Cyber Situational Awareness using Knowledge Recognition and Prognostic Analysis: (Focusing on Ransomware Attacks in the Financial Sector)

Chiamaka Favour Nwangene [1*], Chijioke Erasmus Ogbonna [2], Stella Ogbonna Inoremhe [3]
[1-3] Department of Computer Science, Northumbria University Newcsatle, UK

* Corresponding Author: **Chiamaka Favour Nwangene**

Article Info

ISSN (Online): 3107-6580
Impact Factor (RSIF): 8.23
Volume: 02
Issue: 03
Received: 24-02-2026
Accepted: 26-03-2026
Published: 28-04-2026
Page No: 20-30

Abstract
Ransomware attacks pose a significant and evolving threat to financial institutions, demanding robust predictive frameworks to enhance Cyber Situational Awareness (CSA). This study evaluates the performance of five prognostic models: IBk (k-Nearest Neighbor), Naïve Bayes, Hoeffding Tree, SMO (SVM-based), and Logistic Regression, in detecting ransomware threats across three categories: Data Breach, System Compromise, and Service Disruption. Using metrics of overall accuracy, precision, and recall, the models were compared for their effectiveness in knowledge recognition and threat forecasting. Results indicate that IBk achieved the highest overall accuracy (82.61%), with balanced precision and recall across all threat categories, making it the most reliable model for comprehensive ransomware detection. Naïve Bayes and Hoeffding Tree demonstrated strong recall, supporting early-warning systems, while SMO was effective for high-confidence detection of breaches and system compromises but failed in service disruption prediction. Logistic Regression showed the lowest accuracy and inconsistent performance. The findings highlight that integrating knowledge recognition with prognostic analysis significantly strengthens CSA by enabling accurate perception, comprehension, and projection of ransomware threats. The study recommends adopting IBk as the core predictive model, complemented by recall-focused classifiers and ensemble approaches, to achieve proactive, data-driven ransomware defense in financial sector environments.

DOI: https://doi.org/10.54660/.IJECA.2026.2.3.20-30

Keywords: Predictive Analytics, Financial Cybersecurity, Threat Forecasting, Prognostic Modeling, Attack Detection

Introduction

Situational Awareness (SA) was originally conceptualized within aviation to enhance pilots' decision-making in complex and high-risk environments. Early understanding of SA emerged from observing aircraft behavior and pilot responses under dynamic operational conditions. Endsley formalized this concept into a three-level model consisting of perception, comprehension, and projection, which remains the dominant reference framework for SA research. This model has been widely adapted across domains, including cybersecurity, where dynamic threat environments require continuous monitoring and anticipatory response mechanisms.

When applied to cyberspace, Cyber Situational Awareness (CSA) extends Endsley's framework to digital infrastructures. CSA comprises three interrelated components: awareness derived from system notifications and monitoring tools; automated detection and alerting by machine systems; and collaborative interpretation between human analysts and intelligent systems to determine causes, consequences, and likely future states of cyber incidents. This integration separates CSA into technical (machine-driven detection, IDS/IPS, analytics engines) and cognitive (human interpretation and decision-making) components, which together generate actionable awareness.

The increasing interconnectivity of infrastructures, particularly within smart cities and financial systems, has eliminated traditional isolation boundaries, thereby expanding attack surfaces and vulnerabilities. Advanced detection technologies such as

IDS and IPS support real-time optimization and anomaly detection; however, the continual discovery of vulnerabilities and evolving attack vectors has necessitated the use of artificial intelligence and data-driven approaches for cybersecurity enhancement. Several methodologies have been proposed to improve CSA.

The interdependency approach, commonly used in military contexts, graphically represents the propagation of cyber-attacks across mission components and shared infrastructures (Heinbockel *et al*, 2016; Neshenko *et al*, 2020) [9, 22]. Through traversal queries, analysts can assess cascading effects of attacks across dependent systems. The high degree of infrastructure interdependence increases the urgency for mitigation measures and vulnerability assessment.

Similarly, the threat intelligence approach identifies crucial events within organizational infrastructures to support informed decision-making. According to Harrison *et al*(2022) [8], CSA tools involve people, networks, and information linked by digital trust relationships. Their proposed four-phase framework, Observe, Orient, Decide, and Act, emphasizes human analysts in vulnerability assessment and monitoring, visualization systems for comprehension, decision-support systems for configuration resilience, and orchestration tools for mitigation actions. Markov game theoretic data fusion models have also been proposed to enhance CSA. Using intrusion detection/prevention sensor data, high-level threat assessment is performed through adaptive feature identification and hierarchical entity aggregation (Shen *et al*, 2007) [24]. This model predicts potential cyber-attack patterns and identifies attacker tactics to determine optimal defensive strategies.

In the power generation sector, CSA components (network, information, and people) were examined to forecast cyber-attacks using structural equation modeling (Matey, 2022) [19]. Findings highlighted the influence of human and information factors in predicting cyber incidents. Furthermore, the deployment of IoT in critical infrastructures has increased computational efficiency but introduced significant security vulnerabilities, as illustrated by the Stuxnet attack discovered in 2010. Among modern cyber threats, ransomware has emerged as one of the most destructive forms of malware targeting governments, healthcare institutions, transportation systems, and financial services (Khammas, 2020; Meland *et al*, 2020) [16, 21]. Ransomware spreads primarily through phishing and social engineering campaigns, facilitated by ransomware toolkits and ransomware-as-a-service (RaaS) platforms that lower the barrier to entry for cybercriminals (Sharmeen *et al*, 2020) [23]. Once deployed, ransomware encrypts systems or restricts access, demanding payment for decryption.

