning brings a smarter, faster, and more adaptive way to protect online banking systems. While rule-based systems are still useful for simple checks, machine learning is the future of fraud prevention—able to evolve, scale, and respond to new challenges in real time.

## VII) Challenges & Future Directions:

Online banking has made financial transactions faster and more convenient. However, it also opens doors for fraud, which continues to grow in scale and complexity. To fight this, banks are turning to machine learning (ML) — a powerful tool that helps detect fraud by spotting unusual behavior in user transactions. While it has brought significant improvements, it's not without challenges, and there's still a long way to go. Here's a breakdown of the current issues and what the future may hold.

- Current Challenges

1. Limited and Unbalanced Data: One of the biggest hurdles is that fraud is relatively rare compared to legitimate transactions. Because of this, ML models often don't get enough fraudulent examples to learn from, making them less accurate. Training on such unbalanced data can lead to missed fraud cases or false alarms.

2. Privacy Concerns: Banking data is extremely sensitive. Due to privacy regulations and ethical concerns, sharing this data (even for research or improvement purposes) is heavily restricted. As a result, developers and researchers often lack access to the real-world data needed to build smarter fraud detection models.

3. Rapidly Changing Fraud Tactics: Fraudsters constantly update their methods to bypass security systems. This creates a challenge for ML models, which might perform well initially but become less effective over time if not regularly updated to adapt to new fraud patterns.

4. False Positives: Sometimes, ML systems incorrectly flag legitimate customer activities as fraud. These false positives not only disrupt customers but also put pressure on fraud investigation teams, increasing operational costs and hurting customer satisfaction.

5. Complexity in Real-World Use:

While building a machine learning model in a lab setting is manageable, deploying it into a live banking environment is far more complex. It needs to be fast, reliable, and work well with existing systems, all while meeting strict regulatory requirements.

6. Lack of Transparency: Many advanced ML models work in ways that are hard to explain, especially to non-technical users. This can be a problem in banking, where it's often necessary to explain why a transaction was flagged as fraudulent — both to customers and regulators.

7. Security Threats to the Model: Ironically, ML models themselves can become targets. Skilled fraudsters may study how these systems work and attempt to trick them by mimicking normal behavior — a tactic known as an "adversarial attack."

- Future Directions: Despite these obstacles, machine learning continues to evolve and improve. Here are some of the promising developments that could shape the future of fraud detection:

1. Advanced and Hybrid Models: More advanced forms of machine learning, like deep learning and hybrid systems (which combine multiple types of models), are being explored. These methods can spot complex patterns that simple models might miss, leading to better fraud detection.

2. Federated Learning: To address data privacy concerns, federated learning allows multiple banks to work together by training models on local data without actually sharing the data itself. This helps improve fraud detection across the industry while keeping customer information private.

3. Better Techniques for Rare Data: New approaches are being developed to deal with the problem of imbalanced data. For instance, creating synthetic fraud samples or using

anomaly detection techniques can help improve the model's ability to recognize fraud, even when the data is limited.

4. Models That Learn Continuously: The future may involve systems that learn in real time. As soon as a new type of fraud is detected, the model can adjust itself immediately. This would make fraud detection far more dynamic and responsive.

5. Collaboration Across Borders: As fraud becomes more global, it's important for banks and cyber security experts to collaborate internationally. Sharing knowledge and fraud patterns (without compromising privacy) could help build more robust and prepared systems.

6. Explainable AI (XAI): There's a growing focus on making machine learning models more transparent. With explainable AI, banks will be better equipped to understand why a decision was made, making it easier to comply with regulations and maintain customer trust.

7. Integration with Biometrics and Behavior Analysis: Future systems may combine ML with biometric tools (like fingerprint or facial recognition) and behavior tracking (like typing speed or device use patterns). These extra layers of verification can make fraud detection even more accurate and harder to bypass.

- Final Thoughts

Machine learning has already transformed how online banking fraud is detected, offering faster and more accurate results than traditional methods. But it's not a perfect solution yet. Challenges like limited data, evolving fraud tactics, and privacy concerns still stand in the way. Looking ahead, the future will likely focus on building smarter, safer, and more adaptable systems. By improving technology, encouraging collaboration, and making AI models more understandable, the banking industry can stay one step ahead in the fight against fraud — protecting both their systems and their customers.

