AN INTELLIGENT MODEL FOR ASSESSING CYBER RISKS BASED ON USER BEHAVIOR IN INFORMATION SYSTEMS

Fayzi Bekkamov

International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan
E-mail: fayzi@mail.ru

K E Y W O R D S

User behavior analytics, cyber risk assessment, machine learning, anomaly detection, insider threat, adaptive weighting, security decision support

A B S T R A C T

Modern information systems generate heterogeneous traces of user activity, while conventional security controls frequently evaluate events independently and rely on static thresholds. This study proposes an intelligent model for assessing cyber risks from user behavior by integrating behavioral deviations, access context, resource sensitivity, anomaly severity, and historical compliance into a normalized risk score. The model separates feature extraction, adaptive weighting, risk aggregation, and security-action recommendation. A contextual weighting mechanism is introduced to avoid assigning identical importance to behavioral indicators under different access conditions. The analytical assessment uses controlled user-activity scenarios to examine the sensitivity of the model to changes in behavioral and contextual factors. The results show that the proposed structure produces monotonic risk growth when high-impact indicators change in an adverse direction and reduces the probability of excessive responses to isolated low-impact anomalies. The model is intended as a decision-support layer that complements, rather than replaces, conventional authentication, access control, logging, and endpoint security mechanisms.

1. Introduction

Information systems increasingly depend on continuous user interaction with digital resources, services, databases, cloud platforms, and administrative interfaces. Consequently, security-relevant evidence is not limited to isolated technical events: the sequence, frequency, timing, location, and context of user actions can provide important information about emerging cyber risk. User behavior analytics (UBA) is specifically concerned with identifying meaningful deviations from established behavioral patterns and can support the detection of insider threats and other anomalous activity [1]. Research on insider-threat detection also shows a shift from static rules toward adaptive data-driven approaches that model temporal and contextual behavior [2].

Traditional security monitoring often treats authentication failures, unusual resource access, large downloads, and abnormal working hours as separate alerts. Such an approach can produce fragmented evidence and may generate either excessive alerts or insufficient escalation. The key methodological problem is therefore not only anomaly detection, but also the transformation of heterogeneous behavioral evidence into an interpretable risk level that can support a proportionate security decision. Previous behavioral studies have demonstrated that malicious and benign users may exhibit distinguishable patterns in information-search and data-collection activities [3], while recent work emphasizes the importance of combining technical and behavioral factors in insider-threat mitigation [4].

The purpose of this study is to develop a structured intelligent model that converts user behavior and access context into a normalized cyber-risk score and maps the score to an adaptive security response. The novelty of the proposed approach lies in the joint treatment of behavioral

(cc) BY

deviation, resource sensitivity, access context, anomaly severity, and historical compliance rather than evaluating each event in isolation. The model is designed to be transparent: every component contributes to the final score through an explicit weighting mechanism, which makes the resulting decision traceable for security analysts.

The main contributions are: a five-dimensional representation of user behavior; a context-dependent softmax weighting mechanism; a gated decision mapping that converts the score into a graded action; and an analytical evaluation combining controlled scenarios, sensitivity analysis, and a population experiment. Section 2 reviews related work, Section 3 describes the model, Section 4 reports the analysis and discussion, and Section 5 concludes.

2. Related Work

Research on insider threats and user and entity behavior analytics (UEBA) spans more than two decades. Early work on the analysis and detection of malicious insiders [8] framed the problem as the identification of misuse by authorized users, whose actions are difficult to separate from legitimate work when only technical privilege is considered. Empirical studies of insider-threat behavior indicate that behavioral cues carry information that purely technical monitoring does not capture [6], and the ATT&CK-based analytics process of Strom et al. [7] shows how detection logic can be derived systematically from documented adversary behavior. A review of neural approaches to UEBA [5] documents a shift toward learned representations of user activity, but also points to training difficulties and limited interpretability. Hunker and Probst [9] review definitions of insider threats, and the CERT Guide to Insider Threats [10] relates recurring incident patterns, such as IT sabotage, fraud, and theft of intellectual property, to observable precursors. These sources motivate the use of context rather than raw event counts and the inclusion of a history component, because precursors tend to accumulate over time.

