Kasun Jayawardena, 2023, 11:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Machine Learning-Based Risk Scoring in Enterprise Security Frameworks

Kasun Jayawardena
University of Moratuwa, Sri Lanka

Abstract- In the contemporary digital landscape, enterprise security has transitioned from a perimeter-centric defense model to a data-driven, risk-aware paradigm. Traditional risk assessment methodologies, which often rely on static qualitative heat maps and manual vulnerability scoring, are increasingly unable to keep pace with the velocity and sophistication of modern cyber threats. This review article explores the emergence and integration of Machine Learning (ML)-based risk scoring within enterprise security frameworks. By leveraging advanced algorithms—ranging from supervised ensemble methods to unsupervised anomaly detection and deep learning—organizations can now generate dynamic, real-time risk scores for users, devices, and network entities. These scores facilitate a "Zero Trust" architecture by providing the granular intelligence necessary for automated access decisions and incident prioritization. This article categorizes current ML methodologies, including the use of Random Forests for vulnerability prioritization and Recurrent Neural Networks (RNNs) for behavioral risk modeling. We examine the critical role of feature engineering in synthesizing telemetry data from diverse sources such as Endpoint Detection and Response (EDR) systems, Identity and Access Management (IAM) logs, and Threat Intelligence feeds. Furthermore, the review addresses the challenges of model interpretability, data bias, and the necessity for Explainable AI (XAI) in security operations. By synthesizing recent academic research and industrial case studies, this paper provides a strategic roadmap for the implementation of predictive risk scoring. The findings suggest that ML-based scoring significantly reduces the "Mean Time to Respond" (MTTR) by filtering noise and highlighting high-probability threats, thereby fortifying the enterprise's overall resilience.

Keywords: Risk Scoring, Machine Learning, Enterprise Security, Predictive Analytics, Vulnerability Management.

I. INTRODUCTION

The architecture of the modern enterprise has undergone a radical transformation, moving away from centralized data centres toward highly distributed, cloud-native, and hybrid environments. This expansion has created an exponentially larger attack surface, where a single misconfiguration or compromised credential can lead to a catastrophic breach. Historically, enterprise risk management was a periodic, manual process. Security teams would conduct annual or quarterly audits, assigning static risk values to assets based on their perceived importance. However, in an era where thousands of new vulnerabilities are disclosed every month and "living-off-the-land" attacks bypass traditional signatures, these static models are fundamentally obsolete. The fundamental problem is one of scale and speed; human analysts cannot manually correlate the millions of alerts generated by security tools to determine which represent a true existential risk to the business. This gap has necessitated the move toward Machine Learning (ML)-based risk scoring, a proactive approach that utilizes mathematical models to quantify risk dynamically and continuously.

ML-based risk scoring serves as the "brain" of the modern Security Operations Center (SOC). It functions by ingesting massive streams of telemetry data from across the enterprise—network traffic, user login patterns, file integrity changes, and cloud API calls—and applying statistical models to identify patterns associated with malicious intent. Unlike traditional rule-based systems that trigger an alert based on a simple "if-then" logic, ML models can weigh hundreds of variables simultaneously to produce a normalized risk score, typically on a scale of 0 to 100 or 0.0 to 1.0. This score represents the probability that a specific entity or event is malicious or compromised. This intelligence allows security teams to move from a reactive posture to a "predictive" one. Instead of chasing every minor

alert, analysts can focus their limited resources on the "top 1%" of high-risk entities. This prioritization is critical in an environment where "alert fatigue" is a leading cause of analyst burnout and missed detections.

The integration of ML into risk frameworks also facilitates the transition to a Zero Trust Architecture (ZTA). In a Zero Trust environment, the philosophy is "never trust, always verify." However, continuous verification is only possible if there is an automated mechanism to assess trust in real-time. ML-based risk scoring provides this mechanism. By constantly recalculating the risk associated with a user's session, the system can dynamically adjust access permissions. If a user's risk score spikes because they are accessing sensitive data from an unusual location, the system can automatically trigger a Multi-Factor Authentication (MFA) challenge or terminate the session entirely. This section sets the stage for the rest of the review, outlining the evolutionary pressures that have made ML an indispensable component of enterprise security. We will explore how the fusion of big data analytics and algorithmic intelligence is not just an incremental improvement but a fundamental shift in how we define, measure, and mitigate risk in the digital age.

