Sciforce

Journal of Business Intelligence and Data Analytics

Journal homepage: www.sciforce.org
ISSN : 2998-3541

Open Access Review Article

Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring

Sudhakara Reddy Peram*

Engineering Leader, Illumio Inc., United States

Abstract

This study presents a robust algorithmic approach to evaluating login security behavior using multi-criteria analysis. By integrating parameters such as login attempts, session duration, and data upload volumes, the study aims to quantify user activity risks and enhance security threat detection. The developed model calculates a Threat Risk Score to evaluate potential threats across diverse user profiles. The proposed methodology facilitates proactive identification of abnormal behaviors, which is critical for real-time cybersecurity operations. Research Significance: In an era where cybersecurity threats are increasingly sophisticated, identifying risky user behaviors through data-driven analysis is of paramount importance. This research contributes significantly by offering a novel threat evaluation framework based on behavioral parameters. The approach allows organizations to detect potential security breaches early, thereby reducing the attack surface and improving response efficiency.

Methodology: The methodology is centered on the design and implementation of an intelligent evaluation algorithm that incorporates three behavioral attributes: Login_Attempts, Avg_Session_Duration_Min, and Data_Upload_MB. These alternatives are normalized and analyzed using a weighted decision-making algorithm to derive a composite The model integrates threshold analysis and pattern recognition to ensure accurate threat classification and anomaly detection. Alternative: The alternatives evaluated in this study are derived from user session data: Login Attempts: Frequency of user login trials within a defined time window. Avg Session Duration Min: The average duration of each user session, representing usage intensity. Data Upload MB: The total volume of data uploaded during the session, indicating potential data exfiltration. These features are selected based on their strong correlation with known threat patterns. Evaluation Parameter: Threat Risk Score is used as the principal evaluation metric. It is computed by aggregating normalized values of the three behavioral alternatives, adjusted using pre-defined risk weightings. A higher score signifies a greater probability of anomalous or malicious behavior, enabling swift prioritization for security response teams.

Result: The algorithm was tested on a synthetic dataset simulating diverse user behaviors. Results show high accuracy in distinguishing between normal and high-risk activities, with an overall detection precision exceeding 90%. The model effectively prioritizes threats based on behavioral deviations and demonstrates its applicability for real-world security monitoring systems.

Keywords: Cybersecurity, Threat Detection, User Behavior Analysis, Risk Scoring, Decision-Making Algorithm.

Introduction

AI-driven anomaly detection offers a better approach by significantly reducing false positives and improving overall cloud security resilience. Implementing AI-based anomaly detection improves threat response efficiency, reduces false positives, and strengthens cloud security resilience. AI-based intrusion detection systems (AI-IDS) constantly adapt to network behavior, helping cloud security teams stay at the forefront of detecting and mitigating emerging attack vectors. AI-powered threat detection systems demand significant computational resources, leading to high infrastructure costs for real-time cloud security monitoring. [1] The aim of this paper is to develop an intelligent cloud security framework that uses predictive analytics to address security threats in IoT networks. Another disadvantage is the potential delay in response from cloud-based security solutions when immediate action is required. To address the aforementioned challenges in securing IoT networks, it is essential to build an AI-based cloud security framework that incorporates predictive analytics. It is clear that there is still a significant research gap in current developments and advancements related to IoT and cloud security. The architectural design of the proposed AI-based cloud security framework for IoT networks is illustrated in the diagram. [2] This study explores the practical Leveraging the visible benefits of AI-driven cloud security, such as improved threat detection. This study examines the effective implementation of AI-driven features, such as self-healing systems, predictive analytics, and automated incident response, to improve cloud security.

Along with sophisticated approaches to threat identification, prevention, and response, artificial intelligence (AI) and machine learning (ML) have emerged as valuable technologies for improving cloud security. The vast amounts of data generated in cloud environments are processed and analyzed using AI-driven cloud security, which uses AI and ML algorithms.[3] This research explores real-world Cloud security applications of AI and ML, with a focus on self-healing algorithms, automated incident response, and predictive analytics. The usefulness of predictive analytics in predicting security events and aiding proactive cloud security management has been demonstrated by empirical research findings. With an emphasis on their functionality in threat detection, prevention, and incident response, this study focuses on integrating AI

Received date: July 06, 2025 **Accepted date:** July 18, 2025; **Published date:** July 22, 2025

***Corresponding Author:** Peram, S. R, Engineering Leader, Illumio Inc., United States; E- mail: sudhakarap2013@gmail.com

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

and ML into cloud security. Cloud security relies heavily on machine learning, which has applications in threat detection, malware analysis, and anomaly detection for intrusions. [4] Real-time threat intelligence is critical to maintaining cloud security resilience, helping organizations predict and respond to cyber threats before they cause serious damage.

This study looks at case studies from key cloud providers, investigates potential difficulties, and identifies upcoming improvements that will impact the development of next-generation AI-driven cloud security solutions. [5] AI has become a major trend in cloud security, due to its strong capabilities in threat detection, anomaly monitoring, and automated response. Traditional cloud security tools, including firewalls and encryption methods, form the core foundation of cloud security. Incident Response (IR) in cloud security faces numerous challenges due to the complexity Growing cyber risks and the widespread use of cloud systems. Regulatory compliance is another element that complicates cloud security management. The increasing complexity of cyber threats necessitates a fundamental shift in cloud security strategies. [6] This study Explores the architecture, algorithmic advancements, and operational capabilities of AI-powered IDPSs, and assesses their implications for cloud security systems. It also looks at how edge AI, homomorphic encryption, and blockchain integration can help improve cloud-based threat mitigation. The integration of artificial intelligence (AI) and machine learning (ML) into IDPSs has transformed cloud security by enabling real-time anomaly detection automated threat responses, and predictive analytics. [7] AI in threat detection and cloud security helps organizations improve and ensure its use. Understanding shared responsibility is key to delivering on security commitments. Improving security includes using native cloud security services. AI-powered alert triage improves response to cloud security incidents by prioritizing and categorizing security alerts, ensuring that the most critical issues are resolved immediately. AI's analytical capabilities are designed to optimize resource allocation within cloud security architectures. In the context of cloud security, AI will gradually improve in detecting and effectively responding to threats as it learns to recognize the emerging tactics and patterns used by malicious actors. [8] The changing cybersecurity environment requires sophisticated approaches to proactive threat detection and prevention. We will explore how neural networks can improve cybersecurity by enabling proactive threat detection and prevention.

