International Journal of
Business Management and Entrepreneurship

Personalized Fraud Detection in Online Banking Using Sequential Pattern Mining and Machine Learning

Faezeh Kordlu[1], **Mahdi Joneidi Jafari** [2]

Received: **Accepted:**

A B S T R A C T

With the rapid expansion of online banking and the increase in financial transactions on digital platforms, fraud detection in these systems has become a major challenge for financial institutions. Traditional fraud detection methods, which are typically based on predefined rules, are unable to cope with new threats due to the complexity and diversity of frauds and the rapid changes in user behavior. In this paper, a personalized approach for fraud detection in online banking using sequential pattern mining and machine learning is proposed. In this approach, each user's financial transactions are analyzed as time sequences to identify the normal behavioral patterns of each user and extract suspicious deviations. Advanced algorithms like PrefixSpan are used for extracting sequential patterns, and machine learning algorithms such as Random Forest and SVM are employed to classify transactions as fraudulent or non-fraudulent. Experimental results show that the proposed model outperforms traditional methods, with higher accuracy and better ability to detect complex frauds. This research can contribute to improving fraud detection systems in online banking and enhancing the security of financial transactions.

Keywords:

Fraud detection, online banking, sequential pattern mining, machine learning

[1] *Affiliation:* ***Master's student in Information Technology Engineering-Electronic Commerce, Comprehensive*** University the Islamic Revolution , Tehran, Iran; **Emil Address:** Kord.faezeh@cuir.ac.ir

[2] **Affiliation**: Assistant Professor, Faculty of Industrial Engineering and Management, Shahab Danesh University, Qom, Iran; **Emil Address:** joneidi@shdu.ac.ir

[1] Corresponding Author: Faezeh Kordlu

1. Introduction

The rapid expansion of online banking services has transformed the financial industry. According to a report in one of your papers, the widespread use of mobile and online payment services has led to a significant increase in transaction volumes. While these changes have provided greater convenience for users, they have also created more opportunities for fraud. This shift towards digital banking, particularly with the adoption of mobile payments, has created a larger attack surface for cybercriminals, who exploit advanced techniques such as phishing, smishing, and even AI-based fraud to take advantage of users. (George, Alam, and Hasan 2023; León et al. 2020).

With the increasing adaptation of fraudsters to new banking technologies, the complexity of fraudulent activities has also risen. Rule-based fraud detection systems, initially used by many financial institutions, have become inadequate in the face of evolving fraud methods. While these rule-based systems are transparent and understandable, they lack flexibility and, particularly when fraudsters manipulate predefined rules, generate high rates of false positive alerts(George, Alam, and Hasan 2023). In addition, these systems are unable to cope with new fraud methods that involve new types of transactions or cross-border financial activities(Ismail et al. 2015)(León et al. 2020). Fraud detection must keep pace with changes in banking behavior. The shift from static rule-based systems to dynamic and more adaptable approaches that utilize machine learning and artificial intelligence has become essential(Fariha et al. 2025). For example, some institutions now use hybrid deep learning models that combine algorithms such as LSTM and convolutional networks to identify fraudulent transactions by analyzing sequences and patterns of individual transactions(Kim, Jung, and Kim 2022)(Compagnino et al. 2025).

In the past, traditional methods for fraud detection in online banking were primarily based on "rule-based" fraud detection systems. These systems applied a set of fixed rules to identify suspicious transactions. For example, a rule-based system might apply rules such as "transactions above a certain amount must be reviewed" or "transactions from unknown accounts are suspicious." These systems were popular due to their simplicity and transparency, and they offered easier interpretability compared to more complex models(George, Alam, and Hasan 2023). However, rule-based methods quickly faced challenges. The main issue with these systems is that they are unable to update or manage their rules rapidly and generally cannot adapt to new fraud patterns that are constantly created by cybercriminals. These systems operate based on a set of fixed assumptions, making them inefficient when faced with new fraud methods that use more sophisticated techniques, such as phishing and social engineering attacks. Additionally, since these rules are applied universally to all users, they cannot identify individual user-specific behavioral patterns, resulting in a high likelihood of errors in fraud detection(George, Alam, and Hasan 2023; Ismail et al. 2015).

Machine learning models used for fraud detection typically operate without considering the sequence and behavioral traits of users. These models usually analyze transaction data independently, without taking into account the sequence and trends in user behavior. While these models may be able to identify general fraud patterns, they perform poorly in detecting complex and personalized fraud that is associated with changes in user behavior. In such models, no attention is paid to individual-specific trends or the precise sequence of users' transactional behaviors, which can lead to misidentification or failure to detect fraud specific to individual users(Ismail et al. 2015; Kim, Jung, and Kim 2022) . These traditional methods, whether rule-based or simple machine learning models, are not capable of accurately and effectively detecting fraud in the current environment, where frauds have become highly

complex and diverse. As a result, there is an increasing need for approaches that can detect specific behavioral patterns and transaction sequences. To tackle complex and diverse fraud, fraud detection systems must be able to identify personalized behavioral patterns and dynamically respond to changes in user behavior.

Despite significant advances in the use of machine learning and artificial intelligence-based methods for fraud detection, there are still gaps in this area that require further attention and research. One of the biggest challenges in current research is the lack of sufficient focus on the transaction sequences and individual behavioral patterns of each user. Many existing models, even machine learning models, are unable to identify complex fraud patterns in individual user behavior because they typically analyze transactions independently, without considering the sequence and behavioral characteristics of the users(Kim, Jung, and Kim 2022).