II. MATHEMATICAL FOUNDATIONS OF ALGORITHMIC RISK ASSESSMENT

The efficacy of an ML-based risk scoring system is fundamentally rooted in its underlying mathematical and statistical frameworks. At its core, risk scoring is often treated as either a classification problem or a regression problem. In a classification context, the model seeks to categorize an entity as "malicious," "suspicious," or "benign." In a regression context, the model predicts a continuous value that represents the intensity of the risk. To achieve this, enterprises typically employ a variety of supervised learning algorithms. Decision Trees and their ensemble counterparts, such as Random Forests and Gradient Boosted Machines (e.g., XGBoost), are highly popular due to their ability to handle non-linear relationships between features and their relative transparency compared to deep learning models. These algorithms are particularly effective at vulnerability scoring, where they can weigh factors like exploit availability, system criticality, and network exposure to predict the likelihood of a specific CVE (Common Vulnerabilities and Exposures) being targeted.

However, supervised learning requires labeled data—examples of "bad" and "good" behavior—which can be difficult to obtain in the rapidly shifting landscape of cyber-attacks. This has led to the increased adoption of Unsupervised Learning and Semi-Supervised Learning for risk scoring. Clustering algorithms, such as K-Means or Isolation Forests, are used for anomaly detection. These models establish a baseline of "normal" behavior for every user and device on the network. When an entity's behavior deviates significantly from this baseline—measured by statistical distance—the risk score is elevated. This is crucial for detecting zero-day threats or insider threats where no prior "malicious" label exists. Furthermore, Deep Learning architectures, specifically Autoencoders, have shown great promise. An Autoencoder learns to compress and then reconstruct normal network traffic; if it encounters traffic it cannot reconstruct accurately (high reconstruction error), it assigns a high risk score to that packet or session.

The "scoring" part of the process involves a normalization layer where the raw output of these algorithms is translated into a business-readable format. This often involves Bayesian inference, where prior knowledge about the threat landscape is combined with new evidence from the telemetry data to update the risk score. This section explores the technical intricacies of these models, discussing how loss functions are optimized to minimize false positives—the "kryptonite" of security ML. We also analyze the role of "Graph Theory" in risk scoring. By representing the enterprise as a graph of nodes (users, devices, apps) and edges (communications), ML models can identify "high-centrality" nodes that, if compromised, would represent the highest risk to the organization. This mathematical depth is what allows ML to transcend simple threshold-based alerting and provide a nuanced, probabilistic view of enterprise danger.

III. DATA ACQUISITION AND PREPROCESSING FOR RISK TELEMETRY

Machine learning models are only as good as the data they ingest, and in the enterprise security domain, data is often siloed, unstructured, and noisy. The first major hurdle in building a predictive risk scoring system is the construction of a robust data pipeline. Enterprise telemetry comes from a vast array of sources: Syslogs from servers, NetFlow data from routers, endpoint telemetry from EDR agents, identity data from Active Directory, and cloud-native logs from AWS CloudTrail or Azure Monitor. Each of these sources has its own format and timestamping convention. Therefore, the data must first undergo "Normalization" and "Standardization." This involves mapping diverse log entries to a common schema, such as the Open Cybersecurity Schema Framework (OCSF), to ensure that the ML model can correlate events across different layers of the technology stack.

Once the data is normalized, the process of "Feature Engineering" begins. This is perhaps the most critical step in the ML lifecycle. A raw log entry saying "User X logged in at 3:00 AM" is not very useful on its own. However, a engineered feature like "Time since last login" or "Is login time outside the 95th percentile for this user?" provides the context the model needs to calculate risk. Security experts must work with data scientists to create features that capture the "tactics, techniques, and procedures" (TTPs) of attackers. For example, to detect lateral movement, features might include the "fan-out" ratio (how many new systems a user connects to in a short window) or the "directory traversal depth." This section deep-dives into the challenges of handling "high-cardinality" data, such as IP addresses or file hashes, and how techniques like "Embeddings" (mapping categorical data to numerical vectors) are used to make this data digestible for neural networks.

