V.I. Zhukov. (2025). Toward automated detection of economically significant anomalies in digital storefronts: evaluation criteria and conceptual system architecture. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 25* (pp. 499-508). ICEB'25, Hanoi, Vietnam, August 21-25, 2025.

Toward Automated Detection of Economically Significant Anomalies in Digital Storefronts: Evaluation Criteria and Conceptual System Architecture

V.I. Zhukov [1,*]

*Corresponding author
[1] Academy of Engineering, RUDN University, Moscow, Russia, PhD student, slava.zhukov@gmail.com

ABSTRACT

E-commerce has become an integral part of our lives, with online stores and marketplaces offering products through Digital Storefronts. The Digital Storefront and user behavior on it form a dynamic system, subject to constant changes that ultimately impact the storefront's economic performance, such as the expected average purchase value or conversion rate. Changes in product assortment, technical issues, and many other factors can cause deviations in performance indicators. In cases of critical deviations, companies are forced to involve analysts to identify the root causes, which can take considerable time and lead to financial losses due to delayed reactions.

This paper addresses the problem of detecting and localizing economically significant anomalies in Digital Storefronts based on user behavior analysis. It reviews the historical development of the field through the lens of Web Mining and identifies key data structures used for modeling behavior. Particular attention is given to evaluation criteria for anomaly detection methods—such as adaptability, interpretability, automation, and sequential awareness—as well as to the limitations of existing method classes. As no single method class meets all criteria, the proposed approach relies on a modular architecture that decomposes the task into specialized subsystems. The paper outlines the conceptual architecture of such a system, integrating time series analysis (e.g., ARIMA, HMM), unsupervised learning, graph-based models (e.g., GCN), and deep learning techniques (e.g., LSTM, Transformers), with a focus on actionable and interpretable outputs for decision support.

Keywords: Digital storefronts, clickstream, anomaly detection, web analytics, user behavior analysis, online commerce.

INTRODUCTION

Electronic commerce (eCommerce) has become an integral part of modern life and represents a significant sector of the global economy. Within the broader scope of eCommerce, a prominent role is played by online stores and marketplaces that facilitate product selection and purchase through Digital Storefronts (i.e., websites and mobile applications).

In the current digital economy, the management of Digital Storefronts has emerged as a critical and timely business objective. Typically, such management efforts are directed toward enhancing the economic performance of the storefront (Hoffman, Novak, & Chatterjee, 1995). Achieving this objective necessitates the implementation of advanced, data-driven technological solutions for the collection, processing, and analysis of relevant data.

The Digital Storefront, along with user interactions within it, constitutes a complex and dynamic system, continuously subject to internal and external changes. These changes directly influence key performance indicators. For example, the assortment breadth and depth are frequently updated, new products are introduced, and users provide feedback in the form of positive or negative reviews. Simultaneously, changes in audience composition and behavior in response to these updates give rise to a multitude of potential factors—often numbering in the tens of thousands — that may contribute to deviations in business performance metrics.

In cases of significant anomalies, organizations are often compelled to engage data analysts to conduct in-depth investigations to determine the root causes. Such analyses may require substantial time and resources, during which the business continues to experience financial losses. Consequently, the development of efficient and timely methods for anomaly detection and root cause analysis is of considerable practical importance.

This paper explores the historical foundations of anomaly detection in user behavior through the Web Mining paradigm and introduces a set of evaluation criteria—including adaptability, interpretability, computational efficiency, and sequential awareness — for comparing detection methods. Given that no single method class fully satisfies all criteria, the paper proposes a modular system architecture in which specialized subsystems apply different techniques (e.g., time series analysis with ARIMA or HMM, graph-based detection using GCN, sequential modeling via LSTM or Transformers). The system is designed to operate

in a shared multidimensional space ⟨U, X, A, T⟩, capturing user characteristics, behavioral sequences, storefront attributes, and temporal context.

The proposed architecture aims to bridge the gap between algorithmic detection and business decision-making by generating interpretable, targeted insights into anomalies affecting storefront performance.

RESEARCH OBJECTIVES AND NOVELTY

The research presented in this paper is motivated by the growing need to detect and explain anomalies in user behavior on Digital Storefronts that have direct economic consequences — such as sharp drops in conversion rate or average order value. While existing anomaly detection literature offers a wide range of techniques, most approaches focus on technical irregularities, fraud, or abstract behavioral deviations, without explicitly linking these to changes in key business metrics.

This study advances the formulation of the anomaly detection problem by placing it within the context of business impact. An anomaly is not simply a deviation from statistical norms, but rather a change in user behavior that requires managerial attention due to its potential influence on storefront performance. Moreover, such anomalies are often caused by changes in storefront-side factors—including assortment updates, promotions, UX modifications, or technical issues. Understanding this causal relationship is crucial for timely and effective response.

Unlike traditional approaches, where interpretability is treated as a secondary feature, this study incorporates it as a functional requirement: the system must not only detect anomalies but also produce actionable explanations grounded in the operational context of the storefront. This perspective reshapes the objective of anomaly detection into a decision-support task aimed at enabling scalable, automated diagnostics in high-volume eCommerce settings..