## VIII) Conclusions:

In today's digital age, online banking has become a part of everyday life—but with its rise comes an increase in fraudulent activities. Hackers and scammers are finding smarter ways to exploit online systems, making it harder for traditional fraud detection methods to keep up. Systems that rely on fixed rules or manual checks are no longer effective enough on their own. This is where machine learning is making a real difference. By analyzing patterns in huge amounts of banking data, machine learning models can spot suspicious behavior much faster and more accurately than older methods. These systems don't just follow rules—they learn from past data, adapt to new fraud techniques, and continue to improve over time. That said, using machine learning to fight fraud isn't without its challenges. Getting access to quality data, keeping customer information safe, and explaining how the system makes decisions are all important concerns. Plus, fraud tactics are always changing, so ML systems need regular updates and monitoring .Even with these hurdles, the future looks
promising. As machine learning technology continues to evolve, we can expect fraud detection systems to become even smarter, quicker, and more reliable. There's also a growing focus on making AI more transparent and easier to understand, which is crucial for building trust with customers and meeting legal requirements.

## IX)Acknowledgement:

I would like to take a moment to express my heartfelt thanks to everyone who supported and guided me throughout the journey of working on this project titled "Machine Learning-Based Detection of Fraud in Online Banking."To begin with, I'm deeply thankful to my mentor, Professors, for their consistent support, helpful advice, and encouragement. Their guidance helped me stay focused and understand the real-world importance of this topic beyond just the technical details. I'm also grateful to JSPM University, Pune for providing the learning environment and resources that made this project possible. Having access to the right tools, materials, and facilities was essential in carrying out my research effectively. A big thank-you goes to my friends and classmates who were always ready to share ideas, give feedback, or just offer a word of motivation when I needed it. Their support made this experience much smoother and more enjoyable. Lastly, I truly appreciate the work of researchers and developers in the fields of artificial intelligence and cyber security, whose studies gave me deeper insight into how machine learning is shaping the future of online banking security. Without the support of these individuals and resources, completing this project would not have been possible. Thank you all.

## X)References:

1. Shen & Kurshan (2020)
They developed a smart system using deep reinforcement learning that can automatically adjust fraud detection settings in banking systems.

2. Vivek et al. (2023)
This research shows how real-time data streams can help detect ATM fraud more quickly using machine learning tools.

3. Journal of Big Data (2022)
A study focused on selecting the best data features using genetic algorithms to improve fraud detection accuracy.

4. Tiwari et al. (2021)
A detailed review of how machine learning techniques like decision trees and logistic regression are used in fraud detection systems.

5. Chy (2024)
This paper highlights the growing role of machine learning in staying ahead of constantly evolving online fraud tactics.

6. Zurafshan et al. (2023)
They introduced a method that uses natural language processing (NLP) to detect fraud in online banking transactions and messages.

7. Ghobadi et al. (2015)
A hybrid model combining fuzzy logic and neural networks was used to create a more accurate fraud detection system.

8. Malini & Pushpa (2019)
This work applies K-Nearest Neighbor and outlier detection methods to identify unusual credit card activities.

9. Kannan & Somasundaram (2017)
Focused on identifying money laundering patterns using machine learning clustering algorithms.

10. Panda (2024)
The study presents how machine learning models can be used to detect fraud in everyday banking applications with practical examples.

11. Al-Hashedi & Magalingam (2021)
A comprehensive review of 10 years of research on data mining techniques used in financial fraud detection.

12. Bonkoungou et al. (2024)

They compared different machine learning methods for credit card fraud and discussed which ones work best in real-world cases.
13. Choi (2018)
Explores how artificial intelligence can be combined with Internet of Things (IoT) devices to improve financial fraud detection.
14. Phiri et al. (2024)
A cross-country study comparing fraud detection approaches in South Africa and Spain, with insights into regional challenges.
15. Geddes et al. (2025)
Introduces explainable AI systems that allow banks to detect fraud more transparently using federated learning.
16. MDPI - Risks Journal (2024)
Provides a deep dive into how banks are using machine learning to assess and manage risks, especially related to fraud.
17. MDPI - AI Journal (2023)
Discusses how artificial intelligence can identify suspicious banking activities and prevent fraud before it happens.
18. Ramos et al. (2024)
Presents a comparison between multiple machine learning models and finds which are most effective at catching fake transactions.
19. Dal Pozzolo et al. (2015)
They used a unique method called Isolation Forest to catch fraud by finding data points that behave abnormally.