Data quality and "Concept Drift" also pose significant challenges. Enterprise environments are dynamic; a new software deployment or a change in work-from-home policy can make old training data irrelevant. If the ML model is not updated, it will produce "stale" risk scores. This section explores the use of "Automated Data Labeling" and "Active Learning," where the system identifies ambiguous cases and asks human analysts for a verdict, which is then used to retrain the model. We also discuss the importance of "Data Enrichment"—pulling in external data such as IP reputation scores, GeoIP information, and threat feeds to provide the model with a "global" perspective on risk. By building a high-fidelity data foundation, enterprises ensure that their risk scores are based on factual evidence rather than statistical noise, leading to a more reliable and trustworthy security posture.

IV. BEHAVIORAL RISK MODELING OF HUMAN AND NON-HUMAN ENTITIES

The most difficult aspect of risk scoring is the human element. Attackers frequently target identities because once they have valid credentials, they can "blend in" with legitimate traffic. Behavioral risk modeling, often categorized under User and Entity Behavior Analytics (UEBA), is the ML-driven answer to this problem. Unlike vulnerability scoring, which looks at "bugs" in software, behavioral scoring looks for "anomalies" in activity. This requires the model to build a "Longitudinal Profile" for every entity. For human users, the AI monitors things like typing speed (keystroke dynamics), the specific sequence of applications they open, and the sensitivity of the data they access. For "non-human" entities, such as service accounts or IoT devices, the model monitors API call patterns and communication heartbeats.

This section explores the use of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks in behavioral modeling. These architectures are designed to process sequences of data, making them perfect for identifying a "chain of events" that indicates risk. For example, a user downloading a large file is not necessarily risky. However, if that user first clears their system logs, then runs an unusual PowerShell script, and then downloads the file, the sequence is highly indicative of data exfiltration. The ML model assigns a cumulative risk score that increases with every suspicious step in the chain. This allow the system to catch "low and slow" attacks that would bypass a simple anomaly detector.

Furthermore, we analyze "Peer Group Analysis" as a method for refining behavioral scores. If an entire marketing team suddenly starts using a new cloud collaboration tool, the risk score for an individual marketing employee doing so should remain low. However, if a developer starts using that same tool, their risk score should increase. This contextual awareness is achieved by clustering users into dynamic peer groups based on their actual behavior rather than just their static job title in Active Directory. This section emphasizes that behavioral risk scoring is about "intent." By quantifying deviations from a dynamic baseline, ML allows the enterprise to identify compromised accounts and malicious insiders before they can execute their final objective. The goal is to move away from "binary" security (allow/deny) toward a "nuanced" security that responds to the subtle shifts in entity behavior.

V. VULNERABILITY PRIORITIZATION AND ASSET CRITICALITY SCORING

Enterprises are currently facing a "vulnerability overload." On any given day, a large organization may have hundreds of thousands of unpatched vulnerabilities across its infrastructure. Traditional prioritization, based on the Common Vulnerability Scoring System (CVSS), is often misleading because it measures the "severity" of a bug in a vacuum, not the "risk" to the specific enterprise. A "Critical" CVSS score on a disconnected lab machine is less of a risk than a "Medium" score on a public-facing web server. ML-based risk scoring solves this by creating an "Enterprise-Specific Risk Score" for every asset and vulnerability. This involves a hybrid model that combines the technical severity of the bug with the "business criticality" of the asset and the "threat intensity" in the wild.

This section examines the use of supervised learning models to predict the "Likelihood of Exploitation." By training on historical data from the "Exploit Prediction Scoring System" (EPSS) and real-time "Threat Intelligence," ML models can identify which vulnerabilities are actually being exploited by threat actors in the current landscape. If a vulnerability is being used in an active ransomware campaign, its risk score is immediately escalated. Simultaneously, the model ingests data from the enterprise's "Configuration Management Database" (CMDB) to determine the asset's importance. Does the machine hold PII? Is it part of a critical financial transaction chain? The final score is a product of these factors: (Threat Probability) x (Technical Severity) x (Asset Impact).

