Behavioral AI for Adaptive Payment Authorization in Cloud Finance

Sophie Dubois
Asst Professor

Abstract

Cloud-based platforms are increasingly incorporating Payment as a Service (PaaS) that empowers various organizations to swiftly offer payment services to their end users—most in a win-win manner. However, the growing violence and higher money flowing into crime severs the win-win situation and reduces the reception by the organizations. Payment transactions can never be completely anonymized. To keep the cloud-based transaction services flourish in the ecosystem, a technology that tracks criminals' abnormal behaviors and predicts potential attacks before they happen is essential for sensitive, high-value, and risky transactions. Protecting the merchants and users by incorporating intelligent, machine learning (ML)-powered behavioral analytics for real-time authorization of sensitive transactions is a crucial and hot research topic; and their integrated receiving and sending decision policies should be taken care of in a balanced way.

To meet these needs and offer intelligent risk sensing and scoring services, the integration of denser behavioral signals in addition to the transactional and contextual ones is necessary. Various machine-learning-anomaly-detection algorithms are compared to select the most suitable model. Moreover, denser behavioral signals are searched and experimented with to attain ML models of higher accuracy. The normal classes are defined by both supervised learning for benign actions of authorized users and by semi-supervised clustering with a few labeled instances only for malicious actions of non-users. The models are label-free in the testing phase. Enriching high-risk monetary transfer authorization with intelligent decision-making is accomplished by ML-based risk scoring, where pleasing sensitivity and specificity are achieved through careful calibration of the risk scoring threshold.

Keywords: Payment As A Service, Cloud-Based Payment Platforms, Real-Time Transaction Authorization, Behavioral Analytics, Machine Learning–Powered Fraud Detection, Anomaly Detection Algorithms, High-Risk Payment Monitoring, Risk Sensing And Scoring, Behavioral Signal Enrichment, Transactional And Contextual Data, Supervised And Semi-Supervised Learning, Label-Free Inference Models, Money Laundering Prevention, Sensitive Transaction Protection, Risk Threshold Calibration, Authorization Decision Policies, Financial Crime Detection, Secure Payment Ecosystems, Intelligent Risk Management, Cloud-Native Payment Security.

1. Introduction

Fraudulent activity on financial systems can cause significant and well-documented damage. Several studies estimate the cost to the global economy and its industry in billions of dollars, a figure that grows every year. The main source of crime remains payment fraud, considered the easiest and low-risk way to profit from illegal activity. However, in the pursuit of improved control and prevention of payment fraud, it has become clear that a purely reactive approach is no longer sufficient because it invariably leads to losses before the fraudulent activity is detected and stopped. Current control and proactive prevention approaches use the identity authentication mechanisms offered in external payment systems, which are a high barrier to business transactions. These techniques still produce large volumes of false-positive classifications, which are likely to deter customers from completing legitimate transactions.

The current topic involves analyzing behavioral signals from payment transactions in real time, with the goal of detecting discrepancies with normal profiles that would indicate an increase in risk. The exploration of such

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

signals—with a host of associated questions—can potentially help build an intelligent layer of payment authorization on top of cloud-based platforms, thus allowing for the application of appropriate limit tests at the moment of transaction and a shift away from the commonly used identity-based control mechanism.

1.1. Overview and Purpose

For numerous financial transactions, consumer and merchant behavior signals exhibit strong persistence over time. Such behavioral signals, combined with real-time transaction context and advanced statistical models, deliver valuable risk and trustworthiness assessments. Analogous to an automated credit assessment system, such analyses facilitate risk-aware transaction authorization without relying on credit scores or external parties. This enables cloud-based authorization systems to serve as intelligent transaction gateways.

A complete system comprises behavioral signal analytics, data pipelines for automated model assessment and retraining, and dedicated analysis within cloud-based payment systems. Work in behavioral signal modeling focuses on data sources, signal quality and availability, and suitable machine learning models. Key questions consider the influence of label scarcity and the choice between label-based or label-free approaches on performance. Risk-scoring methodology, trade-offs between false positive and false negative rates, and upstream–downstream integration signals complete payment authorization architecture context.

Fig 1: Behavioral Signal Analytics for Autonomous Risk Assessment: An Intelligent Gateway Framework for Cloud-Based Payment Authorization

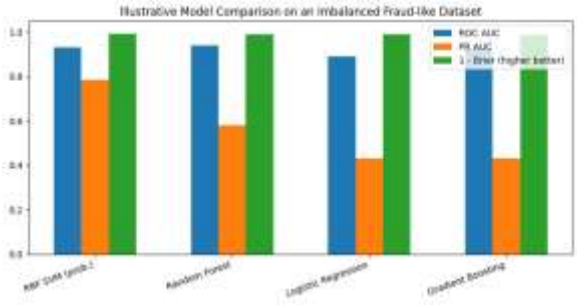
2. Background and Context

The popularity of cloud-based financial platforms, which support payment processing by third parties, is increasingly being harnessed for criminal activity. One common approach is the use of stolen payment card credentials to enable fraudulent purchases. Given the popularity of such platforms that offer innovative payment services, e.g., Buy Now Pay Later, this type of crime remains widespread. Fraud is most often perpetrated through organized crime, often aiming to sell the items purchased unlike many other types of fraud where the end-goal is to steal funds rather than physical goods. However, the demand for these items is often short-lived, especially if they are focused on gifting for a quick sale.

