Journal of Information Security Research

DLINE JOURNALS

Print ISSN: 0976-4143
Online ISSN: 0976-4151

JISR 2026: 17 (2)
https://doi.org/10.6025/jisr/2026/17/2/57-75

Statistical and Predictive Evaluation of Cybersecurity Threat Dynamics and System Response Latency

Martin Lopez Nores
University of Vigo
Spain
mlnores@det.uvigo.es

ABSTRACT

The rapid evolution of cyber threats has exposed the limitations of traditional, reactive cybersecurity frameworks, necessitating data driven and adaptive defense mechanisms. This study investigates the dynamics between threat characteristics and system response latency through a comprehensive statistical and predictive analysis of 1,000 cybersecurity incident records. Employing chi-square tests, correlation analyses, effect size measurements, and Ridge regression modeling, the research evaluates how variables such as threat severity, access level, and detection status influence breach occurrence and response time. Results reveal that threat severity and privilege levels exhibit negligible associations with breach likelihood and response latency, challenging conventional risk prioritization assumptions. Instead, threat detection status emerges as the dominant predictor of response time ($n^2 \approx 0.76$), indicating that current systems operate primarily on reactive, event triggered paradigms rather than proactive, severity aware mechanisms. Furthermore, the analysis identifies critical data quality challenges, including zero inflated response time distributions and timestamp induced feature leakage, which artificially inflate predictive metrics and compromise model generalizability. These findings underscore a significant gap between theoretical cybersecurity frameworks and practical system behavior. The study concludes that enhancing threat-detection accuracy, implementing rigorous data-preprocessing protocols, and integrating meaningful temporal and contextual features are essential for developing robust, proactive cybersecurity architectures. Future research should prioritize real world dataset validation and advanced machine learning techniques to overcome current analytical constraints.

Keywords: Cybersecurity Threat Dynamics, System Response Latency, Predictive Analytics, Statistical Association Analysis, Real-Time Threat Detection, Data-Driven Frameworks, Machine Learning Modelling, Data Quality Assessment

Received: 10 October 2025, Revised 5 January 2026, Accepted 19 January 2026

1. Introduction

The rapid advancement of digital technologies has significantly transformed the cybersecurity landscape, introducing both unprecedented opportunities and complex security challenges. As cyber threats continue to evolve in scale, frequency, and sophistication, the need for real-time threat analysis and adaptive defense mechanisms has become increasingly critical. Traditional cybersecurity frameworks often struggle to cope with this dynamic environment, as they are largely based on static assumptions and fail to adequately address the ever-evolving nature of cyber threats [1]

Over recent decades, the proliferation of interconnected systems, including the Internet of Things (IoT), cloud computing infrastructures, and artificial intelligence (AI)-driven applications, has expanded the attack surface for malicious actors. Consequently, modern cybersecurity threats encompass a wide spectrum, including malware, ransomware, phishing attacks, and advanced persistent threats (APT) [2] These developments have heightened the complexity of threat detection and response, rendering conventional approaches increasingly inadequate [3].

Traditional security mechanisms predominantly rely on signature-based detection systems and reactive response strategies, which are inherently limited in their ability to identify novel or evolving threats [4]. Attackers continuously refine their techniques to evade detection, thereby necessitating the adoption of proactive and adaptive security frameworks. As highlighted by Stallings and Brown [5]. organizations face substantial challenges in maintaining effective cybersecurity defenses while ensuring the protection of sensitive data in an ever-changing threat landscape.

2. Background

2.1 Evolution of Cybersecurity Risk Assessment

Cybersecurity risk assessment has undergone a significant paradigm shift, moving from reactive, remedial approaches to proactive, predictive strategies. This transformation is largely driven by the increasing availability of large-scale data and advancements in analytical techniques.

Predictive analytics has emerged as a key enabler in this transition, leveraging statistical modeling and machine learning techniques to anticipate potential threats before they materialize [6]. By enabling early identification of vulnerabilities and attack patterns, predictive approaches enhance both the timeliness and effectiveness of cybersecurity responses, while also improving resource allocation and decision-making processes [7].

Furthermore, predictive analytics has the potential to fundamentally transform cybersecurity into a proactive discipline, capable of real-time threat anticipation and mitigation. However, its implementation requires substantial investment in data infrastructure, model development, training, and continuous adaptation to accommodate emerging threats [8]. These challenges highlight the need for scalable and flexible analytical frameworks capable of evolving alongside the threat landscape.

2.2 Role of Analytics in Cybersecurity

The increasing reliance on data-driven decision-making has led organizations to adopt advanced analytics capabilities as a means of achieving both operational efficiency and competitive advantage [9, 10, 11]. Analytics,

defined as the systematic computational analysis of data [12], plays a crucial role in extracting actionable insights from complex cybersecurity datasets.

Analytics capability is broadly categorized into three types:

1. Descriptive Analytics – focuses on historical data analysis

2. Predictive Analytics – forecasts future events and risks

3. Prescriptive Analytics – recommends optimal decision strategies

These categories collectively support both retrospective analysis and forward-looking decision-making [13, 14]. However, traditional analytics systems are often affected by latency across data, analysis, and decision-making because data must be stored and processed before insights are generated [15].

