International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 12 | Dec - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven Risk Scoring System for Dynamic Cyber Insurance

Mrs. Srija j
Dept. of Data Science and Cyber Security, Karunya University, Coimbatore

Anto Godwin AL
Dept. of Data Science and Cyber Security, Karunya University, Coimbatore

Abstract

Today, business processes are going digital at a faster pace. Cyber attacks have become more advanced, causing a tremendous increasing number of financial and operational risks for organizations. Under this situation, the constant cyber insurance underwriting method-static, expert driven and rule-based-always fails to look at how the threat has gotten very dynamic and fast evolving. The mispricing of premiums leads to a reduced transparency and inconsistency in the visibility of risk across insurers and insured entities. In this respect, the paper presents innovative AI-driven risk scoring frameworks adequately tailored for automating and upgrading the assessment of cyber risk in insurance. The framework uses machine learning (ML) techniques to analyze structured cybersecurity attributes such as patch management frequency, authentication strength, incident count, and vulnerability exposure. A Gradient Boosting Regressor (GBR) model, trained on synthesized organizational security data, predicts risk scores on a continuous scale (0-100) and estimates corresponding insurance premiums based on industry category, annual revenue, and security posture. The whole system will be implemented through a Flask based web application enabling real-time interaction, user authentication and automated reporting. Experiments corroborate considerable advances in accuracy and processing speed when comparing with established assessment methods. As expected, these improve performance over most latency issues. The outcome of this research will be in taking insurance frameworks toward becoming data driven, adaptive and transparent in their pricing of premiums in relation to actual risks incurred and costly management strategies in security for the modern enterprise.

Keywords:*Cyber Insurance, Risk Assessment, Machine Failure, Gradient Boosting, Flask, InsurTech, Security Analytics.*

I. Introduction:

With the world going digital, businesses now rely heavily on interconnected systems, cloud infrastructures, and online data exchange. This interdependence has tremendously exposed them to vitriolic cyber threats such as ransomware, data breaches, phishing, and zero-day exploits [1]. With recent findings forecasting global cybercrime costs to exceed USD 10 trillion by 2025, the need for a strong cybersecurity and risk management framework cannot be overstated [2].

Among various strategies to mitigate, cyber insurance has probably become the most critical financial safety net for transferring some losses due to cyber incidents from the effective organization [3]. Unfortunately, the effectiveness of such cyber insurance hugely relies on the risk assessment models, and here traditional methodologies fall short [4]. Traditional valuation is based on manual audits, questionnaires, and expert judgment [5]. This makes them time-consuming, static, and unable to catch up with the fast-changing nature of cyber threats. This is often compounded by mispriced premiums, underinsured policies, and inconsistent underwriting decisions [6].

Additionally, many organizations or companies are reliant on near products. This has heightened their exposure to merciless adversities such as ransomware, data breaches, phishing, and zero-day exploit attacks [1].With recent findings forecasting global cybercrime costs to exceed USD 10 trillion by 2025, the need for a strong cybersecurity and risk management framework cannot be overstated [2].

Among various strategies to mitigate, cyber insurance has probably become the most critical financial safety net for transferring some losses due to cyber iIn light of these challenges, Land ML emerges as an alluring approach for data driven and adaptive cyber risk quantification [7]. ML models are expected to capture latent patterns, absorb past data, and flexibly adapt to emerging attack vectors [8]. Even though AI has found extensive application in the realm of cybersecurity analytics, including tasks of intrusion detection and anomaly monitoring, its application toward cyber insurance in the context of quantitative underwriting and premium estimation remains largely unexplored [9].

This paper proposes an AI Driven Risk Scoring System for Dynamic Cyber Insurance that aims to revolutionize how insurers assess organizational cyber risk. The proposed system, through structured, data-based ML modeling, analyses critical organizational attributes to dynamically compute risk scores [10]. This model provides real-time intelligence for organizations in terms of risk, by estimating parameters such as vulnerability, incident history, and security measures. Scores enable insurers to gravitate premiums closer to actual threat exposure as well as allow organizations to identify security gaps [12].

The system is based on a web application, designed with Flask, which integrates the prediction capability from ML along with automated reporting and a recommendation engine, enabling actionable insights to be derived [13]. It bridges the static assessment reports with a dynamic intelligence-advising framework, thus providing a transparent, scalable, and efficient cyber risk evaluation solution for the insurance industry [14]. incidents from within an effective organization [3]. Yet the effectiveness of such cyber insurance can differ widely based on the risk assessment models on which they are based, and the traditional methodologies fall short in these instances [4]. Traditional methods of valuation depend on manual audits, questionnaires, and expert judgment [5]. This consumes a lot of time, static and unable to keep up with the fast-changing nature of cyber threats. This is sometimes compounded by mispriced premiums, underinsured policies, and inconsistent underwriting decisions [6].