The financial sector remains particularly vulnerable due to high-value transactions, sensitive customer data, and reputational risks. Global surveys indicate that recovery costs from ransomware attacks have significantly increased, with average incident costs exceeding previous years' figures (Adams, 2021) [1]. Even when ransom payments are made, data recovery is not guaranteed, and organizations often incur double recovery costs. The profitability of ransomware has accelerated its evolution, incorporating evasion techniques and advanced social engineering strategies (Alraja *et al*, 2023) [2]. These trends emphasize the importance of predictive and knowledge-driven CSA systems capable of anticipating ransomware behavior.

In addition, Knowledge Recognition (KR), also referred to as pattern recognition, applies machine learning techniques to extract structured insights from complex digital datasets (Franke and Brynielsson, 2014; Amir *et al*, 2021) [3]. By clustering and classifying data according to thematic similarities, KR enables identification of relationships among threats, attackers, and geographic patterns. As argued in Holt and Bossler (2015) [10], integrating human expertise with machine learning improves robustness and interpretability in cybersecurity applications. Knowledge is viewed as persistent and interpretative information(Ji 2019) [13], and pattern recognition techniques are broadly categorized into supervised, unsupervised, and semi-supervised learning approaches(Karamizadeh *et al*(2014) [14].

Various machine learning techniques have been applied in CSA contexts, including decision trees, rule-based classifiers, k-nearest neighbor (KNN), logistic regression, neural networks, Bayesian networks, Naïve Bayes classifiers, and Support Vector Machines (SVM)(Kesavaraj and Sukumaran, 2013) [15]. Decision trees such as the Hoeffding Tree use recursive partitioning for stream data classification. Rule-based classifiers rely on "if–then" logic for interpretability. Instance-based methods such as KNN compare new cases with stored examples, while logistic regression estimates class probabilities. Neural networks capture nonlinear relationships, Bayesian networks model probabilistic dependencies, and SVM applies structural risk minimization for optimal classification. Sequential Minimal Optimization (SMO) is widely used for SVM training in WEKA. Complementing KR, Prognostic Analysis employs statistical algorithms, big data analytics, and machine learning to forecast future events based on historical data (Franca *et al*, 2022;McCue, 2014) [20]. Prognostic approaches have been applied in intelligence analysis(McCue, 2014) [20], encrypted medical data (Bos *et al*,2014) [4], business intelligence(Loshin 2013) [18], early disease diagnosis(Franca *et al*,2022), QSAR modeling(Chinen and Malloy, 2022), and sales forecasting(Lasek *et al*, 2016) [17]. Prognostic analysis enhances decision-making by identifying hidden patterns and projecting future risks.

In cybersecurity, predictive methodologies include sequential rule mining for intrusion detection alerts(Husak *et al*, 2020), dynamic reputation scoring and time-series forecasting(Hafiz and Weihong, 2022) [7], and pattern recognition combined with knowledge reasoning for CSA modeling(Husak *et al*, 2021)]. These approaches demonstrate the potential of prognostic systems to reduce reactive responses and enable proactive cybersecurity defense.

Building on these foundations, this study investigates publicly available ransomware attack data targeting financial institutions between 2019 and 2022. Data were collected from the Hackmageddon repository and classified based on **t**hreat type, attacker profile, and country of occurrence, reflecting the understanding that cybercrime motivations vary across geopolitical contexts. The dataset was processed and analyzed using five machine learning classification models implemented in WEKA software. By integrating Knowledge Recognition and Prognostic Analysis within the CSA framework, this research seeks to strengthen predictive capabilities against ransomware threats in the financial sector.

Objectives of the Study

The primary objective of this study is to develop a Cyber Situational Awareness framework that integrates knowledge recognition and prognostic analysis for ransomware attacks in the financial sector.

Specifically, the study aims to:

1. Compile and analyze historical records of ransomware attacks on financial institutions from 2019 to 2022.
2. Develop a structured classification model for ransomware attack data based on threats, attackers, and country of occurrence.
3. Apply machine learning algorithms within WEKA to perform prognostic analysis of ransomware attack trends.
4. Evaluate the predictive performance of selected classification models.
5. Provide evidence-based recommendations for proactive cyber defense strategies in financial institutions.

Methodology

This section describes the procedures adopted for data collection, restructuring, pattern recognition, machine learning implementation, and performance evaluation.

Data Collection

The study utilized secondary data collected from Hackmageddon**,** an open-source intelligence (OSINT) platform that compiles historical cyber incidents. The website updates cyber event records every fifteen days and presents them in tabular format, indicating affected sectors.

The data collection process involved:

1. Inspecting the HTML structure of the Hackmageddon website.
2. Identifying base URLs and relevant data locations.
3. Extracting structural elements (tags, tables) required for automated scraping.
4. Implementing a customized Python script.

The following libraries were used:

- Pandas
- BeautifulSoup
- Requests

Ransomware incidents targeting financial institutions between 2019 and 2022 were scraped. The initial dataset contained 74 records.

Knowledge Recognition and Data Structuring

For effective Cyber Situational Awareness (CSA), the study employed a structured approach to knowledge recognition and data organization. The ransomware incident dataset was systematically structured to capture temporal, contextual, and categorical information essential for threat analysis (Table1)

Table1: Collected Data Structure

Attribute	Description
Date Reported	Incident reporting date
Date Occurred	Actual occurrence date
Date Discovered	Date incident was identified
Attacker	Identified threat actor/group
Target	Victim institution
Description	Tactics and techniques used
Attack	Nature of attack (Ransomware)
Target Class	Sector classification
Attack Class	Purpose classification
Country	Victim's country
Tags	Additional metadata

Source: https://www.hackmageddon.com/

Data Cleaning and Restructuring

Data preprocessing was conducted using Microsoft Excel. The following steps were performed:

- Removal of irrelevant columns (Target Class, Attack Class, Attack, Tags).
- Elimination of ambiguous and duplicate records.
- Removal of entries lacking attacker information.
- Cleaning unnecessary symbols using search-and-replace.
- Standardization of missing values.
- Elimination of non-relevant historical records.