The novelty of the research is threefold:

1. **Problem Formulation**: a new task definition is introduced, where anomalies are defined by their economic relevance and linked to underlying storefront-level factors. This framing connects behavioral deviations with managerial decision-making and elevates interpretability to a functional requirement.
2. **Criteria-based Evaluation Framework**: a domain-specific set of evaluation criteria is proposed—adaptability, interpretability, computational efficiency, sequential awareness, confidence control, and automation. These criteria are applied to a structured comparative analysis of existing method classes, enabling a clear assessment of their strengths, limitations, and applicability.
3. **Modular and Interpretable Architecture**: a conceptual system design is proposed that integrates diverse analytical methods—time series, session-level clustering, deep sequential models, and graph-based representations—into a scalable and explainable architecture tailored to Digital Storefronts.

Based on these foundations, the study pursues the following objectives:

1. To formulate and justify a domain-specific set of evaluation criteria for anomaly detection methods in Digital Storefront environments.
2. To conduct a structured comparative analysis of existing method classes using the proposed criteria, and identify their practical limitations.
3. To propose a modular conceptual architecture for anomaly detection and interpretation, designed for scalability, adaptability, and explainability.

To situate the proposed problem formulation within the broader research landscape, the next section offers a historical perspective on how behavioral data in eCommerce has been analyzed and modeled. This background in Web Mining not only illustrates the evolution of user behavior analytics but also lays the conceptual foundation for the subsequent system design and methodological choices in anomaly detection.

THE EMERGENCE OF THE WEB MINING FIELD AND ITS COMPONENTS

The history of research related to the management of Digital Storefronts can be traced back to the early development of digital sales channels, which followed the widespread adoption of the Internet in the consumer market during the 1990s. The precursors to modern eCommerce were digital catalogs, which presented various products along with their prices. The term “Online Storefront” was first introduced in (Hoffman, Novak, & Chatterjee, 1995).

In the second half of the 1990s, the analysis of user behavior data in online environments began to develop. These behavioral data became known as web logs. Academic research started exploring new data mining techniques specifically tailored for web data (Chen, Park, & Yu, 1996), and the term Web Mining was introduced in (Mobasher, Jain, Han, & Srivastava, 1996). Researchers and practitioners quickly recognized that users leave behind digital footprints when interacting with online storefronts, and analyzing this data could lead to valuable insights for understanding customer preferences and behavior. The digital environment enabled the collection of significantly more information compared to traditional offline interactions.

Under this new umbrella term, a wide range of publications emerged, encompassing both novel mathematical data analysis methods developed for web logs and data storage and processing technologies. It is important to note that at the time, web logs

represented an entirely new type of data, and researchers initially attempted to apply existing techniques, adapting and extending them to address this emerging class of problems. For instance, OLAP cube modeling—a technology already widely adopted in corporate business analytics—was proposed as a method for structuring and analyzing web data in Web Mining tasks (Büchner & Mulvenna, 1998). That work also identified several key user behavior attributes, still relevant today:

- customer;
- session;
- prrodct;
- ettc.

In (Zaiane, Xin, & Han, 1998), additional features were proposed, including page views, time spent on each page, server status, and more.

Research in Web Mining can generally be categorized into three major branches (Kosala & Blockeel, 2000; Sharma, Shrivastava, & Kumar, 2011): Web Content Mining – extraction of content from web pages; Web Structure Mining – extraction of web page structure and hyperlink data; Web Usage Mining – analysis of user interaction patterns. Among these, Web Usage Mining (WUM) is the most pertinent to this study.

By 2000, the WUM research community had already established three key sub-tasks (Srivastava, Cooley, Deshpande, & Tan, 2000): Pre-processing, Pattern Discovery, Pattern Analysis.

It is also worth noting that even at the inception of Web Usage Mining, several practical application areas were clearly defined:

1. Personalization;
2. Storefront modernization (initially focused solely on websites, though mobile applications are now equally relevant);
3. System-level technical enhancements;
4. Business intelligence.

Subsequent research and development efforts in this domain have largely been concentrated within these areas. For instance, even in the early 2000s, WUM was being applied to assess eCommerce efficiency. One example is the study presented in (Spiliopoulou, Pohle, & Faulstich, 1999), which proposed a methodology for evaluating effectiveness by identifying user types through pattern analysis.

The first steps in personalization were marked by studies such as (Joachims, Freitag, & Mitchell, 1997; Ngu & Wu, 1997), which introduced personalized services that recommended potentially interesting links. This was especially relevant at the time, as fully developed search engines were not yet available, and users primarily relied on categorized web directories. An example of personalization in the context of a storefront is found in (Mobasher, Cooley, & Srivastava, 1999), which proposed a method for recommending website pages based on previous user behavior, employing clustering algorithms.

A wide range of machine learning and statistical data analysis methods have since been applied in personalization research (Pierrakos, Paliouras, Papatheodorou, & Spyropoulos, 2003), including: Clustering, Classification, Association Rule Mining, Sequential Pattern Mining, etc.

In parallel with personalization, the task of purchase probability prediction began to gain momentum. The goal of such methods is to predict, based on user behavior, whether a visitor is likely to exit the storefront without making a purchase—thus enabling targeted direct marketing mechanisms, such as pop-ups offering discounts. Examples of such research can be found in (Moe & Fader, 2004; Sismeiro & Bucklin, 2004).