Surveys of anomaly detection distinguish point, contextual, and collective anomalies [11]. Contextual anomalies, which are abnormal only under specific conditions such as time, role, or resource, dominate user-behavior data, so context-free thresholds perform poorly. Unsupervised detectors such as Isolation Forest [12] remain popular because labeled incidents are scarce; deep sequence models have been applied to structured log data to score deviations online [13], and ensembles of unsupervised detectors have been studied for insider threats [14]. These detectors, however, output an anomaly score rather than a decision: the score is not tied to the value of the accessed asset or to the confidence of the evidence. Post-hoc explanation techniques such as SHAP [15] and LIME [16] attribute predictions to input features, but they explain the detector rather than the organizational risk. The present study instead makes the aggregation itself transparent by construction.

Risk-based decision making is central to zero-trust architectures. NIST SP 800-207 [17] describes access decisions that are re-evaluated continuously by a trust algorithm and distinguishes criteria-based from score-based and singular from contextual algorithms, but it does not prescribe how the inputs should be weighted. General risk-assessment guidance expresses risk as a function of the likelihood of a threat event and its impact [18], and ISO/IEC 27005 [19] provides process guidance for information-security risk management.

The gap addressed here lies in the combination of three requirements that are rarely met together: use of context when weighting evidence, an aggregation that is interpretable by construction, and an explicit link between the score and a graded, proportionate action.

3. Materials and Methods

The study follows a model-development and analytical-evaluation design. The input layer consists of security-relevant user activity records. For each observation, the model considers five principal dimensions: behavioral deviation (B), access context (C), resource sensitivity (R), anomaly severity (A), and historical compliance (H). These dimensions can be obtained from authentication logs, access-control records, data-loss-prevention events, endpoint telemetry, and security monitoring systems. The selected variables are normalized to the interval [0,1] so that indicators measured on different scales can be

(cc) BY

combined without allowing a high-magnitude variable to dominate the aggregation.

For an observation i, the baseline risk score is defined as:

$$S_i = wB \cdot B_i + wC \cdot C_i + wR \cdot R_i + wA \cdot A_i + wH \cdot H_i, \quad \sum w_j = 1, 0 \le S_i \le 1$$

where wB, wC, wR, wA and wH are indicator weights. The formulation is deliberately additive because it preserves interpretability and permits direct sensitivity analysis. However, a static weighting scheme is insufficient when the security context changes. For example, unusual access to a low-sensitivity document should not automatically receive the same risk contribution as the same deviation applied to a highly sensitive resource. Therefore, the proposed model uses contextual weight adjustment:

$$w_j^* = \frac{exp(\theta_j \cdot z)}{\sum_k exp(\theta_j \cdot z)}$$

where z represents contextual variables and θ_j controls the influence of context on the j-indicator. The softmax transformation guarantees non-negative weights whose sum equals one. In practice, θ can be estimated from labeled historical incidents or calibrated through expert-defined constraints when labeled data are limited. This structure also permits gradual updating as new validated security outcomes become available.

3.1. Data sources and feature extraction

The model consumes records that are normally produced by identity providers, audit subsystems, data-loss-prevention (DLP) tools, endpoint agents, and network monitors (for example, authentication and audit logs feed B and C, DLP and endpoint events feed A, compliance records feed H, and the asset inventory feeds R). Events are aggregated into fixed observation windows (for example, one session or one working day) and compared with a per-user baseline estimated from a rolling history.

3.2. Behavioral indicators and risk interpretation

Behavioral deviation represents the distance between the observed activity and a user-specific or role-specific baseline. Access context captures variables such as unusual time, location, device, network segment, or privilege context. Resource sensitivity represents the importance of the accessed information, while anomaly severity reflects the technical significance of the observed event. Historical compliance is introduced as a stabilizing factor: repeated policy violations increase risk, whereas a consistent history of compliant behavior reduces the contribution of isolated low-impact deviations. These variables correspond to the principle of User Behavior Analysis in which patterns of human activity are analyzed to identify anomalies rather than monitoring technical events independently [1].

