https://doi.org/10.31891/2307-5732-2026-363-28
УДК 004.9

**SVYRYDOV ARTEM**
Kharkiv National University of Radio Electronics
https://orcid.org/0000-0002-9830-4103
e-mail: SvyrydovArtem@gmail.com
**HUSEINOV RUSTAM**
Kharkiv National University of Radio Electronics
https://orcid.org/0009-0001-3318-2969
e-mail: rustam.huseinov@nure.ua
**VYNNYCHENKO SERHII**
Kharkiv National University of Radio Electronics
https://orcid.org/0009-0007-5200-845X
e-mail: serhii.vynnychenko@nure.ua

# METHOD FOR DETECTING ANOMALIES IN WEBSITE USER BEHAVIOR

*The object of the study is an approach to detecting anomalous user behavior in a web environment based on the analysis of web activity logs. The subject of the study comprises methods for behavioral analysis of user web sessions and their integration with deep learning algorithms for building intelligent anomaly detection systems. The paper addresses the problem of forming and analyzing a dataset constructed from proxy logs that contain information on the temporal characteristics of web sessions, activity duration, and sequences of visited domain names. Particular attention is paid to data preprocessing, including web session construction, extraction of temporal and behavioral features, development of indicators for user-specific characteristic resources, and temporal splitting of the dataset into training and test subsets to prevent information leakage. To model the dynamic nature of user behavior, a sequential data representation is employed, enabling preservation of action order within a session. Given the limited number of anomalous behavior samples and the lack of complete information about possible violation types, the suitability of a one-class anomaly detection framework is justified, in which the model is trained exclusively on data representing the normal behavior of a specific user. To address the stated problem, an LSTM autoencoder is proposed, as it is capable of modeling temporal and behavioral dependencies in web sessions and detecting deviations from the normal behavior profile through reconstruction error analysis. An approach to determining the anomaly threshold based on statistical characteristics of the reconstruction error is also proposed. The effectiveness of the method is evaluated using standard classification performance metrics, including precision, recall, and F1-score. The obtained results confirm the ability of the proposed approach to effectively distinguish the target user's behavior from external activity without relying on prior knowledge of anomaly types. The practical significance of the study lies in the possibility of applying the proposed method to tasks of user behavioral identification, detection of unauthorized access, and enhancement of the security level of web-based information systems.*

***Keywords:*** *anomaly detection, user behavior analysis, normal behavior profile, web sessions, web logs, one-class classification, LSTM autoencoder, deep learning, information security, behavioral user identification.*

**СВИРИДОВ АРТЕМ, ГУСЕЙНОВ РУСТАМ, ВИННИЧЕНКО СЕРГІЙ**
Харківський національний університет радіоелектроніки

# МЕТОД ВИЯВЛЕННЯ АНОМАЛІЙ У ПОВЕДІНЦІ КОРИСТУВАЧА ВЕБСАЙТУ

*Об'єктом дослідження є підхід до виявлення аномальної поведінки користувачів у веб середовищі на основі аналізу журналів веб-активності. Предметом дослідження є методи поведінкового аналізу веб-сесій користувачів та їх поєднання з алгоритмами глибинного навчання для побудови інтелектуальних систем виявлення аномалій. У роботі розглядається задача формування та аналізу набору даних, побудованого на основі проксі-журналів, що містять інформацію про часові характеристики веб-сесій, тривалість активності та послідовність відвіданих доменних імен. Особливу увагу приділено попередній обробці даних, зокрема формуванню веб-сесій, виділенню часових і поведінкових ознак, побудові індикаторів характерних ресурсів користувача та часовому поділу вибірки на навчальну і тестову частини з метою запобігання витоку інформації. Для моделювання динамічної поведінки користувача використано послідовне представлення даних, що дозволяє враховувати порядок дій у межах сесії. З урахуванням обмеженої кількості прикладів аномальної поведінки та відсутності повної інформації про можливі типи порушень, обґрунтовано доцільність застосування підходу виявлення аномалій у one-class постановці, заснованої на навчанні моделі виключно на даних нормальної поведінки конкретного користувача. Для розв'язання поставленої задачі запропоновано використання LSTM-autoencoder, здатного моделювати часові та поведінкові залежності веб-сесій та виявляти відхилення від профілю нормальної поведінки шляхом аналізу помилки реконструкції. Запропоновано підхід до визначення порогу аномальності на основі статистичних характеристик помилки реконструкції. Ефективність методу оцінено за допомогою стандартних метрик якості класифікації, зокрема precision, recall та F1-score. Отримані результати підтверджують здатність запропонованого підходу ефективно відокремлювати поведінку цільового користувача від сторонньої активності без використання інформації про типи аномалій. Практична значущість роботи полягає у можливості застосування розробленого методу для задач поведінкової ідентифікації користувачів, виявлення несанкціонованого доступу та підвищення рівня безпеки веб-інформаційних систем.*

