IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning-Based Technique to Detect SQL Injection Attack

[1]Vishal Sherkhane, [2]Aamir Tadasarkar,[3] Akshay Shivpure, [4]Suresh Rathod, [5]Prof. Mrs Vina M. Lomte
[1]Student, [2]Student, [3]Student, [4]Student, [5]guide
[1]RMD Sinhgad School of Engineering, Pune,
[2]RMD Sinhgad School of Engineering, Pune,
[3]RMD Sinhgad School of Engineering, Pune,
[4]RMD Sinhgad School of Engineering, Pune,
[5]RMD Sinhgad School of Engineering, Pune

Abstract:-

Lack of secure codes implemented in the web apps will lead to cyber- attack because of vulnerabilities. The statistic shows that highest record on the data theft related cyber-attacks are through the SQL injection technique. Hence, an effective SQL injection detection is needed in any web system to combat this threat. In this research, machine learning technique is used where training is provided to the SQL injection detector using a training data and then is evaluated against a testing data. The research relies on the preparation of the training and testing datasets. Training sets are used by the detector to establish the knowledge base and the test set is used to evaluate the performance of the detector. The result of the detection shows that the proposed technique produces high accuracy in recognizing malicious and begin web requests.

Keyword :- Machine Learning, Signature-Based, Knowledge-Based, SQL Injection,

SQL Injection Tools

Introduction

As data is the most critical asset to any organization nowadays, the rise of cyber threat and cyberattack to the organisation's database is increasing. Hackers are the culprit and threat to data privacy and as an example, they could launch an SQL Injection Attack (SQLIA) against vulnerable websites. Furthermore, there are many existing tools that can be used to check a website's vulnerabilities and execute hacking activities automatically. These tools give an attacker more chance of getting into the web system database.

If a web developer has no proper security knowledge, he/she will likely develop codes that contain vulnerabilities. It is hard to implement secure odes to defend websites against such attacks. Because of that, the systems they developed are vulnerable to SQL injection attacks. SQL injection is a type of attack to manipulate the website to disclose sensitive data by injecting malicious SQL queries to the database. Hence, I the SQL injection can be recognised earlier, it can hel security officer or security analyst to terminate the attack.

Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) can be used to detect SQL injections (Patil et al., 2017). In this research, a machine learning technique is used to detect an SQL injection attack by comparing the website access log file with the knowledge-based of malicious features.

Machine learning part will undergo some training and which will then be used to scan the log for classifying where the log is being injected or not. The classification will result in a malicious or benign web request

Literature:-

SQL attact Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Intelligence SQL intelligence is the provision of evidence-based knowledge about existing or potential Intrusion. Benefits of threat intelligence include improved efficiency and effectiveness in security operations in terms of detective and preventive capabilities.

Successful threat intelligence within the cyber domain demands a knowledge base of threat information and an expressive way to represent this knowledge. This purpose is served by the use of taxonomies, sharing standards, and ontologies. This paper introduces the Cyber Threat Intelligence (CTI) model, which enables cyber defenders to explore their threat intelligence capabilities and understand their position against the ever-changing cyber threat landscape.

Probabilistic Threat Detection for Risk Management in Cyber-physical Medical Systems
edical devices are complex cyber-physical systems exposed to numerous security risks and vulnerabilities.

This article presents a dynamic risk management and automated threat mitigation approach based on a probabilistic threat estimation framework. A smart connected pacemaker case study illustrates the approach

Managing cyber threat intelligence in a graph databaseEfforts to cope jointly with the ever-increasing number of breach incidents have resulted in the establishment of the standard format and protocol and given birth to many consultative groups. In addition, various

channels that distribute This paper also proposes a method of supporting the detection provided by existing security equipment with the information saved in the graph database and an effective method of analysis.

Lastly, the paper discusses the advantages that can be expected from saving cyber threat information in the graph database developed using information Collected From the outside.
Cyber Threat Intelligence information free of charge have emerged, and studies on utilizing such channels have spread.

