International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 649-654
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0089

Automated Cybersecurity Risk Assessment for SMEs Using Machine Learning and Public Threat Intelligence

Nithilan Valan[1], Nisha Soms[2], Raghul K R[3], Dinesh K[4]

[1,3]UG Scholar, Dept. of CSE, KPR Institute of Engg. & Tech., Coimbatore, Tamilnadu, India

[2]Associate Professor, Dept. of CSE, KPR Institute of Engg. & Tech., Coimbatore, Tamilnadu, India

[4]UG Scholar, Dept. of MECH, KPR Institute of Engg. & Tech., Coimbatore, Tamilnadu, India

Emails: *22cs103@kpriet.ac.in[1], nishasoms@kpriet.ac.in[2], 22cs123@kpriet.ac.in[3], 22me019@kpriet.ac.in[4]*

Abstract

In the modern digital landscape, small and medium-sized enterprises (SMEs) face increasing cybersecurity risks due to limited financial resources, lack of dedicated security teams, and insufficient visibility into their threat exposure. This paper presents an intelligent and automated Cybersecurity Risk Scoring System designed to quantitatively assess the cybersecurity posture of SMEs based on their publicly accessible digital footprint. The proposed system integrates multiple threat intelligence sources, including VirusTotal, Shodan, Have I Been Pwned, and AbuseIPDB, through a unified backend API developed using Python-based Flask and FastAPI frameworks. Security-related features such as exposed network services, malware indicators, IP reputation, domain characteristics, and breach history are aggregated and analyzed using an XGBoost-based machine learning model to generate a normalized and interpretable risk score. A cross-platform Flutter-based mobile interface enables organizations to visualize domain health, vulnerability exposure, and network anomalies in real time. By automating data collection, analysis, and risk visualization, the proposed approach supports proactive cybersecurity risk management for SMEs while remaining cost-effective and scalable. The system aligns with established cybersecurity best practices and demonstrates the effectiveness of machine learning-driven risk assessment using publicly available threat intelligence data.

Keywords: *Cybersecurity Risk Scoring, FastAPI, Flask, Flutter, Machine Learning, SMEs, Threat Intelligence, Vulnerability Assessment, XGBoost*

1. Introduction

Small and medium-sized enterprises (SMEs) play a central role in the digital economy, but they often face higher cyber risk because they have smaller security budgets, limited in-house expertise, and little insight into the systems and services that are exposed outside the organization [1], [8]. Recent industry reports show that small and medium-sized enterprises are being targeted more often through exposed network services, misconfigured cloud resources, weak authentication, and unpatched vulnerabilities, which can lead to serious financial losses and operational disruption [6], [10], [20]. Even as cyber risks increase, many SMEs still depend on reactive or occasional security reviews, which do not give ongoing and practical awareness of their current risk level. [2], [21]. Traditional cybersecurity risk assessment frameworks - NIST Cybersecurity Framework, NIST SP 800-30, ISO/IEC 27005 and CVSS - supply step-by-step methods for spotting threats and gauging risk [7], [11], [16], [25]. They work well in large firms but each demands a full asset inventory, trained staff and hours of manual review - small plus medium-sized enterprises rarely have those resources - the frameworks are seldom used [3], [4]. Application-level standards like OWASP Top 10 and OWASP API Security Top 10 give useful advice but they do not turn technical flaws into an overall risk figure that a non-expert can act on [17], [18].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 649-654
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0089