Over time, the volume of collected data has grown substantially due to increased internet traffic and broader adoption of Digital Storefronts. At the same time, computational capabilities have improved, enabling the effective application of deep learning techniques, especially LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit) networks. For instance, in a relatively recent study (Cirqueira, Hofer, Nedbal, Helfert, & Bezbradica, 2019), this class of models was shown to be capable of learning hidden behavioral patterns for purchase prediction tasks and is often used in conjunction with classical machine learning methods.

ANOMALY DETECTION IN USER BEHAVIOR

A distinct and essential topic within this review is the task of anomaly detection in user behavior—a research direction that began in the field of computer science in the late 1980s. Historically, this class of problems emerged in the context of security and fraud detection, particularly in identifying malicious or low-quality traffic (Bi, Xu, Wang, & Zhou, 2016). By the mid-1990s, academic literature had already begun exploring web log analysis for anomaly detection, and this continues to be an active area of research to this day.

User behavior plays a central role in determining key economic indicators of a Digital Storefront. Among the most widely used metrics in the industry are the conversion rate (CR)—the proportion of visits that result in a purchase—and the average order value (AOV). While this set of indicators may be extended with additional delayed metrics, such as repurchase rate or customer

retention, our research focuses specifically on real-time performance metrics that reflect the instantaneous selling effectiveness of the storefront.

Identifying the causes of deviations in key performance metrics is often a complex and time-consuming analytical task. This creates a clear need for systems capable of automatically detecting potential drivers behind such changes. For these systems to be practically useful, their outputs must be interpretable by human experts, allowing analysts to uncover behavioral or contextual factors that contribute to fluctuations in economic indicators and implement appropriate adjustments to the Digital Storefront. In this context, the system should generate results that are not only accurate but also actionable and explainable, rather than relying solely on opaque or purely statistical models.

An anomaly may be defined either as a deviation of a time series metric from its expected value, or as the emergence of an object whose characteristics differ significantly from the norm. In the latter case, object features are typically presented in vector or tabular format, allowing for the application of methods such as Support Vector Machines (SVMs).

A specialized subdomain has also emerged: anomaly detection in graphs. Graph-based representations offer a rich framework for modeling structures and processes, such as user navigation paths modeled via Markov chains. In this representation, nodes correspond to meaningful events or actions, while edges represent transition probabilities between them.

Traditional anomaly detection methods were not well-suited for graph data, prompting the development of new approaches. Early methods often relied on expert-driven feature engineering and statistical analysis, which proved insufficient for identifying unknown or unexpected anomalies. In recent years, these expert-based approaches have largely been replaced by machine learning algorithms, particularly deep learning models capable of capturing nonlinear dependencies in graph structures (Ma *et al.*, 2021).

A key point to emphasize is that anomaly detection tasks are frequently solved using unsupervised learning methods, since in most real-world settings, there is no prior knowledge of which events will be anomalous, making it infeasible to train models using labeled examples (Chollet, 2021).

USER BEHAVIOR DATA COLLECTION AND REPRESENTATION

User behavior data on Digital Storefronts is commonly referred to as clickstream data. Clickstream data serves as the primary source for user behavior analysis and, in practice, is represented either as server logs or as structured datasets stored in specialized databases, such as ClickHouse. Technologically, the composition and granularity of clickstream data are determined by specific instrumentation solutions deployed on the Digital Storefront. These solutions define which events are captured and transmitted, as well as how they are structured and stored.

Each individual user action is characterized by a predefined set of parameters, which may vary across different systems. However, it is possible to identify a set of core attributes that are common to most clickstream logging frameworks:

- User ID – a unique identifier associated with the user (anonymized or authenticated);
- Session ID – an identifier linking multiple actions within the same interaction session;
- Timestamp – the exact time at which the action occurred;
- Event name – a label describing the type of action (e.g., “product_view”, “add_to_cart”);
- IP address – used to infer geographical location or detect anomalous access patterns.

This set of parameters may be extended depending on the specific data collection system and its configuration. Moreover, through enrichment with historical data, the clickstream can be supplemented with higher-level attributes such as the date of the last purchase, user classification (e.g., new or returning), customer lifetime value estimates, or behavioral segments.

In the context of our research, it is essential to emphasize that user behavior is represented as a temporal sequence of actions, where each action occurs in a specific order over time.

For analytical tasks, user behavior (denoted as **S**) can, in its simplest form (1), be represented as a combination of a vector of known user attributes (**U**) and a vector of aggregated session characteristics (**C**). These aggregated features may include indicators of whether specific stages of the sales funnel were reached (e.g., product page viewed, cart visited), the categories of products browsed, and session duration, among others. Funnel-related features are typically encoded as binary variables, while attributes such as region, device type, and user type are treated as categorical variables; purchase amount and session time are represented as numerical features. This aggregated representation effectively “collapses” the full sequence of user actions into a single summary object. Such a format is well-suited for ad hoc analysis using standard analytical methods and, when applied to machine learning models, is less demanding computationally compared to more detailed, sequence-based data representations.

