



UDC 004.056:004.8

IRSTI 81.93.29

https://doi.org/10.53364/24138614_2026_42_3_8

Ye.K. Makatov^{1*}, Bishwajeet Pandey², A.N. Ismukanova¹, E.S. Glock¹,
 G.K. Esmagambetova¹

¹Shokan Ualikhanov Kokshetau University; Kokshetau, Kazakhstan

²GL Bajaj Institute of Technology and Management, Greater Noida, India

*E-mail: yerkhanmakatov@gmail.com

RISK RANKING OF AUTHENTICATION EVENTS BASED ON TEMPORAL AND BEHAVIORAL CONTEXT

Abstract. Not all authentication log events require the same analysis priority; therefore, ranking them using temporal and behavioral context is relevant to information security. This study examines temporal and behavioral features for ranking authentication events by relative risk. The objective is to assess the contribution of behavioral representation to ranking quality and robustness, statistical validity, explainability, and calibration, and to test the technical feasibility of behavioral prioritization on real-world logs. Methods included baseline model comparison, representation ablation, chronological train/validation/test splitting, empirical random-ranking comparison, robustness and statistical analyses, explainability, calibration, low-and-slow temporal-horizon sensitivity, and behavioral feature-group ablation; NDCG@100 was the primary ranking metric. In the synthetic experiment, R2_behavioral_core achieved a higher mean NDCG@100 than R0_minimal (0.126952 vs. 0.071314) and outperformed it in 10 of 12 scenario-model combinations. Across 60 paired observations, mean Δ NDCG@100 was +0.046803 with a 95% confidence interval of [0.035444, 0.057751]. Empirical random-ranking analysis provided an additional reference for interpreting the absolute ranking scores: for Logistic Regression with R2_behavioral_core, NDCG@100 was 0.189052 versus a random mean of 0.010045 in the baseline scenario and 0.145514 versus 0.004943 in the rare-attacks scenario. The behavioral advantage remained condition-dependent and weakened under low-and-slow activity. Explainability and calibration were used to characterize score contributions, explanation stability, calibration quality, and ranking preservation.

For 1,760,511 real-world log events, prioritization scores and unique ranks were calculated from five past-only behavioral features, satisfying all 9/9 temporal/representation integrity requirements. Predictive effectiveness was not evaluated because verified security labels were unavailable. The synthetic results support the incremental ranking value of temporal and behavioral context, while the real-world analysis demonstrates the technical feasibility of implementing the corresponding past-only behavioral prioritization procedure. Confirming predictive effectiveness in operational settings requires temporal external validation using logs with verified security labels and semantic field documentation.

Keywords: authentication events, behavioral analysis, risk ranking, temporal validation, security event prioritization, machine learning, explainability.

Introduction.

In modern information systems, authentication logs are an important source of data describing access activities of users and system entities. Since security events differ in their

significance, ranking them by priority for further analysis, alongside their detection and classification, is an important task [1].

Rule-based methods, supervised machine learning, and anomaly detection approaches are used to analyze authentication logs [2]. However, rare security events, class imbalance, label noise, behavior that changes over time, and limited labeled data affect their results [3, 4]. Because individual event features do not fully capture the temporal and behavioral context of prior activity, features derived from historical activity are relevant [2]. Such features require validation that prevents temporal leakage, as the inclusion of future information in the model may lead to overestimated results [5].

When verified target labels are unavailable, supervised predictive evaluation may not be feasible, motivating the use of label-independent approaches to activity-log analysis [6]. Therefore, this study focuses on a multilevel evaluation of authentication event ranking based on behavioral context and on the technical applicability of its core methodological principles to incompletely labeled real-world logs.

The objective of the study is to evaluate the effects of temporal and behavioral features on relative-risk ranking of authentication events, as well as the robustness, statistical validity, explainability, and calibration properties of the results, and to determine the technical applicability of the ranking principles to real-world logs. To achieve this objective, baseline model comparison, representation ablation, robustness and statistical analyses, explainability and calibration assessments, and technical feasibility testing of the behavioral prioritization procedure on real-world data were performed.

The main hypothesis is that strictly past-only temporal and behavioral context provides incremental ranking information beyond a minimal event-level representation, while the magnitude of this contribution may vary across models, experimental scenarios, and temporal aggregation horizons. For real-world logs, an additional research question is whether the same temporal-integrity and behavioral-prioritization principles can be implemented reproducibly within a real-world log structure in the absence of verified target labels.

The scientific contribution of this study is a leakage-controlled, ranking-oriented evaluation framework for quantifying the incremental value of strictly past-only temporal and behavioral context in authentication-event prioritization. The framework integrates representation ablation, chronological temporal validation, empirical random-ranking comparison, robustness and statistical inference, low-and-slow temporal-horizon sensitivity, behavioral feature-group ablation, explainability, and calibration/ranking-preservation analysis.

Research on authentication security is expanding beyond the assessment of individual authentication events toward the analysis of behavioral profiles of users and system entities formed over time. In User and Entity Behavior Analytics, such profiles are constructed from previous actions and historical activity and used to detect deviations [2]. This highlights the importance of historical behavioral context as an additional source of information for evaluating authentication and security events.

Machine learning, anomaly detection, and contextual relationship modeling are used to analyze authentication events. In authentication alerts, accounting for relationships between events and contextual information may help reduce false-positive alerts [7]. Thus, security events should be considered not only individually but also within their contextual structure.

Although most studies focus on anomaly detection or classification, events identified during operational monitoring do not have equal significance. Security event prioritization therefore ranks events by analysis priority based on their characteristics and contextual information, transforming detection results into a prioritized list of security events or alerts [1].

Prioritizing events requires the evaluation of model outputs using ranking metrics. NDCG and precision-oriented measures can be applied to such tasks [8]. Accordingly, this study uses NDCG@100 as the primary endpoint and PR-AUC and Precision@100 as secondary metrics, while examining the sensitivity of NDCG@K at K = 25, 50, 100, 200.

Rare positive events, class imbalance, and label noise are important methodological issues in machine-learning-based security event analysis. Class imbalance may affect model results [3], while label noise may alter model performance and stability [4]. Sensitivity analysis across changes in positive-event prevalence and label noise therefore enables assessment of result robustness to changes in data composition.

