International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 3 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

MACHINE LEARNING FOR WEB VULNERABILITY DETECTION: THE CASE OF CROSS-SITE REQUEST FORGERY

[1] Rathlavath Sindhuja ,[2]Dr.Sukanya K

[1]M. Tech Student, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.

[1]Email : sindhujarathod123@gmail.com

[2]Professor, Department of Computer Science and Engineering, TKR College of Engineering & Technology, Meerpet, Balapur, Saroornagar, Hyderabad - 500097, Telangana.

[2]Email : sukanya.addagatla@gmail.com

ABSTRACT

Web applications have become an essential part of modern digital services, supporting activities such as online banking, e-commerce, healthcare, education, and government operations. However, the increasing dependence on web applications has also led to a rapid rise in cyberattacks that exploit security vulnerabilities, resulting in data breaches, financial losses, and unauthorized access to sensitive information. Traditional vulnerability detection methods, including manual code reviews and rule-based security scanners, often require significant time, expertise, and computational resources, while struggling to identify newly emerging or complex attack patterns. To address these challenges, this project presents a Machine Learning-based approach for Web Vulnerability Detection that automatically analyzes web application features and classifies them as secure or vulnerable. The proposed system involves collecting vulnerability-related datasets, performing data preprocessing to remove inconsistencies, extracting meaningful security features, and training supervised machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes to recognize patterns associated with common web vulnerabilities, including SQL Injection, Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), Command Injection, and File Inclusion attacks. The trained model evaluates incoming data, predicts potential security threats with high accuracy, and enables early detection of vulnerabilities before they can be exploited by attackers. By continuously learning from historical data, the system enhances its capability to detect both known and previously unseen attack behaviors, thereby improving the overall effectiveness of web security. The proposed framework reduces manual effort, increases detection speed, minimizes false alarms, and provides a scalable and cost-effective solution for securing modern web applications.

Keywords: Machine Learning, Web Vulnerability Detection, Cybersecurity, Web Application Security, SQL Injection, Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), Intrusion Detection, Feature Extraction, Vulnerability Classification, Artificial Intelligence, Security Analysis, Threat Detection, Supervised Learning, Network Security.

I. INTRODUCTION

The rapid advancement of internet technologies has made web applications an essential component of modern digital services, supporting various domains such as online banking, e-commerce, healthcare, education, social networking, and government platforms. These applications handle large amounts of sensitive information, making them attractive targets for cybercriminals. The increasing number of web-based attacks has created significant

security challenges, where attackers exploit vulnerabilities in application logic, input validation, authentication mechanisms, and data handling processes. Common web vulnerabilities such as SQL Injection (SQLi), Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), Command Injection, and File Inclusion attacks can result in unauthorized access, data breaches, financial damage, and loss of user trust. Traditional vulnerability detection techniques, including manual code inspection, penetration testing, and signature-based security scanners, are widely used for identifying security flaws; however, these methods require significant time, expert knowledge, and continuous updates to detect newly evolving threats. Furthermore, conventional approaches often fail to identify complex attack patterns and zero-day vulnerabilities due to their dependency on predefined rules and signatures. To address these limitations, Machine Learning (ML) techniques have emerged as an effective solution for automated and intelligent vulnerability detection. Machine learning models can analyze large volumes of security data, identify hidden patterns, and make accurate predictions by learning from previous vulnerability examples. The proposed **Machine Learning for Web Vulnerability Detection** system focuses on developing an automated framework that collects web application data, preprocesses the information, extracts relevant security features, and applies machine learning algorithms to classify web requests as safe or vulnerable. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes are utilized to detect different types of web attacks and improve the accuracy of vulnerability identification. By combining cybersecurity concepts with machine learning capabilities, the proposed system reduces manual effort, improves detection speed, minimizes false alarms, and provides an adaptive approach for handling emerging cyber threats. This intelligent vulnerability detection framework can support developers, security analysts, and organizations in improving the security of modern web applications and protecting valuable digital assets from sophisticated attacks.