This research gap becomes especially apparent as complex and personalized frauds continuously emerge in online banking. In such conditions, fraud detection systems based on machine learning and artificial intelligence models, capable of analyzing and simulating specific user behavioral trends, are urgently needed. One of the major research gaps in this area is that existing models typically cannot identify the subtle differences in individual behavior and, as a result, are unable to detect frauds that are specifically tailored to a particular user(George, Alam, and Hasan 2023).

Another gap in existing research is the lack of personalized models that can accurately simulate the transaction patterns of each individual user. Many current fraud detection systems use a set of common rules and patterns for all users, without considering individual behavioral differences. This issue results in many systems being unable to identify complex and user-specific frauds. Future research in this area should focus on developing personalized systems that can detect fraud patterns based on the transactional behavior of each user(Kim, Jung, and Kim 2022).

Another research gap is the issue of limited access to complete and reliable data for training fraud detection models. Most machine learning models require labeled data for training, which is difficult to obtain in the case of fraud due to the complexity and diversity of fraud methods. Many systems face challenges related to imbalanced data, where fraudulent transactions are significantly fewer in number. This can lead to reduced model accuracy and an increase in false positive alerts(George, Alam, and Hasan 2023; Ismail et al. 2015).

2- **Review of Related Works**

Fraud Detection in Online Banking

In recent years, fraud detection in online banking has become a critical area, as digital transactions have increased and so have fraud threats. Traditional fraud detection systems primarily relied on rule-based methods, which used predefined rules and thresholds to identify suspicious activities. However, these systems have often been ineffective in dealing with complex fraud tactics and the evolving methods of fraudsters. In this regard, the use of machine learning and hybrid models is on the rise, as these models have better adaptability to new fraud patterns. For instance, one study introduced a deep learning framework for detecting fraud in online digital transactions, which showed better accuracy and recall compared to rule-based systems. Another study emphasized the need for hybrid approaches that include supervised models, unsupervised learning, and deep learning, which have demonstrated better performance in handling complex fraud scenarios(Compagnino et al. 2025).

Sequence and Behavioral Modeling

Sequence and behavioral modeling has gained increasing importance in fraud detection, especially as fraud methods have become more sophisticated. Traditional fraud detection methods were unable to model transaction sequences, which led to their inefficiency in identifying complex fraud. Recently, various studies have focused on extracting sequence patterns as a way to model user behavior and identify deviations from typical behavior. For instance, a personalized model for fraud detection using sequential pattern mining was implemented, which outperformed rule-based systems by recording deviations in an individual's transactional behavior, leading to a reduction in false positives(Fariha et al. 2025). Furthermore, integrating behavioral analysis with machine learning models is essential for identifying subtle fraud patterns, especially when transaction data exhibits temporal dependencies(Compagnino et al. 2025).

Limitations of Existing Approaches

Despite advancements in fraud detection systems, significant limitations remain. One of the main issues with current systems, especially rule-based ones, is their inability to dynamically adapt to new fraud patterns. These systems rely on predefined rules, which quickly become obsolete as fraud techniques evolve. The complexity and diversity of modern fraud techniques, including phishing, unauthorized account access, and transaction laundering, make it difficult for these systems to detect them. Additionally, while machine learning models improve in identifying known fraud patterns, they struggle with detecting new and unknown frauds due to data imbalance and the lack of labeled data. Even advanced models, such as deep learning models, face challenges in maintaining accuracy and interpretability, two factors that are crucial for regulatory compliance and consumer trust(Ismail et al. 2015)(George, Alam, and Hasan 2023).

In this section, recent studies on fraud detection in financial transactions are reviewed to identify dominant approaches, the types of data used, and their limitations. The goal is to provide a systematic overview of both conventional and advanced methods to establish the position of our proposed method. The table below summarizes selected papers, the types of data used, key features considered, methods, and their key findings. This review helps us identify the strengths and weaknesses of various approaches and substantiate the framework for selecting the appropriate methodology for our project.

Table 1 - Comparison of Recent Fraud Detection Methods Based on Transaction Sequences

Paper Title (Year)	Method	Data Type	Key Features	Key Results	Short Method	Strengths / Weaknesses
Sequential Pattern Mining… (Sustainability, 2022)	Customized sequential mining; comparison with Markov and rule-based systems	Online banking logs, 1 year	Channel: Yes, Device: Yes/Partial	Better performance than Markov and rule-based systems	Learning common sequences for each user; Alert when new sequence doesn't match	**Strengths**: Personalization and sequential deviations, reduced dependence on manual rules. **Weaknesses**: Raw state modeling, unknown test labels
Multiple	HMM +	Belgium	Amount	+15.1%	Multiple	**Strengths**:

Perspectives HMM… (SAC, 2019)	Random Forest	credit card transactions, 47M transactions	: Yes, Time gap: Yes, Position : Indirect	PR-AUC compared to aggregated features	HMMs for cardholder and terminal; HMM likelihood added to RF	Strong performance, interpretable, sequential model. **Weaknesses**: Heavy feature engineering, HMM assumptions limit model
Credit card fraud detection … (Survey, 2020)	Review: Feature engineering, sequence modeling, HMM, LSTM	Credit card transactions	Normal amount and time gap, other features dataset-dependent	Recommend PR-AUC instead of Accuracy	Summarized challenges : imbalance, dataset shift, missing sequences	**Strengths**: Clear evaluation framework, identifies sequence gaps. **Weaknesses**: No new model, not personalization-focused
Appl. Sci. (2025)	Supervised ML: RF, GBDT, KNN, LR, AdaBoost, SVC	Transaction dataset: N=5000, fraud 2.37%	Not specified	Best: RF Accuracy 0.9838, F1(w) 0.9801, AUPRC 0.6188	Classic evaluation + Recall@K	**Strengths**: Operational PR metrics, top-K. **Weaknesses**: Non-sequential and non-personalized
CCNN-based fraud detection (Mathematics, 2025)	CCNN with time steps T	Kaggle CCFD	Not specified	T=2, Accuracy 0.9998, Recall 1.0	Evaluation with different T, best performance at short horizon	**Strengths**: Strong metrics, explicit time steps. **Weaknesses**: Limited transfer to operational environments, limited interpretability
Edge-attribute GNN… (arXiv, 2019)	GNN with edge features (EdgeProp)	Transaction graph (Ethereum/Tencent)	Amount and time gap Yes	Tencent: Accuracy 0.8698, F1 0.8690	Graph from transactions, node/edge feature extraction, GNN	**Strengths**: Relational/sequential signals, competitive performance. **Weaknesses**: Requires graph, indirect

						personalization
Fraud detection research trends (arXiv, 2025)	Review: Unsupervised + deep learning + hybrid	General fraud detection literature	Not applicable	Emphasized PR-AUC and F1	Trends in methods, evaluation, and interpretability	**Strengths**: Operational summary, useful hybrid. **Weaknesses**: Personalization approach not specified

As shown in Table 1, most existing studies either lack explicit personalization at the user level or do not fully model temporal dependencies, which is the main motivation for the proposed framework in this paper.

Critical Discussion

Transaction Sequence Modeling for Fraud Detection

1. Why Some Methods Perform Better: Key Factors

1.1. The "Appropriate" Behavioral Signal Depends on How Sequences Are Modeled

One of the determining factors in the performance of fraud detection systems is how transaction sequences and temporal dependencies are modeled. Some methods, such as Hidden Markov Models (HMM)-based approaches, have managed to address this gap through feature engineering. In these models, unlike aggregated statistics like the sum of transactions in the past 24 hours, which only provide a general picture of user behavior, the order of transactions is also considered. For example, models that account for temporal dependencies can detect patterns such as "small trial transactions followed by a large transaction." Research has shown that incorporating sequence probabilities can introduce dependencies into the classifier's features, improving the PR-AUC metric by up to 15.1% compared to purely aggregated baselines(Ismail et al. 2015; Wu et al. 2025).

1.2. Personalization Increases Sensitivity to Small and User-Specific Deviations

General models typically assume that normal behavior is the same across all users, which can lead to the loss of individual patterns and an increase in false positives/negatives. Personalized models, designed based on each user's frequent sequences, can address this issue and enhance the system's sensitivity to small but meaningful deviations in individual behavior. Instead of using general rules, these methods use each user's specific frequent sequences as the basis for decision-making, increasing the chances of detecting small yet significant anomalies(Kim, Jung, and Kim 2022).

1.3. The Model Type Is Less Important Than the Alignment of the Metric with Practical Goals

Several studies have shown that high accuracy may not necessarily correlate with effective fraud detection, as fraud is rare and most models struggle due to the scarcity of fraudulent data. For example, models based on PR-AUC and Recall@K, as explored in a study from *Appl. Sci.*, show higher accuracy than other metrics and are more significant in real-world conditions. Therefore, a model may perform better not just because of inherent superiority in its type, but because it is optimized for practical goals.(Lucas et al. 2019).

1.4. The Data Type (Logs vs. Transaction Graph) Alters the Meaning of "Sequence"

Models like EdgeProp perform well for detecting structural and relational frauds that are distributed in a network. These models utilize edge features in transaction graphs. In contrast, personalized models that simulate behavioral sequences for each user are more effective for detecting individual behavioral deviations. These differences indicate that sequence modeling should be tailored based on the data type and the type of fraud being addressed(Li et al. 2018).

2. Limitations of Existing Methods

2.1. Limitations in Features and "Coarse" State Representations

Personalized models in online banking that work with state/event logs may overlook precise numerical and contextual features (such as amounts or small time intervals). On the other hand, transaction-level models work with rich features but ignore the effect of transaction order, which can lead to incorrect or weak fraud detection(Sailusha et al. 2020).

2.2. Evaluation Limitations and Lack of Labels

In some models, test sets include fraudulent transactions, but the labels for some data are unknown. This issue prevents standard metrics from being fully evaluated and makes objective comparison difficult. It creates challenges for the accurate validation of systems(Ismail et al. 2015).

2.3. Scalability and Maintenance Costs

- Rule-based systems require continuous manual updates and are fragile as fraud patterns change.
- Personalized models for each user require more storage and computation to maintain individual sequences or patterns for each customer.
- Graph-based neural models require the construction and updating of large graphs and feature computations, which can be costly in data-stream environments(Wu et al. 2025).