Temporal context complements behavioral information by capturing the historical sequence of events and activity dynamics [2]. However, improper data splitting or preprocessing when using temporal features may cause data leakage and compromise the reliability of intrusion detection model evaluation. Leakage-free experimental procedures are therefore essential for reliable evaluation of cybersecurity models. Data leakage during preprocessing may lead to overestimated machine-learning performance and compromise the reliability of model evaluation [5]. In the synthetic experiment, target generation was therefore separated from the behavioral features subsequently used for model training to reduce direct deterministic circularity.

Temporal changes in cybersecurity data, class distributions, and label quality may affect model robustness and generalization [3, 4, 5, 9]. Accordingly, robustness can be evaluated across independent random seeds, temporal folds, positive-event prevalence, label-noise levels, and different K values. In addition, separate evaluation of low-and-slow scenarios, characterized by weak signals distributed over time, enables assessment of model robustness under complex behavioral patterns [10].

When comparing models, statistical validity should be assessed alongside mean performance metrics. Confidence intervals, statistical tests, and effect sizes make it possible to determine the extent to which differences observed across experimental conditions are independent of random variation. Statistical comparisons also require consideration of the independence of observations, because treating dependent results derived from the same data source as independent observations may introduce a risk of pseudo replication [11].

Explainability in security models helps identify which features influence model outputs and supports interpretation of decisions [12]. Permutation-based feature importance can be used to assess the influence of individual features [13], while comparing explanations across experimental conditions characterizes their stability. Feature contribution in this context does not represent a causal relationship; it describes a feature's contribution to the model output.

Calibration is used to assess and adjust the probabilistic properties of model outputs [14]. However, the ordering of original scores may be preserved differently depending on the calibration method. Therefore, calibration quality and preservation of the ranking structure should be evaluated separately in ranking tasks.

The reproducibility of machine-learning research is affected by randomness, training conditions, hyperparameters, software dependencies, and the documentation of experimental artifacts [15]. Recording random seeds, hyperparameters, the software environment, and experimental parameters is therefore essential for reproducing and verifying experimental procedures.

The use of real-world log data may be constrained by data availability, label quality, and the suitability of log fields for analysis. Although activity logs are used in security analysis, the structure and available attributes of real-world logs constrain the procedures that can be applied. Consequently, when approaches developed on synthetic or fully labeled data are applied to real-world logs, their compatibility with the available data structure and their technical feasibility should be assessed separately. When verified target labels are unavailable in real-world logs, supervised predictive evaluation may not be feasible, motivating the use of label-independent approaches to activity-log analysis [6]. In this study, a pre-specified neutral scoring rule was therefore used to rank events by relative analysis priority without interpreting the resulting score as an absolute risk level.

The literature shows that machine learning, behavioral and temporal context, anomaly detection, security event prioritization, and explainability are used in security log analysis, but

these aspects are often investigated separately. Therefore, it remains relevant to evaluate the additional contribution of behavioral representation to relative-risk ranking of authentication events within a unified experimental protocol covering temporal validation, robustness, statistical validity, and explainability, and to examine the technical applicability of these principles to incompletely labeled real-world logs.

Materials and research methods.

The study evaluates the ranking of authentication events by relative risk based on behavioral features using a synthetic dataset and assesses the technical applicability of the proposed approach to real-world log data. The study consists of a synthetic experiment and a technical feasibility assessment using real-world logs. The study comprises two interrelated parts. In the synthetic experiment, proxy labels are available, allowing model performance in relative-risk ranking to be evaluated using quantitative metrics. Because the real-world logs lack verified attack/benign labels and complete semantic documentation of event fields, this part is not treated as external validation of supervised attack detection or predictive performance; its purpose is to assess the technical feasibility of the behavioral prioritization procedure on real-world logs.

The synthetic experiment included baseline, rare attacks, low-and-slow, and noisy labels scenarios. Each scenario comprised 100,000 events over 30 days, with 2,500 users, 5,000 IP addresses, and 3,500 devices simulated. Normal activity was generated as a function of user, time, IP, and device context. Proxy-risk events were not defined using behavioral features calculated later: a latent scenario state was generated first, after which authentication result, IP, and device characteristics were modified probabilistically. This limited direct deterministic circularity between target generation and feature engineering.

Latent risk-event prevalence was 1% in the baseline scenario and 0.5% in the rare attacks scenario. In the low-and-slow scenario, prevalence remained at 1%, while the signal was weakened and proxy-risk events were distributed over time. In the noisy labels scenario, baseline generation was retained and asymmetric noise was introduced into the proxy labels. Pre-specified random seeds were used for the primary scenario generation, and five independent seeds were used for robustness analysis.

Three representations were compared: R0_minimal – hour and is_new_device; R1_temporal_volume – these features plus fails_ip_10m, fails_ip_1h, and fails_user_1h; and R2_behavioral_core – additionally including distinct_usernames_from_ip_1h. All historical behavioral features were calculated using past events only. The current event, other events with the same timestamp, and future events were excluded from historical aggregates; true_label and proxy_label were not used during feature construction. This design was intended to limit temporal and label leakage.

The data were chronologically split into train/validation/test sets at a 60%/20%/20% ratio. Model selection was based on NDCG@100 on the validation set, with PR-AUC used to break ties. The test set was not used to select models, representations, or hyperparameters and was reserved for final evaluation. Temporal stability was additionally assessed using four rolling temporal folds.

Logistic Regression, Random Forest, HistGradientBoosting, Isolation Forest, and a hybrid approach combining the Random Forest supervised score with the Isolation Forest anomaly score were evaluated. Hyperparameters were selected using validation data only and then frozen for subsequent representation ablation, robustness, and statistical analyses. Preprocessing operations requiring learning from the data were fitted only on the training set and applied unchanged to the validation and test sets. Class imbalance was addressed through class weighting in Logistic Regression and Random Forest.

NDCG@100 was used as the primary endpoint, with PR-AUC and Precision@100 as secondary metrics. Ranking robustness was evaluated across five independent random seeds, temporal folds, prevalence changes, and label-noise changes, while sensitivity to K was assessed at K = 25, 50, 100, and 200. To provide a chance-level reference for the absolute ranking scores,

an empirical random-ranking baseline was additionally evaluated by repeatedly randomizing event ordering while preserving the scenario-specific label composition. Observed NDCG@K values were compared with the corresponding random distributions using the random mean, absolute difference, relative lift, and empirical exceedance probability. The low-and-slow scenario was further examined as a separate robustness condition characterized by weak signals distributed over time.