***Ключові слова:*** *виявлення аномалій, поведінковий аналіз користувачів, профіль нормальної поведінки, веб-сесії, веб-журнали, one-class класифікація, LSTM-autoencoder, глибинне навчання, інформаційна безпека, поведінкова ідентифікація користувачів.*

Стаття надійшла до редакції / Received 17.01.2026
Прийнята до друку / Accepted 11.02.2026
Опубліковано / Published 26.03.2026

**Problem statement in a general form and its relation to important scientific and practical tasks**

In the modern digital environment, websites play a key role in providing information services, e-commerce, distance learning, and communication between users and software systems. The continuous growth in the number of Internet users, increasing complexity of web resource architectures, and the expanding volume of data generated during user–website interactions create new challenges in the analysis and processing of behavioral information [1]. Under these conditions, methods of intelligent data analysis capable of automatically detecting atypical or potentially dangerous behavioral scenarios become particularly relevant.

User behavior on a website is characterized by multidimensionality and dynamicity and depends on a range of factors, including visit objectives, technical characteristics of devices, temporal parameters, and individual interaction patterns with the interface [2]. At the same time, anomalies—deviations from typical behavior that do not conform to established patterns—may appear within the overall stream of user actions. Such anomalies can result from malicious activity (e.g., bot attacks, attempts at unauthorized access, fraud), technical failures, or sudden changes in the behavior of legitimate users. Timely detection of these deviations is critically important for ensuring information security, maintaining the stable operation of web resources, and improving the quality of user experience.

Traditional approaches to user behavior analysis, based on static rules or manual monitoring, are inefficient in the context of large-scale data and rapidly evolving behavioral patterns. Consequently, there is growing interest in anomaly detection methods based on statistical techniques, machine learning, and time-series analysis. These methods enable the construction of a "normal" user behavior model and the automatic identification of deviations without requiring a prior description of all possible anomalous scenarios.

This article proposes a method for detecting anomalies in website user behavior that is based on the analysis of behavioral characteristics and takes into account the specifics of user interaction with web interfaces. The proposed approach aims to improve anomaly detection accuracy and can be applied to security monitoring, optimization of website operation, and decision support in web-based systems.

**Analysis of Related Research and Publications**

The problem of anomaly detection in website user behavior is actively investigated in contemporary scientific publications in the fields of data analysis, information security, and web analytics. With the growth of web data volumes and the increasing complexity of behavioral scenarios, considerable attention is paid to the development of methods capable of automatically identifying atypical or potentially dangerous user actions.

Scientific studies examine both classical statistical approaches to anomaly detection and machine learning and deep learning methods that enable modeling of normal user behavior and detection of deviations from it. A separate line of research focuses on the analysis of behavioral characteristics of web sessions, such as action sequences, temporal parameters, and the intensity of interaction with web interfaces.

In [3], the application of advanced machine learning algorithms for anomaly detection in cloud networks is investigated, with an emphasis on supervised, unsupervised, and hybrid approaches. A comprehensive analysis of algorithmic performance is presented, including random forests, support vector machines, autoencoders, and deep neural networks, taking into account accuracy, false-positive rates, and computational costs. The study also addresses real-time anomaly detection using streaming data and the integration of cloud monitoring tools. The results demonstrate that machine learning–based models, when properly trained and continuously updated, significantly improve the detection of both known and zero-day anomalies, providing robust and adaptive cloud security frameworks.

In [4], the authors present a hybrid anomaly detection method called DT-SVMNB, which cascades several machine learning algorithms, including a decision tree (C5.0), a support vector machine (SVM), and a naïve Bayes classifier (NBC), to classify normal and anomalous users in social networks. A set of unique features extracted from user profiles and content is constructed. Using two types of datasets with selected features, the proposed DT-SVMNB machine learning model is trained and evaluated.

Study [5] addresses the problem of web traffic anomaly detection under conditions of increasing web data volumes and heightened cybersecurity requirements driven by the digital transformation of enterprises. The authors justify the use of unsupervised machine learning methods for analyzing large-scale weblog data and propose the Isolation Forest algorithm for automatic separation of anomalous and normal traffic. The research is based on a publicly available e-commerce website weblog dataset, for which a complete data preparation pipeline is implemented, including cleaning, normalization, and feature engineering. The effectiveness of the proposed approach is evaluated by comparing model outputs with expert assessments from cybersecurity specialists. The Isolation Forest model implemented using the Scikit-learn library demonstrates high performance metrics, including accuracy, precision, recall, and F1-score. The results confirm that, with proper data preparation, the Isolation Forest algorithm can serve as an effective and practically applicable tool for web traffic anomaly detection in real-world information systems.

In [6], a multi-perspective approach to detecting anomalies in website user behavior is proposed, aiming to overcome the limitations of traditional methods that consider only request sequences or individual semantic features. The authors analyze user behavior at the session level and combine access sequence analysis with semantic analysis of session content using an enhanced SimHash algorithm and a multi-attention Transformer model. The proposed end-to-end model effectively integrates sequential and semantic characteristics, significantly improving anomaly detection performance. Experimental results indicate high accuracy, stability, and practical applicability of the approach, particularly in real-world scenarios, where the model achieves high precision, recall, and F1-score even with a limited number of anomalous sessions.

