International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net
Vol-20 Issue-01 April 2025

SQLINJECTIONATTACKS: LEVERAGING AI FOR CLASSIFICATION AND RESPONSE STRATEGIES

Sreeya siddha[1],D.Abhitarun Reddy[2], Karigari Bharath kumar[3], A.Bhasha[4]
[1,2,3] UG Scholar, Dept. of IT ,St. Martin's Engineering College,Secunderabad,Telangana,India,500100
[4]Assistant professor, Dept. of IT, St. Martin's Engineering College, Secunderabad, Telangana, India,500100
Siddhasreeya02@gmail.com

Abstract

SQL injection attacks pose a serious threat to databases as they exploit vulnerabilities in the database layer by injecting SQL codes into user databases. The consequences of a successful attack cause unauthorized access to databases and attackers can gain access to sensitive data. So to avoid the data breach and unauthorized access we have to detect whether the executed SQL code at the user side is malicious or not. Even though some methods like parameterized queries, escaping characters and input validation are some traditional techniques to detect SQL injection, they have their own limitations. These methods often rely on manual coding practices and may not identify new attacks. As attackers continuously evolve their techniques to attack and gain access to sensitive data there is a need for advanced solutions that can proactively identify and mitigate SQL injection attacks. AI has the capacity to analyse vast amounts of data, detect patterns, and learn from previous attacks. AI brings significant benefits to the prediction of SQL injection attacks. Its ability to detect anomalies, learn from new attack patterns, recognize complex patterns, reduce false positives, provide real-time protection, and scale to handle large applications makes it an indispensable tool. Here we use count vectorizer to create tokens and give these tokens to a Neural Network Algorithm i.e., Multi Layer Perceptron to detect the malicious SQL code. By leveraging artificial Intelligence, we can detect and mitigate malicious SQL codes swiftly and accurately to ensure the safety of databases.

KEYWORDS: Artificial Intelligence, Natural Language Processing, Prediction, Malicious SQL codes, Machine Learning, Threat Detection, SQL Injection, Neural Networks, count vectorizer, Multi layer Perceptron.

1. INTRODUCTION

SQL Injection is a type of cyber-attack that has been around for a long time. It involves injecting malicious SQL code into an application's input fields, which allows attackers to gain unauthorized access to the application's database. This can lead to severe consequences, such as data breaches and system compromises. In recent years, Artificial Intelligence (AI) and machine learning have become popular in various fields, including cyber security. The idea of using AI to predict SQL Injection attacks emerged to bolster security measures and counter sophisticated attack techniques. By developing AI models that can analyse application input data, we can identify patterns that indicate the presence of an SQL Injection attack. The traditional methods used to prevent SQL Injection attacks rely on simple rule-based approaches or static pattern matching. However, these methods can sometimes be bypassed by well-crafted attacks. This is where AI-based prediction of SQL Injection attacks becomes essential. We need AI-based prediction because cyber attackers continuously evolve their methods, making it challenging to rely solely on traditional approaches. AI-powered systems can process large amounts of data, discover hidden patterns, and adapt to new attack techniques, making them more effective in identifying SQL Injection attacks. The significance of AI-based prediction lies in its ability to enhance detection accuracy. AI models can learn from historical attack data and identify even subtle patterns that might go unnoticed by traditional methods. By doing so, they can reduce false positives, which helps minimize disruptions to legitimate user activities. Additionally, AI can serve as a proactive defence mechanism, continuously monitoring and protecting applications from potential threats, including novel and previously unseen SQL Injection attacks. Artificial Intelligence, particularly machine learning, has shown promise in various cyber security applications due to its ability to analyse vast amounts of data, detect patterns, and make predictions. Overall, the integration of AI-enabled NLP for predicting malicious SQL code represents a significant advancement in cyber security, offering organizations a powerful tool to defend against SQL injection attacks and other database-related threats.

2. LITERATURE SURVEY

SQL Injection is a type of cyber-attack that has been around for a long time. It involves injecting malicious SQL code into an application's input fields, which allows attackers to gain unauthorized access to the application's database. This can lead to severe consequences, such as data breaches and system compromises. In recent years, Artificial Intelligence (AI) and machine learning have become popular in various fields, including cyber security

Khaled elshazlyThe idea of using AI to predict SQL Injection attacks emerged to bolster security measures and counter sophisticated attack techniques. By developing AI models that can analyze application input data, we can identify patterns that indicate the presence of an SQL Injection attack. The traditional methods used to prevent SQL Injection attacks static pattern matching.