$$S = (\{c_1, c_2, \dots c_n\}, U) \; \forall \; c_i \in C \tag{1}$$

However, it is important to note that such an aggregated representation inherently loses information about the sequence of user actions and the time intervals between them. This temporal and sequential information can be critically important for many

analytical and predictive tasks. For more detailed analysis, the session is represented not as a set of aggregated features, but rather as a sequence of user actions x over time T , formally expressed as (2).

$$S = (\{x1^{t1}, x2^{t2}, \ldots xn^{tm}\}, U, A) \; \forall \; x_i \in X, t_j \in T \tag{2}$$

An important aspect of this action sequence is that, throughout the session, the user interacts with various attributes of the Digital Storefront, which ultimately influence their next action and response. These attributes can be broadly categorized into technological (e.g., page load speed), marketing-related (e.g., promotions, discounts), and content/product-related (e.g., product descriptions, images, navigation structure). These storefront attributes introduce an additional dimension to the data model, reflecting the contextual environment in which user actions occur. We denote this set of features as A, representing the attribute space of the Digital Storefront integrated into the behavioral model. The overall structure of this interaction is schematically illustrated in Figure 1.

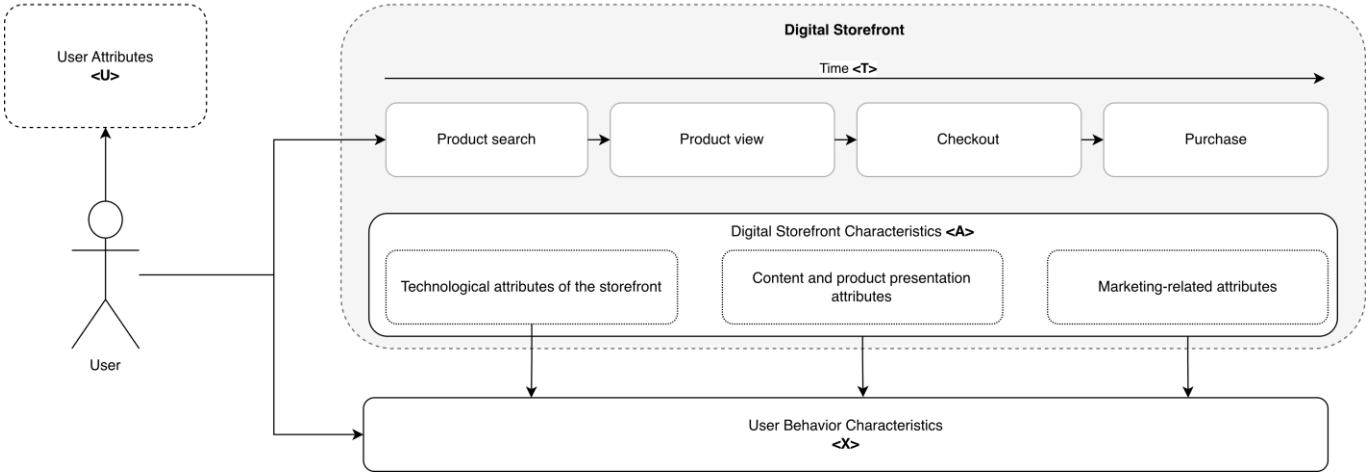

Figure 1: User Interaction Flow with the Digital Storefront

FORMULATION OF EVALUATION CRITERIA AND ANALYSIS OF ANOMALY DETECTION METHOD CLASSES

A key component of this study is the analysis of anomaly detection methods for user behavior in Digital Storefronts. To support this analysis, we define a set of evaluation criteria that are particularly relevant to the problem at hand.

Adaptability to Change
User behavior is inherently dynamic: shifts in segment distributions, seasonal effects, and other sources of variability can cause significant changes in behavioral patterns. What is considered an anomaly today may become the norm tomorrow. Effective methods must therefore be sensitive to evolving baselines and capable of adapting to structural changes in data.

Interpretability
The system under development is intended to assist in identifying critical behavioral shifts that warrant managerial action on the Digital Storefront. This is only possible if the system's outputs are interpretable to human decision-makers. Interpretability must extend not only to the detected anomalies themselves but also to the storefront-related changes that may have caused them. Since decisions are ultimately made by humans, the system must provide clear, evidence-based justifications for its findings.

Computational Efficiency
Modern Digital Storefronts generate millions of user interactions per day. Resource-intensive analytical methods may require substantial computational power to deliver results in a timely manner, which can be economically inefficient. The employed methods must strike a balance between analytical power and computational feasibility.

Sequential Awareness
User behavior is inherently sequential: each action occurs in a temporal context and may influence subsequent actions. Therefore, methods that capture and model the order and timing of events are better suited to reflect the real structure of user behavior.

Confidence Control
The finer the segmentation of users, the greater the likelihood of noise and false positives in anomaly detection. The system should allow for confidence level adjustment, enabling the ranking of anomalies based on their statistical or practical significance.

Automation
The sheer number of potential anomaly patterns makes it impractical to rely on manually defined and constantly updated rule sets. The selected methods must support automated detection processes, minimizing the need for manual configuration and enabling scalable operation.