Machine-learning advances now let systems learn attack patterns from past incidents, exposed infrastructure and threat-intelligence feeds automating much of the analytic work [14], [19], [22]. Researchers have built predictive cyber risk models with supervised learners but also ensembles like XGBoost and neural nets - those models reach higher accuracy than rule based tools [9], [12]. Work aimed at SMEs shows that managers need a plain, single risk score that turns technical results into clear guidance [13], [29]. Many current products, however, run on secret data sets, hide the scoring rules or price themselves beyond the reach of smaller firms [15], [27]. Open-source intelligence platforms like Shodan, VirusTotal besides Censys let anyone look from the outside at which services an organisation has left visible on the internet. They show whether an address has a bad reputation, whether a certificate is set up wrongly and whether malware has been seen coming from that network [5], [23], [28]. Because the check is done from the outside, no software agent has to be installed on the target systems. Earlier studies have proved that this kind of check, carried out through public application programming interfaces, finds weaknesses without active scanning plus therefore suits round-the-clock monitoring [24], [26]. But most researchers still treat each data source separately and have not combined the streams into one machine learning pipeline that produces a single risk score aimed at small but also medium-sized enterprises. To remove that shortcoming, the paper presents an artificial intelligence risk-scoring system built for SMEs. It fuses multiple open source intelligence feeds and uses supervised machine learning to give automated, explainable as well as low-cost assessments. The system gathers public exposure indicators, normalises the features, assigns weights and places every organisation in a clear risk band. An application-programming-interface architecture allows instant scoring, shows the exact arithmetic behind each result or scales without proprietary tools or invasive agents. The contribution is a down-to-earth, SME-focused tool that links standard risk principles, open threat data and artificial-intelligence decision support.

1.1. Problem Statement and Motivation

Small and medium-sized businesses are hit by more and more cyberattacks because some of their services are visible on the open internet, their equipment is often set up incorrectly and attackers have stolen or guessed employee log in details. Those firms rarely have the money or staff to watch their systems every hour of every week. The checks that exist today are done only once in a while, by hand and need a specialist to explain the results - they are too slow, too expensive plus too labour intensive for a company that wants a quick, cheap and automatic answer. Many firms do not know which of their own outward facing assets an attacker could reach until the day those assets are broken into. A method is therefore required that looks at the company from the outside, needs no software agent installed on site, runs on open data and repeats the test again and again while asking the user for only the bare minimum of information. A second problem is that the few scores that do exist are hard to read - the firms cannot decide what to fix first.

1.2. Research Gap and Proposed Solution

Well-known standards but also paid platforms give thorough risk evaluations but they are intricate, expensive and out of reach for most small or medium businesses. Earlier academic work that uses machine learning for cyber risk often depends on private data sets, hides the way the number is calculated or studies only one kind of attack instead of the whole risk to the firm. Studies that rely on open source intelligence usually inspect single warning signs and never roll them into one overall figure. This paper closes those shortfalls. It presents an artificial intelligence system that needs no internal agent, pulls data from many public sources, blends supervised machine learning with a clear step-by-step scoring rule, accepts a domain name, IP address or e-mail address as its only input, examines every security clue that can be seen from the outside, returns a plain language risk figure as well as offers a report that can be downloaded - that advanced risk measurement becomes usable by any small or medium sized enterprise.

2. Method

The proposed system uses an agentless approach to cybersecurity assessment based on observations made from outside the target environment. This methodology supports automated risk assessment with minimal input from the user and sets out steps

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 649-654
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0089

that can be repeated so that the same results can be obtained under the same conditions. The system takes a domain name, IP address, or email address as input. It collects relevant data, runs machine-learning analysis, and assigns a risk score to produce a cybersecurity assessment report.

2.1. Input Specification and Data Acquisition

The system takes one externally observable identifier as input: a domain name, an IP address, or an email address. The type of input determines which open-source intelligence (OSINT) data are gathered from public threat intelligence sources, such as DNS records, IP reputation data, breach exposure databases, and results from service enumeration. This approach removes the need for internal system access or deploying agents, which makes it appropriate for small and medium-sized enterprises that have limited technical infrastructure (Table 1).

Table 1 Supported Input Parameters and Collected OSINT Data

Input Type	Collected Parameters
Domain Name	VirusTotal reports, DNS records, SSL status, exposed services, misconfigurations
IP Address	Open ports, service banners, reputation score, blacklist status
Email Address	Breach exposure, credential leaks, threat intelligence matches

2.2. Feature Extraction and Preprocessing

The OSINT data was cleaned and converted into consistent numerical and categorical variables for machine learning analysis. Missing values are replaced with predefined default entries so that the data remain consistent across different kinds of input. Where needed, features are scaled and variables are encoded so the inputs match the requirements of the trained machine learning model. The analysis uses only indicators that can be checked against public sources, so it does not depend on proprietary or otherwise inaccessible datasets.