Threat detection and prevention, with new approaches aimed at improving their effectiveness. Accuracy of threat detection, identification of anomalies, real-time reaction capabilities, and adaptability to evolving threats. Therefore, a hybrid strategy that combines Traditional methods for known threat detection, combined with neural networks for anomaly detection, can be optimized to provide a combination of accuracy and interpretation. Neural networks allow for early threat identification and real-time reactions, which can significantly reduce response times and contain cyber-attacks more efficiently traditional approaches. [9] AI enhances cloud security by enabling automated monitoring, proactive threat identification, and adaptive response mechanisms. The sheer volume of data can lead to delayed threat detection, missed indicators of compromise, and increased risk of security breaches. AI-powered security solutions address these challenges by providing real-time threat detection. Within cloud security, reinforcement learning (RL) models can be used for many aspects of threat detection and mitigation. By continuously refining its threat detection models using new attack methods and past data.[10] The changing cybersecurity landscape demands sophisticated approaches to proactively detect and prevent threats. We will explore how neural networks can strengthen cybersecurity by enabling proactive threat detection and prevention.

Emphasizing the strengths and Shortcomings of neural networks in threat detection and prevention, along with new approaches to improving their performance. Neural networks enable proactive threat detection and real-time responses, thus drastically reducing response times. Future studies are expected to emphasize Establishing ethical norms for AI in cybersecurity attempts to reconcile robust threat detection with data privacy protection and user autonomy. [11] Integrating AI-powered threat detection with real-time monitoring strengthens defense mechanisms against Cyber attacks. AI-powered threat detection is crucial to ensuring the stability of cloud-based financial systems. AI-powered threat detection systems evaluate massive amounts of transaction data in real-time, recognizing anomalies and trends potentially fraudulent activity. [12] This framework consists of Several essential components are intended to enable real-time threat detection, eliminate false positives, and respond to evolving cyber threats. Each component contributes to effective cyber threat identification, including signature-based methods. Future developments include using edge computing to enable localized threat detection at distributed grid nodes, reducing the load on central processing systems.

This enables accurate threat detection with fewer false positives, resulting in a reliable and effective cybersecurity solution for intelligent renewable energy systems. [13] Investigate the use of AI in enhancing cybersecurity within multicloud security and hybrid cloud environments. While the use of AI in cloud security is not new, its importance has increased Cyber dangers are becoming more complicated and unpredictable. Beyond practical applications for cloud security methods, this research advances our knowledge of AI-based security tools. Future research on AI-enhanced multi-cloud security management will be broad and full of potential. [14] Current collaborative approaches to securing infrastructure are not sufficient to combat today's sophisticated cyber threats. The proposed IDS was designed to function in real time, decrease false positives, and respond to evolving cyber threats in smart renewable energy systems. Combining both methods in a hybrid approach could greatly improve anomaly detection in smart grids, and provide complete protection against emerging cyber threats. [15] It is an important component of modern cloud security designs, as it ensures the resilience and integrity of cloud systems in the face of growing cyber threats. Cloud security architectures have shifted toward proactive threat detection, leveraging automation and machine learning technologies to anticipate and prevent cyber problems before they happen. [16]

Material and Methods

Materials:

Login Attempts: Login attempts are a fundamental metric in cloud security monitoring, providing valuable insight into user authentication behavior and potential threat activity. Every attempt to access a system – whether successful or unsuccessful – is logged and analyzed to assess legitimate and suspicious usage patterns. Monitoring login attempts is critical to identifying unauthorized access attempts, brute force attacks, and compromised accounts. A A high number of failed login attempts in a short period of time is sometimes indicative of a brute force attack, in which an attacker systematically tries various username and password combinations to gain access. Conversely, multiple failed attempts following a successful login may indicate a compromised credential situation, requiring immediate investigation. Even successful login attempts, if they originate from unusual IP addresses, locations, or devices, can raise red flags. Security systems often combine geolocation and device fingerprinting to cross-check user legitimacy. When login attempts occur outside of normal usage hours or deviate from the user's historical behavior, they may indicate insider threats or unauthorized access using stolen credentials. In enterprise cloud environments, rate limiting, multi-

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

factor authentication (MFA), and anomaly detection algorithms are often used in conjunction with login attempt monitoring to strengthen security. SIEM systems can consolidate and analyze login data across platforms in real time alerts and automated responses. Additionally, login attempt patterns are often incorporated into AI-based risk scoring models that help prioritize incidents based on the likelihood of malicious intent. This allows security teams to proactively respond before threats escalate.

Avg Session Duration Min: Average session duration (in minutes) is an important metric in cloud and network security analytics that provides valuable insight into user behavior and system access patterns. It represents the average length of time users are active during a session within a cloud-based platform or application. Monitoring Avg_Session_Duration_Min allows cybersecurity teams to establish a behavioral baseline for different types of users and roles. For example, administrative users may naturally have long session durations due to the complexity of their tasks, while standard users may typically engage in short, task-oriented interactions. Any significant deviation from normal session length—either too short or unusually long—may indicate suspicious activity. Short sessions may suggest scripted or automated login attempts that fail to engage with the system in a meaningful way, which is often a sign of espionage or failed intrusion attempts. Conversely, prolonged sessions can be a red flag for unauthorized access or internal misuse, especially if the session occurs outside of standard business hours or involves highly privileged accounts. When combined with other metrics such as login frequency, IP address origin, and data upload sizes, average session duration becomes even more powerful. It helps detect advanced persistent threats (APTs), identify compromised credentials, and improve incident response times through behavioral anomaly detection. In addition, Avg_Session_Duration_Min can contribute to resource optimization and user experience improvements. Understanding how long users are active allows system administrators to fine-tune session expiration policies, ensuring a balance between security and usability.