Such latency limitations render conventional analytics unsuitable for applications requiring immediate response, such as fraud detection, situational awareness, and cybersecurity threat mitigation. To address these challenges, researchers have introduced the concept of real-time analytics, which enables continuous monitoring and analysis of events as they occur [15, 16].

2.3 Real-Time Threat Detection and Intelligent Systems

Real-time cybersecurity systems leverage advanced technologies such as artificial intelligence, machine learning, and big data analytics to continuously monitor network activity, detect anomalies, and respond to threats in real time. These systems are particularly effective in identifying patterns and outliers within large-scale datasets, which often indicate malicious behavior [17].

AI-driven approaches address the limitations of traditional detection systems by enabling adaptive learning and dynamic response mechanisms. For instance, the system proposed by Khalaf [18] utilizes advanced machine learning algorithms to improve detection accuracy, adaptability, and operational efficiency. Such systems are capable of identifying previously unseen threats, thereby enhancing overall cybersecurity resilience.

In addition, dynamic models such as ThreatResponder, proposed by Zhu, [19] employ instantaneous-state Markov processes to represent attack behaviours and defence strategies. By leveraging real-time data, these models can dynamically adjust system responses and optimize the handling of concurrent attacks.

Similarly, simulation-based studies, such as those conducted by Omamo, [20] emphasize the importance of adaptive policies, real-time response mechanisms, and proactive resource allocation in addressing critical cybersecurity challenges, including malware evolution, phishing attacks, and zero-day vulnerabilities.

2.4 Data-Driven Cybersecurity Frameworks

The integration of data-driven methodologies has significantly enhanced the capability of cybersecurity systems to address complex and dynamic threats. These approaches are applicable across a wide range of domains, including IoT environments, financial systems, and business analytics, where large-scale data analysis is essential for informed decision-making [21, 22, 23, 24].

A comprehensive data-driven cybersecurity framework typically combines:

- Predictive classification models for threat detection
- Association rule mining techniques for uncovering hidden patterns and relationships

Such hybrid approaches are recognized as critical components of modern artificial intelligence systems, enabling both accurate prediction and interpretable insights [25, 26].

By integrating predictive and associative learning techniques, these frameworks facilitate the development of robust, scalable, and intelligent cybersecurity solutions that address both known and emerging threats.

2.5 Research Motivation and Contribution

Despite significant advancements in cybersecurity analytics, several challenges remain unresolved. [27]

Traditional systems continue to struggle with:

- Limited adaptability to evolving threats
- High latency in decision-making processes
- Inability to effectively integrate real-time data streams

This study is motivated by the need to address these limitations by developing a data-driven, real-time analytical framework for cybersecurity threat analysis.

The primary contributions of this work include:

1. Integration of statistical and machine learning techniques for comprehensive threat analysis

2. Evaluation of relationships among key cybersecurity variables, including threat characteristics and response behavior

3. Identification of data quality issues and modeling limitations, such as latency and feature-related biases

4. Provision of insights into the role of real-time analytics in enhancing cybersecurity resilience

By bridging the gap between traditional analytical approaches and modern real-time systems, this research contributes to the development of next-generation cybersecurity frameworks that operate effectively in dynamic, complex environments.

Despite the growing adoption of real-time and data driven cybersecurity systems, there remains a limited empirical understanding of how key threat attributes such as severity, detection status, and access level interact with system response behaviour. In particular, the extent to which response latency is influenced by threat characteristics remains underexplored.

To address this gap, this study is guided by the following research questions:

- RQ1: Does threat severity significantly influence system response time and breach occurrence?
- RQ2: What are the key predictors of response latency in cybersecurity systems?
- RQ3: To what extent do data characteristics affect statistical inference and predictive model performance?

These questions provide a structured foundation for the subsequent analytical framework.

3. Dataset Overview

The dataset employed in this study comprises a structured collection of 1,000 row-level cybersecurity incident records, each characterized by 14 attributes. Each observation corresponds to a distinct event associated with an individual entity identified through unique identifiers, namely *Student_ID* and *Student_Name*. These identifiers are unique across all records, ensuring a one-to-one mapping between rows and entities.

The dataset integrates a heterogeneous mix of variable types, including categorical descriptors, binary indicators, and a response-time attribute. Categorical variables such as *Threat_Type*, *Threat_Severity*, and *Access_Level* capture qualitative aspects of cybersecurity events, while multiple binary indicators—including *Threat_Detected*, *Data_Breach*, *Insider_Threat*, *Phishing_Attempt*, and *Malware_Detected*—encode the presence or absence of specific threat conditions. Additionally, the response-time variable is originally stored as a textual field with millisecond annotations but can be systematically transformed into a continuous numerical representation for quantitative analysis.

A notable strength of the dataset is its completeness, as no missing values are observed across any attribute. While this characteristic facilitates seamless statistical modeling without the need for imputation, it also suggests that the dataset may be synthetically generated or carefully curated, rather than derived from raw operational logs that typically exhibit noise and incompleteness.