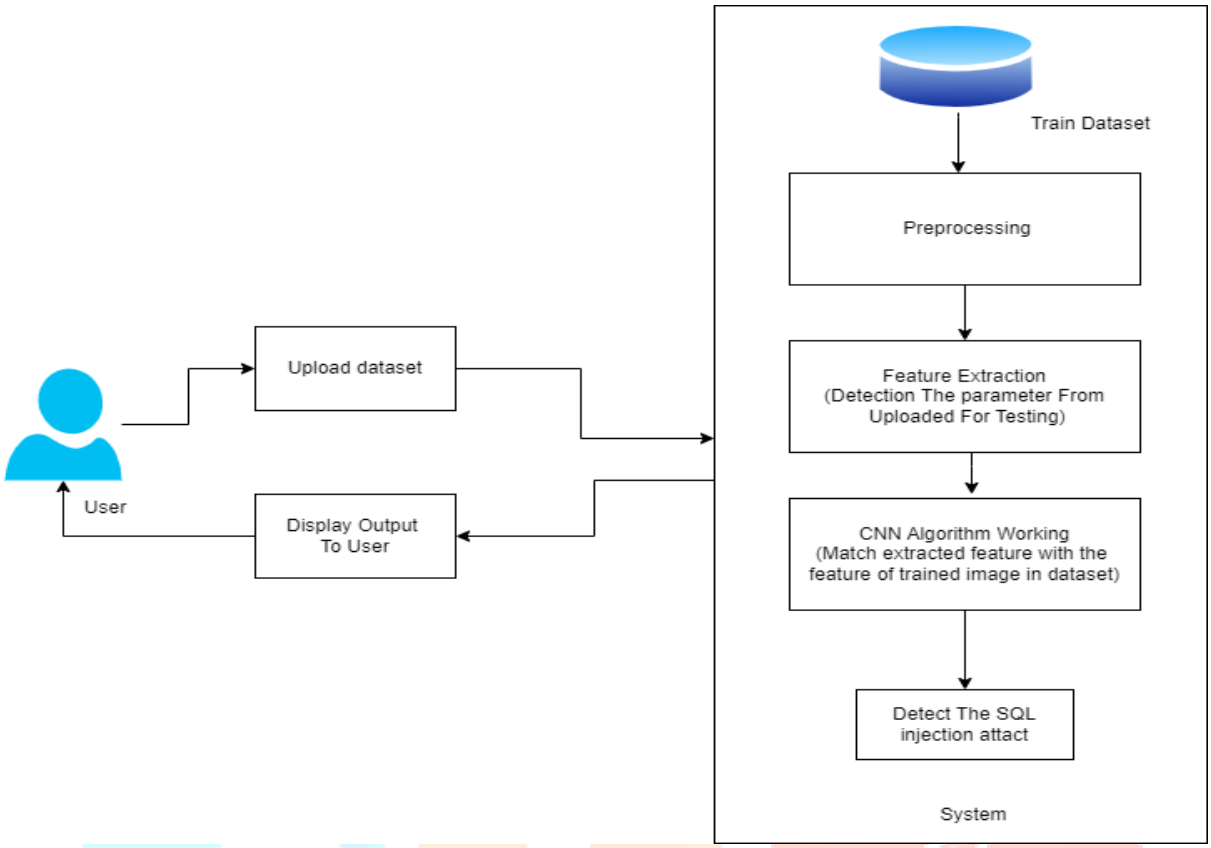
As the market for sharing information professionally is expanding, the need to manage the shared information in various ways in order to achieve better result has arisen. This paper proposes a standardized management structure and method based on the standardized format and a meaning and standard of Cyber Threat Intelligence that can be shared outside when loading OS INT information collected from various channels into the graph database

A Design of IL-CyTIS for Automated Cyber Threat Detection As cyber squabbling has been intensified, the necessity of sharing cyber threat information has increased. Therefore, attempts to develop a technology to upgrade and Machineen the related system will continue. In particular, it is anticipated that automated response and analysis using machine learning will be actively

conducted.

With the security situation in Cyberspace constantly becoming worse, Cyber threat detection has attracted a lot of researching attentions. In this paper, existing detection technologies are firstly reviewed. Secondly, a framework of capturingthe abnormal traffic of botnets is proposed.