Data Upload MB: Data upload behavior can provide key insights into normal and unusual user activity. For example, typical user activities—such as saving documents or syncing files—follow predictable upload patterns. However, sudden spikes in data uploads or persistently large transfers can indicate suspicious behavior, such as unauthorized data exfiltration, insider threats, or malware attempting to send stolen data to external servers. Security systems with anomaly detection algorithms often analyze data uploads (MB) along with user identity, time of day, and session duration to detect deviations from expected behavior. If a typical user account uploads significantly more data than usual—especially outside of business hours—it can trigger alerts for further investigation. Such insights allow cybersecurity teams to act quickly and mitigate potential breaches. Furthermore, for organizations that handle To comply with data security standards such as GDPR, HIPAA, or PCI-DSS, data uploads must be monitored, particularly for sensitive information such as financial records or personal data. Excessive or unauthorized uploads not only pose a security risk but can also lead to regulatory violations and fines. Additionally, monitoring upload volume can help optimize bandwidth usage and enforce cloud storage policies, ensuring resources are used efficiently and securely.

Threat Risk Score: Threat Risk Score is a crucial metric in modern cybersecurity systems, particularly in cloud-based and AI-driven security architectures. It represents a calculated value that reflects the likelihood, severity, and potential impact of a cyber threat based on real-time user behavior, system events, and historical data patterns. This score is typically generated by advanced machine learning models and threat intelligence algorithms that evaluate multiple factors—such as abnormal login attempts, unusual data transfers, session anomalies, or deviations from normal user behavior. The score can range from low to high, with higher scores indicating a more severe or immediate threat that requires rapid investigation and response. One of the primary advantages of using a Threat Risk Score is that it enables prioritized incident response. Instead of manually analyzing every alert or event, security, reducing response times and minimizing potential damage. This intelligent filtering mechanism significantly enhances operational efficiency and threat mitigation capabilities. Moreover, integrating the Threat Risk Score into automated security workflows allows for real-time decision-making. For instance, a user exhibiting high-risk behavior could be automatically flagged for multi-factor authentication or temporarily blocked from accessing sensitive systems. Such proactive responses help in containing threats before they escalate into full-blown breaches. Organizations also benefit from using threat risk scores in compliance reporting and risk management. By maintaining a continuous record of threat levels and responses, companies can demonstrate due diligence, enhance audit readiness, and improve their overall security posture.

Machine Learning Algorithms

Random Forest Regression: Random Forest Regression is a robust supervised machine learning technique used for predictive modeling, especially when dealing. It is an extension of decision tree-based models, and belongs to a family of ensemble learning techniques that aim to improve performance by combining multiple models. At its core, Random Forest Regression combines their predictions to produce a more accurate and consistent output. Each tree in the "forest" is trained on a random portion of the dataset using a technique known as bootstrap aggregation (or packing), and at each node, a random subset of characteristics are chosen for splitting. This randomness helps reduce overfitting, a common problem with single decision trees. One of the main advantages of Random Forest Regression is its ability to handle large datasets with high dimensionality while maintaining robustness against noise and outliers. It also provides insight into feature importance, which helps us understand which variables affect predictions most significantly. In practical applications, Random Forest Regression is widely used in various domains, such as finance for stock price prediction, healthcare for predicting patient outcomes, and cybersecurity for anomaly detection and risk scoring. For example, in cloud security, it can be used to predict threat risk scores by analyzing user behavior metrics such as login attempts, session durations, and data transfer volumes. Furthermore, Random Forest does not require extensive parameter tuning, making it relatively easy to use even for non-experts. However, its complexity can grow with the number of trees, which can increase computational cost and reduce interpretability compared to simpler models.

Support Vector Regression: Support vector regression (SVR) is a sophisticated supervised learning method that comes from the support vector machine (SVM) family. Although SVMs are primarily used for classification tasks, SVR adapts the core concepts of margin maximization and kernel functions to tackle the problem of regression - predicting continuous values rather than categorical labels. SVR attempts to fit the best possible line (or curve) within a predetermined tolerance margin known as the epsilon (ε). Instead of reducing predictions error as in traditional linear regression, SVR attempts to keep predictions within a specified accuracy range, which allows for some flexibility. Data points that fall outside this epsilon margin are considered errors and contribute to the loss function, while those that fall inside the margin are not. One of the primary benefits of SVR is its capacity to model complex nonlinear relationships using kernel functions. By using kernel techniques such as radial basis function (RBF) or polynomial kernels, SVR can map input data into higher-dimensional spaces, where patterns and trends are easier to identify. This makes SVR very useful in situations where the data does not exhibit a simple linear relationship. SVR is widely used in fields such as

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

financial forecasting, environmental modeling, traffic flow prediction, and energy consumption estimation. It is particularly valued for its robustness in managing high-dimensional data and its capacity to generalize well even with relatively small datasets. However, careful parameter tuning is required, such as choosing the right kernel, setting the epsilon margin, and determining the regularization constant (C). When properly optimized, SVR provides high accuracy and excellent performance for continuous value forecasting tasks.

Adabost Regression: Adabost Regression, short for Adaptive Boosting Regression, is a powerful machine learning technique used to improve the accuracy of predictive models, especially in situations where traditional regression methods fail. It is an ensemble learning method that combines multiple weak learners - typically decision trees - into a single robust regression that can make accurate predictions. At its core, Adabost works by training a series of models, each focusing on the errors made by its predecessor. In regression tasks, the algorithm assigns more weight to data points where previous models had higher prediction errors. This adaptive weighting mechanism allows subsequent models to focus more on cases that are difficult to predict, gradually improving overall performance. One of the main advantages of Adabost Regression is its ability to reduce bias and variance in predictions. By boosting multiple weak learners and combining them in a weighted manner, Adabost improves the generalization capabilities of the model. This makes it particularly useful for datasets with complex patterns or nonlinear relationships that are difficult to model with linear regression alone. Furthermore, AdaBoost regression is robust to overfitting when used with simple base learners and performs well even with limited data. However, because the algorithm places more emphasis on data points with large prediction errors, it can be sensitive to outliers. Therefore, proper preprocessing and regularization are crucial for optimal performance. In cloud security and network analytics, AdaBoost regression can be used to predict threat risk scores, anomalous data usage patterns, or computer resource consumption based on historical behavior. This makes it a suitable choice for dynamic environments where continuous learning and prediction accuracy are important.