II. LITERATURE SURVEY

1. Title: A Machine Learning Approach for Web Application Vulnerability Detection

Authors: M. Almomani, K. R. Choo, A. Al-Nemrat

Abstract:

The authors proposed a machine learning-based approach for detecting vulnerabilities in web applications by analyzing security-related features extracted from web traffic and application behavior. The study focused on improving the limitations of traditional vulnerability scanners that depend on predefined rules and signatures. Various machine learning algorithms were evaluated to classify web requests as either normal or malicious. The experimental results demonstrated that machine learning techniques can effectively identify common web attacks and improve vulnerability detection accuracy. The research highlighted the potential of automated learning-based security systems for enhancing web application protection.

2. Title: Deep Learning Based Vulnerability Detection in Source Code

Authors: S. Li, J. Cui, Y. Li

Abstract:

This research introduced a deep learning approach for identifying software vulnerabilities directly from source code. The authors used neural network models to automatically learn complex patterns related to vulnerable programming practices. Unlike traditional static analysis tools, the proposed method reduced dependency on manually designed rules and was capable of detecting hidden vulnerability patterns. The study showed that deep learning models can provide effective solutions for automated

vulnerability analysis and improve the accuracy of software security assessment.

3. Title: A Survey on Machine Learning Techniques for Cyber Security Applications

Authors: A. Buczak and E. Guven

Abstract:

The authors presented a comprehensive survey of machine learning techniques applied in cybersecurity domains, including intrusion detection, malware analysis, and vulnerability identification. The study discussed different supervised and unsupervised learning algorithms and their effectiveness in detecting cyber threats. The survey emphasized that machine learning provides adaptive capabilities compared with traditional security approaches by learning from previous attack data. The research concluded that ML-based cybersecurity systems can improve threat detection and response efficiency.

4. Title: Automated Detection of Web Vulnerabilities Using Machine Learning Algorithms

Authors: R. Shar and K. Tan

Abstract:

The authors proposed an automated framework for detecting web vulnerabilities using machine learning classification techniques. The system analyzed web application characteristics and user input patterns to identify potential security flaws. Different machine learning algorithms were tested to determine their effectiveness in vulnerability classification. The results indicated that machine learning models can significantly reduce manual security testing efforts and provide faster vulnerability identification compared with conventional approaches.

5. Title: Using Machine Learning to Detect SQL Injection Attacks

Authors: H. H. H. Thanh, M. H. Le, and T. Nguyen

Abstract:

This study focused on detecting SQL Injection attacks using machine learning models. The authors analyzed HTTP request patterns and extracted important features to train classification algorithms for identifying malicious queries. The proposed system was able to distinguish between normal user requests and SQL injection attempts with improved accuracy. The research demonstrated that machine learning-based methods can provide an efficient alternative for protecting web applications against database-related attacks.

III. EXISTING SYSTEM

The existing web vulnerability detection systems primarily rely on traditional security techniques such as manual code reviews, rule-based vulnerability scanners, signature-based detection methods, and penetration testing to identify security flaws in web applications. These approaches are capable of detecting known vulnerabilities by comparing application behavior against predefined rules or attack signatures. Popular tools can identify common attacks such as SQL Injection (SQLi), Cross-Site Scripting (XSS), and Cross-Site Request Forgery (CSRF); however, they require frequent updates to maintain their effectiveness against newly emerging threats. Manual security assessments are time-consuming, labor-intensive, and heavily dependent on the expertise of cybersecurity professionals, making them difficult to scale for large and complex web applications. Furthermore, conventional detection methods often generate high false positive and false negative rates, leading to inaccurate results and increased effort during vulnerability analysis. They also struggle to recognize unknown or zero-day attacks because they cannot adapt to evolving attack patterns without manual intervention. As modern web applications continue to grow in complexity and cyber threats become increasingly sophisticated, the limitations of traditional vulnerability detection systems highlight the need for more intelligent, automated, and adaptive security solutions capable of providing faster and more accurate detection.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent and automated **Machine Learning-based Web Vulnerability Detection** framework designed to overcome the limitations of traditional security assessment techniques. The system collects web application and network request data from reliable vulnerability datasets, preprocesses the data by removing inconsistencies and irrelevant information, and extracts meaningful security-related features for effective analysis. Supervised machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes are trained using labeled datasets to accurately distinguish between secure and vulnerable web requests. The trained model is capable of detecting common web vulnerabilities, including SQL Injection (SQLi), Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), Command Injection, and File Inclusion attacks, by identifying hidden patterns and anomalies within the data. Unlike conventional rule-based systems, the proposed approach continuously improves its detection capability through learning, enabling it to recognize both known and previously unseen attack patterns with higher accuracy.