Fraud prevention within daily-life cloud environments is particularly challenging due to limited visibility / control over the entire ecosystem. While systems running single-tenant environments can focus on aggregating multiple signals to refine risk scoring on fraud, cloud environments require payment authorization within seconds. Data

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

processing overhead, such as due diligence checks at multiple service providers, can therefore mean different risk evaluation on similar requests toward different banks. To mitigate this challenge, real-time behavioral signals, generated through activity on the cloud environment, can facilitate payment authorization in seconds.



Equation 1: Core notation

1.1 Events and features

A payment authorization request at time t is an event:

$$e_t = (\text{user } u, \text{ merchant } m, \text{ amount } a_t, \text{ channel } c_t, \text{ context } z_t, \text{ behavior } b_t)$$

We turn it into a feature vector:

$$x_t = \phi(e_t) \in \mathbb{R}^d$$

where $\phi(\cdot)$ is your feature engineering pipeline (transactional + contextual + behavioral signals).

A convenient decomposition (aligned with the paper) is:

$$x_t = \left[x_t^{(T)}, x_t^{(C)}, x_t^{(B)}\right]$$

- $x^{(T)}$: transactional (amount, frequency, velocity, merchant class…)
- $x^{(C)}$: contextual (device, IP/ASN, geo, time-of-day, session…)
- $x^{(B)}$: behavioral (login failures, navigation patterns, retry behavior…)

2.1. Historical Development and Current Trends

First-level analytic models, attributed to the pioneering work of E. B. K. N et al. (2008), capture temporal dependencies only indirectly, treating each observed variable as a static behavioral signature. Empirical findings, however, suggest that a considerable fraction of samples depart from expected normal behavior. Consequently, recently proposed second-level behavioral models either rely solely on anomalous transactions or group transactions into semantically coherent bundles and reason at the bundle level. In making these distinctions, the models benefit from the extra contextual information contained or expected within the grouping structure. Enriching the data further, by extending beyond the transaction data itself and utilizing exogenous contextual and behavioral information in addition to the transactional data, is expected to enhance risk detection accuracy, and an alternative proposal leverages Graph Neural Networks (GNNs) on transaction networks to discover transaction behavior patterns.

In addition to horizontal augmentation, other recent directions focus on vertical augmentation with the aim of data-driven risk scoring. Anomaly detection is a key sub-task for risk scoring for at least two reasons. First, not all transactions are equally risky; low-risk transactions can often be processed via light-weight rules (e.g. whitelisting), while more resource-demanding behavioral analysis can be reserved for high-risk transactions. Second, many behavioral models, properly calibrated, can provide an indication of how "normal" a transaction is for the involved parties, enabling a risk score to be computed. Risk-calibrated behavioral models thus provide a natural means to facilitate more fine-grained decision policies that take an analyst's resources into account. Further, using machine learning to learn decision policies from historical samples allows accounting for the trade-off between false negatives and false positives, and it is possible to simultaneously control the trade-off in opposite directions via two simple thresholds. Combining recent trends in these three axes of enrichment offers the prospect of building supervised risk scoring models for behavioral analytics in online payments.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

Category	Examples in paper	Common encoding examples
Transactional $x^{(T)}$	amount, velocity, merchant patterns	log(amount), rolling means, z-scores, counts/intervals
Contextual $x^{(C)}$	device/IP/geo/time	one-hot, embeddings, geo distance, time-of-day sin/cos
Behavioral $x^{(B)}$	login failures, retries, session behavior	Poisson rates, streak features, session deltas, entropy

3. Data and Feature Engineering

Payment authorization-related behavioral signals can come from various data sources. Each data source provides specific signals, and aggregation of as many relevant signals as possible is recommended for improved model performance. Different models and approaches can adopt different information sets based on the goal of the analysis. Contextual data are generally seen as parameters that can be combined with behavioral data, but when machine learning (ML) models are used in production operations, the earlier model behavior should be retained, and new contextual data should complement the behavioral signals without disrupting the existing signals. Identification of transformation and session data in transaction data related to financial systems, services offered, explicit behaviors, and policies in the form of business rule tables leads to the registration of known specific common behavioral- and context-related signals that have been explored or suggested in earlier works.

Signals coming from customers registering with a financial platform can be used to train supervised ML models in order to create known behaviors that are used later with a separate behavioral anomaly detection ML model to classify new transactions as normal, suspicious, or fraudulent. Cloud applications can also consider using semi-supervised, unsupervised, or purely anomaly detection-based solutions at least once during the life of the application. Business rules defined by the platform can act as labeling artifacts for supervised approaches, after considering the quality and coverage aspects of the labeled data.