2.4. Model Limitations and Assumptions

HMM models capture temporal dependencies but assume Markov and fixed observation models. This feature may overlook complex multi-factor behaviors. Deeper models can simulate more complex dependencies, but they may lose performance when faced with data drift(Li et al. 2018).

4. Methodology

In online banking systems, each user generates a sequence of financial transactions over time. Assume that each transaction is represented as a dataset containing attributes such as transaction amount, time, channel, geographical location, and device information. For each user, transactions are chronologically ordered, forming a sequence of transactions. The objective of this study is to identify fraudulent transactions by analyzing deviations from the user's normal transactional behavior. This task is formulated as a binary classification problem, where each transaction is labeled as either legitimate (0) or fraudulent (1). The proposed methodology adopts a sequence-based learning framework that combines sequential pattern mining and machine learning techniques to enhance fraud detection accuracy.

In fraud detection systems, modeling transactions as sequential data is a powerful approach that enables deeper analysis of user behavior over time. Transactional data naturally exhibits temporal dependencies and sequential patterns, which cannot be effectively simulated by models that analyze transactions independently. Therefore, representing transactions as sequences allows the system to learn behavioral trends, habits, and temporal regularities specific to each user.

Unlike traditional transaction-level models that rely solely on isolated features such as transaction amount or timestamp, sequential modeling takes into account the order and evolution of transactions, providing a more realistic representation of user behavior. By analyzing transaction sequences, the system can model what constitutes normal behavior for each user and identify significant deviations that may indicate fraud(Kim, Jung, and Kim 2022)(Sami, Mir, and Insany 2025). This type of modeling can identify specific behavioral patterns that may not be observable in individual data points. Each transaction sequence includes the following key features:

1. **Amount**
   The transaction amount is a fundamental feature in financial behavior analysis. Previous studies show that fraudulent activities often follow a specific pattern, where attackers perform low-value transactions to test system vulnerabilities before executing larger fraudulent transfers(Kim, Jung, and Kim 2022)(Sami, Mir, and Insany 2025). Modeling the transaction amounts within sequences allows the system to identify such gradual deviations from normal spending behavior.
2. **Time Interval**
   The time interval between consecutive transactions provides valuable behavioral clues. Unusually short or irregular time gaps between transactions may indicate automated or suspicious activity. By incorporating time intervals into transaction sequences, the model can identify unusual transactions and sudden changes in user habits.(Kim, Jung, and Kim 2022)(Khan et al. 2025).
3. **Channel**
   The transaction channel (e.g., mobile banking, ATM, or online payment) is another important behavioral indicator. Fraudulent transactions often occur through channels that differ from the user's usual usage patterns. Therefore, sudden changes in preferred transaction channels may indicate potential fraud(Kim, Jung, and Kim 2022)(Sami, Mir, and Insany 2025).
4. **Geographical Location**
   Geographical information plays a crucial role in fraud detection. Transactions made from unexpected locations, such as a different city or country, may indicate unauthorized access. Including location features in transaction sequences allows the identification of spatial inconsistencies in user behavior.(Sami, Mir, and Insany 2025)(Wu et al. 2025).
5. **Device**
   Device-related information helps identify anomalies that may result from unauthorized access. Fraudsters often use unfamiliar or previously unseen devices to perform fraudulent transactions. Tracking device usage in transaction sequences enhances the system's ability to detect such irregular behaviors(Wu et al. 2025)(Lucas et al. 2019).

In general, fraud often emerges gradually through subtle behavioral changes rather than sudden deviations. Therefore, sequence modeling is particularly effective for identifying personalized fraud patterns, as it captures subtle temporal changes that transaction-level models may overlook(Kim, Jung, and Kim 2022; Sami, Mir, and Insany 2025).

Sequence Construction and Preprocessing

For effective sequence-based analysis, raw transaction data must undergo systematic preprocessing. This step directly impacts the quality of the extracted patterns and the performance of machine learning models in later stages.

1. **Temporal Sorting of Transactions**
   The first preprocessing step involves sorting transactions chronologically for each user. Temporal sorting is crucial because transaction sequences depend on the correct order of events. This step ensures that behavioral patterns, such as transaction amount increases or abnormal transaction spikes, are accurately captured(Kim, Jung, and Kim 2022).
2. **Noise Removal**
   The second preprocessing step involves removing noise from the data. Transaction data often contains noise that can degrade the quality of the model. This noise may include incorrect data, unrelated transactions, or errors entered into the system. Therefore, this step involves identifying and removing or correcting data that unexpectedly contradicts the user's normal patterns and may affect the model's performance. Removing this noise allows the model to focus on higher-quality data and achieve better accuracy in fraud detection(Kim, Jung, and Kim 2022)(Li et al. 2018).
3. **Windowing Strategy (Optional)**

   For users with a long transaction history, sequences may become excessively large. To address this issue, an optional windowing strategy can be applied to divide long sequences into smaller time windows. This approach allows the model to focus on short-term behavioral changes and more effectively simulate sudden deviations(Wu et al. 2025).

4. **Normalization**

   Normalization of features is applied to ensure a consistent scale for numerical features such as transaction amount and time intervals. Normalization prevents features with large numerical ranges from dominating the model and enables improved convergence and stability of machine learning models. Typically, values are scaled to a fixed range, such as [0, 1] (Kim, Jung, and Kim 2022)(Sailusha et al. 2020).