Because the original low-and-slow experiment showed a reduced behavioral advantage, a dedicated temporal-horizon sensitivity analysis was conducted. The canonical historical horizon was compared with $\times 2$, $\times 4$, and $\times 8$ extensions using the frozen Logistic Regression, Random Forest, and HistGradientBoosting configurations across four rolling temporal folds. The purpose was to determine whether ranking performance under temporally distributed weak signals depended on the amount of historical context rather than to select an optimal horizon post hoc. A subsequent behavioral feature-group ablation used the canonical reference horizon to examine the conditional contribution of user failure history, source-IP failure history, calendar-time context, device novelty, and source-IP username diversity. Feature groups were defined from the pre-existing feature implementation and representation specification rather than from validation outcomes.

Explainability analysis was conducted using the pre-specified focal configuration of Logistic Regression with `R2_behavioral_core`. Global coefficients, permutation importance, and event-level feature contributions were calculated, while local additivity, Top-100 explanations, and explanation stability were evaluated across independent seeds and temporal splits. Explanations were interpreted as feature contributions to the ranking score rather than causal effects.

Calibration analysis included raw probability, sigmoid calibration, and isotonic calibration. Calibration followed the original temporal train/validation/test split. The base model was trained only on the training data, while the validation set previously used for model selection served as the calibration split, and the calibrator was fitted on these data. Test data were not used to fit the calibrator. Calibration was evaluated using the Brier score, log loss, ECE-10, and calibration slope/intercept; preservation of the ranking structure was assessed using Spearman correlation, Kendall correlation, Top-100 overlap, and NDCG@100.

Before robustness, statistical, explainability, and calibration analyses, the experimental protocol was fixed, specifying NDCG@100 as the primary endpoint, Logistic Regression as the focal model, and `R2_behavioral_core` as the focal representation. Test results were not used to select this configuration. Random seeds, frozen hyperparameters, protocol parameters, software environment metadata, and key artifact hashes were recorded for reproducibility.

The real-world data consisted of anonymized logs obtained from the `platonus.kgu.kz` platform of Shokan Ualikhanov Kokshetau University. The data were provided under Agreement No. 282 dated July 31, 2025, and processed in accordance with information security and privacy requirements.

The logs first underwent structural and semantic auditing, including checks of file integrity, record structure, timestamp validity, event identifier uniqueness, duplicates, temporal ordering, candidate entity identifier stability, and the availability of source-level fields. Because complete semantic documentation was unavailable, unverified authentication success/failure states, attack/benign labels, or risk classes were not assigned to the records.

The feasibility of calculating the `R2_behavioral_core` features used in the synthetic experiment from the available real-world log fields was first assessed. Because no direct correspondence was available, repeated values of the `subjectHost` field were technically used to characterize candidate source-level activity, although the exact semantics of this field were not verified.

The real-world representation consisted of five past-only behavioral features: `source_events_10m` – the number of events recorded for the same source during the 10 minutes preceding the current event; `source_events_1h` – the number of events for the same source during

the preceding hour; `candidate_entity_events_1h` – the number of events for the same candidate entity during the preceding hour; `distinct_candidate_entities_source_1h` – the number of unique candidate entities observed through the same source during the preceding hour; and `source_changed_from_previous_event` – a binary feature indicating whether the candidate entity's current source differed from its source in the most recent preceding event. The candidate entity was not interpreted as a verified user identifier.

When calculating historical features for each event i , only preceding events satisfying $t_j < t_i$ were used. The current event, other events with the same timestamp, and future events were excluded from historical aggregates. The first four numerical behavioral features were transformed to the $[0,1]$ interval using strictly-past midrank percentiles based on the empirical distribution of historical values available before each event. The `source_changed_from_previous_event` feature was used as a binary 0/1 component.

Because no verified target was available, neither a supervised model nor feature weights were learned from the real-world data. The five behavioral components were assigned equal pre-specified weights of 0.2, and the relative prioritization score for each event i was calculated as

$$S_i = 0.2(P_{10m,i} + P_{1h,i} + P_{entity,i} + P_{distinct,i} + C_i)$$

where $P_{10m,i}$, $P_{1h,i}$, $P_{entity,i}$ and $P_{distinct,i}$ – are the strictly-past empirical percentile values of the corresponding behavioral features, and C_i is the binary `source_changed_from_previous_event` component. Equal weighting was treated as a pre-specified neutral scoring rule to avoid learning unsupported predictive weights from data without verified labels. The prioritization score does not represent attack probability, calibrated probability, a malicious/benign class, or an absolute risk level; it is used to rank events by relative priority based on observed behavioral activity.

Events were ranked deterministically by prioritization score. The technical correctness of the ranking procedure was verified through score–rank consistency, rank uniqueness, missing and duplicate ranks, and score–order violations. Temporal and representation integrity were further controlled using nine contract checks: consistency with the stored representation hash, independent hash recomputation, consistency of feature names and order, enforcement of the strictly past-only principle, exclusion of same-timestamp events from history, prevention of missing candidate entity values from being merged into a single artificial entity, prevention of missing source values from being merged into a single artificial source, a cold-start zero policy for count-based historical features, and a cold-start NULL policy for the previous-event feature. Compliance with these requirements was verified through leakage and representation integrity audits.

Feature-level contribution analysis was performed for the Top-10, Top-100, Top-1000 events. Contribution values were used to characterize the event-level contribution of the five components forming the prioritization score and were not interpreted as causal attribution or evidence of the cause of risk.

Compatibility between the synthetic and real-world data was evaluated not through feature-by-feature correspondence but through adherence to shared methodological principles: preservation of temporal ordering, construction of past-only behavioral context, use of entity/source-level information, calculation of a relative score, deterministic ranking, and feature-level explanation. Accordingly, the real-world component is not a predictive external validation of the synthetic experiment but an assessment of the technical feasibility of the proposed behavioral prioritization procedure within a real-world log structure. The software code used in this study is publicly available in a GitHub repository the link is provided below.

Results and their discussion.

The experiment began with processing the synthetic dataset using the methods described in the Materials and Methods section. In the baseline scenario, no additional distortion was introduced into the data, and the initial performance of machine-learning models in ranking

authentication events by behavioral risk was evaluated. NDCG@100 was used as the primary metric, with PR-AUC and Precision@100 as secondary evaluation metrics.