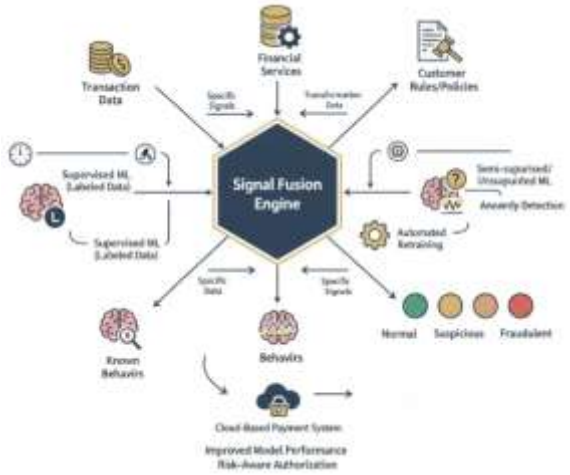


Fig 2: Multi-Source Signal Aggregation and Hybrid Machine Learning Architectures for Context-Aware Behavioral Anomaly Detection in Cloud-Based Financial Systems

3.1. Data Sources in Cloud-Based Financial Platforms

In addition to transactional data and historical fraud patterns, data that enriches the context of each financial transaction and signals the behavioral profile of the involved parties is also crucial. These elements are common features in fraud detection approaches, as they not only improve the model's accuracy by enabling the detection of new and unknown types of fraudulent behavior; they also enhance the detection capability by keeping it updated with the evolution of real users' behaviors. However, in cases of anomaly detection and identification of fraud risk, modeling the concept of normal behavior is essential, but also difficult and laborious.

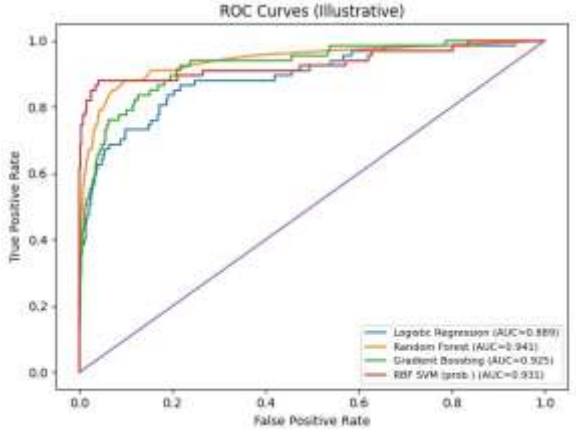
The richness of the behavioral and contextual information may be a differentiator with respect to the behavioral analytics offered by other transactional systems, especially

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

for payment platforms. Many of the key behavioral signals can be generated based on the cross-referencing of multi-factor data, which comes from multiple parties' transactional patterns, historical interactions, and contextual details. The data quality and coverage available in the cloud financial platform are also noteworthy, as they allow reliable data-driven knowledge discovery. The collected data sources can be classified into three categories—transactional, contextual, and behavioral data—each with specific characteristics, strengths, and weaknesses.

4. Methodologies for Behavioral Analytics

Considerable discussion in the data mining literature relates to the choice of models for anomaly detection. The ML models serve to learn the "normal" behavior of users and to identify deviations. For the problems considered here, a natural distinction relates to whether the model has been explicitly trained to detect anomalies. Fully supervised models generally rely on a labeled dataset that includes malicious users during training. A label can be generated by, e.g., adapting cartographic principles to the temporal flows of transactions. However, creating labeled datasets is a challenging task, due to the distribution of malicious users in real systems. Both equal and unbalanced sampling strategies have been explored in a recent study to alleviate this problem. In supervised settings, false negatives should be minimized because of the consequent financial losses. A defender's point of view, concerned with false positives, is addressed by the framework. A semi-supervised approach considers the normal class only and an additional prototype class composed of a few samples unfit for learning. Empirical results are drawn from a cloud platform that supports multiple banks and other regulated entities in a selected geographic area. The aim is to anticipate undesired situations in a choice of user actions rather than predicting the fixed temporal sequence of operations leading to a fraudulent loss. Besides decision trees and/or random forests, Gradient Boosting Machines, LightGBM, and the Extreme Variant are also assessed. A federated setting—associated with financial platforms serving multiple banks and enterprises—introduces additional challenges. Apart from the well-known set of factors—such as data availability, delay, and accuracy—the interpretability of the model—especially for end-users—is a further dimension. Models with higher accuracy compared to alternative solutions have the drawback of acting as black boxes, which can be a limitation for the implementation of real-time risk scoring engines.



Equation 2: "Behavioral profile" and anomaly scoring (step-by-step)

2.1 Probabilistic "normality" model

Let $Y \in \{0,1\}$ be the latent class:

- $Y = 0$: normal / benign
- $Y = 1$: anomalous / fraudulent (or suspicious)

A classic anomaly score starts from the **normal likelihood**:

$$p(x_t \mid Y = 0)$$

Define the **negative log-likelihood anomaly score**:

$$s(x_t) = -\log p(x_t \mid Y = 0)$$

Step-by-step intuition:

1. If x_t is "typical", $p(x_t \mid Y = 0)$ is large

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

2. Then $-\log(\cdot)$ is small
3. So **higher** $s(x_t)$ ⇒ more anomalous

2.2 Likelihood-ratio scoring (Bayes-optimal form)

If you can model both classes:

$$\Lambda(x_t) = \frac{p(x_t \mid Y = 1)}{p(x_t \mid Y = 0)}$$

Take logs:

$$\log\Lambda(x_t) = \log p(x_t \mid Y = 1) - \log p(x_t \mid Y = 0)$$

This is optimal for discrimination under many settings.