The conducted analysis of recent scientific research and publications indicates substantial interest in the problem of anomaly detection in website user behavior, driven by the growth of web data volumes, increasing complexity of behavioral patterns, and rising information security requirements. The reviewed works cover a wide range of approaches—from classical statistical methods to modern machine learning and deep learning algorithms—that effectively model normal user behavior and detect deviations across various application scenarios.

At the same time, the analysis reveals that, despite the high accuracy of individual models, there are limitations related to dependence on data quality, the complexity of interpreting deep learning model decisions, and the need to adapt to dynamic changes in user behavior. This underscores the relevance of further research aimed at developing methods that combine different types of behavioral features, ensure high accuracy and stability of anomaly detection, and can be effectively integrated into real-world web-based cybersecurity systems.

## Research Objectives

The aim of the study is to improve the effectiveness of detecting anomalies in website user behavior by applying an LSTM autoencoder to analyze web session sequences and identify atypical user–website interaction scenarios.

## Presentation of the Main Material

For this study, a dataset of user web activity was collected from the proxy servers of Blaise Pascal University. The data consist of HTTP request logs containing information about user behavior during interactions with web resources. Each log entry includes a user identifier, a session start timestamp, and the domain name of the visited resource. The initial dataset contained approximately $17\times10^6$ records generated by over 3,000 users.

To improve data quality and remove noise, preliminary filtering was performed. In the first stage, blacklist filters were applied to remove advertising and service-related domains. In the second stage, filtering was performed based on HTTP request status codes, allowing the exclusion of unavailable or invalid domains. After these procedures, the dataset was reduced to approximately $4\times10^6$ records.

The cleaned data were grouped by users and segmented into individual web sessions. For each session, a structured description was created, including the sequence of visited domains (up to 10 sites per session), session start time, duration, day of the week, start hour, and other temporal characteristics. Subsequent analysis focused on 150 users with the highest number of requests, ensuring sufficient behavioral data for modeling purposes.

For the purposes of behavioral identification and anomaly detection, all users were divided into two classes: the target user (Alice) and other users. A binary variable target was introduced, where 1 corresponds to Alice and 0 to other users. This approach enables comparative analysis of behavioral patterns and the construction of a model of normal behavior.

Additional features were generated to prepare the data for modeling, including an indicator for the presence of domains characteristic of Alice within a session. The dataset was temporally split into training and test subsets based on a fixed date, preventing information leakage and enabling accurate evaluation of model performance under conditions close to real-world scenarios.

The prepared dataset provides a foundation for applying sequence analysis methods and constructing an anomaly detection model for user behavior using an LSTM autoencoder.

Figure 1 shows the distribution of web sessions by day of the week for the analyzed dataset. The results indicate that the highest user activity occurs on weekdays, with a peak in the middle of the week—Wednesday. High numbers of sessions are also observed on Tuesday and Thursday, reflecting regular and stable user activity during the workweek.

Conversely, weekends show a marked decrease in session numbers, particularly on Sunday, likely due to changes in user routines and reduced use of web resources during non-working hours. This clearly defined weekly cyclic pattern is an important characteristic of normal user activity and should be considered when building anomaly detection models. Deviations from the typical distribution of activity by day of the week may serve as indicators of atypical or potentially anomalous user behavior.

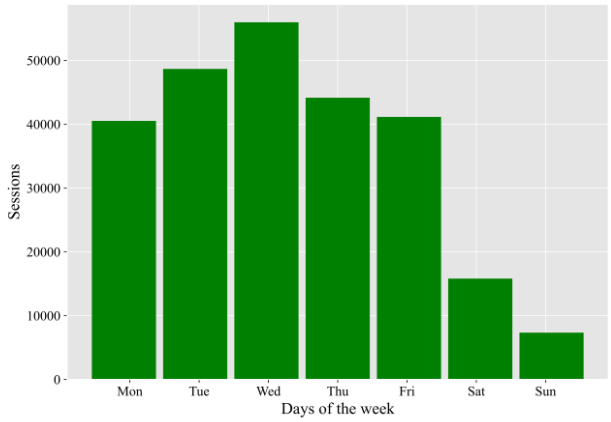


**Figure 1. Distribution of web sessions by day of the week**

In Figure 2, the distribution of web sessions by day of the week for the target user Alice is shown. The analysis shows that Alice's behavior exhibits pronounced unevenness throughout the week. The highest number of sessions occurs on Monday, indicating high user activity at the beginning of the workweek. Significant activity is also observed on Tuesday and Thursday, while Wednesday shows a noticeably lower number of sessions compared to other weekdays.

On Friday, activity decreases, and during the weekend (Saturday and Sunday), the number of sessions is minimal. This distribution reflects a clear dependence of user behavior on the work schedule and allows characterization of Alice's typical weekly activity pattern. The observed patterns can be used as an important indicator of normal behavior when building anomaly detection models, as significant deviations from this distribution may indicate atypical or potentially anomalous activity.