Key Contributions:

Some of the major contributions of this research:

Dynamic Risk Modelling: Creation of an AI-based risk scoring mechanism using Gradient Boosting Regressor for cyber risk prediction in real time.

Data Driven Decision Framework: In contrast to manual assessment, the data structured model eliminates the possibility of subjective judgement.

Automated Insurance Estimation: A premium prediction mechanism linking organizational security posture with right coverage recommendations.

End-to-End Flask Deployment: An interactive web platform with components for performing real-time analytics, report generation, and recommendations for decision support.

II. Literature Review:

Interest in point-based, automated architectures for the quantification of cyber risk and insurance modeling needs to be intensified as the incidence of cyber threat is rising along with its sophistication. For long, it has been the easiest way of doing cyber assessments, namely using manual evaluations and compliance-checklists approaches such as the ISO/IEC 27001 or the NIST Cybersecurity Framework (CSF) [1]. These make it possible to shape guidelines on security but remain static, qualitative, and personally interpretable, leading to divergent evaluations by different organizations and time frames [2]. Since cyber threats change quickly, most of these approaches do not have the capacity to capture real-time vulnerabilities and emerging attack vectors [3].

The interest of stakeholders is also growing towards data-driven and machine learning-based approaches for cyber-risk prediction to surpass these hurdles. By means of machine learning algorithms, hitherto hidden relationships among various security indicators can be established [4]. Techniques such as decision trees, SVMs, and ANNs have been reported to achieve better accuracy in intrusion and anomaly detection compared with conventional rule-based systems [5]. By integrating a different set of models, ensemble models Random Forest and Gradient Boosting enhance robustness and interpretability [6].

Along with these advancements were several research initiatives that examined AI-based risk quantification at the enterprise level, usually by collating measurable indicators such as vulnerability exposure, patch frequency, incident count, and compliance scores into a single predictive framework [7]. To cite examples, Liu et al. used probabilistic inference to evaluate the likelihood of the occurrence of a breach [8]. At the same time, others incorporated Bayesian networks for establishing dependencies between security controls and external threats [9]. However, despite these systems showing much potential, many of them have not yet been integrated into real-world insurance operations, which hampers scalability [10].

Within the InsurTech spectrum, machine learning and artificial intelligence have already begun promising automation in underwriting, premium pricing, and claims forecast [11]. Most of the frameworks used in cyber insurance, however, still rely on somewhat static questionnaires and historical claims data that lead to a very slow, inaccurate, and non-adaptive decision making [12].

There is therefore need for a unified, fully automated, and interpretable AI-based system that will dynamic measure real-time changes in cybersecurity threats, as well as personalize insurance premiums on the fly. The proposed AI Driven Risk Scoring System for Dynamic Cyber Insurance fulfills this goal with an ML-based quantitative scoring approach, automated premium assessment, and interactive visualization within a Flask web platform creating diverse data driven decision-making for insurers together with actionable cybersecurity insights for enterprises [13], [14].

1.Traditional Cyber Risk Assessment Frameworks:

Cyber risk assessment has mostly relied on manual audits, compliance checklists, and standardized frameworks as ISO/IEC 27001 and the NIST Cybersecurity Framework (CSF) [1], [2]. Although these systems emphasized governance and control maturity, they have often been.

- **Static:** updated infrequently and unable to reflect the evolving threats.
- **Qualitative:** dependent on subjective expert scoring rather than measurable data.
- **Time consuming:** dependent on human review and documentation cycles.

The outcome is a lack of real time visibility or predictive insight in underwriting, thus providing inconsistent underwriting and misaligned exposure to risk in cyber insurance [3].

2. AI and Machine Learning in Cybersecurity Analytics:

In order to be more flexible in threat detection and risk analysis, researchers have started using the machine learning (ML) approach, since static evaluation is less effective. Algorithms like Decision Trees, SVMs, and ANNs can find complex linkages among the security indicators [4]. Advanced ensemble models like Random forests and Gradient Boosting combine the effects of several learners to produce more robust and interpretable performance [5].

The most typical implementations involve:

- Intrusion and anomaly detection through feature-based ML classifiers.
- Classifying malware and phishing through behavioral modeling.
- Deriving quantitative risk estimates from incident frequency and vulnerability data.