V. SYSTEM ARCHITECTURE

The system architecture of the proposed **Machine Learning for Web Vulnerability Detection** framework consists of multiple interconnected phases that work together to automatically identify and classify security vulnerabilities in web applications. The architecture begins with the **Data Collection Layer**, where web application data, HTTP requests, source code information, and vulnerability datasets containing examples of both normal and malicious activities are collected from different sources. The collected data is then passed to the **Data Preprocessing Layer**, which performs operations such as data cleaning, handling missing values, removing irrelevant information, normalization, and transformation of raw data into a suitable format for machine learning analysis. After preprocessing, the system moves to the **Feature Extraction Layer**, where important security-related characteristics such as input patterns, request parameters, code behavior, URL structures, and attack signatures are identified and converted into meaningful numerical representations. These extracted features are provided to the **Machine Learning Model Layer**, where different supervised learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes are trained using labeled vulnerability datasets. During the training phase, the models learn the patterns associated with various web attacks including SQL Injection, Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), Command Injection, and File Inclusion attacks. Once trained, the optimized model is deployed in the **Vulnerability Detection Layer**, where it analyzes new web application inputs and predicts whether they are safe or vulnerable. The prediction results are forwarded to the **Alert and Reporting Layer**, which generates vulnerability reports containing detected threats, affected components, and possible security risks, allowing administrators and developers to take appropriate preventive actions. The architecture also includes a continuous learning mechanism where new vulnerability data can be added periodically to improve model performance and adapt to emerging cyber threats.