We also discuss the role of "Reachability Analysis" in risk scoring. Modern ML models can ingest network topology data to determine if a vulnerable service is actually reachable from the internet. If the vulnerability is "Deep" in the network and protected by multiple layers of firewalls, its risk score is lowered. This allows the IT and security teams to focus on the "Top 100" vulnerabilities that represent the most immediate danger. This section highlights how ML-based prioritization transforms vulnerability management from a "game of whack-a-mole" into a strategic exercise in risk reduction. By focusing on the "Critical Path" of an attacker, enterprises can achieve a higher level of security with fewer patches, significantly reducing the operational burden on IT staff and minimizing system downtime.

Integrating Risk Scores into Zero Trust Architectures Zero Trust is the dominant security philosophy of the decade, but it is often misunderstood as a product rather than a process. At its core, Zero Trust requires a "Policy Decision Point" (PDP) that can make real-time access decisions based on the current risk level. This is where ML-based risk scoring becomes the "engine" of Zero Trust. In a ZTA, the risk score of the user, the device, and the requested resource are combined into a "Transaction Risk Score." If this score exceeds a certain threshold defined by the organization's policy, the request is denied or "stepped up." This section explores how these scores are integrated into identity providers (IdPs) and secure access service edge (SASE) platforms to enable "Adaptive Access Control."

The beauty of ML-driven Zero Trust is its fluidity. Traditional access control is "Binary and Persistent"—once you log in, you are "trusted" for the next 8 hours. In contrast, ML-driven access is "Continuous and Granular." The risk score is recalculated every time a new action is taken. This

section analyzes the "Trust Decay" model, where the confidence in a user's identity naturally decreases over time or with every new network jump. To maintain access, the user's behavior must continue to align with the "low-risk" profile. We also discuss "Automated Remediation" triggered by risk scores. For example, if a device's risk score spikes because it is missing a critical security patch, the ZTA can automatically move that device to a "Quarantine VLAN" until it is remediated.

This integration also extends to "Least Privilege Enforcement." By analyzing the historical risk scores and access patterns of a user, an ML model can suggest that certain permissions be revoked because they are never used and only serve to increase the "blast radius" of a potential compromise. This section emphasizes that Zero Trust cannot function without the "Intelligence" provided by ML risk scoring. Without it, Zero Trust is just a set of static, manual rules that are too rigid for a modern business. With ML, Zero Trust becomes a dynamic, invisible shield that protects the organization without hindering the productivity of legitimate users. We conclude this section by looking at the "Feedback Loop"—how the results of Zero Trust decisions (e.g., a blocked login that was actually legitimate) are used to retrain the ML model to improve its future scoring accuracy.

VI. EXPLAINABILITY AND TRANSPARENCY IN SECURITY AI

One of the primary obstacles to the adoption of ML in enterprise security is the "Black Box" problem. If an ML model assigns a high risk score to a senior executive and blocks their access to a board meeting, the security team must be able to explain why. In a high-stakes environment, "The AI said so" is not an acceptable answer. This has led to the rise of "Explainable AI" (XAI) within risk scoring frameworks. XAI aims to make the internal logic of complex models transparent and interpretable for human analysts. This section explores the various techniques used to achieve this, such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations).

These XAI tools work by identifying which specific "features" contributed most to a particular risk score. For instance, an analyst looking at a high-risk alert could see a breakdown: "40% due to unusual login location, 30% due to spike in outbound data, 20% due to use of administrative tools." This level of transparency is vital for "Analyst Trust." If the AI provides a reasoning path, the analyst can quickly verify its findings and take action. If the reasoning is flawed—perhaps the "unusual location" is just the executive on a business trip—the analyst can "white-list" the event and lower the score. This section analyzes the trade-off between "Model Complexity" and "Interpretability." While a deep neural network might be 2% more accurate than a Random Forest, the Random Forest is much easier to explain. For many enterprises, the ability to explain a decision is more valuable than a marginal gain in accuracy.

We also discuss the "Regulatory and Compliance" aspect of XAI. Frameworks like GDPR in Europe grant individuals a "right to an explanation" for automated decisions that affect them. If an ML-based risk score is used to deny an employee access to a resource, the company may be legally required to provide the reasoning. This section highlights how XAI serves as a bridge between the mathematical world of data science and the operational world of cybersecurity. By making risk scores "Human-Readable," enterprises can integrate AI into their workflows without losing control. Transparency also helps in identifying "Model Bias." If the AI is consistently assigning high risk scores to certain departments or devices due to flawed training data, XAI will make this pattern visible, allowing the data science team to correct the bias and ensure a "Fair and Accurate" scoring system.