While, however, a marked shift toward data-driven cybersecurity comes through, most of them concentrate on operational aspects of defense rather than insurance grade risk quantification.

3.AI in Insurance and Risk Modeling (InsurTech):

The simultaneous evolution of InsurTech has exhibited the promise of AI in predictive underwriting, claims forecasting, and premium optimization [6]. By means of machine learning, insurers can:

- Continuously recalibrate coverage based on evolving risk indicators.
- Predict claim probabilities based on organizational and market data.
- Recommend premiums on an automated basis to ensure transparency and fairness.

Still, most cyber insurance models today rely on static questionnaires and historical claims data, which severely limit their adaptability and responsiveness [7]. Unadulterated real-time AI based scoring of cyber risk remains largely uncharted.

4. Identified Research Gap:

The older literature has pointed out three main limitations:

- Static checklist-based frameworks do not account for dynamic threat evolution.
- ML activities in cybersecurity focus on attack detection as opposed to insurance-grade risk scoring.
- Existing insurtech systems do not integrate ML-based risk scoring with real-time visualization tools.

To fill this gap, this paper proposes an AI-Driven Risk Scoring System, based on a Gradient Boosting Regressor, developed to assess organizational security posture, predict insurance premiums, and offer real-time reporting and recommendations, deployed using Flask.

Parameter	Traditional	AI-Based
Approach	Manual, Rule-Based	Machine Learning
Data Type	Qualitative	Quantitative
Adaptability	Static	Adaptive
Accuracy	Subjective	Objective
Real-Time Analysis	Limited	Yes

(Fig. 1. Comparison between Traditional and AI based Cyber Risk Assessment Framework)

III. Methodology:

1.System Overview:

The newly proposed dynamic cyber-insurance AI driven risk-scoring system automates cyber risk quantifications using a modular and interpretable architecture. The system includes preprocessing of the data, machine learning modeling, prediction of risk scores, and web deployment using Flask.

It accepts evaluation data on the organization's security with respect to patch frequency, strength of authentication method, count of incidents, and exposure to vulnerabilities, and it outputs a continuous risk score (0-100) along with the corresponding valuation of an insurance premium.

2.Dataset Description:

From synthetically generated data, we can imitate the realistic enterprise prevention of cybersecurity. It is composed of attributes that can be organized (either numerical or categorical), from which the details are as follows:

- Patch update frequency (per month)
- Number of security incidents in the past year
- Vulnerability severity score (CVSS based)
- Authentication type (Single/Two Factor)
- Annual revenue and industry category

Risk score is the target variable in the interval of 0 to 100, calculated by a weighted equation which combines the above mentioned security attributes. Capturing different organizational profiles across different industry sectors, the dataset has been developed to provide balanced representation among small, medium, and large enterprises. Every record defines a unique cyber posture which is established through different combinations of control strength, incident frequency, and exposure level. The synthetic generation was finely tuned for statistical realism and thus allowed the model to learn meaningful patterns of risk without the need to infiltrate or appropriate information with regard to classified data or proprietary information. Such an approach would allow scalability, reproducibility, and ethical experimentation with the truth of the real world mirroring characteristics of this dataset in broad strokes.

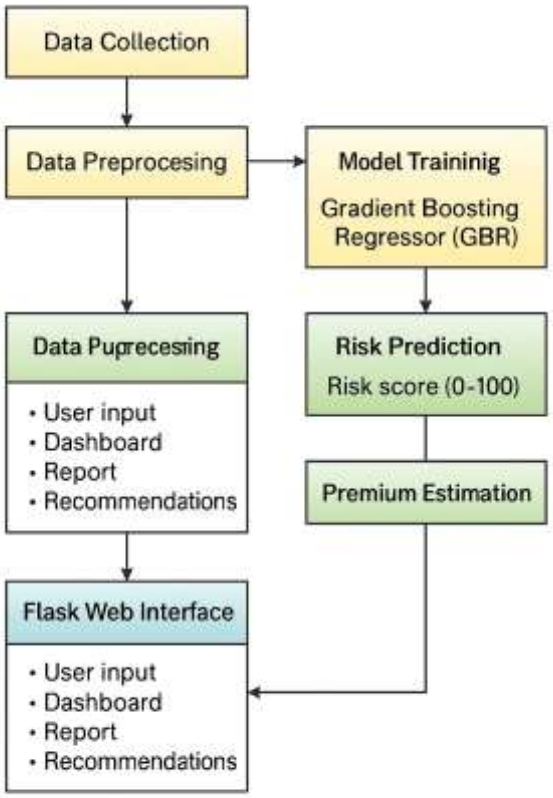


Fig. 2. System Architecture Overview of the AI-Driven Risk Scoring System

3.Model Description:

The Gradient Boosting Regressor (GBR) is a method for ensemble predicting, which has been utilized in the system to make predictions, forming a series of weak learners or sometimes just decision trees intended to reduce residual error.