Table 1 – Ranking performance of models in the baseline scenario on the synthetic dataset

Model	PR-AUC	Precision@100	NDCG@100
Logistic Regression	0.0867	0.2100	0.1891
HistGradientBoosting	0.0793	0.1800	0.1704
Hybrid	0.0811	0.1700	0.1607
Random Forest	0.0795	0.1100	0.1496
Isolation Forest	0.0210	0.0100	0.0075
<i>Note – compiled by the authors.</i>			

The baseline experiment on the synthetic dataset showed differences among the models in their ability to place proxy-risk events at higher ranking positions. Logistic Regression achieved the highest value for the primary ranking metric, whereas Isolation Forest performed below the supervised learning models. These results characterize the initial relative performance of the models under baseline conditions and do not imply the superiority of any single model across all scenarios.

To evaluate the effect of the feature set on ranking quality, three representations were compared: R0_minimal, representing minimal context; R1_temporal_volume, containing temporal and activity-volume features; and R2_behavioral_core, containing the core behavioral features. Representation ablation was conducted across four scenarios using three supervised models: Logistic Regression, Random Forest, and HistGradientBoosting. Isolation Forest, as an unsupervised anomaly-detection method, and Hybrid, as a composite scoring approach, were excluded from the representation ablation and evaluated within the baseline model comparison. NDCG@100 was used as the primary evaluation metric.

Table 2 – Effect of the feature set on ranking quality in the synthetic dataset

Feature set	Description	Mean NDCG@100
R0_minimal	Minimal feature set	0.071314
R1_temporal_volume	Temporal and activity-volume features	0.087488
R2_behavioral_core	Core behavioral features	0.126952
<i>Note – compiled by the authors.</i>		

The ranking benefit of behavioral context was not uniform across experimental conditions. To interpret the absolute NDCG values appropriately, the observed scores were additionally compared with an empirical random-ranking reference rather than treated as classification accuracy. For Logistic Regression with R2_behavioral_core, NDCG@100 was 0.189052 versus a random mean of 0.010045 in the baseline scenario, 0.145514 versus 0.004943 in the rare-attacks scenario, and 0.194898 versus 0.010336 under noisy labels. These results indicate that the principal synthetic rankings were substantially above chance-level ordering, while the magnitude of the behavioral advantage remained dependent on the experimental condition.

The robustness of the results across different experimental conditions was then examined. To avoid relying on a single data split or random initialization, the analysis covered multiple independent seeds, temporal splits, different prevalence levels of proxy-risk events, and label-noise conditions. Sensitivity to the selected K value and robustness under the low-and-slow scenario were also evaluated.

Table 3 – Quantitative assessment of robustness and sensitivity on the synthetic dataset

Evaluation dimension	Experimental scope	Main quantitative result
Seed robustness	5 seeds $\times$ 4 scenarios	Mean NDCG@100: 0.0686–0.1500; SD: 0.0275–0.0527
Temporal robustness	16 folds (4 $\times$ 4 scenarios)	Mean NDCG@100: 0.0591–0.1400; SD: 0.0235–0.0576
Sensitivity to proxy-risk event prevalence	5 prevalence levels	Mean NDCG@100: 0.1088–0.2104
Sensitivity to label noise	5 noise levels	Mean NDCG@100: 0.1041–0.1500
Sensitivity to K	K = 25, 50, 100, 200	Mean NDCG@K: 0.0609–0.1846
Paired R0–R2 comparison	60 paired observations	Mean Δ NDCG@100 = +0.0468; 95% CI [0.0354, 0.0578]
Low-and-slow robustness	3 models $\times$ 5 seeds	LR: R0 0.0633 $\rightarrow$ R2 0.0686; Δ = +0.0053; R2 win rate = 80%
<i>Note – compiled by the authors.</i>		

The robustness and sensitivity analyses showed that the contribution of behavioral representation varied across experimental conditions and was not uniform across scenarios. The low-and-slow condition was the clearest limitation, indicating that weak signals distributed over time are more difficult to rank using the reference temporal context. This observation motivated a dedicated temporal-horizon analysis to determine whether the limitation was related to the amount of historical information available to the model.

A statistical analysis was performed to compare the R2 behavioral feature set with the R0 minimal feature set. Sixty paired observations were used to determine whether the previously observed improvement in ranking quality could be attributed to random variation. The analysis included the NDCG@100 difference, a bootstrap confidence interval, the Wilcoxon test, a paired permutation test, and effect-size estimation.

Table 4 – Statistical comparison of the R2 and R0 feature sets on the synthetic dataset

Statistical measure	Result
Number of paired observations	60
Mean Δ NDCG@100 (R2 – R0)	+0.046803
95% bootstrap confidence interval	[0.035444, 0.057751]
Wilcoxon p-value	1.785×10^{-8}
Paired permutation test p-value	$\approx 1.0 \times 10^{-5}$
Rank-biserial effect size	0.8361
Pseudoreplication detected	No
<i>Note – compiled by the authors.</i>	

The mean increase in NDCG@100 for R2 relative to R0 was positive, and the 95% confidence interval did not include zero. The Wilcoxon and permutation tests also supported the statistical significance of the difference between the two feature sets. The rank-biserial measure indicated a large effect size.

The explainability of the model's ranking results and the calibration properties of its outputs were also evaluated. Explainability analysis focused on identifying the contribution of individual features to the risk-ranking score generated by the model. Explanation stability was additionally assessed across different seeds and temporal splits.