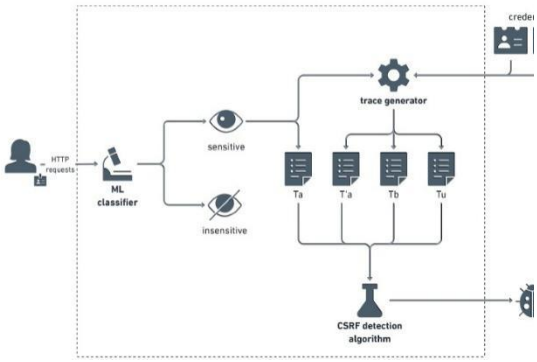

Fig 5.1: System Architecture

VI. IMPLEMENTATION

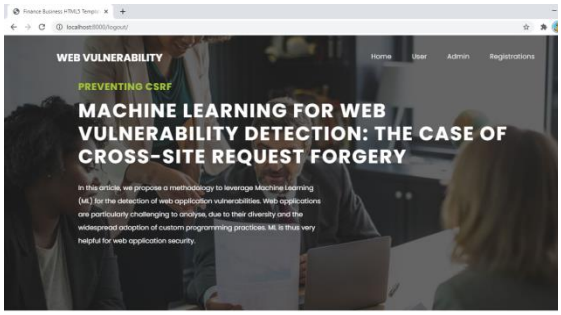

Fig 6.1 Index page

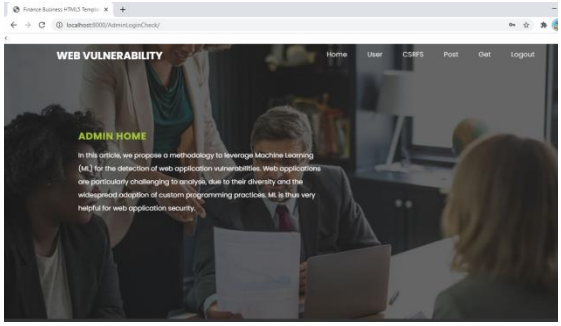

Fig 6.2: Admin home

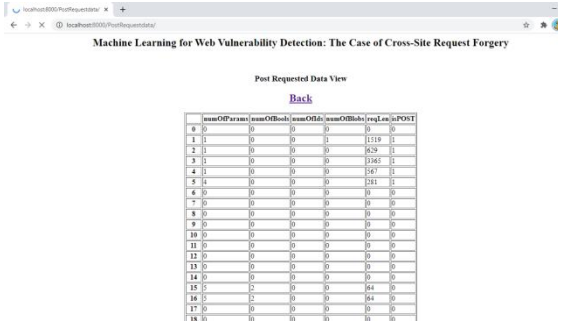

Fig 6.3: Upload dataset

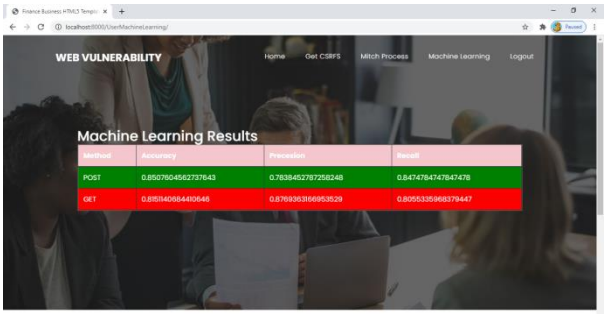

Fig 6.4: Model training

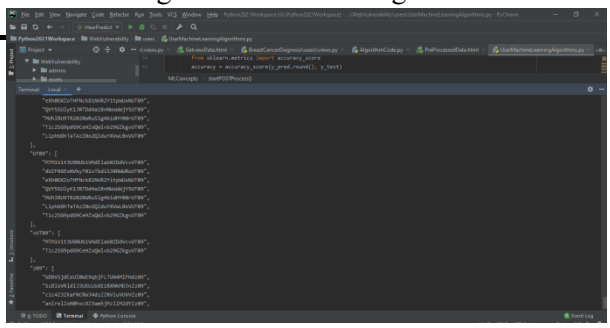

Fig 6.5: Performance

VII. CONCLUSION

The proposed **Machine Learning for Web Vulnerability Detection** system provides an intelligent and automated approach for identifying security vulnerabilities in modern web applications. Traditional vulnerability detection methods often depend on predefined rules, manual testing, and signature-based analysis, which makes them less effective against complex and evolving cyber threats. By integrating machine learning techniques, the proposed system can analyze web application data, extract important security features, and classify potential vulnerabilities with improved accuracy and efficiency. Machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes enable the system to learn patterns from existing vulnerability datasets and detect malicious activities such as SQL Injection, Cross-Site Scripting (XSS), Cross-Site Request Forgery (CSRF), and other web-based attacks. The system reduces manual effort, minimizes false detection rates, and provides faster vulnerability analysis compared to conventional approaches. Furthermore, its

ability to learn from new data makes it adaptable to emerging security threats. The proposed framework offers a scalable, reliable, and cost-effective solution for enhancing web application security across various domains, including banking, healthcare, e-commerce, education, and cloud-based services. Overall, the integration of machine learning with cybersecurity provides an effective mechanism for proactive vulnerability detection and strengthens the protection of web applications against future cyberattacks.