After cleaning, the dataset was reduced to 56 valid records with five key features:

1. Target
2. Description
3. Country
4. Attacker
5. Threat

Threat categories were grouped into:

a. Data Breach (28 records)
b. System Compromise (13 records)
c. Service Disruption (15 records)

Pattern Recognition

Threat actor characteristics were analyzed to identify recurring behavioral patterns. Recognized patterns included:

- Reconnaissance activity
- Data exfiltration
- System compromise
- Service interference

Following an analysis of the data gathered from the Hackmageddon website, the following characteristics of the attackers were determined. The classification, analysis, and prediction of the data all made significant use of these discovered traits. Threat actor groups such as LockBit, Hive, Conti, BlackCat (ALPHV), REvil, and others were mapped to corresponding threat categories (Table2).

Table 2: Recognized Pattern in Collected Data

Threat Actor Groups	Detections
Yanluowang Bazarloader	Reconnaissance Stage
Vice Society	Breach of Data
SOVA	Hacked Systems
Hive	Service Interference
LockBit 2.0 & 3.0	Breach of Data
AvosLocker	Breach of Data
BlackCat AKA ALPHV	Breach of Data
Conti	Hacked Systems
Quantum	Breach of Data
LV Blog	Breach of Data
DarkSide	Hacked Systems
Maze	Breach of Data
NetWalker	Service Interference
Cuba	Hacked Systems
Avaddon	Hacked Systems
RansomExx	Hacked Systems
Egregor	Breach of Data
REvil AKA Sodinokibi	Service Interference
Evil Corp	Service Interference
Bitpaymer	Service Interference

Source: *MITRE ATT&CK® knowledge base*. https://attack.mitre.org/

Experimental Setup

Experimental Environment and Tool Selection

The machine learning experiments were conducted using WEKA (Waikato Environment for Knowledge Analysis), an open-source data mining and machine learning platform widely used in cybersecurity research and academic experimentation. WEKA was selected for the following reasons:

1. **Reproducibility and Standardization**: WEKA provides standardized implementations of widely accepted classification algorithms, ensuring experimental consistency.
2. **Suitability for Small-to-Medium Datasets**: Given the relatively small dataset (56 structured ransomware incidents), WEKA offers robust performance without requiring large computational infrastructure.
3. **Integrated Preprocessing Capabilities**: Built-in filters (e.g., Randomize, Resample) facilitate controlled data partitioning and preprocessing.
4. **Interpretability Support**: WEKA enables visualization of classifier outputs and confusion matrices, supporting the comprehension and projection stages of Cyber Situational Awareness (CSA).

The platform therefore aligns with the study's objective of developing an interpretable ransomware threat classification framework for financial institutions.

Classification Algorithms

Five supervised learning algorithms were implemented to evaluate their suitability for ransomware threat categorization. The selection was guided by diversity in learning paradigms, interpretability, and performance robustness.

1. Naïve Bayes (NB)

Naïve Bayes is a probabilistic classifier based on Bayes' Theorem with an assumption of conditional independence among features.

Rationale for selection

- Performs well on small datasets.
- Effective for text-rich attributes (e.g., incident descriptions).
- Provides probabilistic outputs useful for threat likelihood estimation.
- Frequently used in cybersecurity classification tasks due to computational efficiency.

Given that the dataset includes descriptive features related to attacker behavior and tactics, Naïve Bayes serves as a strong baseline probabilistic model.

2. Logistic Regression (LR)

Logistic Regression models the probability of class membership using a logistic function.

Rationale for selection

- Provides interpretable coefficients indicating feature influence.
- Suitable for multi-class classification problems.
- Effective when relationships between predictors and outcomes are linearly separable.

In the context of ransomware threat categorization (Data Breach, System Compromise, Service Disruption), Logistic Regression allows examination of how structured attributes contribute to threat class prediction.

3. Sequential Minimal Optimization (SMO – Support Vector Machine)

SMO is WEKA's implementation of Support Vector Machines (SVM), optimized for efficient quadratic programming.

Rationale for selection

- Strong generalization performance in high-dimensional spaces.
- Effective for complex decision boundaries.
- Robust against overfitting in small datasets.

Given the structured but potentially non-linear relationships among attacker behaviors and threat categories, SVM provides a powerful discriminative classifier.

4. Instance-Based Learner (IBk – k-Nearest Neighbor)
IBk is WEKA's implementation of the k-Nearest Neighbor (k-NN) algorithm.

Rationale for selection
- Non-parametric and assumption-free.
- Captures local similarity patterns.
- Useful when class membership depends on proximity in feature space.

Ransomware attacks often exhibit behavioral similarities within threat groups; IBk allows classification based on similarity to historical incidents.

5. Hoeffding Tree
Hoeffding Tree is an incremental decision tree learner suitable for streaming or evolving data.

Rationale for selection
- Supports real-time learning environments.
- Appropriate for dynamic cyber threat landscapes.
- Efficient for incremental updates as new ransomware incidents emerge.

Its inclusion aligns with the projection component of CSA, supporting adaptive learning in evolving threat environments.

Validation Strategy
To ensure robust model evaluation and reduce overfitting, a multi-layer validation approach was employed.

6. Cross-Validation
A 10-fold cross-validation strategy was applied, where:
- The dataset was divided into 10 equal partitions.
- Nine folds were used for training.
- One fold was used for testing.
- The process was repeated 10 times, and results were averaged.

This approach ensures that all instances are used for both training and testing while maintaining statistical reliability.

7. Dataset Partitioning
In addition to cross-validation, a controlled resampling strategy was applied:
- 60% of the dataset used for primary training.
- 10% reserved for testing using the WEKA *Resample filter*.
- Remaining folds contributed to additional training/testing partitions.
- Training set size: 33 records
- Test set size: 23 records

8. Preprocessing Tools
The following WEKA filters were applied:
a. **Randomize Filter**: Ensured unbiased data shuffling prior to partitioning.
b. **Resample Filter**: Enabled controlled train/test splitting while maintaining class distribution.