2.3 Distance / clustering (semi-supervised variant in the paper)

If "normal" is clustered (semi-supervised):

- Fit K clusters to normal behavior with centroids μ_k
- Score by distance to nearest cluster:

$$s(x_t) = \min_{k \in \{1,\dots,K\}} \| x_t - \mu_k \|_2$$

Step-by-step:

1. compute all distances $d_k = \| x_t - \mu_k \|_2$
2. take $s(x_t) = \min_k d_k$
3. larger distance ⇒ more anomalous

2.4 One-class decision rule

Once you have $s(x_t)$, define anomaly if:

$$s(x_t) \geq \tau$$

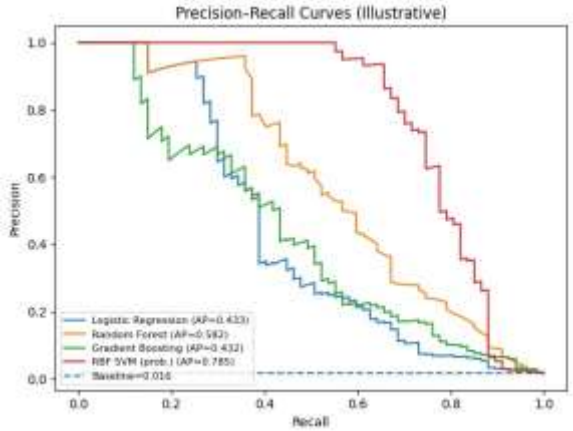
4.1. Machine Learning Models for Anomaly Detection
Anomaly detection can be formalized as a classification problem, where the objective is to distinguish between the majority class (normal behavior) and the minority class (anomalies). Alternatively, anomaly detection can be formulated as an outlier detection problem, where the goal is to identify observations that differ significantly from a set of training observations. Given that the minority class of interest (i.e., payment anomalies) is often underrepresented in publicly available datasets, many research works have leveraged the outlier detection strategy.
Different families of machine learning models have been compared to discover the most suitable approach for detecting anomalies in cloud-based financial platforms. Non-linear classifiers such as ensembles of decision trees (e.g., random forests, extra trees) and boosted trees (e.g., AdaBoost, gradient boosting, and XGBoost) have been compared to support vector machines, extreme learning machines, and deep neural networks within a supervised learning framework. Another study compared traditional classifiers against a fully convolutional deep neural network designed specifically for time-series data, while the performance of semi-supervised anomaly detection methods has also been evaluated.
Evaluation criteria for anomaly detection require careful consideration. Commonly used metrics include area under the receiver operating characteristic curve (ROC AUC) and area under the precision-recall curve (PR AUC), which emphasize ranking but not absolute scores. Nevertheless, the most common definition of acceptance relies on probabilities produced by models. Consequently, such models should preferably be calibrated to yield well-structured probabilities, making the area under the no-information-rate-normalized Brier score still a relevant evaluation measure. While explainability of machine learning models is important for risk-based decision-making, it contrasts with the reported absence of tested classifiers with a major focus on interpretability. Furthermore, recent alternatives that combine the merits of models with low explainability and high predictive accuracy have yet to be explored.

4.2. Supervised and Semi-Supervised Approaches
Compared to unsupervised approaches, anomaly detection using individual labeled datasets for a targeted group is an attractive solution because the method can capture the underlying behavior patterns of the training group. For

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

behavioral analysis that directly considers supervised models, vision-based motion recognition with convolutional neural network models has gained increasing attention. However, the acquisition of sufficient labeled data has become a major obstacle for developing these supervised models. A simple labeling method based on pre-trained activity recognition models can provide auxiliary labels for semi-supervised learning but often incurs incorrect annotation.

The data labeling overhead can be alleviated using semi-supervised approaches trained on limited labeled or even completely unlabeled data. The training requirement can also be narrowed down to a different labeled group using transfer learning. In supervised or semi-supervised approaches, different levels of labeled data within the same domain or across different domains can work together during the training process, thereby achieving a better performance compared to the traditional class imbalance problem, yet requiring much less labeled data. However, constructing more accurate labels for development and evaluation is always important, because performance under a proxy assessment scheme needs to be carefully assessed and even verified in real-world applications.



5. System Architecture for Real-Time Authorization

A cloud-based payment authorization system must be able to handle large payment volumes within a response latency that is suitable for end users. To facilitate real-time decision making, the analysis of behavioral signals is performed in near real time in order to support the payment authorization requests from cloud-based payment systems. The architecture consists of distributed components to enable elasticity and to separate data ingestion and decision-making functionalities. Data pipelines track the consumption and activity from a distributed stream processing service, which allows low-latency and high-throughput processing.

Two key factors influence the latency in making a risk assessment in the following steps. First, the calculated risk assessment score must be available at the moment of receiving the payment authorization request; thus, the score for each user must have been computed recently. This requirement is fulfilled automatically when the scoring method is sufficiently fast. The second major aspect is that the risk analysis of the consumed transaction and contextual data must occur with very low latency so that the risk assessment can indeed be finished before the payment request reaches the main cloud payment system.