We also identify the main classes of anomaly detection methods that will be analyzed in this study:

Rule-Based Methods
These approaches define anomalies as actions or events that fall outside of manually predefined thresholds. For example, exceeding a fixed value for a particular metric—such as session duration, number of operations, or transaction amount—would be flagged as anomalous.

Statistical Methods
In this class, anomalies are defined as deviations from the norm, where the norm is estimated based on statistical properties of the data distribution, such as the mean and standard deviation.

Density-Based Methods
The core idea behind these methods is that anomalies lie in regions of low data density. They are particularly effective in identifying rare or isolated observations in high-dimensional spaces. Notable examples include Local Outlier Factor (LOF) and DBSCAN.

Time Series–Based Methods
These techniques identify anomalies as deviations from the expected temporal patterns, such as trends, seasonality, or autocorrelation. They are well suited for detecting sudden shifts or unexpected spikes in behavioral data over time.

Classical Machine Learning Methods
These are primarily unsupervised learning techniques that assume anomalies are rare and differ significantly from normal behavior. Supervised learning approaches are used less frequently, and typically only when the data is labeled and the behavior of objects is relatively stable, allowing for classification into “normal” and “anomalous” categories.

Deep Learning Methods
This class of techniques is used when working with complex and high-dimensional data. Their main advantage lies in the ability to uncover nonlinear, latent patterns through the use of deep neural networks, making them particularly powerful for modeling intricate user behavior.

The results of the comparative analysis of these method classes are presented in Table 1.

Table 1: – Comparison of classes of anomaly detection methods according to the developed criteria

Method Class	Adaptability to change	Interpretability	Computational efficiency	Sequential awareness	Confidence control	Automation
Rule-based methods	Low	High	High	Low	Low	Low
Statistical methods	Medium	High	High	Low	Medium	Medium
Density-based methods	Medium	Medium	Low	Low	Medium	Low
Time series–based methods	Medium	Medium	High	Low	Medium	High
Classical ML methods	Medium	Medium	High	Low	High	Medium
Deep learning methods	High	Low	Low	High	Medium	High

The results of the analysis demonstrate that no single class of methods fully satisfies all of the defined evaluation criteria. Rule-based methods can be excluded from further consideration, as they are not suitable for practical use in real-world, dynamic environments. An effective anomaly detection system can only be constructed through decomposition of the overall problem into smaller, well-defined sub-tasks, with each sub-task addressed using the most appropriate class of methods based on its specific requirements and constraints.

CONCEPTUAL ARCHITECTURE OF AN ANOMALY DETECTION SYSTEM

It is difficult to envision a single model or neural network capable of simultaneously detecting behavioral changes, identifying their underlying causes, and presenting the results in an interpretable form suitable for human experts. Moreover, it is highly desirable that such a system report only significant deviations that truly warrant attention and response. The system must also

operate efficiently, which presents a major challenge given that a modern Digital Storefront may process millions of user visits per day — making full and deep behavioral analysis a computationally intensive task.

As shown in Table 1, no single class of anomaly detection methods fully satisfies all the evaluation criteria. Therefore, the proposed solution is to design a modular architecture in which the anomaly detection system is decomposed into specialized subsystems, each responsible for a specific stage of the detection and interpretation pipeline. Figure 2 illustrates this decomposition into subsystems.

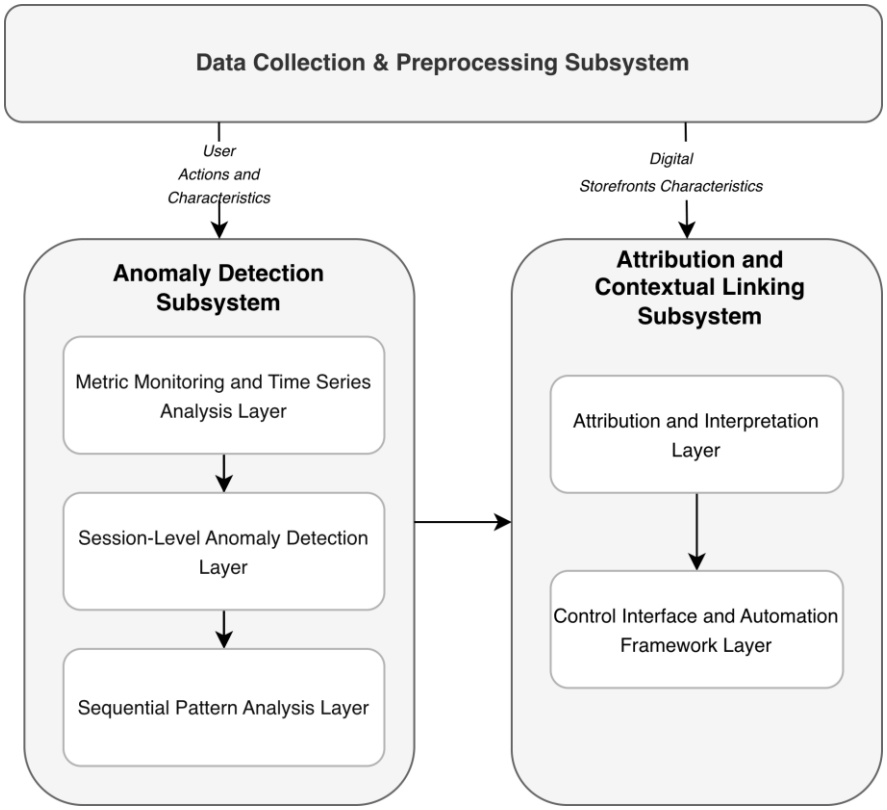

Figure 2: Conceptual Architecture of an Anomaly Detection System

All subsystems operate within a shared multidimensional data space defined as ⟨U, X, A, T⟩, where: U represents the space of users (e.g., user attributes, segments); X represents the space of user action sequences; A represents the space of storefront attributes as experienced by users during interaction; T denotes time, which defines the temporal state of the user, their actions, and the corresponding storefront characteristics at each point in time.