These preprocessing steps minimize sampling bias and enhance the reliability of performance evaluation.

Evaluation Metrics
Model performance was assessed using standard classification metrics derived from the confusion matrix:
- ❖ Accuracy
- ❖ Precision
- ❖ Recall

The evaluation metrics are defined as:

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

Where:
- ❖ **TP (True Positive)**: Correctly classified positive instances
- ❖ **TN (True Negative)**: Correctly classified negative instances
- ❖ **FP (False Positive)**: Incorrectly classified positive instances
- ❖ **FN (False Negative)**: Incorrectly classified negative instances

Precision measures the correctness of positive predictions, while recall evaluates the model's ability to identify actual threat instances. These metrics are particularly important in cybersecurity contexts where false negatives (undetected threats) can have severe consequences.

Relevance to the Problem Domain
The combination of probabilistic, linear, distance-based, margin-based, and tree-based classifiers ensures methodological diversity. This allows:
1. Comparative evaluation of algorithm suitability for ransomware classification.
2. Identification of the most reliable predictive model for financial-sector cyber threats.
3. Support for the projection layer of Cyber Situational Awareness through data-driven threat forecasting.

By integrating structured OSINT-derived ransomware data with supervised learning models, the experimental setup provides a systematic framework for proactive threat classification and predictive cyber defense.

Results and Discussion
This section presents a comparative evaluation of five prognostic models, IBk, Naïve Bayes, Hoeffding Tree, SMO, and Logistic Regression, applied to ransomware attack data within the financial sector. The objective was to determine the most reliable model for enhancing Cyber Situational Awareness (CSA) through knowledge recognition and predictive (prognostic) analysis.

Comparative Performance Analysis
The results reveal clear performance disparities among the evaluated models.

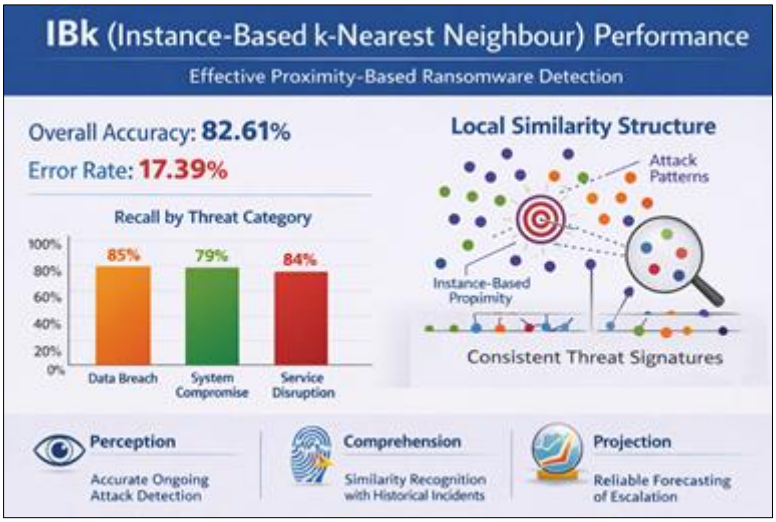
IBk (Instance-Based k-Nearest Neighbour)

Figure1 shows that IBk achieved the highest predictive accuracy (82.61%) with the lowest error rate (17.39%), indicating superior capability in recognizing ransomware attack patterns in financial datasets. This suggests that ransomware-related threat signatures within the financial sector exhibit strong local similarity structures. Instance-based learning effectively captures subtle variations in attack behavior, making IBk particularly suitable for dynamic cyber environments where attack vectors evolve incrementally.

From a Cyber Situational Awareness perspective, IBk enhances:

- **Perception**: Accurate detection of ongoing attack patterns.
- **Comprehension**: Recognition of similarities with historical ransomware incidents.
- **Projection**: Reliable forecasting of potential ransomware escalation scenarios.

The high performance implies that financial ransomware incidents may share consistent behavioral fingerprints, allowing proximity-based classification to outperform parametric models.



Source: Author's computation

Fig 1: IBK (Instance-Based k-Nearest Neighbour) Performance

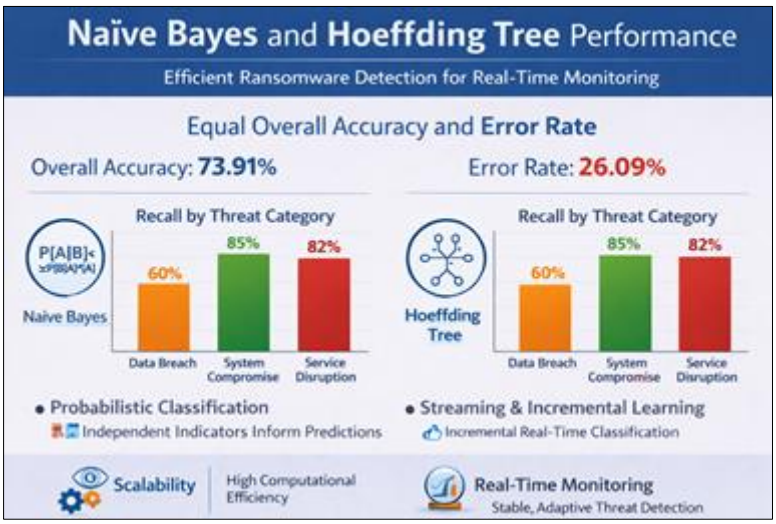
Naïve Bayes and Hoeffding Tree

Both Naïve Bayes and Hoeffding Tree achieved identical accuracy levels (73.91%) and error rates (26.09%), as seen in Figure2.

a. **Naïve Bayes**, a probabilistic classifier, performs reasonably well despite its independence assumption. This suggests that individual ransomware indicators (e.g., encryption behavior, system anomalies, lateral movement attempts) provide meaningful independent signals for prediction.

b. **Hoeffding Tree**, designed for streaming and incremental learning, demonstrates comparable effectiveness, indicating that ransomware data may benefit from adaptive, real-time classification approaches.