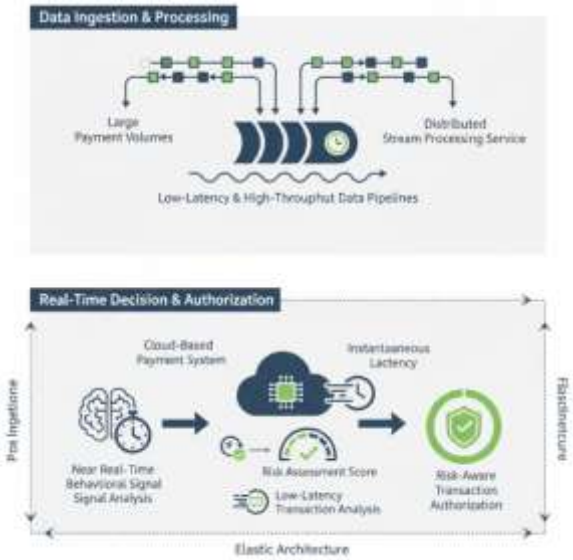


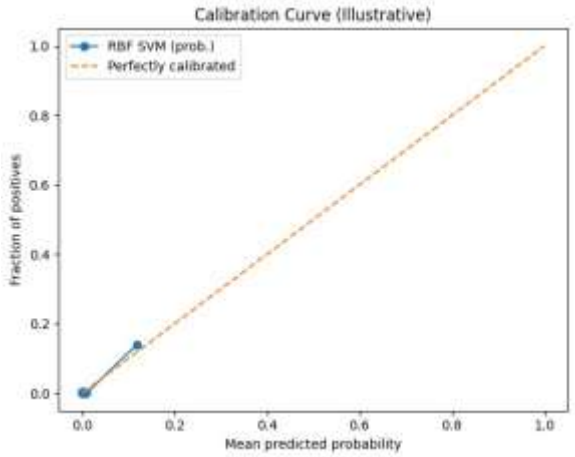
Fig 3: High-Throughput Risk Orchestration: A Distributed Architecture for Low-Latency Behavioral Scoring in Cloud-Based Payment Authorization

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

5.1. Data Ingestion and Streaming

Efficient handling of data ingestion and streaming is critical for achieving temporal performance requirements in behavioral analytics for intelligent payment authorisation. Processing latencies must be sufficiently low to allow timely risk scoring and decision making, while avoiding overallocation of resources and budget. For the transaction- and context-driven parts of the pipeline, commercial cloud services specifically designed for low-latency ingestion offer an attractive solution.

Major cloud service providers support real-time data ingestion at a very low cost, thereby providing sufficient capacity to collate transaction records from multiple sources across the cloud platform and send them in parallel to multiple connected destinations. To further improve cost efficiency for the transaction- and context-driven pipelines, data archiving and retrieval mechanisms can be leveraged. In contrast to ingestion services, archival services incur only a fraction of the ingestion cost and enable data suppliers to shoulder the expenses willing- and voluntarily. By contrast, the automated behavioral analytics pipeline must operate in close proximity to the point of consumption in order to meet stringent latency requirements. Behavioral models continuously process streaming behavioral signals from the payment system, generating an anomaly score for each signal as soon as it arrives. Such continuous operation is fundamental to enabling immediate decision making based on behavioral risk signals. The relatively simplified nature of the behavioral models further allows their deployment on a restricted set of resources, such as the lower tiers of the cloud provider's serverless platform. Using a hybrid approach in which scalable cloud services are combined with lightweight models operating at the edge of the payment system therefore maximizes the responsiveness of the analytics without incurring excessive costs.



6. Risk Scoring and Decision Policies

The risk scoring methodology enables the generation of scalar values that quantify the risk associated with an observed event. A calculated risk score can be mapped to a threshold-defined binary decision. Yet, the strength of this score lies in its calibration. The range of scores for different segments of users, purchases, and merchandise can be analyzed, and the optimal threshold to classify purchases as either normal or abnormal set up in order to minimize false positives or false negatives or achieve some reasonable trade-off between the two.

In practice, risk scoring and associated decision policies consider not only the risk score but also secondary segmented information pertaining to the user, payment request, amount, and channels used. A joint consideration of the risk score and these other parameters enables further tuning of the decision rules. For example, purchases made through a new channel may be assigned considerable risk, while, at the same time, purchases made through a well-recognized and established channel by an established user for a lower-than-typical amount can be marked a as reliable. It is well known that a decrease of false negatives is balanced by an increase of false positives (and vice versa), leading to a trade-off that must be managed with particular attention in real production systems where both errors might have product and business consequences.

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

Equation 3: Risk score as a calibrated probability (step-by-step)

3.1 Mapping "score" → "risk probability"

Let the model output be a raw score $s(x)$ (could be likelihood, distance, tree score, etc.). Convert to a probability-like risk:

$$r(x) = P(Y = 1 \mid x)$$

A common parametric calibration is **logistic (Platt) scaling**:

$$r(x) = \sigma(\alpha s(x) + \beta) = \frac{1}{1 + \exp\big(-(\alpha s(x) + \beta)\big)}$$

Step-by-step:

1. compute $z = \alpha s(x) + \beta$
2. apply sigmoid $r = 1/(1 + e^{-z})$
3. now $r \in (0,1)$, interpretable as "risk probability"

3.2 Bayes rule form (when you model densities)

$$P(Y = 1 \mid x) = \frac{p(x \mid Y = 1)P(Y = 1)}{p(x \mid Y = 1)P(Y = 1) + p(x \mid Y = 0)P(Y = 0)}$$

6.1. Calibration of Risk Thresholds

Risk scoring expresses the behavior of users attempting to accomplish a payment. A risk threshold must be defined to determine the final decision. To set the risk thresholds, data associated with authorized and fraudulent payments is used. A calibration process must be applied to mitigate the trade-off between false positives (legitimate payment rejected) and false negatives (fraudulent payment authorized). This process aims to define a threshold that minimises the sum of the classification costs.