The system comprises the following key components:

Data Collection and Preprocessing Layer
This layer ingests raw clickstream data, session metadata, and Digital Storefront attributes. It performs cleaning, enrichment, and transformation of the data into structured formats. Both aggregated and sequential representations of user behavior are generated to support downstream analytical modules.

Metric Monitoring and Time Series Analysis Module
This module continuously monitors key economic indicators of the Digital Storefront — such as conversion rate (CR) and average order value (AOV) — to detect statistically significant deviations from expected trends, seasonal patterns, or baselines. It employs classical and probabilistic time series modeling techniques to capture the temporal dynamics of these metrics.

Among the most widely used approaches are ARIMA (AutoRegressive Integrated Moving Average) models, which are effective for detecting deviations from established trends under assumptions of linear stationarity. In addition, Bayesian classifiers (Yeo, Kim, Koh, Hwang, & Lipka, 2017) provide a probabilistic framework for incorporating prior knowledge and uncertainty into anomaly detection, allowing the system to evaluate the likelihood of a given deviation under various contextual assumptions.

Furthermore, Hidden Markov Models (HMMs) (Ding, Li, & Chatterjee, 2015) are applied to capture underlying latent states that influence observed performance metrics, making it possible to identify regime shifts or sudden changes in user engagement behavior. These models are particularly useful when performance metrics exhibit non-observable structural transitions over time.

By integrating multiple time series modeling paradigms, this module enhances the system's ability to detect both gradual drifts and abrupt shifts, ensuring timely identification of anomalies that may impact storefront performance.

Session-Level Anomaly Detection Module
This component, based on unsupervised machine learning and density-based clustering methods (e.g., DBSCAN, LOF), identifies user sessions whose behavior deviates from typical clusters. The feature representation includes both aggregated indicators and sequence-aware descriptors.

The purpose of this module is to detect sessions that differ from standard behavioral trajectories within a given context (e.g., time window, user segment, or storefront structure). These sessions are not interpreted directly but serve as input for subsequent detailed analysis aimed at identifying structural shifts in interaction patterns with the storefront.

Sequential Pattern Analysis Module
Utilizing deep learning architectures such as RNN (Toth, Tan, Di Fabbrizio, & Datta, 2017), LSTM, GRU (Koehn, Lessmann, & Schaal, 2020) or Transformer-based models (Grigoraș & Leon, 2023), this module captures the temporal structure of user interactions, enabling the detection of anomalies in action sequences that deviate from established behavioral patterns. In addition to traditional sequence modeling, this component may incorporate graph-based representations of user behavior, where nodes correspond to actions or events, and edges represent transitions or temporal dependencies.

Within this framework, anomaly detection methods range from classical graph-based techniques to neural network–based approaches, such as Network Representation Learning (Yu *et al.*, 2018), Graph Convolutional Networks (GCN) (Zheng, Li, Li, Li, & Gao, 2019), and LSTM applied to graph sequences (Teng, Yan, Ertugrul, & Lin, 2018). Classical techniques typically involve computing aggregate structural indicators—often based on matrix approximations or embeddings — and flagging anomalies when such indicators exceed predefined thresholds. For example, matrix decomposition methods have been successfully used for anomaly scoring in prior studies (Tong, Papadimitriou, Sun, Yu, & Faloutsos, 2008; (Rossi, Gallagher, Neville, & Henderson, 2013), although they tend to be computationally intensive when applied to large-scale data (Thudumu, Branch, Jin, & Singh, 2020).

By combining sequence modeling with graph-based anomaly detection, the module enhances its ability to identify complex, context-dependent deviations in user behavior that may not be evident from aggregated metrics alone.

Attribution and Interpretation Layer
This layer aggregates detected anomalies and cross-references them with storefront attributes (A), allowing the system to link anomalies to potential causal changes (e.g., a new promotion, content update, or technical issue). It generates interpretable outputs for analysts, including ranked lists of anomalies, heatmaps of affected segments, and suggested hypotheses.

To enhance the clarity and actionability of these outputs, a large language model (LLM) component may be employed to synthesize the anomaly context, contributing factors, and potential implications into human-readable summaries. The LLM can transform structured analytical results into concise textual explanations and decision-support suggestions, helping analysts quickly grasp the relevance and urgency of each detected anomaly in business terms.

Control Interface and Automation Framework
A user interface allows analysts and business users to monitor detected anomalies, adjust detection sensitivity, validate findings, and trigger mitigation actions. This layer also supports automation pipelines, enabling real-time detection and alerting without human intervention where appropriate.