2.3. Machine Learning–Based Risk Analysis

A supervised machine learning model analyzes the extracted features to estimate the cybersecurity risk for the given input. The model was trained on labeled cybersecurity exposure data to detect patterns that suggest higher risk levels. The model produces a probability-based indicator of risk. A rule-based scoring method then converts this indicator into a final risk score so that the results are easy to interpret and can be compared across assessments.

2.4. Risk Scoring and Report Generation

The final cybersecurity risk score is calculated by combining the machine learning model's output with weighted security indicators, including service exposure, reputation flags, and prior breach history. The score is assigned to predefined risk categories so that readers without technical training can interpret it more easily. The system generates a PDF assessment report that summarizes the risks identified, the factors contributing to them, and the current security posture. The report is stored securely in an Amazon S3 bucket so it can be accessed and retrieved later.

2.5. System Output

The system displays the calculated risk score and a brief assessment summary in a web-based interface. After processing is complete, users may review the results and download the PDF report. This methodology provides an automated, repeatable, and low-cost process for assessing cybersecurity risk in small and medium-sized enterprises (Figure 1).

3. Results and Discussion

3.1. Results

We evaluated the proposed system to test how well it classifies cybersecurity risk levels using threat intelligence data that can be observed from external sources. The experimental design tested whether a machine learning–assisted risk scoring method could correctly place cases into High, Medium, or Low risk categories based on real-world features derived from OSINT. The evaluation dataset included labeled samples that captured different degrees of exposure, configuration errors, and evidence of threats. Table 2 summarizes the risk classification model results by reporting precision, recall, and F1-score for each risk category, along with the overall accuracy of the classification. In this study, the system reached an overall accuracy of 91%. A value of 0% suggests consistent performance in automated cybersecurity risk assessment. The high-risk category recorded a precision of 0, meaning that none of the cases

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 649-654
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0089

predicted as high risk were true positives in this evaluation. The model achieved a score of 94, while its recall was 0. The score of 97 indicates that the model performs well in identifying critical security exposures. The medium-risk category showed a balanced level of performance. In contrast, the low-risk samples had high precision but lower recall, which suggests the model classified cases as low risk only when it was fairly certain (Table 2).

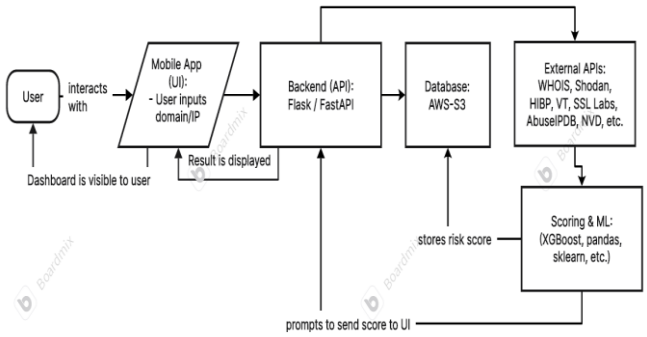

Figure 1 System Architecture of Cyber Risk Scoring System

Table 2 Risk Classification Performance Metrics

Risk Category	Precision	Recall	F1-Score
High	0.94	0.97	0.96
Medium	0.83	0.86	0.84
Low	1.00	0.25	0.40
Overall Accuracy	–	–	91.0%

To examine the model's classification behavior in more detail, we generated a confusion matrix, which is reported in Table 3. In the matrix, correct classifications appear on the diagonal, while misclassifications appear in the off-diagonal cells across categories. Most high-risk inputs were identified correctly, with few cases incorrectly placed in the medium-risk category. In medium-risk cases, the model sometimes labeled instances as high-risk. From a security perspective, this is acceptable because it errs on the side of caution rather than understating risk.

Table 3 Confusion Matrix for Risk Classification

Actual \ Predicted	High	Low	Medium
High	66	0	2
Low	0	1	3
Medium	4	0	24

Figure 2 presents the system's output and shows the final risk assessment produced after the input data were processed. The interface shows the calculated risk score, its category, and the indicators used to support it, and it allows users to download the assessment report as a PDF.