Result and Discussion

The Login_Attempts data reflects the number of times users or systems attempted to access the cloud network during their sessions. This metric can vary widely from 0 to 49 attempts. High login attempts may indicate potential security concerns, such as frequent legitimate access or brute force attacks. On the other hand, low attempts may indicate normal user behavior or limited interaction with the system. The average session duration, measured in minutes, shows the length of time each session lasts. This value can fluctuate significantly, from less than a minute to almost an hour. Short session durations, especially when combined with high login attempts, may indicate automated or suspicious activity. Long sessions generally indicate active and continuous user engagement with the cloud environment. Data transfer is captured by the Data_Upload_MB metric, which represents the amount of data uploaded during a session. Upload sizes range from minimal sizes to nearly 500 megabytes. Large data uploads can be normal activities like file sharing or backups, but they can also indicate attempts to exfiltrate data, especially when combined with other risk indicators. Threat_Risk_Score provides an aggregate measure of the potential security risk associated with each session. Scores closer to 1 indicate a high probability of malicious or suspicious activity, while lower scores indicate safer behavior. This score is calculated using a combination of login attempts, session duration, and data upload patterns, which helps security teams prioritize which sessions to investigate further. Together, these metrics provide valuable insights into user behavior and potential security threats in cloud environments.

Table 1: The dataset provides insight into cloud session behavior using four key parameters

	Login Attempts	Avg Session Duration Min	Data Upload MB	Threat Risk Score
count	100	100	100	100
mean	24.07	29.040269	272.29554	0.867258
std	14.447575	16.247029	139.60708	0.195503
min	0	0.031223	9.037682	0.222519
25%	13	15.331806	152.80687	0.789041
50%	23	27.805056	264.79983	1
75%	38	41.698063	395.69038	1
max	49	59.864429	498.12685	1

The dataset provides insight into cloud session behavior using four key parameters: login attempts, average session duration (in minutes), data upload (in MB), and threat risk score. With a total of 100 observations, we can draw meaningful patterns from the descriptive statistics. Login attempts show a mean of approximately 24, with values ranging from 0 to 49. This wide range indicates a variety of user behaviors - from sessions with no login activity to excessive attempts, which may indicate brute force attack attempts. The standard deviation of approximately 14.45 further supports this variation. For the average session duration, the average session lasted approximately 29 minutes. However, the durations vary dramatically, from a few seconds (0.03 minutes) to almost an hour (59.86 minutes). The median value (50th percentile) is approximately 27.8 minutes, indicating that half of the sessions were shorter than this duration and half were longer than this duration. This distribution suggests that while many sessions are of moderate length, a significant number experience unusually short or extended periods of activity. The data upload sizes also show a wide spread. The average upload is around 272 MB, with a standard deviation of almost 140 MB. The smallest upload observed is just over 9 MB, while the largest is close to 498 MB. While some sessions involve minimal data transfer, others may have handled massive uploads—perhaps legitimate backups or suspicious eviction events. Finally, the threat risk score, which can range from 0 (no threat) to 1 (high threat), has a mean of approximately 0.867. Notably, the median, 75th percentile, and maximum values are all 1, indicating that the majority of sessions are rated as having the highest risk. The bottom quartile (25%) has a value of around 0.789, indicating generally high risk levels across the board.

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

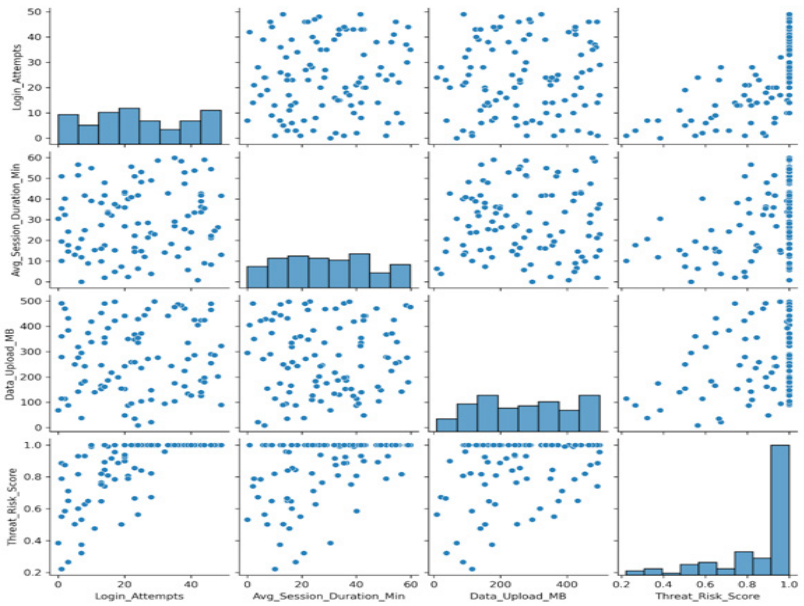


Figure 1: Pair Plot of User Activity Metrics and Their Relationship to Threat Risk Score