Such a system ensures a comprehensive and adaptive approach to anomaly detection by combining statistical rigor, machine learning generalization, and interpretability. Its modular structure supports scalability and incremental development, while its interpretability ensures practical value for business decision-making in dynamic eCommerce environments.

ARCHITECTURAL RATIONALE

The architecture proposed in this study is modular in nature, reflecting the complex, multi-layered structure of the anomaly detection problem in Digital Storefront environments. Instead of relying on a single end-to-end model, the system decomposes the analysis into specialized components, each responsible for solving a subtask at a particular level of granularity—from aggregated metric analysis to behavioral pattern identification within individual user sessions. This approach is justified both technically and methodologically.

From a technical standpoint, modularity enables a progressive narrowing of the analytical space: in the initial stage, the time series analysis module detects deviations in business metrics (e.g., conversion rate or average order value). The system then proceeds to identify anomalous user sessions—those that exhibit abnormal values in aggregated behavioral indicators compared to what is observed under typical operating conditions. Only after this filtering step is the deep behavior analysis module activated, enabling detailed examination of user actions and their structure within the selected sessions. In this way, modularity not only

allows for the application of specialized methods at each stage, but also effectively reduces the dimensionality of the problem by eliminating the need to process and interpret the full volume of behavioral data at once.

From an applied perspective, the modular architecture ensures interpretability and actionability of outputs. The transition from aggregated signals to behavioral units enables the generation of explanations tied to understandable business objects (e.g., user segments, device types, product categories). This, in turn, facilitates the creation of reports and alerts that are readily interpretable and can serve as a basis for managerial decision-making.

Moreover, a key feature of the architecture is its focus on economically significant anomalies. Not every statistical deviation warrants a response: the system prioritizes anomalies that have a tangible impact on key business indicators of the storefront and, therefore, may require attention from analysts and product teams. This substantially reduces the volume of noise and increases the practical usefulness of the system.

Finally, the proposed architecture reflects an established practice in analytical diagnostics commonly followed by product and marketing teams. When an anomaly is detected at the aggregate level, analysts typically aim to localize it to specific user segments or sessions, and then examine behavioral characteristics and potential causal factors on the storefront side. The architecture reproduces this sequence of actions and formalizes it into a structured, systematic workflow, making the anomaly diagnosis process more robust, reproducible, and suitable for automation.

CONCLUSION

This paper addresses the problem of detecting and interpreting economically significant anomalies in user behavior on Digital Storefronts. The analysis demonstrates that user behavior is influenced by a wide range of dynamic and interrelated factors, including technological conditions, marketing stimuli, and content presentation. These factors impact key performance indicators such as conversion rate and average order value, and their fluctuations may indicate underlying behavioral shifts or operational issues.

A comparative review of anomaly detection methods reveals that no single class of methods fully meets all critical evaluation criteria — such as adaptability to change, interpretability, computational efficiency, sequential awareness, confidence control, and support for automation. This limitation underscores the need for a system-level solution that integrates multiple analytical paradigms.

To address this challenge, a conceptual architecture is proposed based on modular decomposition. The system operates within a multidimensional behavioral space ⟨U, X, A, T⟩, which captures user attributes, action sequences, storefront characteristics, and temporal context. The architecture includes specialized subsystems for time series monitoring (e.g., using ARIMA, Bayesian methods, and Hidden Markov Models), session-level anomaly detection (e.g., via unsupervised and density-based methods), sequential pattern analysis (e.g., LSTM, GRU, Transformers), and graph-based anomaly modeling (e.g., GCN, matrix decomposition).

A key emphasis is placed on generating interpretable and actionable results that can support timely decision-making by analysts and storefront managers. The proposed system thus offers a foundation for scalable, adaptive, and explainable anomaly detection in high-throughput eCommerce environments. Future work may include real-world validation on large-scale datasets, development of causality modules, and integration into live monitoring and recommendation infrastructures.

REFERENCES

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. *Data mining and knowledge discovery*, *29*, 626-688. https://doi.org/10.1007/s10618-014-0365-y

Bi, M., Xu, J., Wang, M., & Zhou, F. (2016). Anomaly detection model of user behavior based on principal component analysis. *Journal of Ambient Intelligence and Humanized Computing*, *7*, 547-554. https://doi.org/10.1007/s12652-015-0341-4

Büchner, A.G., & Mulvenna, M.D. (1998). Discovering internet marketing intelligence through online analytical web usage mining. *ACM Sigmod Record*, *27*(4), 54-61. https://doi.org/10.1145/306101.306124

Chen, M.S., Park, J.S., & Yu, P.S. (1996). Data mining for path traversal patterns in a web environment. In *Proceedings of 16th International Conference on Distributed Computing Systems* (pp. 385-392). IEEE. https://doi.org/10.1109/ICDCS.1996.507986

Chollet, F., & Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

Cirqueira, D., Hofer, M., Nedbal, D., Helfert, M., & Bezbradica, M. (2019). Customer purchase behavior prediction in e-commerce: A conceptual framework and research agenda. *In International Workshop on New Frontiers in Mining Complex Patterns* (pp. 119-136). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-48861-1_8