Table 5 – Evaluation of model explainability and calibration on the synthetic dataset

Evaluation dimension	Measure	Result
Global explainability	Logistic Regression coefficients	Baseline largest absolute coefficients: <code>distinct_usernames_from_ip_1h</code> = 0.8414; <code>is_new_device</code> = 0.6301
Feature importance	Permutation importance	Baseline largest Δ NDCG@100: <code>is_new_device</code> = 0.1597; <code>distinct_usernames_from_ip_1h</code> = 0.1204
Local explainability	Local additivity	Absolute error between model score and reconstructed score: $0-4.44 \times 10^{-16}$
Top-100 explanations	Feature-level contributions	Baseline Top-100: <code>is_new_device</code> dominant in 97/100 events (97%); fails <code>user_1h</code> in 3/100 (3%)
Seed explanation stability	5 independent seeds	Mean Spearman ρ : baseline 1.0000; rare attacks 0.9657; noisy labels 0.8914; low-and-slow 0.7257
Temporal explanation stability	4 temporal folds; 24 scenario $\times$ feature combinations	Sign consistency = 1.00 in 22/24 (91.67%); 0.75 in 2/24 (8.33%)
Low-and-slow explainability	Top-100 feature contributions	<code>is_new_device</code> : mean absolute contribution 0.9672, dominant in 62/100 (62%); fails <code>user_1h</code> : 0.8883, dominant in 38/100 (38%)
Calibration data split	Temporal train $\rightarrow$ calibration $\rightarrow$ test	Temporal ordering preserved; test events were not used for calibration fitting
Raw probability	RAW, baseline	Brier = 0.141207; Log loss = 0.450438; ECE-10 = 0.317984
Sigmoid calibration	Baseline	Brier = 0.009779; Log loss = 0.048392; ECE-10 = 0.000738
Isotonic calibration	Baseline	Brier = 0.009761; Log loss = 0.048250; ECE-10 = 0.000559
Calibration slope/intercept	Baseline	Sigmoid: slope 1.070599, intercept 0.308024; Isotonic: slope 0.998574, intercept 0.029386
Ranking preservation after sigmoid	Spearman, Kendall, Top-100, NDCG@100	ρ = 1.000; τ = 1.000; Top-100 overlap = 1.00; NDCG@100: 0.189052 $\rightarrow$ 0.189052
Ranking change after isotonic	Spearman, Kendall, Top-100, NDCG@100	ρ = 0.9440; τ = 0.8505; Top-100 overlap = 0.39; NDCG@100: 0.189052 $\rightarrow$ 0.141684
<i>Note – compiled by the authors.</i>		

The analysis showed that the features contributing to the model's ranking score could be characterized at both global and individual-event levels. Evaluating explanations across seeds and temporal splits made it possible to assess whether they depended on a single experimental run. Separate analysis of the low-and-slow scenario characterized explanation stability and the structure of feature contributions under temporally distributed activity.

Calibration analysis was conducted while preserving the temporal split, and test data were not used for calibration fitting. These results characterize the calibration properties of the model output under the synthetic experimental conditions.

The real-world dataset consisted of four source files containing a total of 1,760,511 events. All eventTime values were parsed successfully. The dataset contained 735,172 unique timestamps. A total of 1,427,741 events (81.098%) shared their timestamp with at least one other event, and

the largest same-timestamp group contained 197 events. No duplicate eventID values were detected; therefore, events were sorted deterministically by eventTime and eventID.

The dataset contained 8,668 candidate entity identifiers recurring over time. Of these, 8,660 (99.908%) occurred in multiple events, 8,399 (96.897%) occurred on multiple days, and 8,121 (93.689%) occurred across multiple source files. The median number of events per identifier was 81, the median number of active days was 19 days, and the median observation span was 63.853 days. The subjectHost field contained 1,458,615 non-empty IPv4-formatted values, while eventTypeID contained 34 distinct codes.

No documentation describing the semantics of the event fields was provided with the source data. Consequently, it was not possible to verify whether the candidate entity identifier represented an actual user or account identifier, to establish the exact semantics of subjectHost, or to determine the meanings of the 34 eventTypeID codes. No reliable session or device identifier was identified.

Because verified target labels were unavailable, predictive temporal validation and label-dependent metrics such as NDCG@100, PR-AUC, and Precision@K were not calculated for the real-world data.

The eventTime field could be used to establish temporal ordering, eventID to uniquely identify each event, the candidate entity identifier to link the historical activity of the same object, and repeated subjectHost values to technically characterize candidate source-level activity. The consistent recurrence of identifiers across multiple events and days provided sufficient history for calculating temporal behavioral patterns. These properties formed the basis for evaluating the technical feasibility of the behavioral prioritization procedure on the real-world logs.

Compatibility between the synthetic experiment and the real-world data was determined to be partial. The original R2_behavioral_core feature set used in the synthetic dataset could not be transferred unchanged to the real-world logs. Therefore, a separate real-native behavioral representation was constructed according to the available fields and semantic limitations of the real-world data.

Table 6 – Technical feasibility results of the behavioral prioritization procedure on real-world log data

Evaluated measure	Result
Real-world events processed	1,760,511
Real-native behavioral features	5
Temporal/representation integrity contracts	9/9
Events with calculated scores	1,760,511/1,760,511
Score range	0.12439–0.99326
Mean score	0.48346
SD	0.15362
Median	0.50000
P95	0.74795
P99	0.79126
Unique ranks	1,760,511/1,760,511
Missing / duplicate ranks	0/0
Score-order violations	0
Verified attack/benign labels	0
<i>Note – compiled by the authors.</i>	

Feature-level explainability was also evaluated for the highest-ranked events. Feature contributions were calculated for the Top-10, Top-100, Top-1000 events.

Table 7 – Feature-level contribution results for the highest-ranked events in the real-world log data

Selection	source changed from previous event	distinct candidate entities source 1h
Top-10	10/10 (100%)	0/10 (0%)
Top-100	99/100 (99%)	1/100 (1%)
Top-1000	995/1000 (99.5%)	5/1000 (0.5%)
<i>Note – compiled by the authors.</i>		

This measure is a relative prioritization score calculated from observed behavioral activity features to prioritize events for further security analysis. It does not represent attack probability, a malicious/benign class, or a diagnostic risk level.

The results demonstrated the additional informational value of temporal and behavioral context for ranking authentication events by relative risk. The R2_behavioral_core configuration outperformed R0_minimal in 10 of 12 scenario–model combinations. Across 60 paired observations, the mean $\Delta\text{NDCG}@100$ was +0.0468, with a 95% confidence interval of [0.0354, 0.0578] that did not include zero. The Wilcoxon and paired permutation tests supported the statistical significance of the difference, while the rank-biserial correlation indicated a large effect size. These findings support the main hypothesis at the overall experimental level; however, because R2 did not outperform R0 under every condition, its advantage cannot be considered universal.