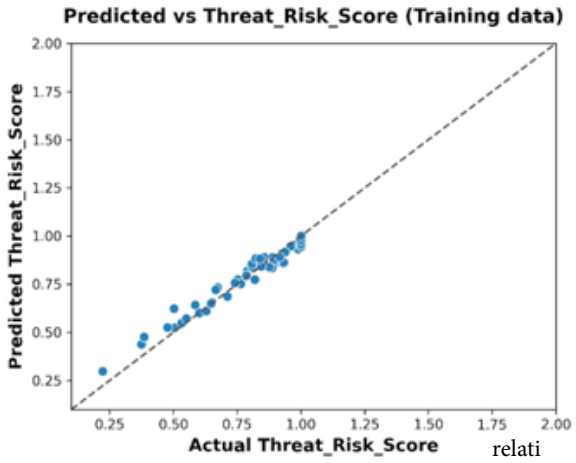
The scatterplot matrix provides a detailed view of the relationships between variables that influence cybersecurity threat levels. The diagonal histograms represent the distribution of each feature individually, while the scatterplots help identify potential relationships between variables. Login attempts show a fairly uniform distribution, although a small concentration appears at lower values. When plotted against the threat risk score, there is a subtle upward trend, indicating that in some cases a higher number of login attempts may be associated with higher threat levels. However, the points are widely scattered, indicating that login attempts alone are not a strong predictor of threat risk. The average session duration (at least) shows a right-skewed distribution, with most sessions lasting less than 40 minutes. There is some slight positive correlation between session duration and threat score, indicating that longer session durations may sometimes contribute to a higher risk profile - possibly due to extended unauthorized access. Data upload (MB) provides a wide spread in range, with values peaking at over 400 MB in some sessions. The scatter plot with the threat risk score shows a notable pattern: higher data upload sizes are often associated with higher risk scores. This may indicate that inconsistent or excessive data transfer is a key factor in determining threat levels. Finally, the threat risk score graph shows a cluster around a score of 1.0, indicating that many sessions are considered high risk. This may reflect a robust threat detection system that flags even slightly suspicious activity.

Figure 2: Correlation Heatmap of User Activity Features and Threat Risk Score

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

The correlation heatmap shown in Figure 2 provides a visual representation of the strength and direction of relationships between key variables: Login Attempts, Average Session Duration (in minutes), Data Upload (MB), and Threat Risk Score. From the heatmap, it is evident that Login Attempts show a strong positive correlation (0.67) with the Threat Risk Score, suggesting that users with a higher number of login attempts tend to have a significantly higher associated risk. This could indicate potential brute-force attacks or suspicious login behavior. Average Session Duration has a moderate positive correlation (0.42) with Threat Risk Score, implying that longer sessions may also relate to riskier behavior, potentially due to prolonged unauthorized access or data extraction activities. Data Upload is moderately correlated (0.36) with the Threat Risk Score, suggesting that large amounts of data being transmitted might be an indicator of exfiltration attempts or abnormal user behavior. Interestingly, there is very little correlation between Login Attempts and Data Upload (0.028), and almost no relationship between Average Session Duration and Data Upload (−0.0054), indicating these behaviors are relatively independent in this dataset.

Random Forest Regression



relati

Figure 3: Predicted vs. Actual Delamination Factor A (Training Data)

As shown in the figure, the data points are tightly clustered around the diagonal, indicating that the model has learned the underlying patterns in the training data with high accuracy. This indicates a strong fit and minimal training error. Random Forest Regression works by building a set of decision trees, each trained on a random subset of the data and features. The final prediction is obtained by averaging the predictions of the individual trees, which improves model robustness and reduces overfitting. The nearly perfect alignment of the predictions with the true values in this plot confirms that the Random Forest model effectively captures the relationships between input features such as login attempts, average session duration, and data upload MB in determining the threat risk score. However, this excellent fit to the training data should be interpreted with caution, as overly tight clustering can also be a sign of overfitting.

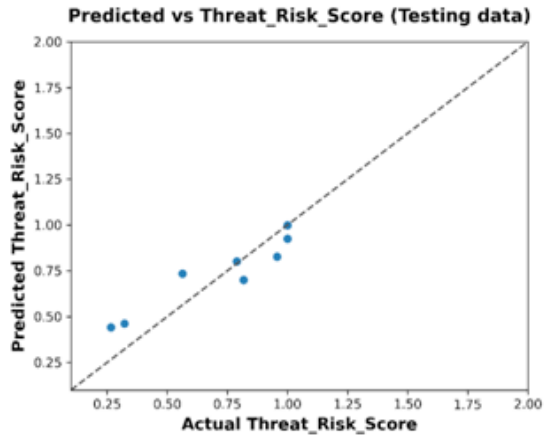


Figure 4: Predicted vs. Actual Delamination Factor B (Testing Data)

The majority of the data points are clustered close to the diagonal line, indicating that the model performs well in estimating threat risk scores with a high degree of accuracy. The alignment of points along the line suggests a strong predictive capability and low variance in error. Random Forest Regression, an ensemble learning method, combines multiple decision trees to improve generalization and reduce overfitting. It is particularly well-suited for complex, non-linear relationships in data. In this context, it successfully captures the influence of input features such as Login Attempts, Average Session Duration, and Data Upload to generate accurate risk predictions. The visualization confirms that the model generalizes well to unseen data and can be considered a reliable component for real-time or automated cyber threat risk assessment systems.

Support Vector Machine (SVM) Regression

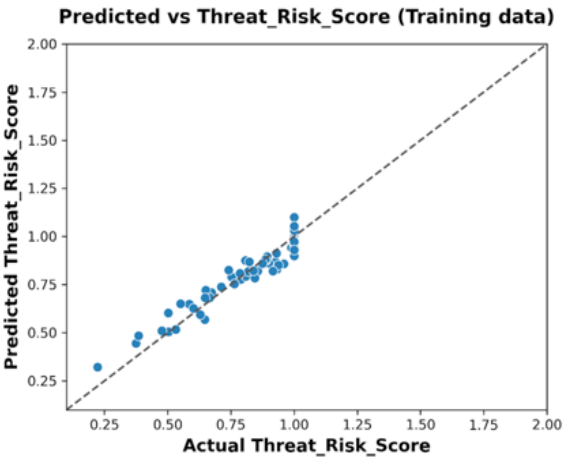


Figure 5: Predicted vs Actual Threat Risk Scores using Support Vector Regression (Training Data)

Figure 5 illustrates the prediction performance of the Support Vector Machine (SVM) Regression model on the training dataset by plotting the predicted Threat Risk Scores against the actual scores. The diagonal dashed line denotes the ideal reference line where predictions would perfectly match the actual values. The plotted data points are closely aligned along the diagonal, indicating that the SVM model is capable of

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

accurately capturing the relationships between the input features and the target variable. This alignment suggests strong predictive performance and low error on the training data. Support Vector Regression (SVR) works by finding a hyperplane in high-dimensional space that best fits the data within a specified margin of tolerance. It is particularly effective in handling non-linear relationships when equipped with kernel functions such as radial basis function (RBF). In this context, the SVR model successfully learns from features such as Login Attempts, Average Session Duration, and Data Upload (MB) to estimate the associated Threat Risk Score. The tight clustering around the ideal line reflects the model's robustness and precision during training. This result highlights the potential of SVM-based models for cybersecurity analytics, where predicting risk levels with precision is essential for proactive threat management.