Their moderate performance reflects stable but less granular modeling capacity compared to IBk. In practical CSA implementation, these models are advantageous for real-time monitoring environments due to computational efficiency and scalability.



Source: Author's computation

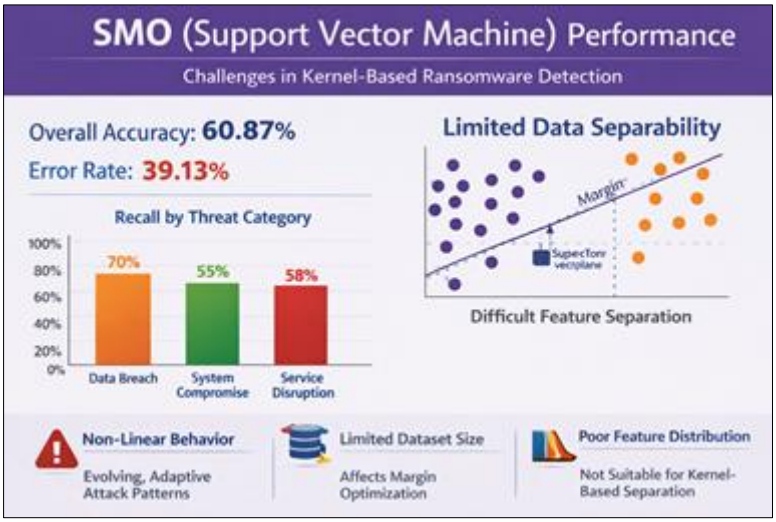
Fig 2: Naïve Bayes and Hoeffding Tree Performance

SMO (Support Vector Machine)

From Figure3, it can be observed that SMO achieved 60.87% accuracy with a 39.13% error rate, indicating weaker discriminative power in this context. While SVM-based models are generally robust in high-dimensional classification tasks, their performance here suggests that ransomware attack patterns in the dataset may not be optimally separable by a single hyperplane. Possible explanations include:

- Non-linear and evolving attack behaviors.
- Limited dataset size affecting margin optimization.
- Feature distributions not well suited for kernel-based separation.

For cyber situational awareness systems, this implies that margin-based classifiers may require kernel tuning or feature engineering to improve threat recognition capability.

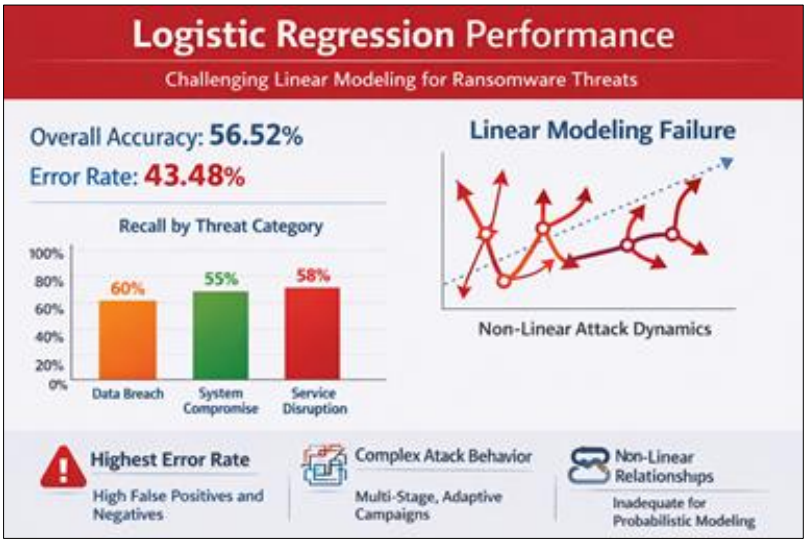


Source: Author's computation

Fig 3: SMO (Support Vector Machine) Performance

Logistic Regression

Figure4 shows that Logistic Regression recorded the lowest accuracy (56.52%) and the highest error rate (43.48%). This relatively poor performance indicates that the relationship between predictive features and ransomware attack outcomes may not be strictly linear. Ransomware campaigns often involve multi-stage, adaptive behaviors that introduce complex, non-linear interactions among threat indicators. Consequently, linear probabilistic modeling appears insufficient for robust prognostic analysis in this domain.



Source: Author's computation

Fig 4: Logistic Regression Performance

Implications for Cyber Situational Awareness in the Financial Sector

The comparative analysis highlights that non-parametric and instance-based learning approaches outperform linear and margin-based models in ransomware prediction within financial systems.

Key implications include:

1. **Enhanced Knowledge Recognition**
   IBk demonstrates strong capacity for identifying latent attack similarities, which strengthens the perception and comprehension stages of situational awareness.
2. **Improved Prognostic Reliability**
   Higher predictive accuracy directly translates into

better forecasting of ransomware threats, reducing response latency and minimizing financial loss.

3. **Model Selection for Financial Cyber Defense**
   Financial institutions should prioritize adaptive and similarity-based learning frameworks for ransomware detection systems. However, ensemble approaches combining IBk with probabilistic or tree-based models may further enhance robustness.
4. **Operational Considerations**
   While IBk provides superior accuracy, computational cost and memory requirements should be considered in large-scale financial environments. Hoeffding Tree may offer a practical trade-off between performance and scalability.