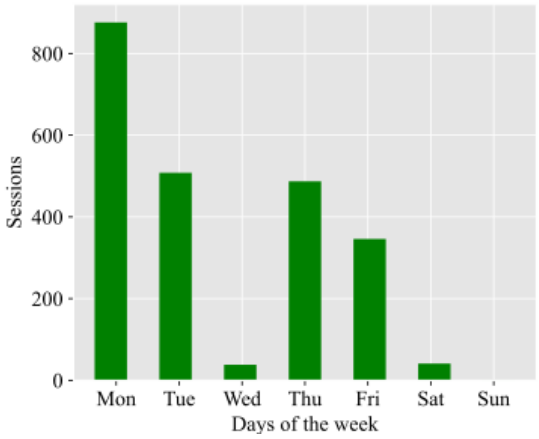


**Figure 2. Distribution of web sessions by day of the week for the user Alice**

Figure 3 shows the distribution of web sessions by day of the week for other users. The results show a clearly defined weekly periodicity in behavior, characteristic of most users. Peak activity occurs in the middle of the workweek, with the highest number of sessions on Wednesday, while Monday and Tuesday also exhibit high session counts.

In the latter half of the week, starting from Thursday, activity gradually decreases, and during the weekend (Saturday and Sunday), there is a sharp decline in session numbers. This distribution reflects a typical workweek schedule and represents a generalized pattern of normal behavior for the user group. Comparing this distribution with the corresponding pattern for the user Alice allows identification of individual behavioral characteristics and can serve as an additional criterion for anomaly detection within behavioral analysis.

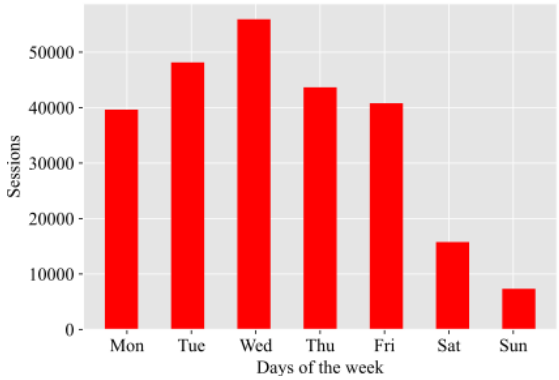


**Figure 3. Distribution of web sessions by day of the week for other users**

In Figure 4, the distribution of web sessions by hour of the day for the user Alice is shown. The results indicate a clearly defined temporal activity pattern characteristic of this user. Most sessions are concentrated during the daytime and early afternoon periods, with peak activity occurring between 16:00 and 17:00. An additional concentration of sessions is observed in the late morning, around 12:00–13:00.

Activity is minimal or almost absent during the early morning hours (before 9:00) and in the evening after 18:00. This distribution reflects a consistent daily routine of web resource usage and can be interpreted as the user's typical work schedule. The identified temporal patterns are an important feature of Alice's normal behavior and can be used in building anomaly detection models, as deviations from this schedule may indicate atypical or potentially anomalous activity.

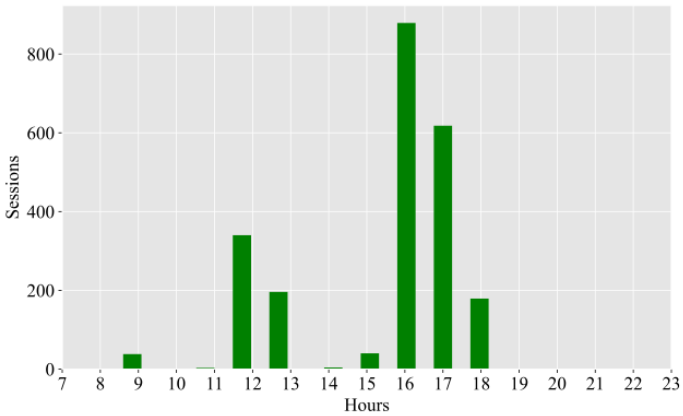


**Figure 4. Distribution of web sessions by hour of the day for the user Alice**

In Figure 5, the distribution of web sessions by hour of the day for other users (excluding Alice) is shown. The results demonstrate a more uniform and temporally extended activity profile compared to the target user. The highest session intensity is observed in the morning and late-morning hours, with a peak between approximately 9:00 and 11:00. In the early afternoon (13:00–16:00), activity remains relatively high, gradually decreasing in the later part of the day.

Starting from 17:00, the number of sessions drops significantly, and during the evening and night hours (after 18:00), the majority of users exhibit minimal activity. This distribution reflects a generalized typical pattern of web resource usage and corresponds to a standard workday schedule. Comparing this profile with Alice's temporal activity schedule allows identification of individual behavioral differences and can serve as an informative feature in constructing models for detecting anomalous user behavior.