The findings are consistent with studies showing that historical activity and behavioral context provide additional information for characterizing the activity profiles of users or system entities and assessing deviations from them [2]. Prioritizing security events according to contextual characteristics is also considered a practical approach to identifying events requiring further analysis [1]. The methodological contribution of this study is the controlled evaluation of the incremental ranking value of strictly past-only behavioral context within an integrated protocol covering representation ablation, temporal leakage control, chronological validation, empirical chance-level comparison, robustness, statistical inference, low-and-slow temporal-horizon sensitivity, feature-group ablation, explanation stability, and calibration/ranking preservation.

The robustness analysis assessed stability across random seeds, temporal folds, positive-event prevalence, label noise, and K values, consistent with previous studies on the effects of data composition and temporal variation on model performance [3, 4, 9]. Because the original low-and-slow configuration showed only a small R2 advantage for Logistic Regression, temporal-horizon sensitivity was examined separately. Mean $\text{NDCG}@100$ for Logistic Regression increased from 0.059091 at the reference horizon to 0.090420, 0.108764, and 0.086422 at the $\times 2$, $\times 4$, and $\times 8$ horizons, respectively. For Random Forest, the corresponding values increased from 0.052101 to 0.075489, 0.079994, and 0.065853. HistGradientBoosting showed a non-monotonic response: $\text{NDCG}@100$ decreased at the $\times 2$ and $\times 4$ horizons but increased from 0.079763 to 0.097131 at $\times 8$. Thus, extending historical context improved low-and-slow ranking for some model–horizon combinations but did not provide a universal solution. Because the analysis was based on four rolling temporal folds, these results are interpreted as sensitivity evidence rather than definitive inferential proof. Behavioral feature-group ablation further indicated a model-dependent contribution structure, with no single feature group universally accounting for low-and-slow ranking performance. Explainability analysis characterized the features contributing to the ranking score at both global and event levels and assessed their stability across seeds and temporal splits, consistent with studies emphasizing the importance of explainability in cybersecurity models [12]. Feature contributions were not interpreted as causal effects. Explanation stability decreased in the low-and-slow scenario. In the calibration analysis, sigmoid calibration fully preserved the ranking structure, whereas isotonic calibration, despite improving calibration metrics, reduced Top-100 overlap to 0.39 and changed $\text{NDCG}@100$ from 0.189052 to 0.141684. These results show that improved calibration does not necessarily imply preservation of the original ranking structure.

Therefore, calibration quality [14] and ranking preservation were evaluated as distinct properties in this study.

In the real-world logs, prioritization scores and unique ranks were calculated for all events using five real-native past-only behavioral features. All 9/9 temporal/representation integrity contracts were satisfied, and no score-order violations were detected. These results establish the technical feasibility, temporal integrity, and reproducibility of the behavioral prioritization procedure within the available real-world log structure. They do not establish attack-detection accuracy, predictive ranking validity, or the maliciousness of highly ranked events because verified attack/benign ground truth was unavailable. The connection between the synthetic and real-world components was based not on feature-by-feature correspondence but on shared methodological principles: temporal ordering, past-only behavioral context, entity/source-level information, relative scoring, deterministic ranking, and feature-level explanation. Because verified labels were unavailable, predictive optimal weights were not learned for real-world scoring; instead, the five components were assigned equal pre-specified weights of 0.2. The main limitations are that predictive ranking performance was evaluated using synthetic data and proxy-risk labels, verified target labels were unavailable in the real-world logs, the semantics of the candidate entity and source fields were not fully verified, and the predictive optimality of the real-world scoring components was not assessed. Therefore, synthetic performance cannot be directly generalized to an operational environment. Future research should conduct temporal predictive external validation using real-world authentication logs with semantic field documentation and verified security labels.

Conclusion.

The study showed that strictly past-only temporal and behavioral context provides measurable incremental information for ranking authentication events under the labeled synthetic experimental conditions. R2_behavioral_core generally improved ranking relative to the minimal representation, and empirical random-ranking analysis provided a chance-level reference showing that the principal observed rankings were substantially above random ordering. The effect was nevertheless condition-dependent: dedicated low-and-slow analysis showed sensitivity to both model choice and temporal horizon, while feature-group ablation indicated a model-dependent contribution structure. The scientific contribution is a leakage-controlled and ranking-oriented evaluation of behavioral context that combines representation ablation, chronological temporal validation, empirical random-ranking comparison, robustness and statistical inference, low-and-slow temporal-horizon sensitivity, feature-group ablation, explainability, and calibration/ranking preservation. This framework characterizes not only whether behavioral context improves ranking, but also the experimental conditions under which its contribution strengthens, weakens, or becomes model-dependent.

Because verified security labels were unavailable in the real-world logs, predictive effectiveness was not evaluated. However, the technical feasibility of the core behavioral prioritization procedures was demonstrated. Generalizing the findings to predictive performance in operational environments requires subsequent external validation on authentication logs containing verified labels and semantic documentation of event fields.

Acknowledgments. *The authors express their sincere gratitude to the Department of Information and Communication Technologies of Sh. Ualikhanov Kokshetau University for the scientific and methodological support provided during the course of this research.*

The software code used in this study is publicly available in the GitHub repository: <https://github.com/ItBilim/behavioral-risk-ranking>.