All of these preprocessing steps have a direct impact on the quality of the patterns. When transaction data is effectively processed, machine learning algorithms can extract better patterns from it, increasing the model's accuracy in fraud detection. Specifically, noise removal and feature normalization can significantly reduce the false positive and false negative rates, making the model more suitable for detecting complex and personalized fraud. Additionally, temporal sorting and the use of a windowing strategy help the model analyze transaction sequences more accurately, enabling better differentiation between normal and abnormal behaviors(Wu et al. 2025)(Kim, Jung, and Kim 2022).

In online fraud detection systems, each user has their own unique behavioral patterns that differ from other users. Therefore, using personalization in behavioral pattern extraction for fraud detection is essential. By modeling each user's behavior separately, it is possible to

simulate their unique behavioral patterns, leading to more accurate detection of specific and personalized fraud. In contrast, in non-personalized systems that use a general pattern for all users, the likelihood of mistakenly identifying a legitimate transaction as fraud increases. Instead of relying on a global behavioral model, the proposed framework learns behavioral patterns individually for each user(Wu et al. 2025)(Kim, Jung, and Kim 2022).

Definition of Normal Behavioral Patterns

Normal behavioral patterns are transaction sequences that a user follows under typical conditions. These patterns are extracted from historical transaction data and reflect normal spending levels, preferred transaction times, commonly used channels, and usual locations. Modeling these personalized patterns allows the system to establish a behavioral baseline for each user(Kim, Jung, and Kim 2022; Wu et al. 2025).

Extraction of Frequent Patterns

To extract normal behavioral patterns, Frequent Sequential Pattern Mining algorithms are used. These algorithms help us identify frequent patterns from a user's transaction sequences. One popular algorithm for this purpose is PrefixSpan, which, through techniques like common prefixes, enables the fast and accurate extraction of frequent sequential patterns. The PrefixSpan algorithm is particularly efficient for large-scale sequential data. By extracting frequent sequences from transactions, it allows the modeling of each user's normal behavior and accurately simulates its variations.

Extraction of Frequent Patterns

To extract normal behavioral patterns, Frequent Sequential Pattern Mining algorithms are used. These algorithms help us identify frequent patterns from a user's transaction sequences. One popular algorithm for this purpose is PrefixSpan, which, through techniques like common prefixes, enables the fast and accurate extraction of frequent sequential patterns. The PrefixSpan algorithm is particularly efficient for large-scale sequential data. By extracting frequent sequences from transactions, it allows the modeling of each user's normal behavior and accurately simulates its variations.(Kim, Jung, and Kim 2022; Lucas et al. 2019).

Difference Between Normal and Suspicious Behavior

After identifying frequent behavioral patterns, new transaction sequences are compared with these patterns. Transactions or subsequences that deviate significantly from the existing norms are flagged as suspicious. For example, a high-value transaction performed through an unfamiliar channel or location can be identified as abnormal. Additionally, sequential pattern mining allows the system to continuously update users' behavioral models, as user habits change over time(Kim, Jung, and Kim 2022),(Lucas et al. 2019).

Feature Engineering from Sequential Patterns

Feature engineering plays a crucial role in transforming sequential patterns into meaningful inputs for machine learning classifiers.

1. **Extraction of Behavioral Features**
   From the extracted transaction sequences, statistical and behavioral features such as

the average transaction amount, transaction frequency, and typical time intervals are derived. These features summarize the user's behavior and provide discriminative signals for fraud detection models(Kim, Jung, and Kim 2022),(Lucas et al. 2019).

2. **Behavioral Change Metrics**
   Features that capture behavioral changes provide crucial information, especially for fraud detection. Sudden deviations in transaction patterns, such as an unexpected increase in spending or unusual transaction timing, are quantified using sequential analysis techniques and added to the feature set.(Lucas et al. 2019),(Li et al. 2018).
3. **Channel and Location-Based Features**
   Changes in transaction channels or geographical locations are encoded as categorical or transition-based features. These features help identify unauthorized access patterns that differ from the user's historical behavior.(Kim, Jung, and Kim 2022; Lucas et al. 2019).
4. **Temporal Summary Features**
   Temporal features such as time of day, day of the week, and intervals between transactions enhance the system's ability to model users' regular habits and identify anomalies during unusual times(Kim, Jung, and Kim 2022; Lucas et al. 2019).

Fraud Classification Based on Machine Learning

After feature extraction, machine learning classifiers are used to distinguish between legitimate and fraudulent transactions. This study employs a supervised learning approach, where labeled transaction data is used for training.
The main algorithms considered are:

- Random Forest, which uses an ensemble of decision trees to capture complex feature interactions.
- Support Vector Machine (SVM), which creates an optimal decision boundary between fraudulent and legitimate transactions.

Traditional transaction-level features and behavioral features extracted from sequences are fed as inputs to these models. The output of each classifier is a binary fraud label indicating whether the transaction is legitimate or suspicious.

Personalized Modeling Strategy vs. Global Modeling

The proposed framework emphasizes personalized modeling, where behavioral patterns are learned individually for each user. Unlike global models that assume behavior is the same across users, personalized models capture individual transaction habits more accurately. While global models have faster training and simpler implementation, personalized models perform better in detecting subtle and user-specific fraud patterns. This trade-off is analyzed in the empirical evaluation section.

