International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09si01.016 ISSN: 2581-4621

Machine Learning-Based Predictive Model for Web Application Vulnerability Detection

Manisha Marella[1], Purelli Sanjana Reddy[2], Akhila Kadimetla[3]
Department of Information Technology
G. Narayanamma Institute of Technology and Science Hyderabad

Abstract:
Web application security is a critical concern as cyber threats continue to evolve, exposing organizations and individuals to significant risks. This study introduces a machine learningbased predictive model that leverages supervised learning techniques to detect and classify vulnerabilities in web applications by analysing feature patterns from labelled datasets. The system identifies threats such as SQL injection and Cross-Site Scripting (XSS), providing an automated approach to early threat detection. A web-based tool has been developed to enable users to assess website security by entering a URL, with the backend processing web requests and responses using the trained machine learning model. Performance evaluation metrics, including accuracy and precision, demonstrate improved detection rates and reduced false positives compared to existing methods. By integrating machine learning into web security analysis, this research enhances automated vulnerability detection, serving as a valuable resource for developers and security professionals to strengthen the security posture of web-based applications.
Keywords: Web application security, machine learning, vulnerability detection ,SQL injection, Cross-Site Scripting(XSS).

Introduction

Web operation security has come a critical concern as cyber pitfalls continue to evolve posing significant pitfalls to both associations and individualities. The adding reliance on web- grounded platforms for communication, tractions, and data storehouse has made them high targets for cyberattacks. Traditional security measures, similar as hand- grounded discovery and rule-grounded approach, frequently fail to descry new and sophisticated attacks. These conventional ways calculate on predefined patterns and autographs, making them ineffective against zero day vulnerabilities and arising cyber pitfalls as attacks continuously upgrade their styles, there is an critical need for adaptive and intelligent security results that can descry and alleviate vulnerabilities stoutly.

Machine literacy ways offer a promising results by enabling automated trouble discovery through pattern recognition and anomaly discovery. ML- grounded models can assay vast quantum of data to identify attack patterns and classify implicit pitfalls efficiently. This study introduces a machine literacy- grounded prophetic model to descry and classify vulnerabilities in web applications.by using supervised literacy ways, the model analyses feature patterns uprooted from datasets to identify common web pitfalls, including SQL injection andcross-site scripting(XSS). an interactive web- grounded tool has been developed, allowing druggies to assess website security by entering a URL, which is analysed using the trained ML model. The model's capability to learn from literal data and fete vicious exertion patterns enhances early trouble discovery, minimizes false cons and improves the overall security posture of web- grounded operations.

Literature Survey

The existing models contain several key drawbacks that limit the effectiveness of traditional vulnerability detection methods. One major limitation is the dependency on predefined payloads and rile-based detection, which makes these systems inflexible in adapting to new and evolving attack patterns. Approaches relying on static analysis often generate false positives due to

DOI:10.46647/ijetms.2025.v09si01.016 ISSN: 2581-4621

overestimating risks, while those dependent on dynamic analysis may struggle to detect issues that arise only under specific conditions. Additionally, risk assessment models, though useful for classifying threats, lack of ability to dynamically account for uncertainties and real-world variations attack methodologies.

Another significant challenge is the inability of some models to handle complex and dynamic web applications effectively. Many scanners struggle with modern frameworks that rely heavily on client-side scripting and asynchronous requests. Furthermore, some detection systems rely on historical data for vulnerability prediction, which can lead to outdated or inaccurate results if the data is not consistently updates or fails to account for emerging threats.

Performance issues also prevalent among traditional vulnerability detection mechanisms. Some methods introduce high computational overhead due to exhaustive scanning techniques, which can slow down the system and impact real-time security assessments. Others focus primarily on SQL Injection and Cross-Site Scripting (XSS) but lack of comprehensive coverage of other security risks. Additionally, web crawlers and scanning mechanisms may miss certain vulnerabilities due to limited data collection techniques or an inability to adapt to different web environments.

Our project overcomes these limitations by leveraging a machine learning-based predictive model that dynamically detects and classifies vulnerabilities based on feature patterns rather that relying solely on rule-based scanning. By training on a diverse dataset, our system can generalize better and detect emerging threats, reducing dependency on predefined payloads. Specifically, our project consists of three specialized models: SQL Injection, Web Vulnerability, and Cross Site Scripting (XSS) detection, each designed to identify and mitigate key security threats in web applications.

Furthermore, our solutions is designed to be scalable and adaptable, allowing it to assess various types of web environments more effectively that traditional scanners. By integrating automation, machine learning and multi-model detection, our system provides a more comprehensive security analysis. This approach ensures that security professionals and developers receive accurate insights into vulnerabilities, significantly enhancing the security posture of web-based applications.

Methodology

The proposed system follows a structures approach that includes data collection, preprocessing, feature engineering, model training ,and evaluation. The key steps are as follows:

1. Dataset collection:

• publicly available datasets containing labelled web vulnerabilities are used .

• Data sources include penetration testing reports, web vulnerability databased, security research publications and real-world attack data collected from security logs.

• The dataset comprises various vectors, including injection-based attacks, scriptbased attacks, and authentication.

2. Data preprocessing: • Cleaning and normalization of raw data to remove inconsistencies and noise .

• Tokenization and vectorization of textual data for model input ,ensuring efficient processing.

• Handling imbalanced datasets using oversampling, under sampling and synthetic data generation techniques to improve model generalization.

feature scaling and encoding categorical variables to enhance model interpretability and accuracy.