In Figure 6, the distribution of web session durations for the user Alice is presented in a logarithmic scale. The use of a logarithmic transformation reduces the skewness of the distribution and visually highlights both short and long sessions. The results show that the vast majority of sessions have relatively short durations, with a concentration in the range of small to medium logarithmic duration values.

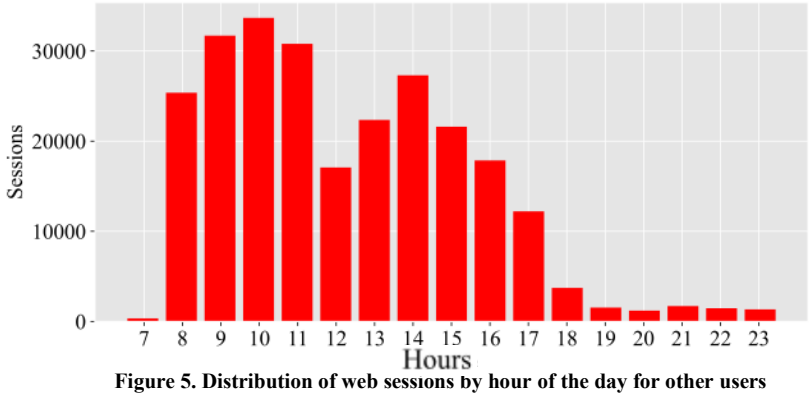


**Figure 5. Distribution of web sessions by hour of the day for other users**

At the same time, the distribution exhibits a long right-hand tail, corresponding to rare but significantly longer sessions. This shape is typical for user behavioral data and reflects a stable pattern of Alice's normal activity. The observed session duration characteristics can serve as an informative feature in building anomaly detection models, as significant deviations from this distribution (e.g., unusually short or excessively long sessions) may indicate atypical or anomalous behavior.

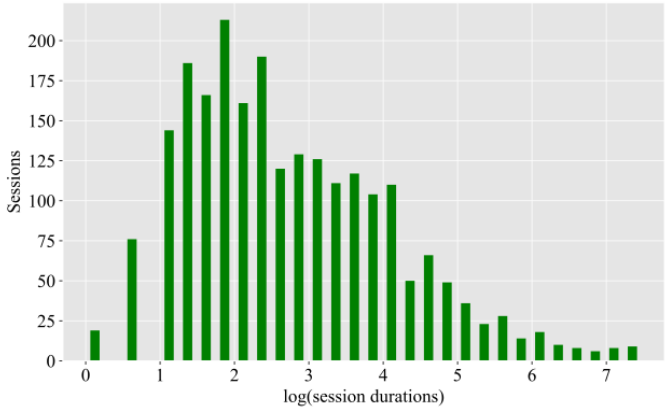


**Figure 6. Distribution of the logarithmic web session durations for the user Alice**

In Figure 7, the distribution of logarithmic web session durations for other users (excluding Alice) is shown. As with the target user, a logarithmic transformation of session durations was applied to accurately represent the wide range of values and reduce the impact of occasional long sessions. The resulting distribution is noticeably skewed, with a concentration of values in the range of small to medium logarithmic durations, reflecting the typical behavior of most users.

At the same time, other users exhibit a broader distribution of session durations and a higher number of sessions with medium to elevated duration values compared to Alice. The presence of a long right-hand tail indicates a significant number of long sessions, which is characteristic of generalized group behavior. Comparative analysis of session duration distributions for Alice and other users allows identification of individual behavioral features and can serve as an informative characteristic in building anomaly detection models, particularly those based on an LSTM autoencoder.

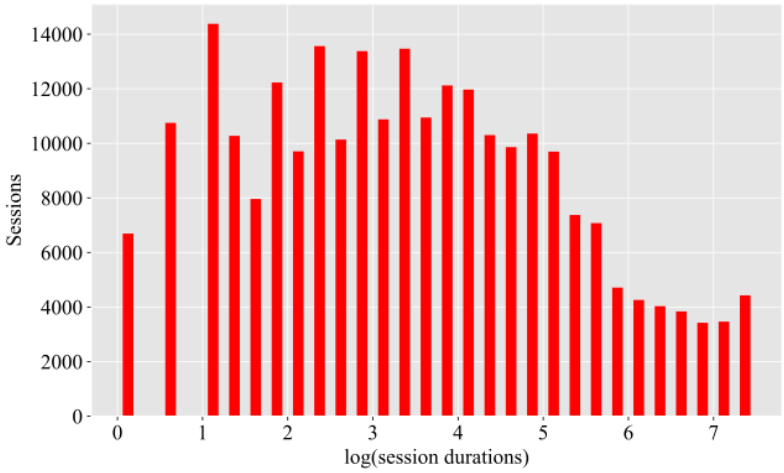


**Figure 7. Distribution of logarithmic web session durations for other users**

In the next stage of the study, the data were prepared for model training and evaluation. To ensure the correctness of the experiment and prevent information leakage between datasets, a temporal split was applied based on a threshold date. All sessions that started before this date were assigned to the training set, while sessions that started on or after the threshold date were assigned to the test set. Additionally, for each session, a feature day was created, representing the day of the year, which allows analysis of user activity across calendar days.