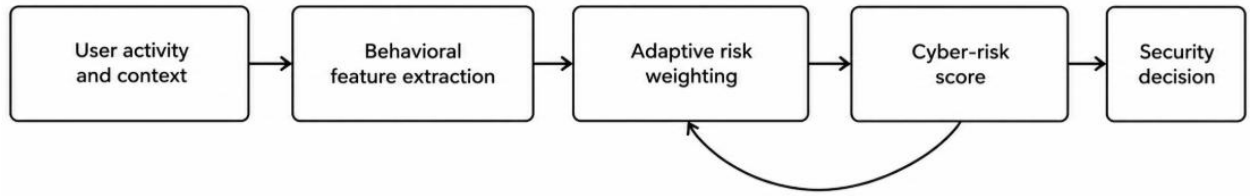


Figure 1. Proposed workflow for intelligent cyber-risk assessment based on user behavior.

Raw measurements are winsorized at the 1st and 99th percentiles and rescaled to [*0,1*]. Behavioral deviation is computed from the multivariate distance between the current feature vector □□ (for example, number of distinct resources, volume transferred, and number of privileged operations) and the baseline mean □ and covariance □:

$$d_i = \sqrt{(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}, \quad B_i = 1 - exp\left(-\frac{d_i}{\tau}\right)$$

where $\tau > 0$ sets how quickly the score saturates. Access context is the weighted share of context attributes (time, location, device, network segment, privilege context) that are inconsistent with the user's profile:

(cc) BY *dtai.uz*

$$C_i = \frac{\sum_m v_m \cdot \delta_m(i)}{\sum_m v_m}, \quad \delta_m(i) \in \{0,1\}$$

where v_m is the importance of attribute m. Resource sensitivity maps the classification level of the accessed resource to a number; in the illustrative setting below, public, internal, confidential, and restricted resources receive $0.10, 0.35, 0.70,$ and 1.00. Anomaly severity combines the severities a_k of the technical events in the window by a noisy-OR rule, so that several moderate events can jointly reach a high value while a single weak event cannot, and historical compliance is an exponentially discounted record of validated violations u_t:

$$A_i = 1 - \prod_k (1 - a_k), \; H_i = 1 - exp(-\lambda \cdot \sum_t \gamma^{T-t} \cdot u_t), \; 0 < \gamma < 1, \; \lambda > 0$$

3.3. Contextual weighting and calibration

The context vector is $z = (1, R_i, P_i)^T$, where $P_i \in [0,1]$ is the privilege level of the account. The access-context indicator C is deliberately excluded from z: if C acted both as an indicator and as a context variable, an increase in C could redistribute weight away from its own term and break monotonicity. Table 1 gives the illustrative parameter matrix used in Section 4. The values were set by hand from expert constraints (higher sensitivity raises the weights of resource sensitivity and behavioral deviation; higher privilege moderately raises all weights).

Table 1.

Illustrative parameter matrix of the contextual weighting mechanism.

Indicator □	**□$_j$ (constant)**	**□$_j$ (resource sensitivity R)**	**□$_j$ (privilege level P)**
Behavioral deviation (B)	0.0	0.9	0.3
Access context (C)	−0.4	0.4	0.2
Resource sensitivity (R)	−0.2	1.5	0.5
Anomaly severity (A)	0.1	0.7	0.4
Historical compliance (H)	−0.3	0.0	0.1

When labeled incidents are available, can be estimated by minimizing a regularized cross-entropy loss between a logistic transformation of the score and the incident label, with sign and magnitude constraints derived from expert judgment; this estimation is not performed here. The sensitivity of the score can be stated explicitly. For an indicator outside the context vector (B, C, A, H), $\frac{\delta S_i}{\delta x_j} = v_j^* > 0$, so the score cannot decrease when such an indicator changes adversely while the context is fixed. For resource sensitivity, which is both an indicator and a context variable, the derivative contains a redistribution term:

$$\frac{\delta S_i}{\delta R_i} = w_R^* + \sum_k x_k \cdot w_k^* \cdot \left(\theta_{kR} - \sum_l w_l^* \cdot \theta_{lR} \right)$$

whose sign is not guaranteed in general. For the values of Table 1 it was verified numerically over a dense grid, and no decrease was observed; in a calibrated deployment this check should be an acceptance test.