However, these methods can sometimes be bypassed by well crafted attacks. This is where AI-based prediction of SQL Injection attacks becomes essential. We need AI-based prediction because cyber attackers continuously evolve their methods, making it challenging to rely solely on traditional approaches

yaseerfouad AI-powered systems can process large amounts of data, discover hidden patterns, and adapt to new attack techniques, making them more effective in identifying SQL Injection attacks. The significance of AI-based prediction lies in its ability to enhance detection accuracy. AI models can learn from historical attack data and identify even subtle patterns that might go unnoticed by traditional methods. By doing so, they can reduce false positives, which helps minimize disruptions to legitimate user activities. Additionally, AI can serve as a proactive defense mechanism, continuously monitoring and protecting applications from potential threats, including novel and previously unseen SQL Injection attacks

Alghawazi et al applied techniques from different areas to detect and deterrence of SQL injection attacks, for which to improve the detect ability of the attack, is not a new area of research but it is still relevant. Artificial intelligence and machine learning techniques have been tested and used to control SQL injection attacks, showing promising results. The main contribution of this paper is to cover relevant work related to different machine learning and deep learning models used to detect SQL injection attacks.

Uwagbole et al explored the generation of data set containing extraction from known attack patterns including SQL tokens and symbols present at injection points. Also, as a test case, this work build a web application that expects dictionary word list as vector variables to demonstrate massive quantities of learning data. The data set is pre- processed, labelled and feature hashing for supervised learning. This paper demonstrated a full proof of concept implementation of an ML predictive analytics and deployment of resultant web service that accurately predicts and prevents SQLIA with empirical evaluations presented in Confusion Matrix (CM) and Receiver Operating Curve (ROC).

Gandhi et al proposed a hybrid CNN-BiLSTM based approach for SQLI attack detection. The proposed CNN-BiLSTM model had significant accuracy of 98% and superior performance compared to other machine learning algorithms. Also, paper presented a comparative study of different types of machine learning algorithms used for the purpose of SQLI attack detection. The study showed the performance of various algorithms based on accuracy, precision, recall, and F1 score with respect to proposed CNN-BiLSTM model in detection of SQL injection attacks.

Ali studied the top 10 security threats identified by the OWASP are injection attacks. The most common vulnerability is SQL injection and is the most dangerous security vulnerability due to the multiplicity of its types and the rapid changes that can be caused by SQL injection and may lead to financial loss, data leakage, and significant damage to the database, and this causes the site to be paralyzed. Machine learning is used to analysed and identified security vulnerabilities. It used classic machine learning algorithms and deep learning to evaluate the classified model using input validation features.

Sharma et al used various classification algorithms to determine whether a particular code is malicious or plain. Some of the neural network and machine learning algorithms are Naive Bayes classifier, LSTM, MLP, and SVM which can be used for the detection of SQL Injection attacks..

Roy et al penetrated the logical section of the database. If the database has a logical flaw, the attackers send a new type of logical payload and get all of the user's credentials. Despite the fact that technology has advanced significantly in recent years, SQL injections can still be carried out by taking advantage of security flaws.

Falor et al and existing techniques for the detection of SQL like Naïve Bayes, Decision trees, Support Vector Machine, and K-nearest neighbour. This work have also analysed the performance of Convolutional Neural Networks (CNN)

Tripathy et al investigated the potential of using machine learning techniques for SQL injection detection on the application code or not. The results showed that these algorithms can distinguish normal payloads from malicious payloads with a detection rate higher than 98%.

Hubskyi et developed a neural network model for identifying SQL injection attacks based on HTTP request analysis. The model allowed classifying URL values by attributing them into one of two classes: attack or normal activity. An additional advantage is the provision of a quantitative identification value which describes the predicted accuracy of SQL injection determination.

Tang et al presented a high accuracy SQL injection detection method based on neural network. This work first acquired authentic user URL access log data from the Internet Service, designed eight types of features, and train an MLP model. The accuracy of the model maintained over 99%. Meanwhile, this work compared and evaluated the training effect of other machine learning algorithms (LSTM, for example), the results revealed that the accuracy of this method is superior to the relevant machine