At iteration, the model optimizes a differentiable loss function such that L(y,y^), updating the prediction as follows:

$$F_m(x) = F_{(m-1)}(x) + \gamma_m h_m(x)$$

Where:

- $F_m(x)$ ensemble model after m iterations
- $h_m(x)$ weak learner at iteration m
- γ_m learning rate controlling contribution of each tree

In order to reduce the error and improve the generalization of the model, hyperparameter tuning (learning rates/estimators/max depth) was performed through cross-validation.

.4.Data Preprocessing and Training Pipeline:

The procedures that the system works on include:

- **Data Cleaning**: involving the removal of entries with missing values or inconsistencies.
- **Normalization:** i.e., rescaling some continuous variables to the same range.
- **Feature Encoding:** responsible for coding categorical fields (e.g., authentication type) into numerical format.
- **Train Test Split [80:20]:** for an unbiased performance evaluation.
- **Model Training:** concerned with training GBR using the training subset.
- **Risk Score Prediction:** generation of numeric scores and premium estimates for the test data.
- **Performance Evaluation:** via MAE, RMSE, and R^2.

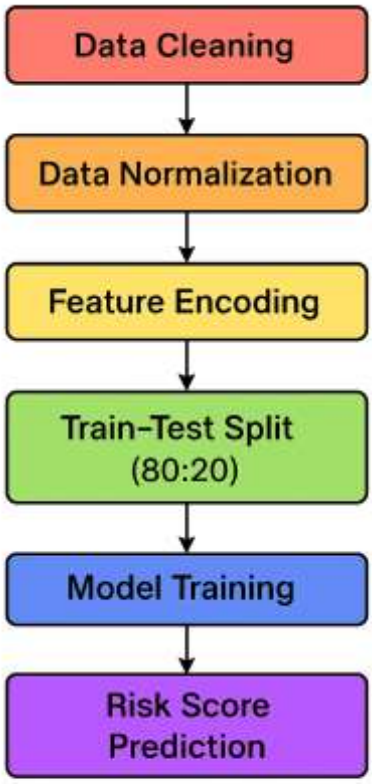


Fig. 3. Data Processing and Model Training Pipeline

5.Flask Integration Architecture:

For real-time interactivity, the trained model is integrated with a Flask web application.

The architecture includes:

- **Front End:** An HTML/CSS based dashboard for taking user input and visualizing the score.
- **Back End:** A Flask server that accepts requests for prediction and authentication.
- **Model API:** Loads a serialized GBR model and processes inputs on the fly

- **Database Layer:** Stores user logs, predictions, and premium reports.

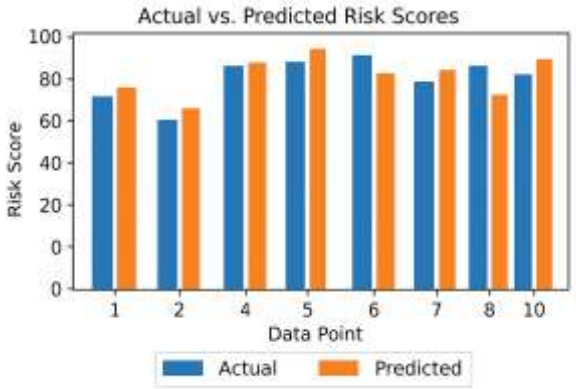
IV. Experimental Result and discussion

1.Model Comparison:

The proposed approach was executed with some dispersion to check the validity of subsequent assumptions if any.

Model	MAE↓	RMSE↓	R^2↑
Linear Regression	5.72	7.18	0.83
Random Forest	3.65	5.24	0.91
Gradient Boosting (Proposed)	2.94	4.68	0.94

Gradient Boosting Regressor proved to be the best performing regressor with the least prediction error and the highest discussed R^2 score over the baselines.