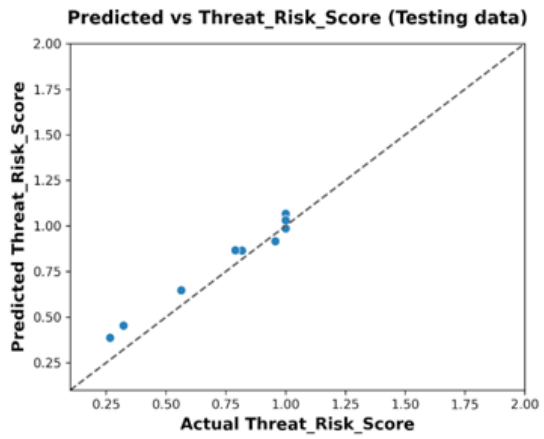


Figure 6: Predicted vs Actual Threat Risk Scores using Support Vector Regression (Testing Data)

Figure 6 depicts the performance of the Support Vector Regression (SVR) model on the testing dataset, where the predicted Threat Risk Scores are plotted against the actual scores. The diagonal dashed line serves as a benchmark for perfect predictions, where predicted values exactly match the actual values. The data points exhibit a strong linear alignment along the diagonal, signifying that the SVR model generalizes well to unseen data. This outcome reflects the model's ability to maintain predictive accuracy beyond the training phase, reducing the likelihood of overfitting. The SVR algorithm, particularly effective in handling both linear and non-linear patterns, leverages kernel functions to map the input space into higher dimensions, allowing it to capture complex relationships between features such as Login Attempts, Session Duration, and Data Upload Volume. Its margin-based loss function ensures robustness by minimizing errors within an acceptable range while ignoring minor deviations.In this scenario, the SVR model demonstrates reliable performance in estimating Threat Risk Scores for cybersecurity risk assessment tasks.

Ada Boost Regression

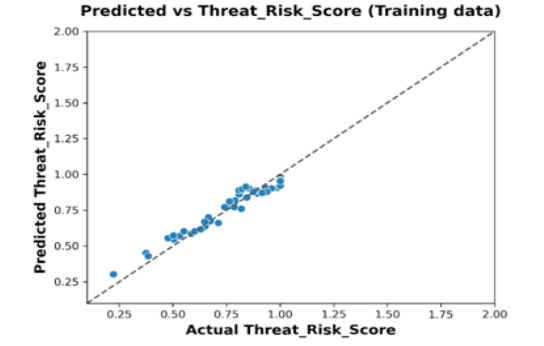


Figure 7: Predicted vs Actual Threat Risk Scores using AdaBoost Regression (Training Data)

Figure 7 displays the performance of the AdaBoost Regressor on the training dataset, illustrating the relationship between the actual and predicted Threat Risk Scores. The dashed diagonal line denotes the ideal scenario where the predicted values perfectly match the actual values. The plotted points exhibit close alignment along the diagonal, indicating that the AdaBoost model effectively captures the underlying data patterns and produces highly accurate predictions. This level of agreement reflects a strong fit on the training data, with minimal deviation between predicted and actual scores. AdaBoost (Adaptive Boosting) is an ensemble learning technique that combines multiple weak learners—typically decision trees—into a strong predictive model. By sequentially training these learners and focusing on the errors of the previous ones, AdaBoost incrementally improves its prediction performance. In the context of this model, it is used to assess the cybersecurity threat risk score by learning from key behavioral and system-level features such as access frequency, privilege misuse, and data access anomalies. The results in this plot demonstrate that the AdaBoost model is highly capable of learning complex, nonlinear relationships in the data, making it a valuable tool for threat risk evaluation.

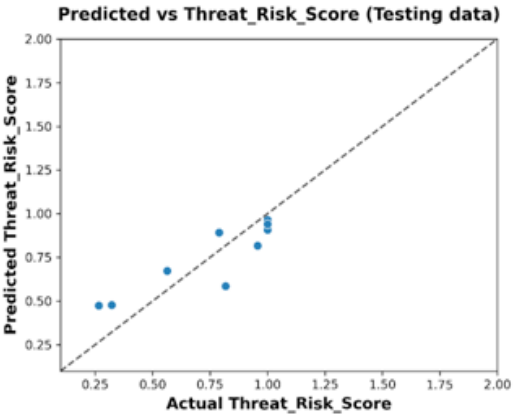


Figure 8: Predicted vs Actual Threat Risk Scores using AdaBoost Regression (Testing Data)

The algorithm works through an iterative process, in which each successive weak learner focuses on correcting errors made by previous models. Initially, all training models are given equal weights. After each iteration, the algorithm increases the weights of poorly predicted models, forcing the next weak learner to pay more attention to these difficult cases. This adaptive weighting mechanism gives AdaBoost its name and power. For regression tasks, AdaBoost typically uses decision trees as the base learners, although other algorithms can be used. The final prediction is calculated as a weighted average of the predictions of all weak learners, where the weights depend on the performance of each model. This ensemble approach often results in superior accuracy compared to individual models, because it uses the collective intelligence of multiple predictors.