In terms of distributional properties, key categorical variables demonstrate a relatively balanced structure. Threat severity is distributed across three levels—low, medium, and high—in near-equal proportions, ensuring that no single category disproportionately influences the analysis. Similarly, the distribution of threat types spans five categories, including insider threats, phishing, unauthorized access, data breaches, and malware or spyware, each contributing comparably to the overall dataset. Access levels are also evenly represented across user, administrator, and guest roles, further supporting unbiased comparative analysis.

In contrast, the binary threat indicators exhibit moderate class imbalance, with the majority of observations corresponding to the absence of specific threat conditions. Positive class proportions for variables such as data breaches, insider threats, phishing attempts, and malware detection typically range between approximately 18% and 22%. Although this level of imbalance is not extreme, it remains an important consideration for both statistical inference and predictive modeling.

The response-time variable displays distinctive distributional characteristics that warrant careful interpretation. After conversion to a numeric format, response times range from 0 to 499 milliseconds, with a mean of approximately 150.6 ms and a median of 103.5 ms. However, a substantial proportion of observations at least 25% exhibit a response time of zero milliseconds, resulting in a pronounced zero-inflated distribution.

This concentration of zero values is atypical in real-world systems and suggests potential data artifacts, such as default logging values or measurement inconsistencies. Such behavior has important implications for subsequent analyses, particularly in explaining the absence of meaningful correlations between response time and other variables.

Preliminary structural analysis further reveals notable patterns of dependence and independence among variables. While several relationships appear weak or statistically insignificant, a strong association is observed between phishing attempts and malware detection, with a zero-frequency cell in their contingency table. This anomaly indicates the possibility of underlying constraints in data generation or labeling processes rather than naturally occurring system behavior.

Overall, the dataset provides a controlled and well-structured environment for examining cybersecurity threat dynamics, system response behavior, and breach outcomes. Its balanced categorical distributions and complete records support robust analytical exploration, while its structural peculiarities such as zero-inflated response times and potential labeling constraints highlight important considerations regarding data realism and interpretability.

3.2 Data Preprocessing and Feature Engineering

To ensure compatibility with statistical testing and predictive modeling frameworks, the raw dataset underwent systematic preprocessing and feature engineering. Categorical and textual variables were transformed into structured numerical representations through a targeted encoding strategy. Threat severity was mapped to an ordinal scale (1 = Low, 2 = Medium, 3 = High, 4 = Critical) to preserve its inherent hierarchy, while access privileges were similarly ordinalized to reflect escalating authorization levels from guest to user and administrator roles. Key binary threat indicators, including data breaches, insider threats, phishing attempts, and malware detections, were recoded as dichotomous variables (0 or 1) to denote absence or presence. Additionally, response time metrics, originally recorded as formatted strings (e.g., 406 ms), were parsed and converted to continuous numeric values. This transformation pipeline standardized the feature space, ensuring that all variables met the mathematical assumptions required for subsequent quantitative analysis and machine learning algorithms.

These dataset characteristics directly influence the choice of analytical methods. Specifically, the dominance of categorical variables necessitates the use of non-parametric statistical techniques, while distributional irregularities such as zero-inflated response times require careful interpretation of both statistical and predictive modeling outcomes.

4. Analysis

Following feature engineering, a correlation matrix was computed to evaluate potential linear and monotonic relationships among the encoded variables. The analysis revealed negligible interdependencies among most feature pairs. Most notably, the association between threat severity and system response time yielded Pearson and Spearman correlation coefficients of approximately “0.0185 and “0.0199, respectively, indicating no meaningful linear or rank-order dependency. Across the full feature set, the absence of strong linear correlations suggests that the engineered variables operate largely independently in terms of direct numerical association. Given the predominantly categorical nature of the underlying data, traditional correlation metrics were

recognized as providing only preliminary, approximate insights into variable relationships. Consequently, these findings underscore the necessity of employing dedicated categorical association tests, which offer more statistically robust and interpretable measures of dependence for this type of cybersecurity dataset.

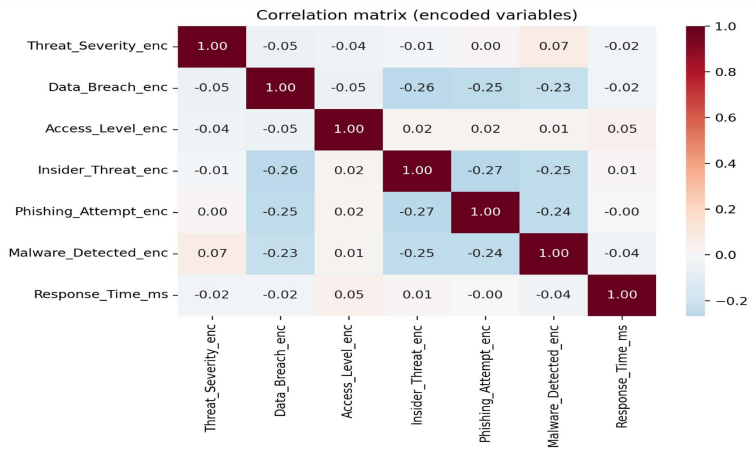

Figure 1. Correlation Matrix (Encoded Variables)

4.1 Categorical Association Analysis

To evaluate statistical dependence among categorical variables, Chi-square tests of independence were conducted, complemented by Cramer's V for effect size estimation. This methodological combination provides a robust framework for assessing non-linear associations that standard correlation metrics cannot adequately capture.