System architecture:-



Algorithm :-

SVM:-

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

Types of SVM

SVM can be of two types:

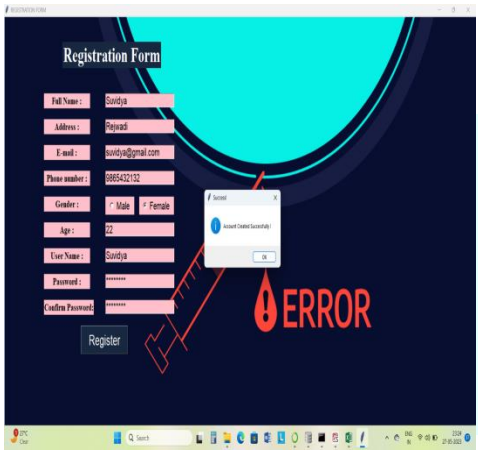
- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Result:-

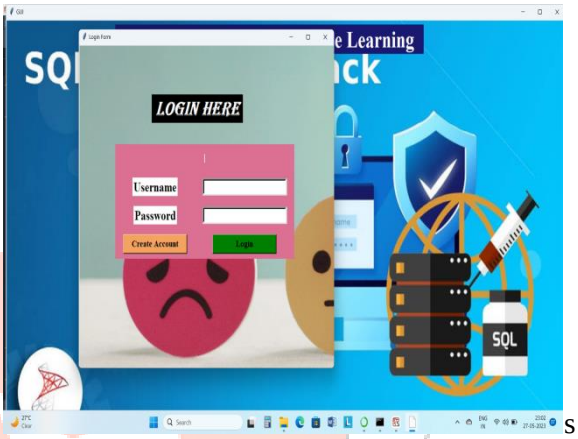
Implementation steps :

- Login
- Registration
- View and Authorize Users

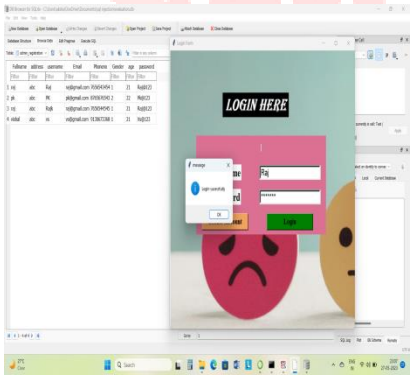
  In this module, the admin can view the list of users who all registered. In this,the admin can view the user's details such as, user name, email, address and admin authorizes the users.
- View Charts Results
- View All Products Search Ratio,View All Keyword Search Results,View AllProduct Review Rank Results.