VII. ADVERSARIAL MACHINE LEARNING AND MODEL ROBUSTNESS

As enterprises weaponize ML for defense, threat actors are weaponizing it for offense. The field of "Adversarial Machine Learning" (AML) focuses on how attackers can "fool" or "bypass" ML models. In the context of risk scoring, an attacker might use "Evasion Attacks" to subtly alter their behavior so that it always stays just below the detection

threshold. For example, if they know the ML model flags any outbound data transfer over 500MB, they will exfiltrate 400MB of data every day for a week. They might also use "Poisoning Attacks," where they inject malicious data into the model's training set to "teach" it that their specific type of attack is "normal." This section explores the various methods used to build "Robust" risk models that can withstand adversarial pressure. This includes "Adversarial Training," where the model is intentionally exposed to "poisoned" or "adversarial" examples during its training phase so it learns to recognize them. We also discuss "Gradient Masking" and "Defensive Distillation" as techniques to make it harder for an attacker to "reverse-engineer" the model's decision boundaries. However, the section emphasizes that there is no "Silver Bullet" for AML. It is a continuous "Arms Race" between the defender and the attacker.

We also analyze the risk of "Model Inversion," where an attacker probes the risk scoring API to figure out what data is being used for training, potentially revealing sensitive information about the enterprise's infrastructure. To counter this, enterprises are adopting "Differential Privacy" and "Secure Multi-Party Computation." This section highlights that "Security for AI" is just as important as "AI for Security." A risk scoring system that can be easily bypassed is worse than no system at all because it provides a "False Sense of Security." Enterprises must adopt a "Red Teaming" approach to their ML models—constantly attacking their own algorithms to find the blind spots. This section concludes that the most resilient risk scoring frameworks are those that use a "Diversity of Models"—combining different algorithms that look at different data—so that an attacker who bypasses one model is still caught by another.

VIII. OPERATIONALIZING RISK SCORES IN THE SOC

The final challenge of ML-based risk scoring is not technical, but operational. A high-accuracy risk score is useless if it is not integrated into the daily workflow of the Security Operations Center (SOC). "Operationalization" involves taking the output of the ML model and turning it into a "Ticket" or an "Action." This requires a tight integration between the risk scoring engine and the "Security Orchestration, Automation, and Response" (SOAR) platform. This section explores how enterprises are using "Risk-Based Alerting" (RBA) to transform their SOC operations. Instead of receiving 10,000 individual alerts, analysts receive a "Narrative" centered around a high-risk entity.

We discuss the concept of "Automated Triage." For entities with a "Low-to-Medium" risk score, the SOAR platform can perform automated investigative steps—such as scanning the device for malware or checking the user's recent emails—and only escalate to a human if the risk is confirmed. For "High-Risk" entities, the system can trigger an immediate "Containment" action, such as disabling the user's account or isolating the host. This section analyzes the "Human-in-the-Loop" (HITL) model, where the AI provides the "Evidence" and the "Score," but the final decision to "Pull the Plug" on a critical system remains with a human analyst. This ensures that the enterprise maintains a balance between "Automation Speed" and "Human Judgment."

Furthermore, we explore the use of "Risk Dashboards" for executive reporting. ML-based scores allow CISO (Chief Information Security Officers) to present a "Temperature Map" of the organization's risk to the Board of Directors. Instead of talking about "Firewall Blocks," they can talk about "Aggregate Risk Reduction" over time. This section highlights how ML-based scoring "Democratizes" security data, making it understandable for non-technical stakeholders. We also look at the "Skill Gap" in the SOC. Operationalizing ML requires analysts who understand both "Security" and "Data Science." Enterprises are investing in training programs to create a new generation of "Cyber Data Analysts" who can interpret risk scores and tune the models. This section concludes that the successful SOC of the future will be "Algorithmic," where ML does the heavy lifting of data correlation, allowing humans to focus on the high-value tasks of threat hunting and strategic defense.