VIII. FUTURE SCOPE

The future scope of the Machine Learning for Web Vulnerability Detection system focuses on improving its intelligence, accuracy, and adaptability to handle advanced cybersecurity challenges. In the future, deep learning techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformer-based models can be integrated to detect more complex and previously unknown vulnerabilities with higher accuracy. The system can be enhanced by incorporating real-time web traffic monitoring and continuous vulnerability scanning capabilities to provide immediate threat detection and response. Advanced feature extraction techniques can be applied to analyze source code, API requests, and user behavior patterns for better vulnerability identification. The framework can also be extended with automated threat intelligence and continuous learning mechanisms, allowing the model to update itself with newly discovered attack patterns and zero-day vulnerabilities. Integration with cloud security platforms and Security Information and Event Management (SIEM) systems can improve large-scale deployment and monitoring capabilities. Additionally, explainable artificial intelligence (XAI) techniques can be introduced to provide detailed reasons behind vulnerability predictions, helping developers understand and fix security issues effectively. With these improvements, the proposed system can evolve into a more powerful, adaptive, and proactive cybersecurity solution capable of protecting modern web applications against emerging cyber threats.

IX. REFERENCES

[1] A. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[2] M. Almomani, K. R. Choo, A. Al-Nemrat, and A. Dehghantanha, "Cybersecurity Threat Intelligence: Challenges and Opportunities," *IEEE Access*, vol. 6, pp. 23056–23070, 2018.

[3] S. M. Shar, L. K. Tan, and K. Tan, "Web Application Vulnerability Detection Using Machine Learning Techniques," *Proceedings of the International Conference on Security and Privacy*, pp. 1–8, 2019.

[4] F. Yamaguchi, N. Golde, D. Arp, and K. Rieck, "Modeling and Discovering Vulnerabilities with Code Property Graphs," in *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 590–604, 2014.

[5] R. Russell, L. Kim, L. Hamilton, T. Lazovich, and J. Harer, "Automated Vulnerability Detection in Source Code Using Deep Learning," *IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pp. 1–10, 2018.

[6] S. Li, J. Cui, and Y. Li, "Deep Learning-Based Software Vulnerability Detection: A Survey," *IEEE Access*, vol. 8, pp. 182698–182710, 2020.

[7] H. H. Thompson, "Application Security Testing Using Machine Learning Approaches," *IEEE Security & Privacy Magazine*, vol. 15, no. 6, pp. 72–79, 2017.

[8] M. Vieira, N. Antunes, and H. Madeira, "Using Web Security Scanners to Detect Vulnerabilities in Web Applications," in *Proceedings of the IEEE International Symposium on Dependable Computing*, pp. 1–10, 2009.

[9] P. W. Fong, "A Privacy Preservation Model for Web Applications Using Security Analysis Techniques," *IEEE Transactions on Dependable and Secure Computing*, vol. 7, no. 3, pp. 247–260, 2010.

[10] A. Sharma and P. K. Singh, "Machine Learning Approaches for Cybersecurity Intrusion Detection: A Review," *IEEE Access*, vol. 9, pp. 141–159, 2021.

[11] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[12] C. Kruegel and G. Vigna, "Anomaly Detection of Web-Based Attacks," in *Proceedings of the 10th ACM Conference on Computer and Communications Security*, pp. 251–261, 2003.

[13] Y. Li, S. Wang, and T. Xie, "Vulnerability Detection in Source Code Using Machine Learning Models," *IEEE Transactions on Software Engineering*, vol. 47, no. 6, pp. 1120–1135, 2021.

[14] N. Idika and A. Mathur, "A Survey of Malware Detection Techniques," *Purdue University Technical Report*, 2007.

[15] E. Cambiaso, G. Papaleo, G. Chiola, and M. Aiello, "Network Traffic Analysis and Machine Learning for Cybersecurity Threat Detection," *IEEE Access*, vol. 7, pp. 133103–133117, 2019.