The relationship between threat severity and data breach occurrence was first examined. The analysis yielded a Chi-square statistic of 2.4298 with 2 degrees of freedom, resulting in a non-significant *p*-value of 0.2967. The corresponding Cramer's V of 0.0493 indicates a negligible effect size, suggesting that threat severity alone is not a reliable predictor of breach likelihood. This conclusion is further supported by the underlying contingency distribution, which shows relatively balanced counts of breaches and non-breaches across severity levels: high-severity events comprised 53 breaches and 263 non-breaches; medium-severity events comprised 65 breaches and 271 non-breaches; and low-severity events comprised 75 breaches and 273 non-breaches.

Similarly, the association between access level and insider threat occurrence was evaluated. The test produced a Chi-square value of 4.0136 (df = 2) with a *p*-value of 0.1344, which was not statistically significant. The calculated Cramer's V of 0.0634 indicates a weak association, suggesting that insider threats in this dataset are not strongly contingent on user privilege levels.

In contrast, the analysis of phishing attempts versus malware detection revealed a highly significant relationship. The Chi-square statistic was 55.4956 (df = 1), with an extremely small *p*-value of 9.37×10^{-14} . The Cramer's V of 0.2356 denotes a moderate effect size, underscoring a substantial dependency between these two variables. The contingency distribution illustrates this pattern clearly: in the absence of phishing attempts, 614 instances showed no malware detection, while 182 showed malware detection; conversely, when a phishing attempt was recorded, 204 cases showed no malware detection, and 0 showed malware detection. The presence of this zero-frequency cell for concurrent phishing attempts and malware detections warrants critical examination. Rather than reflecting a natural cybersecurity phenomenon, this structural absence likely stems

from constraints in dataset generation or from specific logging and system artefacts. Such anomalies raise important concerns about the realism of the data and the validity of statistical independence assumptions, suggesting that the observed strong dependence may be an artefact of the data collection methodology rather than a genuine operational relationship.

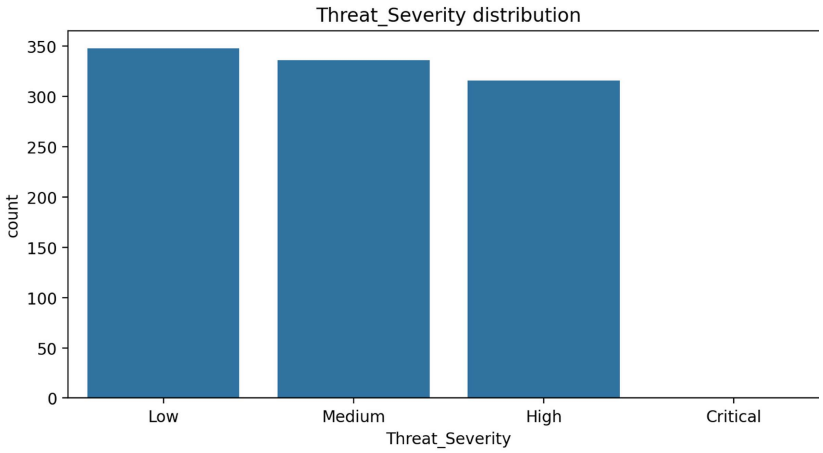

Figure 2. Threat Severity Distribution

Contingency Table:

Severity	Breach	No Breach
High	53	263
Medium	65	271
Low	75	273

Contingency Table:

Phishing	Malware = 0	Malware = 1
0	614	182
1	204	0

5. Response Time Analysis

While the previous section established the absence of strong associations among key categorical variables, it is equally important to examine whether system performance measured through response time is influenced by these factors. Accordingly, the analysis now shifts from inter variable relationships to system behavior evaluation, with a specific focus on response latency.

5.1 Relationship Between Response Time and Threat Severity

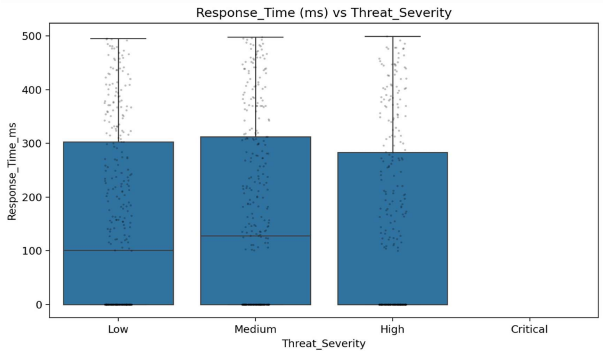

Figure 3. Response Time vs Threat Severity

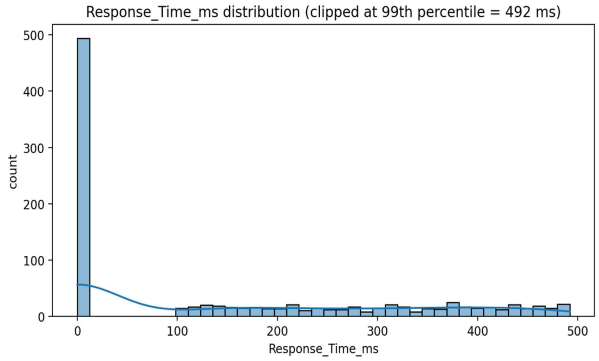

Figure 4. Response Time Distribution

The relationship between system response time and threat severity was examined using both Pearson (linear) and Spearman (rank-based) correlation coefficients. The computed values were:

- Pearson correlation: - 0.0185
- Spearman correlation: - 0.0199

Both coefficients are extremely close to zero, indicating no statistically meaningful relationship between threat severity and response time. This absence of correlation suggests that system response latency is neither linearly nor monotonically influenced by the severity classification of threats.

From an operational standpoint, this finding is significant. In a well calibrated cybersecurity system, higher-severity threats would typically trigger faster or prioritized responses. However, the observed independence implies that response mechanisms are not severity-aware, or that severity labeling does not directly influence automated response pipelines.

The corresponding visualizations (Figures 3 and 4) further reinforce this conclusion, showing overlapping distributions of response times across severity categories without discernible trends or gradients.

5.2 Distributional Characteristics and Data Quality Concerns

A deeper examination of the response time distribution reveals a critical anomaly:

• Median response time for high severity threats = 0 ms

This result is counterintuitive and raises important concerns regarding data validity and measurement reliability. Two plausible explanations emerge:

1. Zero-Inflated Distribution:

A substantial proportion of observations have zero response times, resulting in a highly skewed distribution. Such zero inflation can distort statistical measures, particularly measures of central tendency like the median.

1. Data Collection or Logging Issues:

2. The presence of zero response times for high-severity threats may indicate:

o Missing or improperly recorded timestamps

o Default placeholder values

o System logging inconsistencies

These issues can significantly mask underlying behavioral patterns and reduce the interpretability of statistical analyses. Consequently, any inference regarding response efficiency must be treated with caution unless the data is cleaned or validated.

Overall, the findings from this section emphasize that response time behavior is not governed by threat severity, and that data quality limitations may obscure meaningful system dynamics.

6. Predictors of Response Time

6.1 Dominant Predictor: Threat Detection Status

To identify the primary drivers of response time, effect size analysis was conducted using eta-squared (n^2). The results indicate that:

• Threat_Detected variable: $n^2 \approx 0.76$

This represents a strong effect size, establishing threat detection status as the most influential predictor of response time.

Interpretation

This finding suggests that system responses are fundamentally event-driven rather than anticipatory. In other words:

• Response mechanisms are triggered only after a threat is detected

• There is limited evidence of proactive or pre-emptive response behavior

This reactive architecture implies that detection acts as a gatekeeping variable, determining whether a response is initiated at all. Consequently, improving detection accuracy and latency may yield more substantial performance gains than refining severity classification.

6.2 Timestamp and Data Leakage

A critical issue identified in the modeling process is the presence of data leakage arising from timestamp encoding:

- Eta-squared for timestamp: ≈ 0.99
- Unique values: 988 out of 1000 observations

The near-perfect explanatory power of the timestamp variable is not indicative of genuine predictive strength, but rather a consequence of high cardinality and near-unique identifiers introduced through one-hot encoding.

Implications of Leakage

This encoding strategy leads to:

- Model memorization instead of generalization
- Artificially inflated performance metrics
- Loss of interpretability and real-world applicability

Such leakage undermines the validity of the predictive model, as it captures dataset-specific patterns rather than generalizable system behavior.

Recommended Mitigation

To address this issue, raw timestamps should be excluded or transformed into meaningful temporal features, such as:

- Hour of day
- Day of week
- Weekend vs weekday indicator

These transformations preserve temporal structure while avoiding high-dimensional sparsity and leakage.

6.3 Secondary Predictors

Beyond threat detection status, other variables exhibit minimal explanatory power in predicting response time. Features such as threat severity, access level, and threat type demonstrate negligible effect sizes, reinforcing earlier findings from correlation and Chi-square analyses.

This consistency across statistical and predictive analyses indicates that traditional categorical descriptors

do not significantly influence system response behavior. Instead, the system appears to operate based on event trigger mechanisms rather than contextual threat attributes.

These findings highlight a critical limitation in current cybersecurity frameworks, where response strategies are not sufficiently informed by threat characteristics, potentially limiting prioritization efficiency in high risk scenarios.

7. Model Performance Evaluation

To assess the predictive capability of the proposed modeling framework, a Ridge regression model was implemented with 5-fold cross-validation, ensuring robustness against overfitting and variance in training testing splits.

The model achieved the following performance metrics:

• Mean Absolute Error (MAE): ≈ 56.5 ms

• Root Mean Squared Error (RMSE): ≈ 85.3 ms

7.1 Performance Interpretation

These error values indicate moderate predictive performance when evaluated against the full response time range (0–499 ms). While the model captures general trends in the data, its predictive precision remains limited.