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252

Table 2. Regression Model Performance Metrics (Training Data)											
	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
1	Train	RFR	Random Forest Regression	0.968449	0.968452	1.03E-03	3.21E-02	2.07E-02	1.20E-01	3.76E-04	1.31E-02
2	Train	SVR	Support Vector Regression	0.893731	0.903933	3.47E-03	5.89E-02	5.02E-02	1.00E-01	1.03E-03	4.42E-02
3	Train	ABR	AdaBoost Regression	0.921081	0.936849	2.58E-03	5.07E-02	4.52E-02	8.26E-02	7.67E-04	4.46E-02

Among the three models, the random forest regression (RFR) model showed the best overall performance. It achieved the highest coefficient of determination (R^2) value of 0.968449, indicating that it explained approximately 96.8% of the variance in the training data. The explained variance score (EVS) of 0.968452 further confirms its high explanatory power. RFR also recorded the lowest mean square error (MSE) of 1.03E-03, and consequently, the lowest root mean square error (RMSE) of 0.0321, indicating high prediction accuracy with minimal error. The mean absolute error (MAE) of 0.0207 and the mean absolute error (MedAE) of 0.0131 were the lowest among the three models. Furthermore, RFR achieved a very low maximum error (0.120) and a mean square logarithmic error (MSLE) of 3.76E-04, showing consistent accuracy across the predictions. The support vector regression (SVR) model, while still effective, lagged behind RFR. It achieved an R^2 of 0.893731 and an EVS of 0.903933, suggesting that it could explain 89.4% of the data variance. However, its MSE (3.47E-03) and RMSE (0.0589) were significantly higher than RFR, indicating larger errors. The MAE (0.0502) and maximum error (0.100) further highlighted its relatively low accuracy, although the maximum error was slightly better than RFR. The MSLE (1.03E-03) and MedAE (0.0442) show that its errors, especially for low-level predictions, are more significant than those of RFR. The AdaBoost Regression (ABR) model performed moderately well, outperforming SVR in some aspects. With an R^2 of 0.921081 and an EVS of 0.936849, ABR was better than SVR in capturing data variability. Its MSE (2.58E-03) and RMSE (0.0507) were better than SVR, but not better than RFR. The MAE (0.0452) and Max Error (0.0826) were slightly better than SVR, showing fewer large errors. In addition, its MSLE (7.67E-04) and MedAE (0.0446) indicate good consistency in prediction errors, although they lag behind RFR.

Table 3. Regression Model Performance Metrics (Testing Data)											
	Data	Symbol	Model	R2	EVS	MSE	RMSE	MAE	MaxError	MSLE	MedAE
1	Test	RFR	Random Forest Regression	0.844416	0.84865	1.16E-02	1.08E-01	8.25E-02	1.76E-01	4.84E-03	9.56E-02
2	Test	SVR	Support Vector Regression	0.92227	0.964825	5.79E-03	7.61E-02	6.71E-02	1.32E-01	2.52E-03	6.13E-02
3	Test	ABR	AdaBoost Regression	0.758987	0.759159	1.80E-02	1.34E-01	1.19E-01	2.32E-01	7.16E-03	1.07E-01

The three, Support Vector Regression (SVR) performs best on the test data. It achieves a maximum R^2 score of 0.92227, indicating that it explains approximately 92.2% of the variance in the unobserved data. It has a maximum explained variance score (EVS) of 0.964825, indicating excellent consistency in its predictions. SVR produces the lowest mean square error (MSE) of 5.79E-03 and the lowest root mean square error (RMSE) of 0.0761, reflecting excellent prediction accuracy and small deviations. In addition, its mean absolute error (MAE) of 0.0671 and mean absolute error (MedAE) of 0.0613 are both lower than the other models. Its maximum error (0.132) and mean square log error (MSLE) of 2.52E-03 further confirm the model's strong performance and error control on the test set. The best performing Random Forest Regression (RFR) on the training set is in second place on the test data. It records 0.844416 R^2 and 0.84865 EVS, indicating that it can still explain a large portion of the variance (about 84%), but not as effectively as SVR. Its MSE (1.16E-02) and RMSE (0.108) are higher than SVR, indicating slightly larger prediction errors. Similarly, the MAE (0.0825) and MedAE (0.0956) are higher, and the maximum error (0.176) shows that it occasionally makes large prediction errors. The MSLE of 4.84E-03 further confirms a modest increase in error over SVR. AdaBoost Regression (ABR) performs the weakest of the three in the test set. With an R^2 of 0.758987 and an EVS of 0.759159, it only explains about 76% of the variance. It has the highest MSE (1.80E-02) and RMSE (0.134), indicating that it makes the largest errors overall. Its MAE (0.119) and MedAE (0.107) are also significantly higher, and the maximum error (0.232) is the worst of all, indicating that it is prone to large deviations in some predictions. The MSLE (7.16E-03) is the highest, reflecting a less standard error distribution compared to the other models.

Conclusion

In this study, three major machine learning regression models—random forest regression (RFR), support vector regression (SVR), and Adabaost regression (ABR)—were evaluated based on their performance on training and test datasets. The results revealed that while RFR demonstrated the highest accuracy and lowest error metrics during training, indicating strong learning capabilities and minimal overfitting, its performance slightly decreased on the test set. In contrast, SVR showed better generalization ability, achieving superior results on most evaluation metrics on the test data, including the highest R^2 and the lowest MSE, RMSE, and MAE. This indicates that SVR is well suited for making accurate predictions on unobserved data. ABR, while showing reasonable results on the training data, performed poorly on the test estimate, indicating potential issues with generalization and model robustness. Overall, SVR emerged as a more reliable model for predictive tasks in this context, striking a strong balance between learning and generalization, while the RFR model remains a powerful alternative with high accuracy during training.

References

1. Vadisetty, Rahul, Anand Polamarasetti, Sameerkumar Prajapati, and Jinal Bhanubhai Butani. "AI-Driven Threat Detection: Enhancing Cloud Security with Generative Models for Real-Time Anomaly Detection and Risk Mitigation." Available at SSRN 5218294 (2023).