Proposed Conceptual Framework for Sequence-based and Personalized Fraud Detection

The proposed framework for identifying fraudulent transactions in online banking is designed to model each user's transactional behavior as a temporal sequence to detect deviations from learned behavioral patterns. Unlike traditional transaction-based models or general models, this sequence-based framework combines sequential pattern extraction and supervised fraud

classification into an integrated pipeline. Figure 1 illustrates the main steps of the framework, which are explained further below.

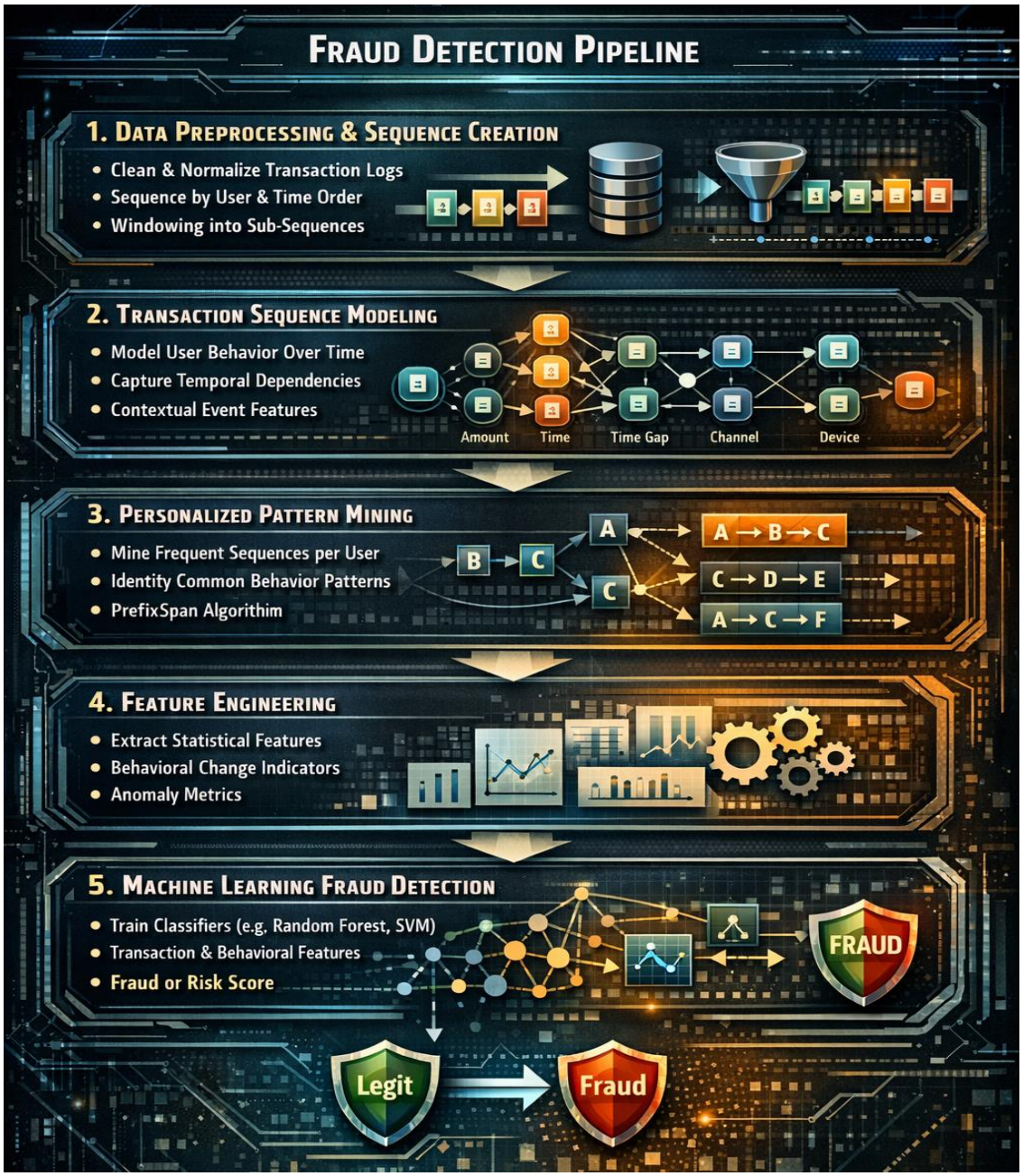

Figure 1: Personalized Fraud Detection Pipeline in Online Banking

1) Data Preprocessing and Sequence Construction

Tasks Performed:

Raw transaction logs are first preprocessed to ensure their integrity and suitability for sequence analysis. Transactions are grouped by user ID and ordered chronologically to create individual transaction sequences. Data cleaning operations are performed to remove noisy or inconsistent records, such as duplicate transactions, missing timestamps, or corrupted values. Numerical features (such as transaction amount or time intervals between transactions) are

normalized, and, if necessary, a windowing strategy is applied to divide long histories into shorter subsequences.

Importance:
Sequence-based fraud detection relies on the correct chronological order of events. Any disruption in the order or scale of transactions can reduce the quality of the extracted behavioral patterns(Kim, Jung, and Kim 2022; Lucas et al. 2019).

The windowing strategy helps the framework focus on recent behavioral dynamics and increases sensitivity to sudden changes in user behavior(Wu et al. 2025).