A key observation is the discrepancy between MAE and RMSE. The higher RMSE suggests the presence of large prediction errors (outliers), which disproportionately influence squared error based metrics. This behavior is consistent with the previously identified zero-inflated distribution, where a substantial number of observations cluster at or near zero, while others span higher response values.

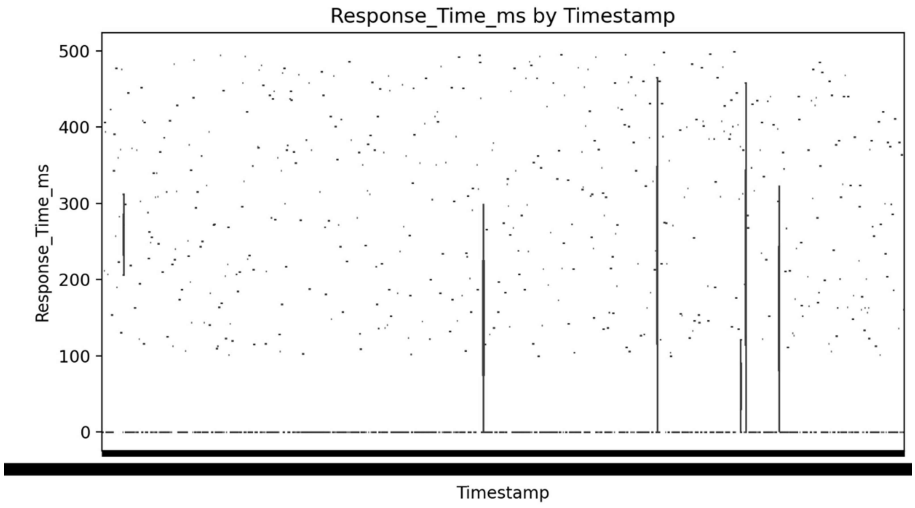

Figure 5. Response Time by Timestamp

7.2 Impact of Data Artefacts on Model Performance

Model performance must be interpreted in light of underlying data characteristics:

- Zero-heavy response time distribution introduces bias toward lower predictions
- Timestamp leakage artificially inflates predictive accuracy prior to feature correction
- Weak explanatory variables (e.g., severity, access level) limit model generalizability

Figure 5 (Response Time by Timestamp) visually demonstrates how timestamp encoding introduces spurious predictive strength, enabling the model to effectively memorize observations rather than learn meaningful patterns.

7.3 Implications for Predictive Modeling

The results highlight a critical distinction between apparent performance and true generalization. While numerical metrics may suggest acceptable accuracy, their validity is compromised by:

- Feature leakage
- Data imbalance
- Limited feature informativeness

Consequently, reliable predictive modeling in cybersecurity contexts requires careful feature validation, distributional analysis, and leakage prevention prior to model deployment.

8. Additional Visual Insights

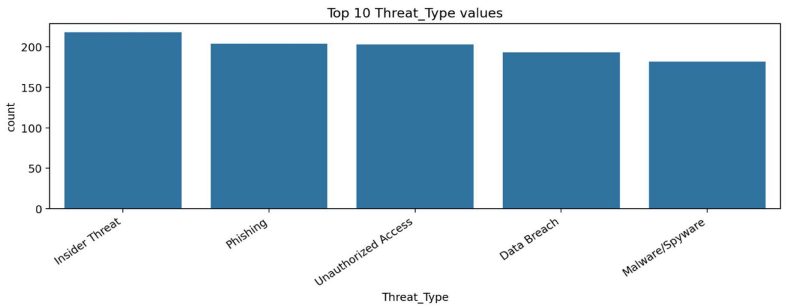

Figure 6. Top 10 Threat Types

To complement the statistical and modeling analyses, additional visualizations were examined to provide contextual understanding of the dataset.

8.1 Distribution of Threat Types

Figure 6 illustrates the Top 10 Threat Types, highlighting the most frequently occurring categories within the dataset. This distribution provides important insights into:

- System exposure patterns, indicating which threats are most prevalent
- Potential bias in data collection, where certain threat categories may be overrepresented

• Operational priorities, guiding where detection and mitigation resources may be most needed

The dominance of specific threat types suggests that the dataset may not fully represent the diversity of real-world cybersecurity incidents. Such imbalances can influence both statistical inference and model behavior, reinforcing the need for balanced and representative datasets.

8.2 Role of Visual Analytics

These visual insights play a crucial role in:

• Validating statistical findings

• Identifying anomalies and skewed distributions

• Supporting interpretability of model outputs

In particular, the alignment between visual patterns and quantitative results strengthens confidence in the observed lack of severity-driven response behavior and the dominance of detection based triggers.

9. Key Findings and Implications

The combined statistical and predictive analyses reveal a consistent pattern: traditional cybersecurity indicators, such as threat severity and access level, exhibit limited explanatory power in both association testing and response time prediction.

This convergence of evidence suggests that the system under study operates on a reactive paradigm, in which response actions are primarily triggered by detection events rather than informed by contextual threat attributes.