3.4. Security decision mapping

Risk scoring is useful only if it is connected to an operational response. The proposed model divides the normalized score into four decision bands: normal, monitoring, enhanced authentication, and restrictive action. An important design principle is proportionality. A low score should not trigger disruptive controls, while a high score combined with sensitive-resource access should support rapid containment. Accordingly, the score is interpreted together with resource sensitivity and confidence in the anomaly evidence.

$$D_i = f(S_i, R_i, Q_i)$$

where D is the recommended security decision and Q is the confidence of the detection evidence. This formulation prevents a single weak indicator from being interpreted as conclusive

(cc) BY

evidence and allows organizations to implement different response policies according to operational requirements.

In the illustrative configuration the thresholds are $\tau_1 = 0.25$, $\tau_2 = 0.50$, and $\tau_3 = 0.65$, defining the bands normal ($S < \tau_1$), monitoring ($\tau_1 \leq S < \tau_2$), enhanced authentication ($\tau_2 \leq S < \tau_3$), and restrictive action ($S \geq \tau_3$). Two gating rules implement proportionality. If the confidence of the evidence is low ($Q < 0.35$), the decision is capped at monitoring. The restrictive band is granted only if $Q \geq 0.60$ and the resource is sensitive ($R \geq 0.50$) otherwise the decision is lowered to enhanced authentication.

3.5. Analytical evaluation

The analytical evaluation was constructed around controlled scenarios rather than a claim of deployment in a live enterprise environment. Each scenario changes one or more normalized indicators while holding the remaining variables constant. The purpose is to verify three properties of the model: (1) monotonicity, meaning that increasing adverse indicators should not reduce the risk score; (2) contextual sensitivity, meaning that the same behavioral deviation should have a stronger effect when the accessed resource is more sensitive; and (3) decision proportionality, meaning that small deviations should not immediately produce the strongest security action.

Two analyses were designed. The first is a set of seven controlled scenarios (S1 to S7) in which indicator values are specified directly. Static weights are set equal to the contextual weights at a neutral reference context ($R = 0.5$, $P = 0.5$), so that the comparison isolates the effect of context. The second is experiment: 20,000 observations were generated with a fixed random seed, 5% from an anomalous and 95% from a benign profile. For benign observations B, C, and A follow Beta(1.5, 8), H follows Beta(1.2, 8), and R follows Beta(2, 3); for anomalous observations B follows Beta(5, 2.5), C and A follow Beta(4, 3), H follows Beta(2, 4), and R follows Beta(3, 2). The confidence Q follows Beta(2, 4) and Beta(4, 2), and the privilege level P follows Beta(2, 4) and Beta(3, 3), for benign and anomalous observations, respectively. These distributions are assumptions made for the analysis; they do not describe a real organization, and the labels denote only the generating profile, not real incidents.

Table 2.

Representative behavioral indicators and their analytical role in cyber-risk assessment.

Indicator	Operational example	Normalized range	Risk contribution
Behavioral deviation (B)	Deviation from user/role baseline	0–1	Higher deviation increases risk
Access context (C)	Unusual time, location or device	0–1	Raises risk when context is inconsistent
Resource sensitivity (R)	Sensitivity/criticality of accessed data	0–1	Amplifies consequences of access
Anomaly severity (A)	Severity of detected technical event	0–1	Raises risk for technically significant events
Historical compliance (H)	Prior policy violations / compliance history	0–1	Persistent non-compliance increases risk

4. Results and Discussion

4.1. Scenario-based analysis

Table 3 reports the scenarios. Scenarios S2 and S3 differ only in the sensitivity of the accessed resource (0.15 versus 0.90). Under contextual weighting the score rises from 0.292 to 0.551 (an increase of 0.258), whereas under static weights the same change raises it by 0.218; the contextual weight of resource sensitivity grows from 0.219 to 0.353, while that of historical compliance falls from 0.146 to 0.076. The identical behavioral anomaly therefore leads to monitoring on a low-sensitivity resource and to enhanced authentication on a highly sensitive one.

(cc) BY

Table 3.