To enhance the recognition of behavioral characteristics of the target user, indicator features related to Alice's "characteristic sites" were created. First, using the training set, domain visit statistics were calculated: for each domain name, the mean value of target (1 for Alice, 0 for other users) was determined. Domains for which this proportion exceeded the threshold of 0.06 were included in the "Alice sites" set. For each session, a binary feature alice_site was generated, taking the value 1 if at least one of the sites (site1…site10) belonged to this set, and 0 otherwise. Similarly, extended indicators (alice_site_2, alice_site_3, alice_site_4) were added to capture the presence of "typical" resources within a session, thereby increasing the informativeness of the behavioral description.

The next step involved filtering sessions based on activity characteristics. For each day of the year, the total number of sessions with alice_site=1 was calculated, and "informative" days (good_days) were defined as those where this sum exceeded the threshold of 70, while "weak" days (bad_days) were those with values not exceeding 70. Similarly, selection rules were applied based on hours of the day: typical activity intervals (good_hours) were identified as 12, 13, 16, and 17, while atypical hours (bad_hours) were 7, 8, 10, 11, 14, and 19–23. Rare hours (9, 15, 18) were treated separately, applying additional filtering based on the minutes of session start times (splitting into permissible and non-permissible minute intervals). As a result, reduced datasets train_df_short and test_df_short were created, containing predominantly sessions with characteristic temporal and behavioral features, as well as sets of identifiers for "good" and "bad" sessions for further analysis.

For preparing the data for machine learning, a final set of features was constructed. This set includes behavioral indicators (alice_site, alice_site_2, alice_site_3, alice_site_4) as well as basic session characteristics: number_of_sites, session_len, weekday, start_hour, end_hour, time_index, and session_id. Categorical variables (weekday, start_hour, end_hour, number_of_sites) were transformed using one-hot encoding with the first category dropped to avoid multicollinearity. After encoding, records with missing values were removed. Feature matrices Xtrain and Xtest were then created by excluding the target column, while label vectors ytrain and ytest were taken from the corresponding column. Additionally, dummy columns corresponding to atypical or non-informative hours were removed, as well as auxiliary variables (time_index and alice_site_4).

After preparing the data and constructing an informative set of behavioral features, it becomes possible to proceed directly to building the anomaly detection method. The proposed approach is based on forming a profile of the normal behavior of a specific user and then identifying deviations from this profile. This method does not require information about all possible types of anomalies and implements a one-class formulation of the problem, which is particularly relevant for user behavioral analytics.

As a tool for modeling normal behavior, a recurrent autoencoder based on Long Short-Term Memory (LSTM-autoencoder) is used, capable of effectively capturing temporal and sequential dependencies in user behavior. The model

is trained exclusively on sessions of the target user, which are treated as examples of normal behavior. After training, deviations between the input and reconstructed session features are used as a criterion for anomaly detection.

Thus, the method allows for the automatic identification of sessions that do not conform to the established behavioral profile of the user and can be applied to tasks such as behavioral identification, access control, and detection of unauthorized account usage.

Below is a detailed description of the method's stages, illustrating the overall logic of constructing a user's normal behavior profile and the mechanism for detecting deviations, according to the structural scheme shown in Figure 8.

Stage 1: Data Collection from Web Logs. At this stage, data are collected from web logs (proxy logs), which contain records of the form

$$\{session\_id, t, domain\}, \quad (1)$$

where session_id is the session identifier, t is the timestamp of the request, and domain is the domain name of the visited web resource. These data reflect the sequence of user actions in the web environment.

Stage 2: Data Preprocessing

At this stage, data cleaning, web session formation, and extraction of temporal features are performed. A web session is defined as a time-ordered sequence of requests from a single user. For each session, additional temporal characteristics are calculated, including the day of the week, start hour, and session duration. The result of this stage is a set of structured sessions:

$$S = \{S_1, S_2, \dots, S_N\}.$$

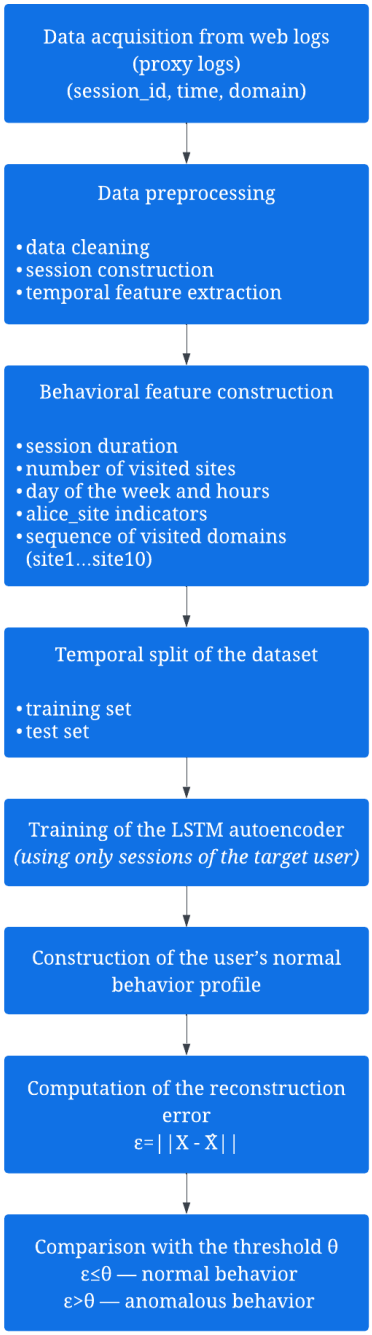


**Figure 8. Structural Diagram of the Method for Detecting User Anomalous Behavior Based on Building a Profile of Normal Behavior**

Stage 3: Feature Extraction. Each session $S_i$ is represented by a vector of behavioral features

$$X_i = \left(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)}\right), \quad (3)$$

These features take into account session duration, number of sites, day of the week, hours of activity, indicators of the user's characteristic sites, as well as the sequence of visited domains site1…, site10. In this way, the input feature space for the model is formed.

Stage 4: Temporal Split of the Dataset. To prevent data leakage, the dataset is split by a temporal threshold $T$ into training and testing parts:

$$\mathcal{S}_{train} = \{S_i \mid t_i < T\}, \quad \mathcal{S}_{test} = \{S_i \mid t_i \geq T\} \quad (4)$$

The training set contains only historical data, while the test set is used to evaluate the method.

Stage 5: Training the LSTM-Autoencoder. At this stage, the LSTM-autoencoder is trained exclusively on the sessions of the specific user, which are treated as examples of normal behavior. The autoencoder performs a mapping

$$f_\theta: X \to \hat{X}, \quad (5)$$

where $X$ represents the input behavioral features, $\hat{X}$ is their reconstruction, and, θ are the model parameters. Training consists of minimizing a loss function, for example, the mean squared error (MSE):

$$\mathscr{L} = \mathbb{E}\left[\|X - \hat{X}\|^2\right]. \quad (6)$$

Stage 6: Constructing the User's Normal Behavior Profile. As a result of training, a profile of the user's normal behavior is formed, reflecting the characteristic temporal and structural patterns of their activity. This profile is defined by the parameters of the trained autoencoder and the distribution of reconstruction errors on the training data.

Stage 7: Computing the Reconstruction Error. For each new session, the reconstruction error is calculated:

$$\varepsilon_i = \|X_i - \hat{X}_i\|, \quad (7)$$

which serves as a quantitative measure of deviation from the user's normal behavior profile.

Stage 8: Threshold Comparison and Decision Making. At the final stage, the reconstruction error is compared with a threshold value θ, determined based on the training data:

$$\begin{cases} \varepsilon_i \leq \theta, & normal\ behavior, \\ \varepsilon_i > \theta, & anomalous\ behavior. \end{cases} \quad (8)$$

Thus, sessions that do not conform to the user's normal behavior profile are identified as anomalous.

Table 1 presents the architecture of the LSTM-autoencoder used for the available system log data.

Table 1

**Architecture of the LSTM-Autoencoder for Building a User Behavior Profile**

| Layer | Layer Type | Output Size | Number of Parameters | Purpose |
|---|---|---|---|---|
| sites | Input | (None, 10) | 0 | Sequence of visited sites |
| embedding | Embedding | (None, 10, 32) | 975 808 | Vector representation of domains |
| not_equal | Mask | (None, 10) | 0 | Masking of missing values |
| lstm | LSTM | (None, 64) | 24 832 | Extraction of temporal dependencies |
| num | Input | (None, 5) | 0 | Numerical behavioral features |
| concatenate | Concatenate | (None, 69) | 0 | Feature concatenation |
| dense | Dense | (None, 64) | 4 480 | Non-linear transformation |
| latent | Dense | (None, 32) | 2 080 | Latent space |
| dense_1 | Dense | (None, 64) | 2 112 | Start of decoding |
| recon_num | Dense | (None, 5) | 325 | Reconstruction of numerical features |

Tables 2 and 3 present the main results of the anomalous behavior detection method.

Table 2

**Confusion Matrix**

| Actual class \ Predicted | Normal (Alice) | Anomaly (not Alice) |
|---|---|---|
| **Normal (Alice)** | **639** (TN) | **63** (FP) |
| **Anomaly (not Alice)** | **29 895** (FN) | **56 857** (TP) |

Table 3

**Key Performance Metrics**

| Metric | Value |
|---|---|
| Precision | **0.999** |
| Recall | **0.655** |
| F1-score | **0.791** |
| Threshold | **0.00537** |
| Number of sessions in test | **87 454** |
| Detected anomalies | **56 920** |

The obtained results indicate a high effectiveness of the proposed method in detecting anomalous user behavior based on the profile of normal activity. The confusion matrix shows that the model correctly identifies the majority of

sessions that do not conform to Alice's behavioral profile (56,857 true positive detections), while the number of false positives is minimal (63 cases).