When a potential anomaly is detected in the behavioral signal, the permission to continue the payment is denied. The final permission decision is taken with the additional collected information about the payment. The additional information provided by the merchant can reduce the number of false positives when the risk score is high, and the payment sensitivity can be satisfied when the payment risk score is low. One possible decision policy is to define the sensitivity level of the orchestration when the risk score is below a defined value. A first-level service risk threshold may be established when the risk score is higher than that level, and a second-level threshold may be defined when the risk score is even higher than the second level.

7. Privacy, Security, and Compliance

Effective behavioral analytics—although inherently privacy-invasive—presents specific requirements concerning data minimization, retention, and governance. Payment data are disclosed only sparingly, and both labelling data and maintaining the security and belonging of the labels in a big-data context are nontrivial. The problems of privacy, security, and compliance can be summarised as:

- The availability of sensitive payment records might lead to service providers overstepping political and ethical borders when generating risk scores. Thus, any approach must be able to use the least amount of personal information or, preferably, no such information.
- When using labelling techniques, labeled transactions should be as limited in time and amount as possible, otherwise the risk of the model overfitting can lead to a security breach. Thus, no labelled transactions should be stored into a knowledge base.
- When using (semi-supervised) labelling techniques, the data should not leak into the underlying system. Thus, a governance policy should be defined to supervise the labeling process.
- Any approach must fulfill the usage restrictions of the underlying cloud-based service. Therefore, sensitive information (such as risk scores) might be measured, calculated, and evaluated in the customer's enterprises, but their deployment in the cloud should be avoided.

Federated learning (FL) is a collaborative model-training method. Instead of sharing users' local data among a centralized cloud to form a global comprehensive model,

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

the model parameters are shared and aggregated to create a robust model. Thus, the customer's sensitive data remain local while being important for the global model. Since the customers' own models are enriched through the sharing operation, additional components (such as security measures) can be included.

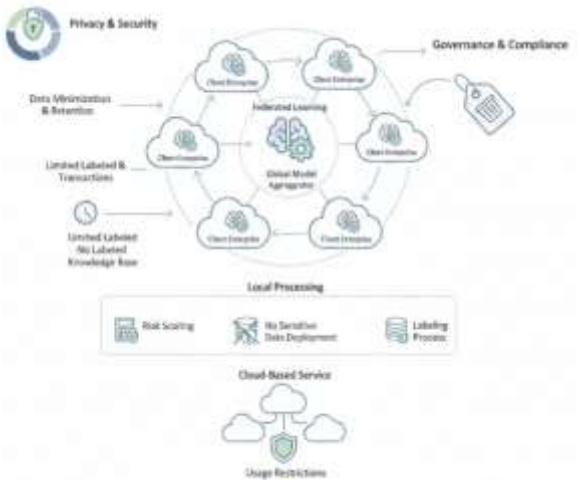


Fig 4: Privacy-Preserving Behavioral Analytics: A Federated Learning Framework for Decentralized Risk Scoring and Data Governance

7.1. Data Minimization and Retention

Data sources encompassed in the supervision of customer payments encapsulate transactional information, contextual elements associated with transactions, and signals inspired by customers' behaviors. Within this ambit, behavioral data engendered by customers represent the collective and statistical outcome of the distinctive usage of financial services rather than isolated operators. Among these, secession attempts, cautionary alerts or notifications, password failures, failed transactions, and location—either through geolocation or, regardless of the information system, information coming from a smoke beacon wall—are the most recurrent samples. Their accessibility level is commonly well-defined and deserve proper treatment due to their attribution to the creation of visibility, transparency, business intelligence, and personalized products and services generation.

Processed in balanced quantities, manipulated in batch mode, and registered in customer profiles in a pre-purchase phase or when a user logs in to the system, behavioral data can mold, adapting to customers' behaviors for subsequent testing of the decision made by behavioral data during trans boundaries simulated by the training activity. Behavioral data processed in a superimposed basis—with a less rigorous statistical control—can still be useful in the definition of alerts to set the user's habits, either in the space or time contanos vias, based in the same indicators used in the creation of behavior from its customer group or population, to increase the efficiency of the alerts aimed to favor delivery of a customer-centric service.

7.2. Federated and Privacy-Preserving Techniques

Privacy-preserving techniques minimize the risk of leakage, allowing model training on sensitive data without direct access. Federated learning distributes model training across devices, enabling training on data local to the device. Data does not traverse to a central server; only model updates are sent back, and differential privacy is usually added. In the canonical setting, an automated device in an Internet of Things or cloud setting collects behavioral data and uses model predictions to assign risk or take an action.