Scenario analysis: indicator values, scores (static and contextual weights), and recommended decisions.

ID	Scenario	B / C / R / A / H	Q	S static	S ctx.	Decision (contextual)
S1	Routine activity	0.10/0.10/0.30/0.05/0.05	0.80	0.141	0.130	Normal
S2	Isolated anomaly, low-sensitivity resource	0.55/0.50/0.15/0.20/0.05	0.50	0.289	0.292	Monitoring
S3	Same anomaly, high-sensitivity resource	0.55/0.50/0.90/0.20/0.05	0.50	0.507	0.551	Enhanced authentication
S4	Bulk access to sensitive data, unusual location	0.80/0.80/0.90/0.70/0.30	0.80	0.753	0.780	Restrictive action
S5	Repeated policy violator, moderate anomaly	0.40/0.30/0.50/0.40/0.80	0.60	0.458	0.458	Monitoring
S6	Privileged account, converging indicators	0.90/0.90/0.95/0.90/0.60	0.90	0.884	0.902	Restrictive action
S7	Technical alert without behavioral deviation	0.10/0.10/0.40/0.80/0.00	0.30	0.352	0.344	Monitoring

4.2. Sensitivity and monotonicity

Figure 2 shows the score as a function of behavioral deviation for $R = 0.15$ and $R = 0.90$, with the other indicators fixed at the values of S2 and S3. All curves are increasing. For $R = 0.15$ the contextual score rises from 0.163 to 0.398 as B grows from 0 to 1, and for $R = 0.90$ from 0.418 to 0.659.

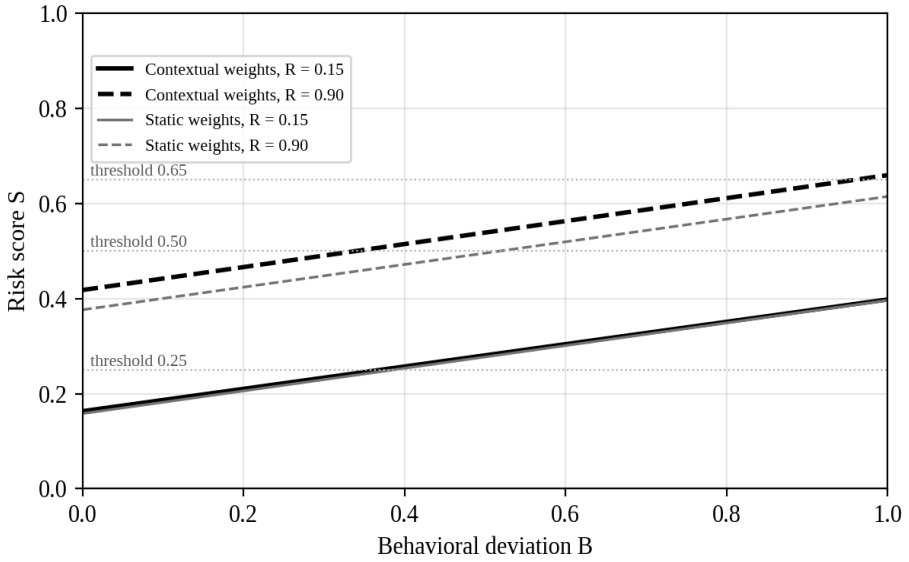


Figure 2. Risk score as a function of behavioral deviation for low- and high-sensitivity resources.

The contextual weights behave as intended. For the low-sensitivity resource the weights are close to each other and the weight of historical compliance (0.146) is relatively high; for the sensitive resource the weight of resource sensitivity rises to 0.353, and in the privileged and sensitive context (S6) to 0.396, while the weights of access context and compliance history fall. When potential impact is high, the score is

(cc) BY

therefore determined mainly by asset sensitivity and by behavioral and technical evidence.

4.3. Experiment

Table 4 shows the decision distributions under three configurations. With static weights, 74.52% of anomalous profiles are escalated to enhanced authentication or restrictive action; contextual weighting without gating raises this to 81.21%, and the full model with gating gives 76.54%. The share of benign observations reaching these bands stays below 0.1% in all configurations (0.01%, 0.07%, and 0.04%), and no benign observation receives the restrictive action.