3. Model Training and selection

• Various supervised learning algorithms such as Random Forest, Support Vector machine(SVM),Decision Tress and Deep Learning models are evaluated for effectiveness.

• The best-performing model is selected based on key performance metrics, including accuracy, precision, recall and F1-score.

4. Web-based tool development

• A user-friendly web interface is developed where users can enter a URL for security assessment.

• The backend processes web requests and responses ,extracting relevant features for classification and vulnerability assessment.

DOI:10.46647/ijetms.2025.v09si01.016 ISSN: 2581-4621

• The trained ML model predicts potential vulnerabilities and provides risk assessment feedback to users.
• Security recommendations and mitigation strategies are offered based on the detected vulnerabilities to help users secure their applications

5. Performance evaluation • The system is tested against datasets to validate its effectiveness.
• Metrics such as accuracy,precision,recall,F1-score and false positives rate are analysed to compare performance with vulnerability detection methods.
• Comparative analysis with traditional rule-based and signature-based detection systems is performed to highlight the advantages of ML-based approaches.

This methodology ensures a robust and automated vulnerability detection system, leveraging machine learning to enhance web application security. The integration of ML-based analysis within a web-based tool provides an accessible and effective
means for security professionals and developers to identify and web vulnerabilities proactively.

Results

The prepared machine learning show was assessed for its viability in identifying web application vulnerabilities. The key discoveries are as takes after :
1. Best approval execution accomplished was 81.75 exactness amid show training.
2. Exactness and review values demonstrated ahigh discovery rate with negligible wrong positives, illustrating the strength of the model.
3. The classification show accomplished an F1-score of 0.973,highlighting its solid prescient ability.
4. Comparative examination with conventional approaches appeared a noteworthy advancement in helplessness discovery precision, diminishing untrue positives and negatives.

Conclusion

The inquire about illustrates the adequacy of machine learning in upgrading web application security by identifying and classifying vulnerabilities. The proposed ML – based show essentially makes strides upon conventional security measures by giving higher precision, diminished untrue positives, and mechanized defenselessness appraisal. By leveraging directed learning strategies, the show successfully recognizes security dangers, counting SQL infusion and XSS assaults, advertising an progressed approach to proactive web security.

The integration of a web-based instrument advance guarantees openness for security experts and designers, empowering real-time security evaluations. The execution assessment measurements highlight the model's vigor, with solid approval scores and relapse investigation showing tall relationship with real-world powerlessness information.

Future work will center on growing the dataset, joining extra assault sorts, and refining the demonstrate to adjust to advancing cyber dangers. Moreover, coordination fortification learning and irregularity location methods can advance progress prescient capabilities. This think about contributes to the field of cybersecurity by illustrating how machine learning can be viably utilized to improve the security pose of web applications.

References

[1] Karishma Rahman and Clemente Izurieta, " A Mapping Consider of Security Defenselessness Disclosure Approaches for Web operations ".
[2] Haibo Chen, Junzuo Chen, Jinfu Chen, Shang Yin, Yiming Wu and Jiaping Xu, " An Programmed Helplessness Scanner for Web operations ".
[3] Wang Gaolong and Li Yongzhen, " Plan and execution of a web operation helplessness disclosure framework ".
[4] Ivan Kovacevic, Mihael Marovic, Stjepan Gros and Marin Vukovic, " Anticipating Vulnerabilities in Web operations predicated on Site Security Show ".
[5] Chandershekhar Sharma, S.C. Jain and Arvind K Sharma, " inconvenience predicated

DOI:10.46647/ijetms.2025.v09si01.016 ISSN: 2581-4621

Quantitative Investigation of SQLIA on Web operation Database ".
[6] Wei, K., Muthuprasanna, M., & Kothari, S.(2006, April). avoiding SQL infusion assaults in put away methods. In Australian Computer program Designing Conference(ASWEC' 06)(pp. 8- pp). IEEE.
[7] Clarke- Salt, J.(2009). SQL infusion assaults and defense. Elsevier.
[8] Shar, L. K., Briand, L. C., & Tan, H. B. K.(2014). Web operation powerlessness prediction utilizing crossbred program investigation and machine information. IEEE Bargains on tried and true and secure computing, 12(6), 688- 707.
[9] Khalid, M. N., Farooq, H., Iqbal, M., Alam, M. T., & Rasheed, K.(2019). Anticipating web vulnerabilities in web operations predicated on machine information. In Cleverly Innovations and Applications To begin with Worldwide Conference, INTAP
[10] Gupta, S., & Gupta, B. B.(2017). Cross-Site Scripting(XSS) assaults and defense instruments sort and state- of- the- craftsmanship. Worldwide Diary of Framework Confirmation Building and Administration, 8, 512- 530.
[11] Dina Hussein, Dina M. Ibrahim, Mona Alsalamah. A Review Study on SQL Injection attacks,Prevention, and Detection.
[12] Mehul Singh, Prabhishek Singh, Pramod Kumar. An Analytical Study on Cross-Site Scripting.
[13] Hsing-Chund Chen, Aristophane Nshimiyimana, Cahya Damarjati, Pi-Hsien Chang. Detection and Prevention of Cross-Site Scripting Attack with Approaches.
[14] Kashish Gaur,Manoj Diwakar, Kaamya Gaur, Prabhishek Singh, Tanya Sachadeva, Neeraj Kumar Pandey. SQL Injecton Attacks and Prevention.
[15] Sajjad Rafique, Mamoona Humayun, Bushra, Ansar Abbas, Muhammad Akhtar, Kamil Iqbal. Web Application Security Vulnerabilities detection approaches: A Systematic mapping Study.