References

1. Liu, J., Zhang, R., Liu, W., Xue, J., & Wang, H. (2022). Context2Vector: Accelerating security event triage via context representation learning. *Information and Software Technology*, 146, 106856. <https://doi.org/10.1016/j.infsof.2022.106856>
2. Martin, A. G., Fernandez-Isabel, A., Martin de Diego, I., & Beltran, M. (2021b). A survey for user behavior analysis based on machine learning techniques: Current models and applications. *Applied Intelligence*, 51(8). <https://doi.org/10.1007/s10489-020-02160-x>
3. Ndichu, S., Ban, T., Takahashi, T., & Inoue, D. (2023). AI-assisted security alert data analysis with imbalanced learning methods. *Applied Sciences*, 13(3), 1977. <https://doi.org/10.3390/app13031977>
4. Yasuda, G., Suko, T., Kobayashi, M., & Matsushima, T. (2023). Asymptotic evaluation of classification in the presence of label noise. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E106.A(3). <https://doi.org/10.1587/transfun.2022TAP0013>
5. Bouke, M. A., & Abdullah, A. (2023). An empirical study of pattern leakage impact during data preprocessing on machine learning-based intrusion detection models reliability. *Expert Systems with Applications*, 230, 120715. <https://doi.org/10.1016/j.eswa.2023.120715>
6. Ianni, M., & Masciari, E. (2023). SCOUT: Security by computing OUTliers on activity logs. *Computers & Security*, 132, 103355. <https://doi.org/10.1016/j.cose.2023.103355>
7. Wang, Y., Hallgren, K., & Larson, J. (2024). A graph-based framework for reducing false positives in authentication alerts in security systems. In *Companion Proceedings of the ACM Web Conference 2024*. <https://doi.org/10.1145/3589335.3648325>
8. Adàlia, R., Sanjuan, G., Margalef, T., & Zamora, I. (2025). The LambdaGap framework for precision-oriented ranking. *ACM Transactions on Information Systems*, 43(4), Article 97, 1–39. <https://doi.org/10.1145/3733235>
9. Rabbi, K. F., Lou, H. H., & Liu, X. (2026). The impact of time-aware design choices in ICS anomaly detection. In *2026 IEEE 16th Annual Computing and Communication Workshop and Conference (CCWC)*. <https://doi.org/10.1109/CCWC67433.2026.11393705>
10. Rios, V., et al. (2024). Enhanced detection of low-rate DDoS attack patterns using machine learning models. *Journal of Network and Computer Applications*, 227, 103903. <https://doi.org/10.1016/j.jnca.2024.103903>
11. Berrar, D., & Lozano, J. A. (2013). Significance tests or confidence intervals: Which are preferable for the comparison of classifiers? *Journal of Experimental & Theoretical Artificial Intelligence*, 25(2), 189–206. <https://doi.org/10.1080/0952813X.2012.680252>
12. Sharma, B., Sharma, L., Lal, C., & Roy, S. (2024). Explainable artificial intelligence for intrusion detection in IoT networks: A deep learning based approach. *Expert Systems with Applications*, 238, 121751. <https://doi.org/10.1016/j.eswa.2023.121751>
13. Altmann, A., Tološi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>
14. Silva Filho, T., Song, H., Perello-Nieto, M., Santos-Rodriguez, R., Kull, M., & Flach, P. (2023). Classifier calibration: A survey on how to assess and improve predicted class probabilities. *Machine Learning*, 112, 3211–3260. <https://doi.org/10.1007/s10994-023-06336-7>
15. Semmelrock, H., Ross-Hellauer, T., Kopeinik, S., Theiler, D., Haberl, A., Thalmann, S., & Kowald, D. (2025). Reproducibility in machine-learning-based research: Overview, barriers, and drivers. *AI Magazine*, 46, e70002. <https://doi.org/10.1002/aaai.70002>

АУТЕНТИФИКАЦИЯ ОҚИҒАЛАРЫН УАҚЫТТЫҚ ЖӘНЕ МІНЕЗ-ҚҰЛЫҚТЫҚ КОНТЕКСТ НЕГІЗІНДЕ ТӘУЕКЕЛ БОЙЫНША РАНЖИРЛЕУ

Аңдатпа. Аутентификация журналдарындағы оқиғалар талдау басымдығы бойынша бірдей емес, сондықтан оларды уақыттық және мінез-құлықтық контекст негізінде ранжирлеу ақпараттық қауіпсіздіктегі маңызды міндет болып табылады. Зерттеудің мақсаты — мінез-құлықтық ұсынудың аутентификация оқиғаларын салыстырмалы тәуекел бойынша ранжирлеу сапасына қосатын үлесін, нәтижелердің тұрақтылығын, статистикалық негізділігін, түсіндірмелілігін және калибрленуін бағалау, сондай-ақ мінез-құлықтық басымдықтау тәсілінің нақты журналдық деректерде техникалық орындалуын тексеру. Зерттеуде бастапқы модельдерді салыстыру, ұсыну нұсқаларын абляциялық талдау, деректерді хронологиялық түрде оқыту, валидация және тестілеу жиындарына бөлу, эмпирикалық кездейсоқ ранжирлеумен салыстыру, тұрақтылық пен статистикалық талдау, түсіндірмелілік пен калибрлеуді бағалау, баяу және біртіндеп жүретін белсенділік жағдайындағы уақыттық горизонтқа сезімталдықты тексеру және мінез-құлықтық белгілер топтарының абляциялық талдауы қолданылды. Негізгі ранжирлеу метрикасы ретінде NDCG@100 пайдаланылды. Синтетикалық экспериментте R2_behavioral_core ұсынуы орташа NDCG@100 көрсеткіші бойынша R0_minimal ұсынуынан жоғары нәтиже көрсетті: тиісінше 0.126952 және 0.071314. R2_behavioral_core 12 сценарий-модель комбинациясының 10-ында R0_minimal-дан басым болды. 60 жұптық бақылау бойынша орташа Δ NDCG@100 +0.046803 құрады, ал 95% сенімділік интервалы [0.035444, 0.057751] болды. Эмпирикалық кездейсоқ ранжирлеу абсолюттік NDCG@100 мәндерін түсіндіру үшін қосымша салыстыру деңгейін қамтамасыз етті. Logistic Regression және R2_behavioral_core үшін бастапқы сценарийде NDCG@100 0.189052 болды, ал кездейсоқ ранжирлеудің орташа мәні 0.010045 құрады; сирек шабуылдар сценарийінде сәйкес мәндер 0.145514 және 0.004943 болды. Мінез-құлықтық контексттің артықшылығы эксперименттік жағдайларға тәуелді болды және баяу әрі біртіндеп жүретін белсенділік кезінде әлсіреді. Түсіндірмелілік пен калибрлеу белгілердің ранжирлеу ұпайына қосатын үлесін, түсіндірмелердің тұрақтылығын, калибрлеу сапасын және ранжирлеу тәртібінің сақталуын бағалау үшін қолданылды.

Нақты журналдық деректердегі 1 760 511 оқиға үшін тек өткен оқиғаларға негізделген бес мінез-құлықтық белгі бойынша басымдық ұпайлары мен бірегей рангтар есептелді. Уақыттық және ұсыну тұтастығына қойылған 9 талаптың барлығы орындалды. Рассталған қауіпсіздік белгілері болмағандықтан, болжамдық тиімділік бағаланбады. Нәтижелер уақыттық және мінез-құлықтық контексттің аутентификация оқиғаларын салыстырмалы талдау басымдығы бойынша ранжирлеуге қосымша ақпарат беретінін көрсетті. Әдістің нақты операциялық ортадағы болжамдық тиімділігін растау үшін рассталған қауіпсіздік белгілері және өрістердің семантикалық құжаттамасы бар журналдық деректерде уақыттық сыртқы валидация жүргізу қажет.