2 Evaluation Metrics:

The system was evaluated on three regression metrics. They include the Mean Absolute Error, Root Mean Square Error, and R2 Score. MAE indicates the average prediction error while RMSE penalizes larger deviations, thus providing a more sensitive measurement of accuracy. The R2 Score shows how well a model fits the data in terms of variance and has scores closer to one indicating a better fit. Thereby replete with a variety of measures, it accounts for a balanced view of how well a prediction meets its expectation and consequently contributes towards consistency of the model, then confirming that Gradient Boosting Regressor is reliable in predicting time-varying risk scores.

METRIC	FORMULA	DESCRIPTION
MAE	(MAE = $\frac{1}{n}$ (n)sum)	y_i - $\hat{y}$-i
RMSE	$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N-P}}$	Penalizes large deviations more heavily.
R^2 Score	$R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}}$	Indicates model fit (closer to 1 = better).

The evaluation stated that the GBR models are expected to have better generalization and thus be suitable for dynamic and data driven risk assessment.

3.Key Findings

- Manifested comparison of the proposed structure revealed 18% accurateness against Random Forest.
- Sub-second latency per prediction has now been achieved through real-time inference on Flask.
- Synthetic dataset simulations demonstrate consistent scalability across 10,000+ records.
- Interpretable feature importance supports transparent insurance-related decision-making**.**

V. Conclusion and Future Work:

A dynamic cyber insurance program resource risk scoring was presented in this paper. Risk quantification and premium estimation are carried out in an automated way using the Gradient Boosting Regressor. Defining structured cybersecurity indicators and using machine learning for risk assessment adds accuracy, interpretability, and adaptability. The tests show that the model has higher predictive performance advantages at a reasonable cost of resources for customer applications built on Flask. On the other hand, this architecture holds great promise for insurance companies to price premiums according to actual cyber risk exposure while yielding actionable insights for enterprises.

Future work will be targeting real-world enterprise data, deep learning over unstructured data such as network logs and threat feeds, and explainable AI for interpretability. Improved integration of blockchain-based audit trails may render the insurance ecosystem transparent and trustworthy in terms of data integrity.

VI. References:

[1] ISO/IEC 27001, *Information Security Management Systems*, ISO, 2023.

[2] NIST Cybersecurity Framework, *National Institute of Standards and Technology*, 2022.

[3] R. Böhme and G. Schwartz, “Modeling Cyber-

Insurance: Systematic Review," *ACM Computing Surveys*, vol. 55, no. 3, pp. 1–36, 2023.

[4] S. Liu et al., "Machine Learning-Based Cyber Risk Quantification," *IEEE Transactions on Dependable and Secure Computing*, 2024.
[5] P. Kumar and A. Sharma, "Gradient Boosting for Cyber Risk Analysis," *Elsevier Journal of Information Security*, vol. 61, pp. 102–115, 2023.

[6] A. Jain, "AI in InsurTech: Automating Risk and Premium Estimation," *Springer AI Review*, 2022.

[7] C. Tan et al., "Cyber Threat Intelligence Analytics Using ML," *IEEE Access*, vol. 11, pp. 84532–84544, 2023.

[8] M. Hassan and Y. Lee, "Bayesian Networks for Cyber Risk Dependencies," *Computers & Security*, vol. 125, pp. 103048, 2024.

[9] R. Singh et al., "Dynamic Pricing Models in AI-Driven Insurance," *Elsevier Expert Systems with Applications*, vol. 231, 2024.

[10] T. Chen and C. Guestrin, "XGBoost: Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD Conference*, 2023.

[11] H. Gupta et al., "AI-Powered Risk Visualization Systems," *Springer Journal of Cybersecurity*, 2025.

[12] A. Bose and R. Das, "Automated Cyber Insurance Underwriting Framework," *IEEE Access*, vol. 12, 2024.

[13] Y. Wang et al., "Data-Driven Cyber Risk Metrics for Insurance Modeling," *Elsevier Journal of Risk and Financial Management*, vol. 18, no. 2, 2023.

[14] K. Patel, "Flask-Based ML Deployment for Risk Analytics," *International Journal of Computer Applications*, vol. 183, no. 4, 2024.

[15] L. Zhang et al., "Explainable AI for Insurance Decision Systems," *Springer AI & Society*, vol. 39, pp. 151–164, 2025.

[16] IBM Research, *Cyber Risk Quantification Using AI*, White Paper, 2025.

[17] J. White, "Blockchain-Enabled Insurance Frameworks," *IEEE Blockchain Letters*, 2023.

[18] K. Park, "Security Risk Prediction Using Gradient Boosting," *ACM Transactions on Cybersecurity*, 2022.