Table 4.

Distribution of recommended decisions in the population (percent of each population).

Configuration	Population	Normal	Monitoring	Enhanced auth.	Restrictive
Static weights, gated	Benign	64.36	35.63	0.01	0.00
Static weights, gated	Anomalous	0.00	25.48	61.36	13.16
Contextual, no gating	Benign	64.67	35.26	0.07	0.00
Contextual, no gating	Anomalous	0.00	18.79	55.41	25.80
Contextual, gated	Benign	64.67	35.29	0.04	0.00
Contextual, gated	Anomalous	0.00	23.46	60.08	16.45

Gating reduces the share of anomalous profiles receiving the restrictive action from 25.80% to 16.45%, because part of them has low evidence confidence or low resource sensitivity and is directed to enhanced authentication instead. The experiment demonstrates the behavior of the mechanisms under stated assumptions; it does not estimate detection accuracy in a real environment.

4.4. Discussion

The analytical structure demonstrates that risk cannot be represented adequately by a single event count. Consider two users who both access a resource outside their normal working period. If the first user accesses a low-sensitivity document from a recognized device and has a stable compliance history, the contextual weighting mechanism assigns limited additional risk. In contrast, if the second user accesses a highly sensitive resource from an unusual location, the resource-sensitivity and access-context terms increase the effective contribution of the anomaly. Thus, the model differentiates identical surface events according to their security context.

A second property is interpretability. Because each factor is explicitly represented in the score, an analyst can trace why a decision was recommended. This is important in operational security environments where automated blocking may have substantial consequences. The model therefore supports a graduated response: monitoring for moderate risk, stronger authentication for elevated risk, and restriction or blocking when high risk is supported by multiple converging indicators. This design is consistent with evidence that legitimate insiders are difficult to distinguish from malicious actors when activity is considered only in terms of technical privilege; identity, past activity, organizational context, and information context are important complementary signals [5].

From a machine-learning perspective, the proposed formulation can serve as either a transparent scoring layer or a feature-engineering stage for supervised and unsupervised models. Tree ensembles can learn nonlinear relationships among the indicators, while the explicit score can remain as a human-readable decision layer. Recent literature indicates that machine-learning approaches are increasingly used for insider-threat detection, but also highlights challenges associated with temporal behavior, contextual dependence, class imbalance, and generalization [2]. Consequently, future implementation should compare the proposed scoring model with logistic regression, random forests, gradient boosting, and anomaly-detection methods using temporally separated validation data.

A rigorous empirical validation should use public insider-threat data produced by the CERT program [20] and compare the model with logistic

(cc) BY

regression, random forests [21], gradient boosting, and Isolation Forest [12] on temporally separated periods, reporting precision, recall, F1, the area under the precision-recall curve, alerts per analyst per day, and time to detection.

In deployment the model can sit between the SIEM and the policy-enforcement layer, with analyst feedback used to adjust baselines, thresholds, and, when enough labels exist. Behavior monitoring also raises privacy and legal issues: only the features needed for the five indicators should be collected, identifiers pseudonymized, access to scores logged, and employees informed in accordance with the applicable personal-data legislation; automatic restrictive actions should be limited to cases meeting the gating conditions, with human review for the others.

4.5. Limitations and threats to validity

An important limitation is that the present evaluation is analytical and scenario-based. It does not establish a deployment-level accuracy estimate, because such a claim would require a sufficiently representative labeled dataset and a reproducible experimental protocol. Public insider-threat research nevertheless provides evidence that controlled behavioral experiments can reveal distinguishable patterns between benign and malicious activity [3], and recent MITRE work further emphasizes the value of data-driven behavioral indicators in real environments [6]. These findings support the proposed direction while also indicating that real-world validation should incorporate privacy, ethics, access control, and organizational governance.

The parameters of Table 1, the thresholds, and the distributions are illustrative assumptions; different values would change the numerical results, although the qualitative properties (monotonicity, contextual amplification, and proportional gating) do not depend on their exact choice. Adaptive adversaries may shift baselines slowly (a low-and-slow strategy), and baseline drift from legitimate role changes can cause false alarms, so baseline updates need governance rules.