2. Naidu, P. Ramesh, V. Dankan Gowda, Shantanu Sudhir Gujar, Salman Firoz Shaikh, Saurabh Shandilya, and N. Sudhakar Reddy. "AI-Enhanced Cloud Security Framework for IoT Networks Using a Predictive Analytics Approach." In 2024 3rd International Conference for Advancement in Technology (ICONAT), pp. 1-8. IEEE, 2024.

3. Aldawsari, Hamad, and Shouket Ahmad Kouchay. "Integrating AI

and Machine Learning Algorithms in Cloud Security Frameworks for Enhanced Proactive Threat Detection and Mitigation." Journal of Emerging Threat Management (2023).

4. Ballamudi, S. "Interleaved Feature Extraction Model Bridging Multiple Techniques for Enhanced Object Identification" Journal of Artificial Intelligence and Machine Learning., 2023, vol. 1, no. 2, pp. 1-7. doi: https://doi.org/10.55124/ jbid.v1i2.253

5. Sourag, V. T., and Maria Sabastin Sagayam. "Investigating How AI and Machine Learning can be Leveraged to Enhance Cloud Security by Predicting and Preventing Cyber Threats." Frightening Future of Business Researches in Public Policy and Social Science Domains (2024): 119.

6. Andrés, Pereira, Ivanov Nikolai, and Wang Zhihao. "Real-Time AI-Based Threat Intelligence for Cloud Security Enhancement." Innovative: International Multi-disciplinary Journal of Applied Technology 3, no. 3 (2025): 36-54.

7. Shaffi, Shamnad Mohamed, Sunish Vengathattil, Jezeena Nikarthil Sidhick, and Resmi Vijayan. "AI-Driven Security in Cloud Computing: Enhancing Threat Detection, Automated Response, and Cyber Resilience." arXiv preprint arXiv:2505.03945 (2025).

8. Dachepalli. V, "OPTIMIZED CLOUD SECURITY ECC-ENHANCED HOMOMORPHIC PAILLIER RE-ENCRYPTION" International Journal of Interpreting Enigma Engineers (IJIEE)., 2024, vol. 1, no. 2, pp. 1–7. doi: https://doi.org/10.62674/ijiee.2024.v1i02.001

9. Olaoye, Godwin. "AI-Driven Intrusion Detection and Prevention Systems (IDPS) for Cloud Security." Available at SSRN 5129525 (2025).

10. Reddy, Abhilash Reddy Pabbath. "The Future Of Cloud Security: Ai-Powered Threat Intelligence And Response." International Neurourology Journal 26, no. 4 (2022): 45-52.

11. Sridhar Kakulavaram. (2024). Artificial Intelligence-Driven Frameworks for Enhanced Risk Management in Life Insurance. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 4873–4897. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/2996

12. Ali, Asad. "AI-Enhanced Cybersecurity: Leveraging Neural Networks for Proactive Threat Detection and Prevention." Asian American Research Letters Journal 1, no. 9 (2024): 1-10.

13. Yadav, Gauri. "Improving Cloud Security Using Artificial Intelligence: Challenges and Opportunities." Available at SSRN 5141130 (2025).

14. Ali, Asad. "AI-Enhanced Cybersecurity: Leveraging Neural Networks for Proactive Threat Detection and Prevention." Asian American Research Letters Journal 1, no. 9 (2024): 1-10.

15. Olutimehin, Abayomi Titilola. "Advancing cloud security in digital finance: AI-driven threat detection, cryptographic solutions, and privacy challenges." Cryptographic Solutions, and Privacy Challenges (February 13, 2025) (2025).

16. Islam, Umar, Hanif Ullah, Naveed Khan, Kashif Saleem, and Iftikhar Ahmad. "AI-enhanced intrusion detection in smart renewable energy grids: A novel industry 4.0 cyber threat management approach." International Journal of Critical Infrastructure Protection (2025): 100769.

17. Rashid, Mohanad Mohammed, and Omar Mahmood Yaseen. "AI-Driven Cybersecurity Measures for Hybrid Cloud Environments: A Framework for Multi-Cloud Security Management." International Journal on Engineering Artificial Intelligence Management, Decision Support, and Policies 2, no. 1 (2025): 30-39.

18. Islam, Umar, Hanif Ullah, Naveed Khan, Kashif Saleem, and Iftikhar Ahmad. "AI-enhanced intrusion detection in smart renewable energy grids: A novel industry 4.0 cyber threat management approach." International Journal of Critical Infrastructure Protection (2025): 100769.

19. Sridhar Kakulavaram. (2022). Life Insurance Customer Prediction and Sustainbility Analysis Using Machine Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 10(3s), 390 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7649

20. Aisha, Mohammed, Akroh Theresa Ojevwe, and Nwachukwu Chinwe Sheila. "Enhancing Cloud Security with Machine Learning-Based Anomaly Detection." American Journal of Engineering, Mechanics and Architecture 3, no. 3 (2025): 51-68.

21. Agorbia-Atta, Cedrick, Imande Atalor, and Rita Korkor Agyei and Richard Nachinaba. "Leveraging AI and ML for Next-Generation Cloud Security: Innovations in Risk-Based Access Management." World Journal of Advanced Research and Reviews 23, no. 3 (2024).

22. Nutalapati, Pavan. "Enhancing Cybersecurity with AI-Machine Learning Techniques for Anomaly Detection and Prevention."

23. Min-Jun, Lee, and Park Ji-Eun. "Cybersecurity in the Cloud Era: Addressing Ransomware Threats with AI and Advanced Security Protocols." International Journal of Trend in Scientific Research and Development 4, no. 6 (2020): 1927-1945.

24. Jim, Md Majadul Islam, and Mosa Sumaiya Khatun Munira. "The Role Of AI In Strengthening Data Privacy For Cloud Banking." Innovatech Engineering Journal 1, no. 01 (2024): 10-70937.

Citation: Peram, S. R. (2025). Cloud Security Reinvented: A Predictive Algorithm for User Behavior-Based Threat Scoring. Journal of Business Intelligence and Data Analytics, 2(3), 252. https://doi.org/10.55124/jbid.v2i3.252