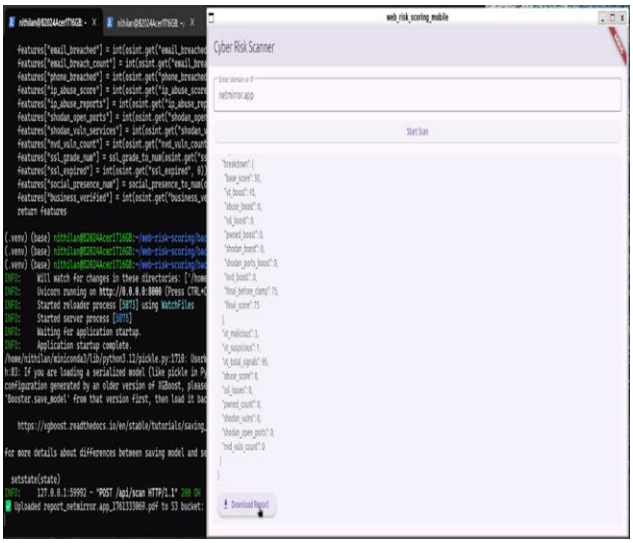

Figure 2 System Output

3.2. Discussion

The results show that the proposed system can identify cybersecurity risks using externally observable data alone, without internal access or agent-based deployment. The observed accuracy supports the feasibility of pairing machine learning predictions with a rule-based risk score for cybersecurity assessment in SMEs. Performance in the High-risk category was strong, since correct identification of critical exposures supports timely mitigation. The lower recall in the Low-risk class suggests a conservative decision rule, where the system tends to label fewer cases as Low-risk. In a security-sensitive setting, this pattern can be appropriate because it reduces the chance of treating a risky case as Low-risk, even if it leads to more Low-

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 649-654
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0089

risk cases being missed. Classifying some medium-risk cases as high-risk aligns with a cautious design choice that may suit SMEs that have limited in-house security expertise. In contrast to traditional security assessment tools that depend on continuous monitoring, proprietary datasets, or manual audits, this approach offers an automated, low-cost method that can be repeated consistently across assessments. The current version does not include analytics based on visual displays, but this does not change the main goal of identifying risks and reporting them. The system's PDF reports still present the results in a clear, simplified form that supports practical decision-making. Taken together, the discussion suggests that the system strikes a workable balance among automation, interpretability, and practical use. Future work could add trend analysis and dashboard visualizations, but the present findings support deploying the system as a basic cybersecurity risk assessment tool.

Conclusion

This paper examines how small and medium-sized enterprises often depend on cybersecurity risk assessment methods that are costly and demand considerable time and expertise, which limits how often they can be carried out. Many current frameworks and tools work well in large organizations, but they are often not practical for SMEs because they depend on internal asset inventories, always-on monitoring systems, and staff with niche technical skills. The findings suggest that externally observable threat intelligence, when combined with supervised machine learning and a structured risk-scoring method, can serve as a practical alternative. The results suggest that automated risk classification based on public exposure indicators can achieve high accuracy while remaining non-intrusive and relatively low cost. The evaluation indicates that the system reliably identifies high-risk entities and applies conservative behavior to lower-risk categories, consistent with security-first principles. The system provides a single, clear risk score and a downloadable assessment report, presenting technical results in practical terms that support decisions by readers without specialist training. The study finds that an AI-assisted, OSINT-driven approach to risk assessment is feasible and practical for SMEs. The current version supports point-in-time assessment rather than longitudinal analysis. The results support the main design choices and provide a basis for later work on trend analysis, visualization, and deployment at scale.

Acknowledgements

The author would like to express sincere gratitude to the project mentor and faculty members of KPR Institute of Engineering and Technology for their continuous guidance, feedback, and academic support throughout this work. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The development and experimentation were carried out using publicly available datasets, open-source tools, and freely accessible threat intelligence platforms.

References

[1]. Arora, A., Hall, D., Pinto, A., Ramsey, D., & Telang, R. (2006). Measuring the risk-based value of IT security solutions. *IT Professional*, 8(1), 35–42. https://doi.org/10.1109/MITP.2006.14

[2]. Birari, H. P., Lohar, G. V., & Joshi, S. L. (2023). Advancements in machine vision for automated inspection of assembly parts: A comprehensive review. *International Research Journal on Advanced Science Hub*, 5(10), 365–371. https://doi.org/10.47392/IRJASH.2023.065

[3]. BSI. (2021). *Cyber security for SMEs: Guidance and best practices*. British Standards Institution.