Түйін сөздер: аутентификация оқиғалары, мінез-құлықтық талдау, тәуекелді ранжирлеу, уақыттық валидация, қауіпсіздік оқиғаларын басымдықтау, машиналық оқыту, түсіндірмелілік.

РАНЖИРОВАНИЕ СОБЫТИЙ АУТЕНТИФИКАЦИИ ПО РИСКУ НА ОСНОВЕ ВРЕМЕННОГО И ПОВЕДЕНЧЕСКОГО КОНТЕКСТА

Аннотация. События в журналах аутентификации различаются по приоритету анализа, поэтому их ранжирование с учетом временного и поведенческого контекста является значимой задачей информационной безопасности. Цель исследования — оценить дополнительный вклад поведенческого представления в качество ранжирования событий

аутентификации по относительному риску, устойчивость результатов, статистическую обоснованность, объяснимость и калибровку, а также определить техническую реализуемость поведенческой приоритизации на реальных журнальных данных. Методология включала сравнение базовых моделей, абляционный анализ представлений, хронологическое разделение данных на обучающую, валидационную и тестовую выборки, эмпирическое случайное ранжирование, анализ устойчивости и статистической значимости, оценку объяснимости и калибровки, анализ чувствительности к временному горизонту и абляцию групп поведенческих признаков. Основной метрикой являлась $NDCG@100$. В синтетическом эксперименте $R2_behavioral_core$ превзошло $R0_minimal$ по среднему $NDCG@100$: 0.126952 против 0.071314, а также в 10 из 12 комбинаций «сценарий–модель». По 60 парным наблюдениям среднее $\Delta NDCG@100$ составило +0.046803 при 95% доверительном интервале [0.035444, 0.057751]. Для Logistic Regression с $R2_behavioral_core$ $NDCG@100$ составило 0.189052 против среднего случайного значения 0.010045 в базовом сценарии и 0.145514 против 0.004943 в сценарии редких атак. Эффект поведенческого контекста зависел от экспериментальных условий и ослабевал при медленной, распределенной во времени активности.

Для 1 760 511 событий реальных журналов на основе пяти поведенческих признаков, рассчитанных исключительно по предшествующим событиям, были вычислены показатели приоритета и уникальные ранги. Все 9 требований временной целостности и целостности представления выполнены. Из-за отсутствия подтвержденных меток безопасности прогностическая эффективность не оценивалась. Полученные результаты подтверждают дополнительную информационную ценность временного и поведенческого контекста для относительной приоритизации событий, однако подтверждение прогностической эффективности требует временной внешней валидации на реальных журналах с верифицированными метками и семантической документацией полей.

Ключевые слова: события аутентификации, поведенческий анализ, ранжирование рисков, временная валидация, приоритизация событий безопасности, машинное обучение, объяснимость.

Сведение об авторах

Макатов Ерхан Каиржанович	магистр педагогических наук, сеньор-лектор, кафедра «Информационно-коммуникационных технологий», Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан, verkhanmakatov@gmail.com , 87071406577
Bishwajeet Pandey	PhD, профессор, Институт технологий и менеджмента GL Bajaj, Грейтер-Нойда, Индия, dr.pandey@ieee.org
Исмуканова Айгерим Наурызбаевна	магистр технических наук, сеньор-лектор, кафедра «Информационно-коммуникационных технологий», Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан, aigera_ismukan@mail.ru ,
Глок Елена Степановна	магистр технических наук, сеньор-лектор, кафедра «Информационно-коммуникационных технологий», Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан, Glok_E@mail.ru
Есмагамбетова Галия Кайратовна	магистр технических наук, сеньор-лектор, кафедра «Информационно-коммуникационных технологий», Кокшетауский Университет имени Шокана Уалиханова, Кокшетау, Казахстан, gal.esm@mail.ru

Авторлар туралы мәлімет

Макатов Ерхан Каиржанович	педагогика ғылымдарының магистрі, сеньор-лектор, "Ақпараттық-коммуникациялық технологиялар" кафедрасы, Шоқан Уәлиханов атындағы Көкшетау университеті, Көкшетау, Қазақстан, verkhanmakatov@gmail.com , 87071406577
Bishwajeet Pandey	PhD, профессор, GL Bajaj технологиялар және менеджмент институты, Грейтер-Нойда, Үндістан, dr.pandey@ieee.org

Исмуканова Айгерим Наурызбаевна	техника ғылымдарының магистрі, сеньор-лектор, "Ақпараттық-коммуникациялық технологиялар" кафедрасы, Шоқан Уалиханов атындағы Көкшетау университеті, Көкшетау, Қазақстан, aigera_ismukan@mail.ru ,
Глок Елена Степановна	техника ғылымдарының магистрі, сеньор-лектор, "Ақпараттық-коммуникациялық технологиялар" кафедрасы, Шоқан Уалиханов атындағы Көкшетау университеті, Көкшетау, Қазақстан, Glok_E@mail.ru
Есмагамбетова Галия Кайратовна	техника ғылымдарының магистрі, сеньор-лектор, "Ақпараттық-коммуникациялық технологиялар" кафедрасы, Шоқан Уалиханов атындағы Көкшетау университеті, Көкшетау, Қазақстан, gal.esm@mail.ru

Information about the authors

Makatov Yerkhan Kairzhanovich	Master of Pedagogical Sciences, senior lecturer, Department of Information and Communication Technologies, Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan, yerkhanmakatov@gmail.com , 87071406577
Bishwajeet Pandey	PhD, Professor, GL Bajaj Institute of Technology and Management, Greater Noida, India, dr.pandey@ieee.org
Ismukanova Aigerim Nauryzbaevna	Master of Technical Sciences, senior lecturer, Department of Information and Communication Technologies, Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan, aigera_ismukan@mail.ru
Glock Elena Stepanovna	Master of Technical Sciences, senior lecturer, Department of Information and Communication Technologies, Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan, Glok_E@mail.ru
Esmagambetova Galiya Kairatovna	Master of Technical Sciences, senior lecturer, Department of Information and Communication Technologies, Shokan Ualikhanov Kokshetau University, Kokshetau, Kazakhstan, gal.esm@mail.ru