3. PROPOSED METHODOLOGY

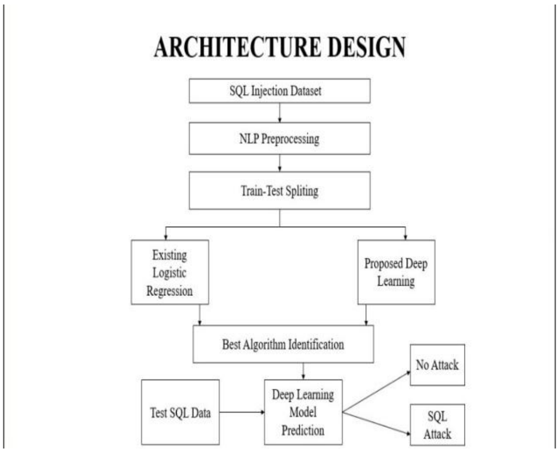


Figure 1: Architecture Design

Step 1: SQLi Dataset The project begins by uploading a SQL Injection (SQLi) dataset that contains SQL statements labeled as either "attack" or "no attack." This dataset is read using pandas, and basic information such as the dataset structure and a preview of the first few rows is displayed.

Step 2: Text Pre processing In this step, data cleaning and preparation are performed. The SQL statements are tokenized, stop words are removed, and other text cleaning steps like stemming or lemmatization might be applied. The dataset is then split into training and testing sets for model training and evaluation.

Step 3: Vectorization Here, the SQL statements are transformed into numerical form using the Count Vectorizer, which converts the text into token counts, creating feature vectors that represent the sentences. This is necessary for machine learning algorithms to process the textual data.

Step 4: The Logistic Regression classifier is used as an existing method to detect SQL injection attacks. The training data is used to fit the model, and the test data is used to evaluate its performance. Accuracy and a confusion matrix are generated to understand the classification results.

Step 5: Simple Neural Network (Proposed Algorithm)A custom neural network (Multilayer Perceptron) is built with layers designed for SQL injections

Step 6: Performance Comparison The performance of both the Logistic Regression classifier and the Simple Neural Network is compared. Accuracy scores and confusion matrices are analyzed to determine which model performs better

Step 7: Prediction with Trained ModelThe trained Simple Neural Network model is used to predict whether new SQL statements are atacks or not. The predictions are displayed, classifying each test data

Disadvantages:

Linear Decision Boundary: Logistic Regression can only separate data that is linearly separable. It struggles with more complex relationships between input features.

Sensitive to Imbalanced Data: When one class significantly outnumbers the other, Logistic Regression may become biased toward the majority class.

Limited Expressiveness: It cannot capture complex, non-linear patterns in data, which limits its performance on more advanced classification tasks.

Architecture:

Input Layer: Takes the feature vector (vectorized SQL statements).
Hidden Layers : One or more layers with several neurons. Each neuron applies a weighted sum and an activation function.

Output Layer: For binary classification, the output layer has a single neuron with a sigmoid activation function, giving a probability of the input being an attack or not.

Advantages:

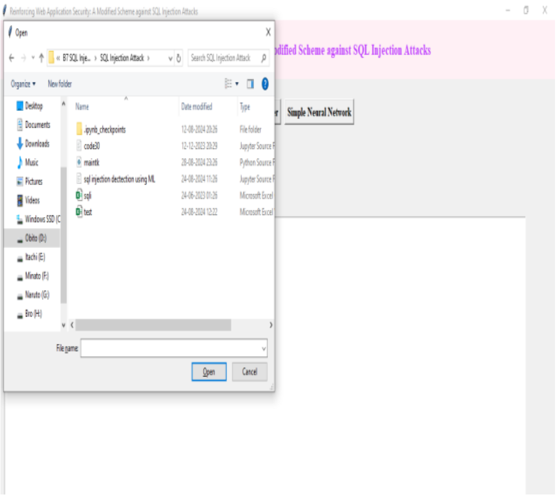
Captures Non-Linear Patterns: Unlike Logistic Regression, neural networks can model complex relationships between inputs and outputs.

Improved Accuracy: Neural networks generally offer higher accuracy, especially for complex tasks like SQL injection detection.

Scalability: As more data and computational resources ae available, neural networks can be scaled up by increasing the number of layers and neurons to improve performance

4. EXPERIMENTAL ANALYSIS

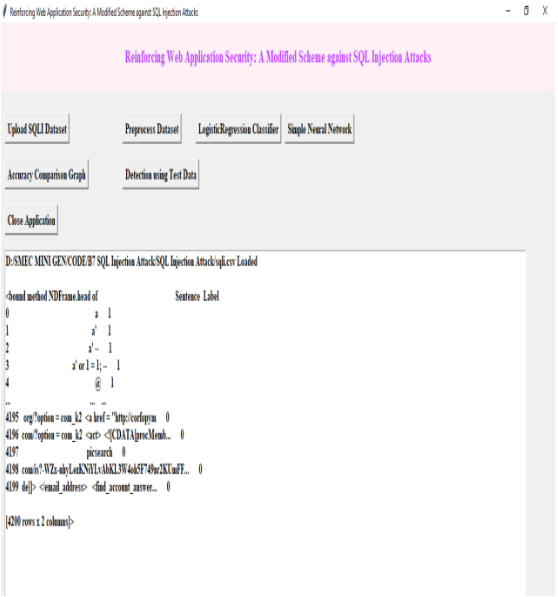
after uploading the dataset we get a description about all the rows and columns in the dataset as show we have 4200 rows and 2 columns each row has unique sql code be predicted either malicious or not.