IX. CONCLUSION

Machine Learning-based risk scoring represents a fundamental evolution in enterprise security, transforming it from a reactive, manual discipline into a proactive, intelligent science. By dynamically quantifying the risk of users, devices, and vulnerabilities, ML allows organizations to navigate the complexities of modern distributed environments with unprecedented precision. As this review has demonstrated, the power of ML-based scoring lies in its ability to synthesize massive telemetry streams, identify subtle behavioral anomalies, and prioritize responses based on actual business impact. However, the path to "Algorithmic Security" is not without its obstacles.

It requires a robust data foundation, a commitment to explainability, and a vigilant defense against adversarial attacks. The move toward Zero Trust architectures further amplifies the necessity of ML, as continuous trust verification is impossible without automated risk assessment. As threat actors continue to adopt AI to automate their attacks, the enterprise's ability to generate accurate, real-time risk scores will be the defining factor in its resilience. Ultimately, ML-based risk scoring is about "clarity in the noise." It provides the situational awareness needed to protect the organization's most critical assets in an increasingly volatile digital world, ensuring that security remains an enabler of business innovation rather than a bottleneck.

REFERENCES

1. Burramukku, N. R. (2015). Real-time detection of network threats using deep packet inspection and telemetry analytics. International Journal of Trend in Research and Development, 2(1), 1–5.
2. Jangala, V. K. (2015). Observability and monitoring of microservices using Splunk and New Relic. International Journal of Engineering Development and Research, 3(3), 1–15.
3. Vangoor, V. K. R. (2016). AI-driven monitoring and alerting systems for enterprise-scale Linux deployments. International Journal of Science, Engineering and Technology, 4(1), 11.
4. Parimi, S. S. (2016). Analyzing the effectiveness of SAP systems in streamlining healthcare supply chains, reducing costs, and improving service delivery.
5. Koukuntla, S. (2018). Event-driven architectures in cloud computing: Tools, patterns, and tradeoffs. International Journal of Trend in Scientific Research and Development, 2(3), 2909–2913.
6. Burramukku, N. R. (2015). Root cause analysis in enterprise networks using correlated telemetry and graph analytics. TIJER – International Research Journal, 2(6), a9–a17.
7. Jangala, V. K. (2016). API gateway security implementation using JWT and Apigee in cloud-native applications. International Journal of Current Science, 6(2), 34–43.
8. Vangoor, V. K. R. (2017). Self-optimizing DevOps pipelines for enterprise infrastructure using machine learning models. International Journal of Trend in Scientific Research and Development, 1(6), 8.
9. Parimi, S. S. R. (2016). Predictive analytics for financial forecasting in SAP ERP systems using machine learning. International Journal of Creative Research Thoughts.
10. Burramukku, N. R. (2016). Secure identity and access management integration for cloud-native network observability platforms. International Journal of Engineering Development and Research.
11. Jangala, V. K. (2018). Database performance tuning strategies for high-volume transaction systems. International Journal of Scientific Development and Research, 3(8), 274–282.
12. Vangoor, V. K. R. (2018). AI-based optimization of automated server deployment using Kickstart and Satellite systems. International Journal of Trend in Research and Development, 5(6), 5.
13. Parimi, S. S. (2018). Exploring the role of SAP in supporting telemedicine services, including scheduling, patient data management, and billing. SSRN Electronic Journal.
14. Burramukku, N. R. (2016). Secure storage and backup architectures for cloud integrated datacenters. International Journal of Science, Engineering and Technology, 4(3).

15. Burramukku, N. R. (2017). End-to-end SD-WAN performance evaluation across private and public transport networks. International Journal of Current Science, 7(1), 56–65.
16. Burramukku, N. R. (2017). Identity-aware network segmentation using NSX and next-generation firewalls. International Journal of Scientific Research & Engineering Trends, 3(5).
17. Parimi, S. S. (2018). Optimizing financial reporting and compliance in SAP with machine learning techniques. SSRN Electronic Journal.
18. Burramukku, N. R. (2018). Evaluating high-availability DHCP architectures: Migration from legacy Linux DHCP to Infoblox grid. International Journal of Scientific Development and Research.