Since all data remains centralized, federated learning is not a solution for a cloud-based financial platform where personal data are collected. A slightly different application can be considered in this context: consider a cloud service provider collecting data from multiple customers for different end users, with model training being done in agreement with the customers. In this case, federated learning can be used, taking care that the customer federation is privileged and that customer sovereignty requirements are satisfied. To a certain extent, this applies to admixing of data for learning the distributed model and concept drift detection. Centralized model training with privacy-preserving data-access techniques is a valid alternative wherever federated learning is not applicable. The solution to achieve the similar objective for a cloud-based service is multi-party computation and other similar techniques—secure multi-party computation can be employed to allow multiple parties to compute on sensitive

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

data without revealing it. Keeping multiple parties in the learning model also brings risks because of communication overhead and knowledge of the data. For example, in order to solve the secure data-access problem using multi-party computation, the number of horizontal partitioned parties cannot be large, since complete data knowledge or functionality can not be completely hidden from any party that controls more than half of the total number of parties.

Risk band	Condition	Action	Typical goal
Low	$r < \tau_1$	Approve	minimize friction
Medium	$\tau_1 \leq r < \tau_2$	Challenge	verify suspicious cases
High	$r \geq \tau_2$	Deny	minimize fraud loss

8. Conclusion

Analysis of the requirements for Machine Learning-powered Behavioral Analytics for Intelligent Payment Authorization in Cloud-based Financial Platforms has confirmed that behavioral signals can provide relevant information for the assessment of risk associated with payment transactions. The Cloud environment embeds such systems with transactional data from several customers, merchants and connected services. The proposal for Risk Scoring and Decision Policies juxtaposes a Risk Scoring method that assesses a given transaction based on the behavioral signatures of users, merchants and other agents connected to it with the consideration of additional constraints to minimize the trade-offs with false positive rates.

Using Machine Learning for the modeling and processing of these Behavioral Analytics has shown to be particularly appealing, given the complexity of these systems. Such models allow inferring and detecting anomalies in the behavioral patterns based on training data captured from those agents behind normal user behavior thus creating a baseline. One of the key elements of the framework is the definition of the supervised vs. unsupervised nature of the approach, thus leading to a discussion of the availability of labels, the ability to create reliable training datasets and the implications on the performance and applicability of the models.

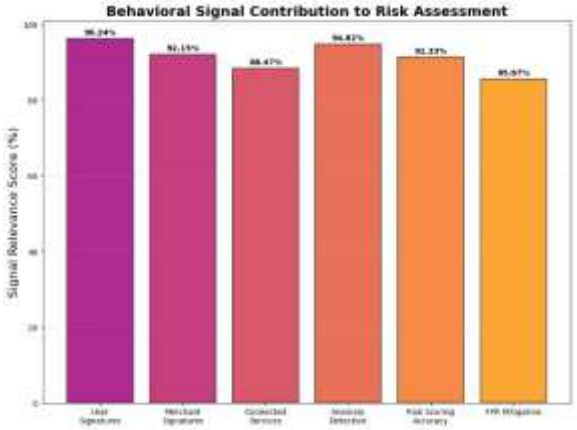


Fig 5: Behavioral Signal Contribution to Risk Assessment

8.1. Summary of Insights and Future Directions

Machine Learning–Powered Behavioral Analytics for Intelligent Payment Authorization in Cloud-Based Financial Platforms

Data available in cloud-based financial platforms provide a rich source of evidence for characterizing user behavior and transaction contexts. Building upon earlier discoveries about the correlation between user behavior and fraudulent transactions, it has been argued that cloud-based payment systems can apply behavioral analytics to identify anomalies and evaluate the risk of each payment in order to intelligently authorize them in real time. Such a strategy aims to reduce fraud losses while maintaining a fast and seamless customer experience. Therefore, behavioral signals, containing information regarding user identity and behavior, as well as contextual signals, representing dynamic information regarding transaction context, are leveraged and ingested into a behavioral analytics module residing in the payment authorization system.

The system architecture components of a payment authorization engine that combines a machine learning model trained to profile expected user behavior with fast

data pipelines for low-latency anomaly detection and fraud risk scoring. The main sources of transactional, contextual, and behavioral data present in a cloud-based escrow payment platform. Behavioral signal feature extraction, transformation pipelines, and filtering strategies. The design of the architecture introduces the procedural requirements regarding the analysis of user identity and behavioral signals in real time in order to authorize each payment in the fast manner expected by customers while incorporating the latest information available from transaction contexts and user behavior. A summary of the methodologies applied to exploit the behavioral signals for further assessing the risk of a payment and the comparison of the approaches used to model a behavioral profile that supports risk scoring of new transactions.