Reinforcing Web Application Security: A Modified Scheme against SQL Injection Attacks

Name	Date modified	Type
.ipynb_checkpoints	12-08-2024 20:26	File folder
code30	12-12-2023 20:29	Jupyter Source F
maintk	28-08-2024 23:26	Python Source F
sql injection dectection using ML	24-08-2024 11:26	Jupyter Source F
sqli	24-06-2023 01:26	Microsoft Excel
test	24-08-2024 12:22	Microsoft Excel

Figure 1: Upload dataset


Reinforcing Web Application Security: A Modified Scheme against SQL Injection Attacks
Reinforcing Web Application Security: A Modified Scheme against SQL Injection Attacks
Upload SQLI Dataset
Preprocess Dataset
LogisticRegression Classifier
Simple Neural Network
Accuracy Comparison Graph
Detection using Test Data
Close Application

```
D:/SMEC MINI GEN/CODE/B7 SQL Injection Attack/SQL Injection Attack/sqli.csv Loaded

<bound method NDFrame.head of                          Sentence  Label
0                                   a    1
1                                  a'    1
2                                 a' --  1
3                          a' or 1 = 1; --   1
4                                   @    1
...                               ...  ...
4195  org/?option = com_k2 <a href = "http://corfopym     0
4196  com/?option = com_k2 <act> <![CDATA[procMemb...     0
4197                         picsearch    0
4198  com/is?-WZx-uhyLezKNiYLvAbKL3W4oh5F749ur2KUmFF...    0
4199  de]]> </email_address>  <find_account_answer...    0

[4200 rows x 2 columns]>
```

Figure 2: Description of dataset

Now we have to pre process the text where we remove all the spaces stop words and convert the sql code in to binary data by using count vectorizer and classifies the data using plot which displays the number malicious or not Also we get data which is spitted in to train data and test data.

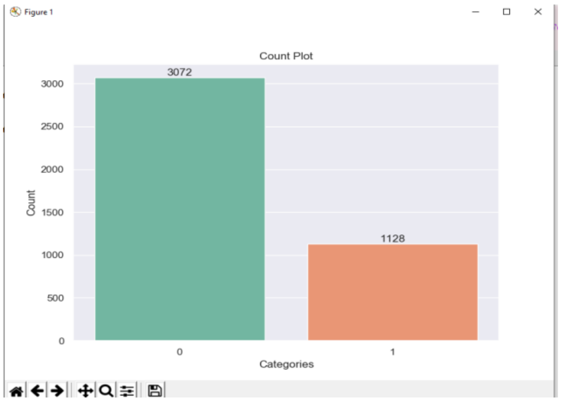
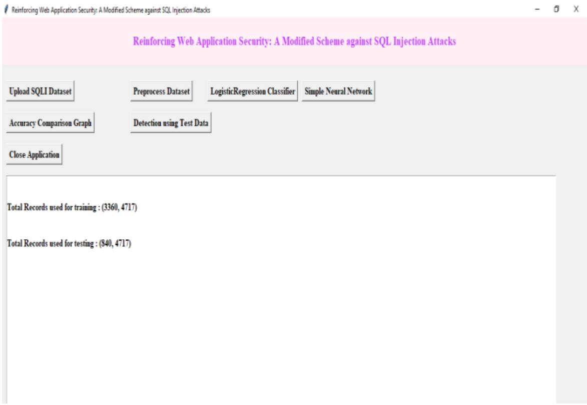


Figure 3: Count Plot for the dataset

Also we get data which is splitted in to train data and test data


Reinforcing Web Application Security: A Modified Scheme against SQL Injection Attacks
Reinforcing Web Application Security: A Modified Scheme against SQL Injection Attacks
Upload SQLI Dataset
Preprocess Dataset
LogisticRegression Classifier
Simple Neural Network
Accuracy Comparison Graph
Detection using Test Data
Close Application

```
Total Records used for training : (3360, 4717)

Total Records used for testing : (840, 4717)
```