Furthermore, structural anomalies in the dataset particularly zero inflated response times and constrained relationships among variables play a significant role in shaping analytical outcomes. These artifacts not only affect statistical validity but also introduce biases in predictive modeling, such as inflated performance due to feature leakage.

Importantly, the findings highlight a disconnect between theoretical cybersecurity frameworks, which emphasize risk prioritization and adaptive response, and practical system behavior, which appears largely event driven and non differentiated.

This gap underscores the need to integrate richer contextual features, improve data quality, and design systems capable of proactive, severity-aware response mechanisms.

9.1 Summary of Statistical Relationships

The empirical analysis reveals that several commonly assumed relationships in cybersecurity contexts are not supported by the data. Specifically, no statistically significant association was observed between threat severity and data breach occurrence, nor between access level and insider threat activity. These findings

suggest that conventional indicators such as severity classification and privilege hierarchy may not independently determine security outcomes. Instead, breach events and insider threats appear to be influenced by more complex, potentially latent factors not captured in the current feature set.

9.2 Significant Dependency and Structural Constraints

A statistically significant relationship was identified between phishing attempts and malware detection, supported by a high Chi-square statistic and a moderate effect size (Cramer's V ≈ 0.2356). However, the presence of a zero-frequency cell in the contingency table indicates a structural irregularity in the dataset. This anomaly suggests that the observed dependency may be partially driven by data-generation or logging constraints, rather than reflecting a true operational relationship. Therefore, caution is warranted when interpreting this result, as statistical significance does not necessarily imply real-world validity.

9.3 Response Behavior Characteristics

The response time analysis demonstrates that system response latency is independent of threat severity, as evidenced by near-zero Pearson and Spearman correlation coefficients. In contrast, threat detection status emerges as the dominant predictor of response time ($n^2 \approx 0.76$). This indicates that the system operates in a reactive mode, where responses are initiated only after a threat has been detected. Such behavior highlights a lack of proactive or severity-aware response mechanisms, which may limit the effectiveness of rapid threat mitigation.

9.4 Data Quality and Modeling Limitations

The study identifies several critical data-related challenges that impact both statistical inference and predictive modeling:

- Timestamp leakage, resulting from high-cardinality encoding, leads to artificially inflated model performance
- Zero-inflated response time distributions, which distort central tendency and error metrics
- Weak predictive strength of engineered features, limiting model generalization

These issues underscore the importance of rigorous data preprocessing, validation, and feature engineering in cybersecurity analytics.

9.5 Practical Implications

From an applied perspective, the findings suggest several important considerations:

- Improving threat detection mechanisms is likely to yield greater benefits than refining severity classification schemes
- Effective temporal feature engineering is essential to avoid leakage and enhance model robustness
- Implementation of data auditing frameworks is critical to ensure analytical reliability

Collectively, these insights emphasize that data quality and system design choices are central to the success of cybersecurity analytics frameworks.

10. Conclusion

10.1 Conclusion

This study presents a comprehensive analysis of cybersecurity data through the integration of statistical association techniques and predictive modeling. The findings demonstrate that analytical outcomes are highly sensitive to data structure, preprocessing decisions, and feature engineering strategies.

Although certain relationships appear statistically significant, deeper examination reveals that many are influenced by dataset artifacts, encoding biases, and measurement inconsistencies. Notably, the results establish that response time is primarily driven by threat detection events rather than by threat severity, indicating a predominantly reactive system architecture.

These observations highlight the need for critical evaluation of both data and modeling assumptions when deriving insights from cybersecurity datasets.

Ultimately, this study demonstrates that the effectiveness of cybersecurity analytics is not solely dependent on advanced modeling techniques, but is fundamentally constrained by data quality, feature representation, and system design principles. Addressing these foundational issues is essential for transitioning from reactive defense mechanisms to intelligent, proactive cybersecurity systems.

10.2 Limitations

Despite its contributions, the study is subject to several limitations:

- Presence of zero-inflated response time data, affecting statistical reliability
- Potential biases arising from synthetic or constrained dataset characteristics
- Feature encoding artifacts, particularly in timestamp representation
- Lack of validation using real-world operational datasets

These limitations may restrict the generalizability of the findings and should be addressed in future work.

10.3 Future Research Directions

To enhance the robustness and applicability of future studies, the following directions are recommended:

1. Data Quality Improvement

Address zero-inflation through appropriate preprocessing techniques and ensure accurate response time measurements.

2. Advanced Feature Engineering

Transform raw temporal data into meaningful features (e.g., hour-of-day, day-of-week) and incorporate system-level behavioral indicators.

3. Enhanced Modeling Techniques
Explore non-linear and ensemble-based approaches, such as decision trees, random forests, and gradient boosting methods, to capture complex relationships.

4. Real-World Validation
Evaluate the proposed analytical framework on real-world cybersecurity datasets to ensure external validity.

5. Proactive System Design
Investigate integrating severity-aware and predictive response mechanisms to shift cybersecurity systems from reactive to proactive.