Contribution to the System:
This step generates clean, time-consistent, user-specific transaction sequences, which form the foundation for subsequent sequence modeling and pattern extraction steps.

2) Transaction Sequence Modeling
Tasks Performed:
Each user's transactions are modeled as a sequence of events, where each event is described by a combination of transaction features such as amount, time interval, channel, location, and device. Instead of independently analyzing transactions, the framework preserves the temporal dependencies between consecutive events.

Importance:
Many traditional systems assume that transactions are independent and overlook temporal dependencies(Lucas et al. 2019). Fraudulent behavior often appears as a gradual deviation, such as small test transactions followed by large, unusual transactions (Kim, Jung, and Kim 2022; Sami, Mir, and Insany 2025).

Contribution to the System:
Modeling the transaction order and temporal context helps in identifying subtle behavioral changes that may not be detectable using non-sequential methods.

3) Extraction of Personalized Frequent Sequential Patterns
Tasks Performed:
Frequent Sequential Pattern Mining algorithms are applied independently to each user's transaction sequences to extract frequent subsequences that define normal behavior. Algorithms like PrefixSpan are used due to their high efficiency and the lack of need for candidate generation.

Importance:
User behavior in online banking is highly diverse, and general models often fail to capture individual norms. Personalizing patterns increases the accuracy of fraud detection(Kim, Jung, and Kim 2022). The extraction of frequent sequential patterns records individual behaviors in an interpretable and updatable manner(Lucas et al. 2019).

Contribution to the System:
This step defines the personalized normal behavior profile. Deviation from these frequent patterns acts as a strong indicator of suspicious activity.

4) Feature Engineering from Sequential Patterns
Tasks Performed:
The extracted patterns are converted into numerical features for machine learning. These features include statistical summaries (e.g., the average transaction amount in frequent patterns), behavioral homogeneity measures, and change indicators that quantify the deviation between new sequences and historical patterns.
Importance:
Combining features extracted from sequences with transaction features enhances the performance of the classifier(Li et al. 2018; Lucas et al. 2019). Behavior Change Indicators provide crucial information, as fraud is often associated with sudden disruptions in normal patterns.
Contribution to the System:
This step bridges the gap between interpretable patterns and predictive models, enabling effective utilization of the extracted behavior for the classifier.

5) Machine Learning-based Fraud Detection
Tasks Performed:
Supervised classifiers like Random Forest or SVM are used to classify transactions as either legitimate or fraudulent. The inputs include transaction features and behavioral features extracted from the sequence, and the output is a fraud label or risk score.
Importance:
Tree-based and margin-based classifiers perform well in fraud detection, especially when enriched with behavioral features(Lucas et al. 2019; Sami, Mir, and Insany 2025).
Contribution to the System:
This step integrates all components into a unified decision-making mechanism, providing accurate, scalable, and operational fraud detection. The framework addresses the limitations of traditional and general models by preserving temporal dependencies, introducing personalized frequent pattern extraction, and combining sequential insights with supervised classification, focusing on each user's behavioral deviations(Kim, Jung, and Kim 2022; Lucas et al. 2019; Sami, Mir, and Insany 2025).

5- Conclusion

In this paper, a conceptual and literature-based framework for personalized fraud detection in online banking, based on sequential pattern mining and machine learning, is presented. The main motivation for this framework stemmed from well-documented limitations in existing fraud detection systems. The proposed framework specifically addresses three recurring gaps in the research literature:
i. The lack of personalization in global fraud detection models,
ii. The weakness in modeling the temporal and sequential behavior of transactions,
iii. The over-reliance on individual statistics at the transaction level.
Building on previous research in sequence modeling and fraud analysis, the proposed framework treats each user's transactional activities as an ordered time sequence and emphasizes extracting frequent sequential patterns unique to each user as a representation of their normal behavior(Kim, Jung, and Kim 2022; Lucas et al. 2019). By considering these patterns as a personalized behavioral baseline, the framework can identify abnormal deviations relative to each user's individual habits, rather than relying on general population

norms. This design directly addresses findings from studies that show global models often fail to capture the heterogeneity of user behaviors, which may result in unnecessary false alerts or failure to detect subtle fraudulent activities(Lucas et al. 2019; Sami, Mir, and Insany 2025). Additionally, the proposed framework combines sequential pattern mining with supervised machine learning, so that the interpretable behavioral knowledge extracted from transaction sequences is transformed into discriminative features for classification. This approach aligns with previous studies that show combining sequence-based features with conventional classifiers can enhance fraud detection capabilities, while maintaining operational flexibility.(Li et al. 2018; Lucas et al. 2019). It is important to emphasize that this research does not claim empirical performance improvement but rather provides a structured methodological perspective that integrates insights from existing research into a coherent chain, suitable for future empirical investigations.

Overall, the main contribution of this paper lies in outlining an integrated conceptual framework that links sequential transaction modeling, personalization, and machine learning, offering a theoretically grounded alternative to the purely global or single-transaction-based approaches currently found in the literature.(Kim, Jung, and Kim 2022; Sami, Mir, and Insany 2025).

6- Future Work

Since this study is conceptual in nature, the proposed framework is intended as a foundation for future empirical validation, rather than a final operational system. The natural next step is to implement and evaluate this framework on real-world online banking datasets, including user transaction histories, temporal data, and contextual features. Such empirical studies would enable the assessment of the framework's feasibility, scalability, and stability in real-world operational conditions.