Ding, A.W., Li, S., & Chatterjee, P. (2015). Learning user real-time intent for optimal dynamic web page transformation. *Information Systems Research*, *26*(2), 339-359. https://doi.org/10.1287/isre.2015.0568

Grigoraș, A., & Leon, F. (2023). Transformer-based model for predicting customers' next purchase day in e-commerce. *Computation*, *11*(11), 210. https://doi.org/10.3390/computation11110210

Hoffman, D.L., Novak, T.P., & Chatterjee, P. (1995). Commercial scenarios for the web: opportunities and challenges. *Journal of computer-mediated communication*, *1*(3), JCMC136. https://doi.org/10.1111/j.1083-6101.1995.tb00165.x

Joachims, T., Freitag, D., & Mitchell, T. (1997). Webwatcher: A tour guide for the world wide web. In *International Joint Conferences on Artificial Intelligence* (1) (pp. 770-777).

Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*, *150*, 113342. https://doi.org/10.1016/j.eswa.2020.113342

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, *2*(1), 1-15. https://doi.org/10.1145/360402.360406

Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q.Z., ... & Akoglu, L. (2021). A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12012-12038. https://doi.org/10.1109/TKDE.2021.3118815

Mobasher, B., Cooley, R., & Srivastava, J. (1999). Creating adaptive web sites through usage-based clustering of URLs. In *Proceedings 1999 Workshop on Knowledge and Data Engineering Exchange (KDEX'99)*(Cat. No. PR00453) (pp. 19-25). IEEE. https://doi.org/10.1109/KDEX.1999.836525

Mobasher, B., Jain, N., Han, E.H., & Srivastava, J. (1996). *Web mining: Pattern discovery from world wide web transactions (pp. 558-567)*. Technical Report TR 96-050, University of Minnesota, Dept. of Computer Science, Minneapolis.

Moe, W.W., & Fader, P.S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science*, *50*(3), 326-335. https://doi.org/10.1287/mnsc.1040.0153

Ngu, D.S.W., & Wu, X. (1997). Sitehelper: A localized agent that helps incremental exploration of the world wide web. *Computer Networks and ISDN Systems*, *29*(8-13), 1249-1255. https://doi.org/10.1016/S0169-7552(97)00055-X

Pierrakos, D., Paliouras, G., Papatheodorou, C., & Spyropoulos, C.D. (2003). Web usage mining as a tool for personalization: A survey. *User Modeling and User-Adapted Interaction*, *13*, 311-372. https://doi.org/10.1023/A:1026238916441

Rossi, R.A., Gallagher, B., Neville, J., & Henderson, K. (2013). Modeling dynamic behavior in large evolving graphs. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining* (pp. 667-676). https://doi.org/10.1145/2433396.2433479

Sharma, K., Shrivastava, G., & Kumar, V. (2011). Web mining: Today and tomorrow. In *2011 3rd International Conference on Electronics Computer Technology* (Vol. 1, pp. 399-403). IEEE. https://doi.org/10.1109/ICECTECH.2011.5941631

Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of marketing research*, *41*(3), 306-323. https://doi.org/10.1509/jmkr.41.3.306.35985

Spiliopoulou, M., Pohle, C., & Faulstich, L.C. (1999). Improving the effectiveness of a web site with web usage mining. In *International Workshop on Web Usage Analysis and User Profiling* (pp. 142-162). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-44934-5_9

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P.N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, *1*(2), 12-23. https://doi.org/10.1145/846183.846188

Teng, X., Yan, M., Ertugrul, A.M., & Lin, Y.R. (2018). Deep into hypersphere: Robust and unsupervised anomaly discovery in dynamic networks. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*.

Thudumu, S., Branch, P., Jin, J., & Singh, J. (2020). A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, *7*, 1-30. https://doi.org/10.1186/s40537-020-00320-x

Toth, A., Tan, L., Di Fabbrizio, G., & Datta, A. (2017). Predicting shopping behavior with mixture of RNNs. In *Proceedings of the 2017 SIGIR Workshop on E-commerce*. *2017*. https://ceur-ws.org/Vol-2311/paper_6.pdf

Yeo, J., Kim, S., Koh, E., Hwang, S.W., & Lipka, N. (2017). Predicting online purchase conversion for retargeting. *In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 591-600). https://doi.org/10.1145/3018661.3018715

Yu, W., Cheng, W., Aggarwal, C. C., Zhang, K., Chen, H., & Wang, W. (2018). Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2672-2681). https://doi.org/10.1145/3219819.3220024

Zaiane, O.R., Xin, M., & Han, J. (1998). Discovering web access patterns and trends by applying OLAP and data mining technology on web logs. In *Proceedings IEEE International Forum on Research and Technology Advances in Digital Libraries-ADL'98* (pp. 19-29). IEEE. https://doi.org/10.1109/ADL.1998.670376

Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. *ACM Sigkdd Explorations Newsletter*, *2*(1), 1-15. https://doi.org/10.1145/360402.360406

Zheng, L., Li, Z., Li, J., Li, Z., & Gao, J. (2019). AddGraph: Anomaly detection in dynamic graph using attention-based temporal GCN. In *International Joint Conferences on Artificial Intelligence* (Vol. 3, p. 7). https://doi.org/10.24963/ijcai.2019/614