References

[1] Prasanthi Vallurupalli. (2022). Real time cybersecurity threat assessment dynamic risk scoring with hybrid data science models. *International Journal of Research Science and Management, 9*(2), 21–25.

[2] Kshetri, N. (2020). *Cybersecurity and cybercrime in the age of digital transformation*. Springer.

[3] Raggad, B. G. (2021). *Cybersecurity and privacy: An introduction*. Springer.

[4] Pfleeger, S. L., Pfleeger, C. P. (2018). *Security in computing* (5th ed.). Pearson.

[5] Stallings, W., Brown, L. (2019). *Computer security: Principles and practice* (4th ed.). Pearson.

[6] Fatima, A., Maurya, R., Dutta, M. K., Burget, R., Masek, J. (2019). Android malware detection using genetic algorithm-based optimised feature selection and machine learning. *In Proceedings of the 42nd International Conference on Telecommunications and Signal Processing (TSP)* (p. 220–223).

[7] Muhammad Danish. (2025). Enhancing cyber security through predictive analytics: Real-time threat detection and response. *International Journal of Advanced Computer Science and Applications, 16*(8).

[8] Chalapathy, R., Chawla, S. (2019). Deep learning for anomaly detection: A survey. *arXiv*. https://arxiv.org/abs/1901.03407

[9] Cao, G., Duan, Y., Cadden, T. (2019). The link between information processing capability and competitive advantage mediated through decision-making effectiveness. *International Journal of Information Management, 44*, 121–131.

[10] Seddon, P. B., Constantinidis, D., Tamm, T., Dod, H. (2017). How does business analytics contribute to business value *Information Systems Journal, 27*(3), 237–269.

[11] Wixom, B. H., Yen, B., Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive, 12*(2), 111–123.

[12] Davenport, T., Kudyba, S. (2016). Designing and developing analytics-based data products. *MIT Sloan Management Review, 58*(1), 83–89.

[13] Ayesha, Naseer., Humza, Naseer., Atif, Ahmad., Sean, B., Maynard, Adil Masood, Siddiqui. (2021). Real-time analytics, incident response process agility and enterprise cybersecurity performance: A contingent resource-based analysis. *International Journal of Information Management, 59*, 102334.

[14] Naseer, H., Maynard, S. B., Desouza, K. C. (2021). Demystifying analytical information processing capability: The case of cybersecurity incident response. *Decision Support Systems, 143*, 113476.

[15] Phillips Wren, G., Lakshmi, S. I., Kulkarni, U., Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. *Communications of the AIS, 37*(1), 448–472.

[16] Russom, P., Stodder, D., Halper, F. (2014). Real-time data, BI, and analytics. *TDWI Best Practices Report*.

[17] Shirkande, S. T., Vaidya, V. P., Sharma, P. K., Aher, V. N., Kolhe, M. R., Meshram, D. A. (2025). Decision systems for real-time threat response in cybersecurity. In V. Bhateja et al. (Eds.), *Innovations in ICT: Sustainability for societal and industrial impact* (Lecture Notes in Networks and Systems, Vol. 1364). Springer.

[18] Khalaf, Noora Zidan., Al Barazanchi, Israa Ibraheem., Radhi, A. D., Shah, Pritesh, Sekhar, Ravi. (2026). Development of real-time threat detection systems with AI-driven cybersecurity in critical infrastructure. *Mesopotamian Journal of CyberSecurity, 5*(2), Article 12.

[19] Zhu, Z., Chen, T., Song, Q., Lu, Y., Zheng, Y. (2025). ThreatResponder: Dynamic Markov-based defense mechanism for real-time cyber threats. In S. Goel et al. (Eds.), *Digital forensics and cyber crime* (Lecture Notes, Vol. 614). Springer.

[20] Omamo, A., Imathiu, J. (2025). Modeling resilience in cybersecurity: A systems dynamics approach to predictive threat response and adaptive security policies. *In Proceedings of the IST-Africa Conference* (pp. 1–6).

[21] Sarker, I. H. (2023). Machine learning for intelligent data analysis and automation in cybersecurity: Current and future prospects. *Annals of Data Science, 10*, 1473–1498.

[22] Fatama Tuz Johoraa, Md Shahedul Islam Khanb, Esrath Kanona, Mohammad Abu Tareq Ronyc, Md Zubaird, Iqbal H. Sarkere. (2024). A data-driven predictive analysis on cyber security threats with key risk factors. *arXiv*.https://arxiv.org/abs/2404.00068.

[23] Sarker, I. (2024). *AI-driven cybersecurity and threat intelligence: Cyber automation, intelligent decision-making and explainability*. Springer.

[24] Shi, Y., Tian, Y., Kou, G., Peng, Y., Li, J. (2011). *Optimisation-based data mining: Theory and applications*. Springer.

[25] Olson, D. L., Shi, Y., Shi, Y. (2007). *Introduction to business data mining* (Vol. 10). McGraw-Hill.

[26] Sarker, I. H. (2023). Multi-aspects AI-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview. *Security and Privacy, 6*, e295.

[27] Holsapple, C., Lee-Post, A., Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems, 64*, 130–141.