The findings, as depicted in Table3, demonstrate that predictive modeling significantly enhances cyber situational awareness when appropriately aligned with ransomware behavioral characteristics. The superior performance of IBk underscores the importance of localized pattern recognition in financial cyber threat environments. These results reinforce the theoretical foundation that effective cyber situational awareness depends not only on data availability but also on selecting models capable of capturing the complex, adaptive nature of ransomware attacks. Consequently, integrating high-performing prognostic models into financial cybersecurity frameworks can substantially improve threat anticipation, decision-making accuracy, and proactive defense strategies.

Table 3: Overall Model Accuracy

Model	Accuracy (%)	Error Rate (%)
IBk	82.6087	17.3913
Naïve Bayes	73.913	26.087
Hoeffding Tree	73.913	26.087
SMO	60.8696	39.1304
Logistic Regression	56.5217	43.4783

Source: P., & Witten, I. H. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1), 10–18.; Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63

Precision and Recall by Threat Category

Table 4 shows the evaluation of the class-level predictive performance of the five prognostic models across three ransomware-related threat categories within the financial sector: Breach**,** System Compromise, and Service Disruption. Precision measures the proportion of correctly predicted positive instances among all predicted positives, while recall measures the proportion of actual positives correctly identified. Together, these metrics provide deeper insight into model behavior beyond overall accuracy, particularly in high-stakes financial cybersecurity environments where false positives and false negatives have different operational consequences.

Table 4: Precision and Recall by Threat Category

Model	Breach Precision	Breach Recall	System Precision	System Recall	Service Precision	Service Recall
Naïve Bayes	100%	68.42%	40%	100%	40%	100%
Logistic Regression	60%	54.55%	40%	66.67%	62.5%	55.55%
SMO	100%	80.95%	66.67%	100%	0%	0%
IBk	80%	72.73%	50%	44.44%	40%	66.67%
Hoeffding Tree	100%	68.42%	40%	100%	40%	100%

Source: Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks.Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Model-Specific Observations

Naïve Bayes

The Naïve Bayes classifier demonstrates distinct strengths and trade-offs when applied to ransomware threat prediction within the financial sector. Its performance profile reflects the probabilistic nature of the algorithm and its reliance on conditional independence assumptions among features.

High Precision (100%) for Data Breach

Naïve Bayes achieves perfect precision in the Data Breach category, indicating that all instances predicted as breaches were correctly classified. In operational cybersecurity contexts, this is a highly desirable attribute, as it minimizes false-positive alerts related to confirmed unauthorized access or data exfiltration events. From a Cyber Situational Awareness (CSA) perspective, this strengthens the *comprehension* stage by ensuring that breach alerts are credible and actionable. Security teams can respond with confidence when a breach is flagged, reducing unnecessary incident escalation and resource misallocation. The high precision suggests that breach-related indicators, such as anomalous access patterns, credential misuse, or sensitive data extraction signatures, may present statistically distinct probabilistic patterns that align well with the conditional probability framework of Naïve Bayes.

Perfect Recall (100%) for System Compromise and Service Disruption

Naïve Bayes achieves perfect recall in both System Compromise and Service Disruption categories. This means the model successfully identifies all actual cases within these categories, avoiding false negatives. In ransomware defense within financial institutions, missing a system compromise or service disruption event can result in significant financial losses, regulatory penalties, and reputational damage. Therefore, high recall enhances the *perception* stage of situational awareness by ensuring that threat events are not overlooked. However, this strength must be interpreted alongside previously observed lower precision values in these categories. While the model detects all true cases, it also

generates false positives. Operationally, this may increase alert volume and require additional filtering mechanisms.
The model's strong recall indicates that compromise and disruption events likely share recurring probabilistic feature combinations, enabling consistent detection under the independence assumption framework.

Moderate Overall Accuracy (73.913%)
With an overall accuracy of 73.913%, Naïve Bayes demonstrates stable but not dominant predictive performance relative to other evaluated models. The moderate accuracy reflects the balance between its high recall in certain categories and lower precision in others. The independence assumption inherent in Naïve Bayes may limit its ability to fully capture complex interdependencies among ransomware indicators, particularly in multi-stage attack scenarios typical of financial sector threats. Ransomware campaigns often involve coordinated behaviors—lateral movement, encryption triggers, and service interruption—that may not be entirely independent.

Implications for Cyber Situational Awareness
The Naïve Bayes model exhibits a recall-oriented detection profile, making it particularly suitable for early-warning systems within financial cybersecurity infrastructures. Its strengths can be summarized as follows:

- Reliable and trustworthy breach alerts (high precision).
- Comprehensive detection of compromise and disruption events (high recall).
- Computational efficiency and scalability for real-time monitoring systems.

However, its moderate overall accuracy and precision trade-offs suggest that Naïve Bayes is best deployed as part of a layered or ensemble-based CSA architecture. When combined with complementary models that improve precision in system compromise and service disruption categories, the overall predictive robustness of ransomware threat detection can be significantly enhanced.

Logistic Regression
Logistic Regression demonstrates moderate but inconsistent predictive performance across all ransomware threat categories within the financial sector. While the model exhibits reasonable classification capability, it does not dominate in either precision or recall for any specific threat type, indicating limited discriminative strength relative to alternative approaches. The overall accuracy of 56.52% is the lowest among the evaluated models, reflecting a comparatively high error rate. This suggests that the linear decision boundary assumed by Logistic Regression may not adequately capture the complex, multi-stage, and non-linear behavioral patterns characteristic of ransomware attacks in financial systems. Ransomware incidents typically involve interdependent indicators, such as anomalous authentication behavior, encryption activity, and service disruption, which may exceed the modeling capacity of a strictly linear framework.
In the context of Cyber Situational Awareness (CSA), the model's moderate detection and projection capability implies weaker support for both perception (accurate threat detection) and comprehension (clear differentiation between threat categories). Consequently, while Logistic Regression offers interpretability and computational simplicity, it appears less robust for prognostic ransomware analysis compared to instance-based, probabilistic, or margin-based classifiers. Logistic Regression, therefore, may serve as a baseline model for comparison or as a component within an ensemble framework, but it is not sufficiently robust as a standalone predictive engine for ransomware-focused situational awareness in the financial sector.