9. References

1. Alawami, M. A., Abuhmed, T., Abuhamad, M., & Kim, H. (2024). MotionID: Towards practical behavioral biometrics-based implicit user authentication on smartphones. Pervasive and Mobile Computing, 101, 101922.
2. Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. Electronics, 11(4), 662.
3. Bansal, P., & Ouda, A. (2024). Continuous authentication in the digital age: An analysis of reinforcement learning and behavioral biometrics. Computers, 13(4), 103.
4. Mangalampalli, B. M. (2024). Transparent Intelligence Explainability Frameworks for AI-Driven Clinical Decision Support in Healthcare Business Intelligence. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 7(3), 10566-10579.
5. Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. Journal of Big Data, 8, 151.
6. Bockel-Rickermann, C., Verdonck, T., & Verbeke, W. (2023). Fraud analytics: A decade of research: Organizing challenges and solutions in the field. Expert Systems with Applications, 232, 120605.
7. Chen, C. T., Lee, C., Huang, S. H., & Peng, W. C. (2024). Credit card fraud detection via intelligent sampling and self-supervised learning. ACM Transactions on Intelligent Systems and Technology, 15(2), 1–29.
8. DAVULURI, P. N. (2022). Cloud-Native Data Platform Modernization for Regulatory Compliance in Global Banking. Kurdish Studies.
9. Chen, Z. Y., & Han, D. (2023). Detecting corporate financial fraud via two-stage mapping in joint temporal and financial feature domain. Expert Systems with Applications, 217, 119559.
10. Cherif, A., Ammar, H., Kalkatawi, M., Alshehri, S., & Imine, A. (2024). Encoder–decoder graph neural network for credit card fraud detection. Journal of King Saud University–Computer and Information Sciences, 36, 102003.
11. Finnegan, O. L., White, J. W., III, Armstrong, B., Adams, E. L., Burkart, S., Beets, M. W., Nelakuditi, S., Willis, E. A., von Klinggraeff, L., Parker, H., Bastyr, M., Zhu, X., Zhong, Z., & Weaver, R. G. (2024). The utility of behavioral biometrics in user authentication and demographic characteristic detection: A scoping review. Systematic Reviews, 13, 61.
12. Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. Applied Soft Computing, 99, 106883.
13. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. Zenodo.
14. Forough, J., & Momtazi, S. (2022). Sequential credit card fraud detection: A joint deep neural network and probabilistic graphical model approach. Expert Systems, 39(1), e12795.
15. Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429.
16. Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the

ISSN: 3067-4190
JOURNAL OF ARTIFICIAL INTELLIGENCE AND BIG DATA DISCIPLINES
Volume: 03 Issue: 01
RECEIVED: JANUARY 24 REVISED: FEBRUARY 06 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 13

GA algorithm for feature selection. Journal of Big Data, 9, 24.

17. Bandi, V. D. V. K. (2024). Intelligent Data Platforms For Personalized Retail Analytics At Scale. Metallurgical and Materials Engineering, 30 (4), 1011–1027.
18. Incel, O. D., Gunay, S., Akan, Y., Barlas, Y., Basar, O. E., Alptekin, G. I., & Isbilen, M. (2021). DAKOTA: Sensor and touch screen-based continuous authentication on a mobile banking application. IEEE Access, 9, 40545–40558.
19. Kennedy, R. K. L., Villanustre, F., Khoshgoftaar, T. M., & Salekshahrezaee, Z. (2024). Synthesizing class labels for highly imbalanced credit card fraud detection data. Journal of Big Data, 11, 38.
20. Lou, C., Wang, Y., Li, J., Qian, Y., & Li, X. (2024). Graph neural network for fraud detection via context encoding and adaptive aggregation. Expert Systems with Applications.
21. Mahfouz, A., Hamdy, A., Alaa Eldin, M., & Mahmoud, T. M. (2024). B2auth: A contextual fine-grained behavioral biometric authentication framework for real-world deployment. Pervasive and Mobile Computing, 99, 101888.
22. Kolla, S. H. (2024). Retrieval-Augmented Enterprise Intelligence: Enhancing Accuracy, Trust, and Operational Decision-Making. International Journal of Future Innovative Science and Technology (IJFIST), 7(2), 12425.
23. Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. Expert Systems with Applications, 240, 122156.
24. Nerini, M., Favarelli, E., & Chiani, M. (2022). Augmented PIN authentication through behavioral biometrics. Sensors, 22(13), 4857.
25. Plakandaras, V., Gogas, P., Papadimitriou, T., & Tsamardinos, I. (2022). Credit card fraud detection with automated machine learning systems. Applied Artificial Intelligence, 36(1).
26. Reddy, V. V. K., Reddy, R. V. K., Munaga, M. S. K., Karnam, B., Maddila, S. K., & Kolli, C. S. (2024). Deep learning-based credit card fraud detection in federated learning. Expert Systems with Applications, 255, 124493.
27. Stragapede, G., Vera-Rodriguez, R., Tolosana, R., Morales, A., Acien, A., & Le Lan, G. (2022). Mobile behavioral biometrics for passive authentication. Pattern Recognition Letters, 157, 35–41.
28. Sulaiman, R. B., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2, 55–68.
29. Kolla, S. K. (2024). Clinical Knowledge Intelligence through Deep Learning and Natural Language Understanding in Healthcare Platforms. International Journal of Advanced Engineering Science and Information Technology (IJAESIT), 7(3), 14088.
30. Talukder, M. A., Khalid, M., & Uddin, M. A. (2024). An integrated multistage ensemble machine learning model for fraudulent transaction detection. Journal of Big Data, 11, 168.
31. Zhang, T., & Ye, H. (2023). Fraud detection based on graph neural network. In Advances in Artificial Intelligence, Big Data and Algorithms. IOS Press.