The high precision value (0.999) indicates that when a session is classified as anomalous, the model almost always makes a correct decision, which is critically important for security systems and behavioral analytics. The recall value (0.655) shows that approximately 65.5% of all anomalous sessions were detected, while some atypical behavior remained undetected and was classified as normal.

The F1-score of 0.791 confirms a balance between precision and recall. Considering the one-class formulation of the task and training the model solely on data from a single user, these results are expected and demonstrate the ability of the LSTM-autoencoder to effectively form an individual behavioral profile and distinguish it from external activity.

## Conclusions and Future Research Directions

The article addresses the problem of detecting anomalous user behavior on a website based on building an individual profile of the user's normal activity. The proposed approach is based on a one-class problem formulation, which involves training the model exclusively on data from the target user without including examples of anomalous behavior during training. Such a formulation is practically relevant for real-world information systems, where the number of anomalous events is limited or unknown in advance.

A method for generating behavioral features from web logs was developed, combining session temporal characteristics, quantitative activity measures, and indicators of visits to user-specific websites. The conducted data analysis revealed stable patterns in the distributions of activity across weekdays, hours of the day, and session durations, confirming the suitability of these features for constructing a behavioral profile.

To model the user's normal behavior, an LSTM-autoencoder was employed, capable of effectively capturing temporal and sequential dependencies in the data. The trained model produced a compact latent representation of user behavior, and the reconstruction error was used as a quantitative measure of deviation from the normal profile. The threshold for decision-making was determined based on the reconstruction error distribution in the training dataset.

Experimental results demonstrated high accuracy in detecting anomalous behavior. In particular, the precision reached 0.999, indicating almost complete absence of false positives, while the F1-score was 0.791 with a recall of 0.655. These metrics confirm the method's ability to effectively distinguish the target user's behavior from the activity of other users in the test set.

The proposed method can be applied to tasks such as behavioral user identification, unauthorized access detection, and the development of behavioral analytics systems in web environments. Future research should focus on optimizing the choice of the decision threshold, expanding the set of sequential features, and integrating the proposed approach with other anomaly detection methods to improve detection completeness.

## References

1. Han, J., Pei, J., & Tong, H. (2022). Data mining: Concepts and techniques. Morgan Kaufmann. [Online resource]. Available at: https://www.sciencedirect.com/book/monograph/9780123814791/data-mining-concepts-and-techniques (Accessed: 22.12.2025)

2. Roy, D. K., & Kalita, H. K. (2025). Anomaly detection in an open set environment using reinforcement learning. IET Information Security, 19(1), 7990749. https://doi.org/10.1049/ise2.7990749.

3. Gudelli, V. R. (2024). Anomaly detection in cloud networks using machine learning algorithms. African Journal of Artificial Intelligence and Sustainable Development, 4(1). https://doi.org/10.5281/zenodo.15271016

4. Rahman, M. S., Halder, S., Uddin, M. A., et al. (2021). An efficient hybrid system for anomaly detection in social networks. Cybersecurity, 4, 10. https://doi.org/10.1186/s42400-021-00074-w

5. Chua, W., et al. (2024). Web traffic anomaly detection using isolation forest. Informatics, 11(4), 83. https://doi.org/10.3390/informatics11040083

6. Wang, L., et al. (2025). A multi-angle semantic feature fusion method for web user behavior anomaly detection. Information, 16(9), 807. https://doi.org/10.3390/info16090807

### Література

1. Han, J., Pei, J., & Tong, H. (2022). Data mining: Concepts and techniques. Morgan Kaufmann. [Електронний ресурс]. Режим доступу: https://www.sciencedirect.com/book/monograph/9780123814791/data-mining-concepts-and-techniques (дата звернення: 22.12.2025)

2. Roy, D. K., & Kalita, H. K. (2025). Anomaly detection in an open set environment using reinforcement learning. IET Information Security, 19(1), 7990749. https://doi.org/10.1049/ise2.7990749.

3. Gudelli, V. R. (2024). Anomaly detection in cloud networks using machine learning algorithms. African Journal of Artificial Intelligence and Sustainable Development, 4(1). https://doi.org/10.5281/zenodo.15271016

4. Rahman, M. S., Halder, S., Uddin, M. A., et al. (2021). An efficient hybrid system for anomaly detection in social networks. Cybersecurity, 4, 10. https://doi.org/10.1186/s42400-021-00074-w

5. Chua, W., et al. (2024). Web traffic anomaly detection using isolation forest. Informatics, 11(4), 83. https://doi.org/10.3390/informatics11040083

6. Wang, L., et al. (2025). A multi-angle semantic feature fusion method for web user behavior anomaly detection. Information, 16(9), 807. https://doi.org/10.3390/info16090807