[4]. Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). A model for evaluating IT security investments. *Communications of the ACM*, 47(7), 87–92. https://doi.org/10.1145/1005817.1005828

[5]. Censys. (2023). *Internet-wide scan data and attack surface management*. Censys Inc.

[6]. ENISA. (2023). *ENISA threat landscape 2023*. European Union Agency for Cybersecurity.

[7]. FIRST.org. (2023). *Common Vulnerability Scoring System v3.1 specification*. Forum of Incident Response and Security Teams. https://www.first.org/cvss/

[8]. Gartner. (2023). *Cybersecurity trends for small and midsize businesses*. Gartner Research.
[9]. Gupta, R., & Singh, A. (2021). Machine learning based cyber risk prediction models: A comparative study. *Journal of Information Security*, 12(3), 145–158.
[10]. IBM Security. (2023). *Cost of a data breach report 2023*. IBM Corporation.
[11]. ISO/IEC. (2022). *ISO/IEC 27005: Information security risk management*. International Organization for Standardization.
[12]. Kim, J., & Lee, H. (2020). Deep learning approaches for cybersecurity risk detection. *Computers & Security*, 93, 101789. https://doi.org/10.1016/j.cose.2020.101789
[13]. Kim, S., & Park, J. (2021). Risk-aware cybersecurity decision models for SMEs. *Journal of Small Business Management*, 59(4), 712–729.
[14]. Meng, W., & Ni, J. (2020). Cyber risk analytics using supervised learning techniques. *IEEE Access*, 8, 18434–18445. https://doi.org/10.1109/ACCESS.2020.2967835
[15]. Microsoft. (2022). *Cybersecurity solutions for small and medium enterprises*. Microsoft Security Whitepaper.
[16]. NIST. (2023). *NIST Cybersecurity Framework (Version 2.0)*. National Institute of Standards and Technology.
[17]. OWASP Foundation. (2023). *OWASP API Security Top 10*. Open Web Application Security Project.
[18]. OWASP Foundation. (2024). *OWASP Top 10 Web Application Security Risks*. Open Web Application Security Project.
[19]. Paul, A., & Deka, B. (2021). Cyber threat intelligence using machine learning: A review. *Journal of Network and Computer Applications*, 177, 102975. https://doi.org/10.1016/j.jnca.2020.102975
[20]. Ponemon Institute. (2023). *The state of cybersecurity in small businesses*. Ponemon Institute LLC.
[21]. Rajan, P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. *International Research Journal on Advanced Science Hub*, 5(11), 389–396. https://doi.org/10.47392/IRJASH.2023.071
[22]. Shaukat, K., Luo, S., Varadharajan, V., & Hameed, I. (2020). Cyber threat detection using machine learning techniques. *IEEE Access*, 8, 21082–21095. https://doi.org/10.1109/ACCESS.2020.2968445
[23]. Shodan. (2024). *Shodan search engine and internet exposure data*. Shodan LLC.
[24]. Singh, R., & Bansal, S. (2023). Passive cyber risk assessment using OSINT techniques. *International Journal of Information Security Science*, 12(2), 98–110.
[25]. Stoneburner, G., Goguen, A., & Feringa, A. (2002). *Risk management guide for information technology systems (NIST SP 800-30)*. National Institute of Standards and Technology.
[26]. Sun, L., Zhang, Y., & Wang, J. (2020). External attack surface analysis using public threat intelligence. *Computers & Security*, 95, 101846. https://doi.org/10.1016/j.cose.2020.101846
[27]. Symantec. (2022). *Cybersecurity solutions overview*. Broadcom Inc.
[28]. VirusTotal. (2024). *Threat intelligence and malware analysis platform*. Google LLC.
[29]. Zhuang, W., Chen, Y., & Li, X. (2022). SME-focused cyber risk scoring models using machine learning. *Journal of Cybersecurity*, 8(1), tyac004. https://doi.org/10.1093/cybsec/tyac004