5. Conclusion

This study proposed an intelligent model for assessing cyber risks based on user behavior in information systems. The model integrates behavioral deviation, access context, resource sensitivity, anomaly severity, and historical compliance into a normalized risk score and connects that score with proportional security decisions. The analytical assessment indicates that contextual weighting is essential for distinguishing low-impact anomalies from combinations of behavior and context that represent materially higher risk. The principal scientific value of the approach is its interpretable integration of heterogeneous behavioral evidence rather than reliance on isolated security events.

Future research should validate the model on public and organizational datasets, optimize adaptive weights using machine-learning techniques, evaluate temporal sequences of user activity, and incorporate explainable AI methods for analyst-facing justification. A further direction is to extend the model from risk assessment to proactive recommendation of security actions, thereby forming an adaptive decision-support mechanism for information-system protection.

References

1. Dahle, N., Tumlin, A., Tran, N., Koutsoukos, X., & Derr, T. (2026). RAD: Rule-augmented relational anomaly detection. In Proceedings of the 35th ACM International Conference on Information and Knowledge Management (CIKM ’26), pp. 112–121. ACM.
2. Hussain, H., & Kiani, A. A. (2026). Explainable behavioral analytics for users and entities using hybrid AI models: A MUEBA-XAI framework integrating LSTM, Isolation Forest, SHAP, and LIME. International Journal of Innovations in Science & Technology, 8(6), 2629–2648.
3. Komka, M. C., Ademuwagun, A. K., Abdullahi, S. M., Mashkur, T. M., & Ifeanyi, F. D. (2026). Proactive insider threat detection in cloud environment using machine learning. Nature Journal of Emerging Sciences, Technologies, and Innovations, 8(4), 397–410.
4. Mishra, A., Vedula, V., Mishra, A., & Sharma, S. (2026). Artificial intelligence for cybersecurity: A scoping survey of

(cc) BY

paradigms, applications, and emerging trends. Algorithms, 19(8), Article 653.
5. Paul, A., Sadiq, B. O., Buhari, D., & Olaniyan, R. (2025). A review of neural networks for enhanced user entity behavior analytics in cybersecurity: Addressing the challenge of vanishing gradient. International Journal of Research and Innovation in Applied Science, 9(12), 154–172.
6. Caputo, D., Stephens, G., Stephenson, B., and Kim, M. (2010). Human Behavior, Insider Threat, and Awareness: An Empirical Study of Insider Threat Behavior. MITRE Corporation.
7. Strom, B., Battaglia, J., Kemmerer, M., et al. (2017). Finding Cyber Threats with ATT&CK-Based Analytics. MITRE.
8. Maybury, M., Chase, P., Cheikes, B., et al. (2005). Analysis and Detection of Malicious Insiders. MITRE/ARDA.
9. Hunker, J., & Probst, C. W. (2011). Insiders and insider threats: An overview of definitions and mitigation techniques. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 2(1), 4–27.
10. Cappelli, D. M., Moore, A. P., & Trzeciak, R. F. (2012). The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes. Addison-Wesley.
11. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15.
12. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 413–422). IEEE.
13. Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. In Proceedings of the AAAI-17 Workshop on Artificial Intelligence for Cyber Security.
14. Le, D. C., & Zincir-Heywood, N. (2021). Anomaly detection for insider threats using unsupervised ensembles. IEEE Transactions on Network and Service Management, 18(2).
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (pp. 4765–4774).
16. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM.
17. Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture (NIST Special Publication 800-207). National Institute of Standards and Technology.
18. Joint Task Force Transformation Initiative. (2012). Guide for Conducting Risk Assessments (NIST Special Publication 800-30, Revision 1). National Institute of Standards and Technology.
19. ISO/IEC 27005:2022. Information security, cybersecurity and privacy protection: Guidance on managing information security risks. International Organization for Standardization.
20. Glasser, J., & Lindauer, B. (2013). Bridging the gap: A pragmatic approach to generating insider threat data. In Proceedings of the IEEE Security and Privacy Workshops (pp. 98–104). IEEE.
21. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

(cc) BY