In future evaluations, it will be essential to use metrics suited to the highly imbalanced nature of fraud detection problems, such as PR-AUC, Recall@K, and false positive rate at fixed alert levels. These metrics have been identified in previous studies as more meaningful indicators than overall accuracy(Kim, Jung, and Kim 2022; Lucas et al. 2019). These metrics provide a better reflection of real-world decision-making needs and enable meaningful comparisons with existing methods.

Additionally, several development paths for the framework can be considered. First, exploring the integration of hybrid models with graph-based approaches for fraud detection, in order to combine personalized sequential patterns with relational risk signals extracted from transaction networks, in line with recent suggestions in graph-based studies(Lucas et al. 2019). Second, studying adaptive or incremental mechanisms in sequential pattern mining to address the phenomenon of concept drift, so that behavioral models can evolve alongside changes in legitimate user habits and fraud strategies(Kim, Jung, and Kim 2022; Sami, Mir, and Insany 2025). Finally, the interpretability of the extracted sequential patterns could be explored in greater depth to support explainable fraud detection—an area that remains a critical requirement in financial applications.

References

Compagnino, Antonio Alessio, Ylenia Maruccia, Stefano Cavuoti, Giuseppe Riccio, Antonio Tutone, Riccardo Crupi, and Antonio Pagliaro. 2025. “An Introduction to Machine Learning Methods for Fraud Detection.” *Applied Sciences* 15(21): 11787. doi:10.3390/app152111787.

Fariha, Nudrat, Md Nazmuddin Moin Khan, Md Iqbal Hossain, Syed Ali Reza, Joy Chakra

Bortty, Kazi Sharmin Sultana, Md Shadidur Islam Jawad, et al. 2025. "Advanced Fraud Detection Using Machine Learning Models: enhancing Financial Transaction Security." *International Journal of Accounting and Economics Studies* 12(2): 85–104. doi:10.14419/c73kcb17.

George, Md Zahin Hossain, Md Khorshed Alam, and Md Tarek Hasan. 2023. "MACHINE LEARNING FOR FRAUD DETECTION IN DIGITAL BANKING: A SYSTEMATIC LITERATURE REVIEW." *ASRC Procedia: Global Perspectives in Science and Scholarship* 03(01): 37–61. doi:10.63125/913ksy63.

Ismail, Mustapha, Mohammed Mansur Ibrahim, Zayyan Mahmoud Sanusi, and Muesser Nat. 2015. "Data Mining in Electronic Commerce: Benefits and Challenges." *International Journal of Communications, Network and System Sciences* 08(12): 501–9. doi:10.4236/ijcns.2015.812045.

Khan, Talha Ahmed, Jagdesh Tulsi, Mansoor Alam, Kushsairy Kadir, Kanwar Mansoor Ali, and M.S Mazliham. 2025. "Analysis and Visualization of Fraud Detection Patterns through Data Mining and Classification Using MLP and Hybrid Deep Learning Model." *Egyptian Informatics Journal* 32: 100829. doi:10.1016/j.eij.2025.100829.

Kim, Junghee, Haemin Jung, and Wooju Kim. 2022. "Sequential Pattern Mining Approach for Personalized Fraudulent Transaction Detection in Online Banking." *Sustainability (Switzerland)* 14(15). doi:10.3390/su14159791.

León, Carlos, Paolo Barucca, Oscar Acero, Gerardo Gage, and Fabio Ortega. 2020. "Pattern Recognition of Financial Institutions' Payment Behavior." *Latin American Journal of Central Banking* 1(1–4): 100011. doi:10.1016/j.latcb.2020.100011.

Li, Xurui, Wei Yu, Tianyu Luwang, Jianbin Zheng, Xuetao Qiu, Jintao Zhao, Lei Xia, and Yujiao Li. 2018. "Transaction Fraud Detection Using GRU-Centered Sandwich-Structured Model." *Proceedings of the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design, CSCWD 2018*: 761–66. doi:10.1109/CSCWD.2018.8465147.

Lucas, Yvan, Sylvie Calabretto, Pierre Edouard Portier, Olivier Caelen, Michael Granitzer, Léa Laporte, and Liyun He-Guelton. 2019. "Multiple Perspectives HMM-Based Feature Engineering for Credit Card Fraud Detection." *Proceedings of the ACM Symposium on Applied Computing* Part F1477: 1359–61. doi:10.1145/3297280.3297586.

Sailusha, Ruttala, V. Gnaneswar, R. Ramesh, and G. Ramakoteswara Rao. 2020. "Credit Card Fraud Detection Using Machine Learning." *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*: 1264–70. doi:10.1109/ICICCS48265.2020.9121114.

Sami, Muhammad, Azka Mir, and Gina Purnama Insany. 2025. "Detection of Bank Transaction Fraud Using Machine Learning." In *The 7th International Global Conference Series on ICT Integration in Technical Education & Smart Society*, Basel Switzerland: MDPI, 34. doi:10.3390/engproc2025107034.

Wu, Yanxi, Liping Wang, Hongyu Li, and Jizhao Liu. 2025. "A Deep Learning Method of Credit Card Fraud Detection Based on Continuous-Coupled Neural Networks." *Mathematics* 13(5): 819. doi:10.3390/math13050819.