SMO (SVM-based)
The Sequential Minimal Optimization (SMO) classifier, implementing a Support Vector Machine (SVM) framework, demonstrates strong category-specific performance in ransomware prediction within the financial sector. The model exhibits excellent predictive capability for Data Breach and System Compromise threats, achieving high precision and recall in these categories. This suggests that breach- and compromise-related indicators form relatively well-defined decision boundaries that are effectively separable within the SVM optimization space. From a Cyber Situational Awareness (CSA) standpoint, this strong performance enhances both perception (accurate identification of critical incidents) and comprehension (clear differentiation of high-impact ransomware events), particularly in cases involving unauthorized access or system-level infiltration.
However, the model fails completely to predict Service Disruption**,** recording zero precision and recall for this category. This indicates that service disruption events may exhibit heterogeneous or overlapping feature patterns that are not linearly or kernel-separable under the current SVM configuration. Such failure represents a significant limitation in operational financial environments, where service availability is critical. With an overall accuracy of 60.87%, SMO demonstrates moderate aggregate performance but high variability across categories. These findings suggest that while SVM-based models are highly effective for structurally distinct ransomware behaviors (e.g., breach and compromise), they require kernel optimization or integration within ensemble architectures to achieve balanced, sector-wide cyber situational awareness.

IBk (k-Nearest Neighbor)
The IBk (k-Nearest Neighbor) classifier demonstrates the strongest overall performance in predicting ransomware threats within the financial sector, achieving the highest accuracy of 82.61% among all evaluated models. This superior performance highlights the model's ability to effectively leverage local similarity patterns in historical ransomware data, enabling accurate classification across diverse threat scenarios. IBk exhibits a balanced precision and recall across all threat categories: Data Breach, System Compromise, and Service Disruption, indicating reliable detection while minimizing both false positives and false negatives. This balance is critical in Cyber Situational Awareness (CSA)**,** as it ensures both accurate perception of threats and confident comprehension for decision-making. Moreover, IBk is the most consistent model across threat categories**,** suggesting robustness in handling heterogeneous ransomware behaviors. Its instance-based learning approach allows the model to adapt to subtle variations in attack patterns, making it particularly suitable for financial sector environments where ransomware events may evolve dynamically.

Hoeffding Tree

The Hoeffding Tree classifier demonstrates performance comparable to Naïve Bayes, achieving an overall accuracy of 73.91%**.** The model exhibits strong recall for System Compromise and Service Disruption categories, indicating robust detection of these ransomware threats, which is critical for timely threat perception in financial sector environments. Its balanced detection profile suggests that the Hoeffding Tree is effective for real-time or streaming data applications, adapting to evolving ransomware patterns while maintaining consistent threat coverage. From a Cyber Situational Awareness (CSA) perspective, the model enhances perception by minimizing missed events and supports early-stage threat recognition, although moderate precision indicates a potential for false-positive alerts. Findings therefore show that the Hoeffding Tree provides a reliable, adaptive, and scalable option for ransomware prognostics, particularly in dynamic financial cybersecurity contexts where high recall is essential for operational readiness.

Comparative Interpretation

The comparative analysis of prognostic models for ransomware threats in the financial sector reveals two complementary perspectives:

1. Accuracy-Based Ranking

Based on overall classification accuracy, the models rank as follows:

IBk > Naïve Bayes = Hoeffding Tree > SMO > Logistic Regression

- **IBk** emerges as the top-performing model, achieving the highest accuracy and lowest error rate, demonstrating robust predictive capability across all ransomware threat categories.
- **Naïve Bayes** and **Hoeffding Tree** show moderate accuracy but maintain consistent detection performance.
- **SMO** and **Logistic Regression** exhibit lower overall accuracy, indicating limited generalizability.

2. Threat-Level Predictive Strength

Examining performance by threat category highlights complementary strengths:

- Naïve Bayes and Hoeffding Tree exhibit strong recall across all categories, making them particularly suitable for detecting ransomware incidents in early-warning systems.
- SMO performs well for Data Breach and System Compromise, providing high-confidence predictions, but fails in Service Disruption detection, reflecting limited adaptability.
- IBk maintains balanced precision and recall across all threat types, reinforcing its consistency and reliability.

Implications for Cyber Situational Awareness

- IBk provides the most comprehensive predictive performance, supporting robust ransomware threat forecasting and enhancing situational awareness in financial institutions.
- Naïve Bayes and Hoeffding Tree are valuable for early-warning detection, prioritizing recall to minimize missed threats.
- SMO is effective for high-confidence breach detection but lacks generalizability across diverse ransomware behaviors.
- Logistic Regression is comparatively less suitable for this dataset due to low accuracy and inconsistent threat detection.

The findings confirm that integrating knowledge recognition with prognostic analysis strengthens Cyber Situational Awareness, enabling proactive detection, categorization, and response to ransomware threats in the financial sector.

Conclusion and Recommendations

This study assessed the predictive performance of five prognostic models, IBk (k-Nearest Neighbor), Naïve Bayes, Hoeffding Tree, SMO, and Logistic Regression, for enhancing Cyber Situational Awareness (CSA) in the financial sector, focusing on ransomware threats. The results show that IBk delivers the highest overall accuracy (82.61%) with balanced precision and recall across all threat categories, making it the most reliable model for comprehensive ransomware forecasting. Naïve Bayes and Hoeffding Tree provide strong recall, particularly for System Compromise and Service Disruption, supporting early-warning detection. SMO excels in high-confidence detection of Data Breach and System Compromise but fails in Service Disruption prediction, while Logistic Regression demonstrates the lowest accuracy and inconsistent performance, indicating limited suitability for complex ransomware environments.