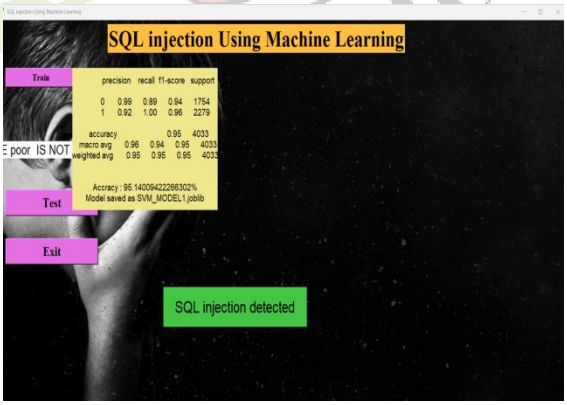
Step 1



Step 2

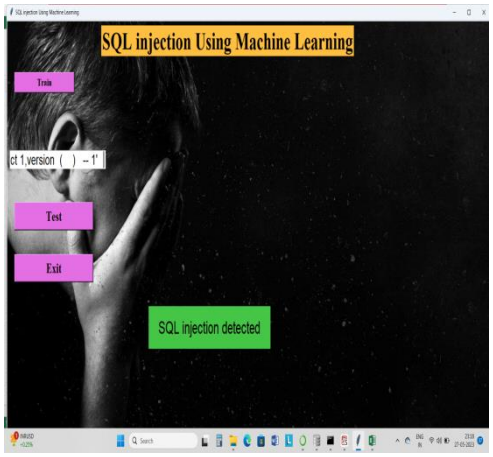


Step 3

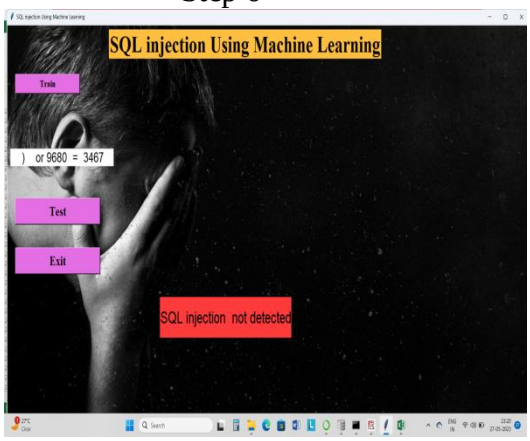


Step 4

Step 5



Step 6



Conclusion :-

This report helps machine learning to detect malicious and benign web requests de- rived from the access log files, which has successfully detected malicious log files. In addition, string matching is used to match the features in the classification phase. The main constraint of SQLIA research is to acquire reputable and suitable dataset on- line. Therefore, data collection is developed in-house, by setting up a simple login website and perform SQL Injection attacks. Fortunately, there are platform such as DVWA that can be used to perform injections to creates datasets. As a result, only a few
samples of SQL injection dataset can be used for training and testing.

This research can be enhanced and improved further by implementing detection in real time, where detection of SQL injections can be discovered and stop faster before any damage inflicts the system. In addition, detecting the web request by session can improve the accuracy of the detector

Future scope:-

As future work, we want to evaluate methods using different web based application
script with public domain to achieve great accuracy in SQL injection prevention
approaches. Integrate SQLiX with nikto HTTP scanner, HTTP scanning proxies, and
with metasploit will helps to detect other web vulnerabilities. Also add feature to dumpvenerable database and database schema.

Reference :-

1 C. R. Srinivasan, B. Rajesh, P. Saikalyan, K. Premsagar, and E. S. Yadav, “A review on the different types of INetwork (Intrusion network),” J. Adv. Res. Dyn. Control Syst., vol. 11, no. 1, pp. 154–158, 2019.

2 G. J. Joyia, R. M. Liaqat, A. Farooq, and S. Rehman, “Internet of Medical Things (IOMT): Applications, benefits and future challenges in healthcare do- main,” J. Commun., vol. 12, no. 4, pp. 240–247, 2017.

3 A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “INetwork for smart cities,” IEEE Internet Things J., vol. 1, no. 1, pp. 22–32, Feb. 2014.

4 E. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, “Android malware detection using Machine learning on API method sequences,” Dec. 2017, arXiv:1712.08996. [Online]. Available: https://arxiv.org/abs/1712.08996

5 S. Jabbar, K. R. Malik, M. Ahmad, O. Aldabbas, M. Asif, S. Khalid, K. Han, and S. H. Ahmed, “A methodology of real-time data fusion for localized big data analytics,” IEEE Access, vol. 6, pp. 24510–24520, 2018.

6 F. Ullah, J. Wang, M. Farhan, M. Habib, and S. Khalid, "Software plagiarism detection in multiprogramming languages using machine learning approach," Concurrency Comput., Pract. Exper., to be published.

7 D.-K. Chae, J. Ha, S.-W. Kim, B. Kang, and E. G. Im, "Software plagiarism detection: A graph-based approach," in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage., Nov. 2013, pp. 1577–1580.

8 Y. Akbulut and O. D¨onmez, "Predictors of digital piracy among Turkish undergraduate students," Telematics Inform., vol. 35, no. 5, pp. 1324–1334,

9 M. ShanmughaSundaram and S. Subramani, "A measurement of similarityto identify identical code clones," Int. Arab J. Inf. Technol., vol. 12, pp. 735–740,Dec. 2015.

10 C. Ragkhitwetsagul, "Measuring code similarity in large-scaled code Cor- pora," in Proc. IEEE Int. Conf. Softw. Maintenance Evol. (ICSME), Oct. 2016, pp. 626–630.