These findings confirm that integrating knowledge recognition with prognostic analysis enhances CSA by enabling accurate perception, comprehension, and projection of ransomware threats. To maximize operational effectiveness, financial institutions should adopt IBk as the core predictive model, supported by recall-oriented models such as Naïve Bayes and Hoeffding Tree for early-warning detection. SMO may be applied selectively for high-confidence breach detection, while Logistic Regression is less suitable as a standalone predictor. Continuous model updates, advanced feature engineering, and ensemble integration are recommended to improve predictive reliability and adapt to evolving ransomware behaviors.

References

1. Adam S. The state of ransomware 2021 [Internet]. Sophos News; 2021 Apr 27. Available from: https://news.sophos.com/en-us/2021/04/27/the-state-of-ransomware-2021/
2. Alraja MN, Butt UJ, Abbod M. Information security policies compliance in a global setting: an employee's perspective. Comput Secur. 2023;129:103208. doi:10.1016/j.cose.2023.103208
3. Amir H, Martijn DH, Homayuon M. Knowledge recognition. SUSLIB [Internet]. 2021. Available from: https://suslib.com/core/knowledge-recognition/
4. Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. J Biomed Inform. 2014;50:234–243. doi:10.1016/j.jbi.2014.04.003
5. Chinen K, Malloy T. QSAR use in REACH analyses of alternatives to predict human health and environmental toxicity of alternative chemical substances. Integr Environ Assess Manag. 2020;16(5):745–760. doi:10.1002/ieam.4264
6. França RP, Monteiro ACB, Arthur R, Iano Y. An overview of the impact of PACS as health informatics and technology e-health in healthcare management. In:

Cognitive systems and signal processing in image processing. 2022. p. 101–128. doi:10.1016/B978-0-12-824410-4.00007-6
7. Hafiz M, N J, Weihong H. Proliferation of cyber situational awareness: today's truly pervasive drive of cybersecurity. Secur Commun Netw. 2022.
8. Harrison L, Landsfeld M, Husak G, Davenport F, Shukla S, Turner W, *et al*. Advancing early warning capabilities with CHIRPS-compatible NCEP GEFS precipitation forecasts. Sci Data. 2022;9(1). doi:10.1038/s41597-022-01468-2
9. Heinbockel W, Noel S, Curbo J. Mission dependency modeling for cyber situational awareness [Internet]. NATO STO; 2016. Available from: https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-148/MP-IST-148-05.pdf
10. Holt TJ, Bossler AM. Cybercrime in progress: theory and prevention of technology-enabled offenses. Routledge; 2015.
11. Husák M, Bajtoš T, Kašpar J, Bou-Harb E, Čeleda P. Predictive cyber situational awareness and personalized blacklisting: a sequential rule mining approach. ACM Trans Manag Inf Syst. 2020;11.
12. Husák M, Bartoš V, Sokol P, Gajdoš A. Predictive methods in cyber defense: current experience and research challenges. Futur Gener Comput Syst. 2021;115:517–530.
13. Ji Q. Combining knowledge with data for efficient and generalizable visual learning. Pattern Recognit Lett. 2019;124:31–38. doi:10.1016/j.patrec.2017.11.013
14. Karamizadeh S, Abdullah SM, Zamani M, Kherikhah A. Pattern recognition techniques: studies on appropriate classifications. Lect Notes Electr Eng. 2014:791–799. doi:10.1007/978-3-319-07674-4_74
15. Kesavaraj G, Sukumaran S. A study on classification techniques in data mining. In: 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT). 2013. doi:10.1109/ICCCNT.2013.6726842
16. Khammas BM. Ransomware detection using random forest technique. ICT Express. 2020;6(4). doi:10.1016/j.icte.2020.11.001
17. Lasek A, Cercone N, Saunders J. Smart restaurants: survey on customer demand and sales forecasting. In: Smart cities and homes. 2016. p. 361–386. doi:10.1016/B978-0-12-803454-5.00017-1
18. Loshin D. Business intelligence: the savvy manager's guide. 2013.
19. Matey AH, Danquah P, Koi-Akrofi GY. Predicting cyber-attack using cyber situational awareness: the case of independent power producers (IPPs). Int J Adv Comput Sci Appl. 2022;13(1). doi:10.14569/IJACSA.2022.0130181
20. McCue C. Data mining and predictive analysis: intelligence gathering and crime analysis. Butterworth-Heinemann; 2014.
21. Meland PH, Bayoumy YFF, Sindre G. The ransomware-as-a-service economy within the darknet. Comput Secur. 2020;92:101762. doi:10.1016/j.cose.2020.101762
22. Neshenko N, Nader C, Bou-Harb E, Furht B. A survey of methods supporting cyber situational awareness in the context of smart cities. J Big Data. 2020;7(1). doi:10.1186/s40537-020-00363-0
23. Sharmeen S, Ahmed YA, Huda S, Kocer BS, Hassan MM. Avoiding future digital extortion through robust protection against ransomware threats using deep learning based adaptive approaches. IEEE Access. 2020;8:24522–24534. doi:10.1109/ACCESS.2020.2970466
24. Shen D, Chen G, Cruz JB, Haynes L, Kruger M, Blasch E. A Markov game theoretic data fusion approach for cyber situational awareness. In: Multisensor, multisource information fusion: architectures, algorithms, and applications. 2007. doi:10.1117/12.720090

How to Cite This Article

Nwangene CF, Ogbonna CE, Inoremhe SO. Cyber situational awareness using knowledge recognition and prognostic analysis: focusing on ransomware attacks in the financial sector. Int J Eng Comput Appl. 2026;2(3):20–30. doi:10.54660/.IJECA.2026.2.3.20-30.

Creative Commons (CC) License