Figure 4:Data splitting

Now train the model using logistic regression classifier

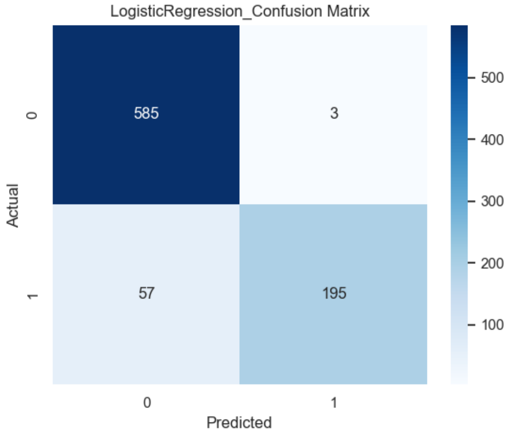


Figure 5: Neural Network confusion matrix

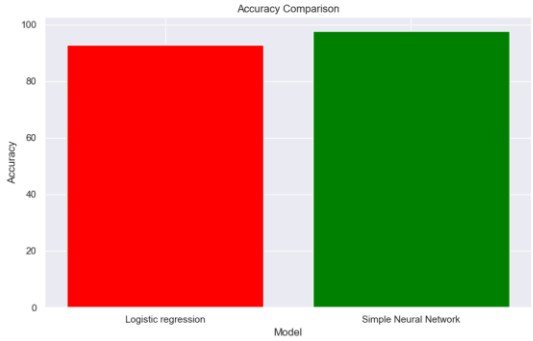
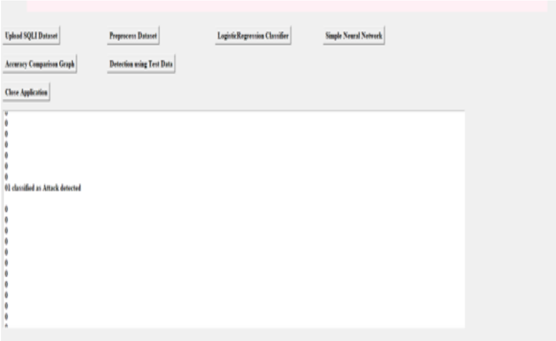


Figure 6: Accuracy comparison

Now test the dataset whether it is attacked or not


Upload SQLI Dataset
Preprocess Dataset
LogisticRegression Classifier
Simple Neural Network
Accuracy Comparison Graph
Detection using Test Data
Close Application

```
0
0
0
0
0
0
01 classified as Attack detected

0
0
0
0
0
0
0
0
0
0
0
```

Figure 7: Output for dataset

5. CONCLUSION

In conclusion, SQL injection attacks pose a significant threat to web applications, potentially leading to unauthorized access, data breaches, and complete compromise of the application and underlying database. While traditional methods such as input validation and parameterized queries offer some level of protection, they have limitations and may not cover all attack vectors. Therefore, this work implemented ANN to proactively identify and mitigate SQL injection attacks. It can detect anomalies and learn from new attack patterns, which enables it to recognize complex attack vectors that traditional methods might miss. Furthermore, it can reduce false positives, provide real-time protection, and scale to handle large applications efficiently. These capabilities make AI an indispensable tool in defending against SQL injection attacks.

REFERENCES

[1]Martins, N.; Cruz, J.M.; Cruz, T.; Abreu, P.H. Adversarial Machine Learning Applied to Intrusion and Malware Scenarios: ASystematic Review. IEEE Access 2020,8,35403–35419.

[2]Mishra, P.; Varadharajan, V.; Tupakula, U.; Pilli, E.S. A Detailed Investigation and Analysis of using Machine Learning Techniques for Intrusion Detection. IEEE Commun.Surv.Tutor.2018,21,686–728.

[3]Yan, R.; Xiao, X.; Hu, G.; Peng, S.; Jiang, Y. New deep learning method to detect code injection attacks on hybrid applications.J.Syst.Softw.2018,137,67–77.

[4]Aliero, M.S.; Qureshi, K.N.; Pasha, M.F.; Ghani, I.; Yauri, R.A. Systematic Review Analysis with SQLIA Detection and PreventionApproaches. Wirel. Pers. Commun.

[5]Alghawazi, Maha &Alghazzawi, Daniyal &Alarifi, Suaad. (2022). Detection of SQL Injection Attack Using Machine Learning Techniques: A Systematic Literature Review. Journal of Cybersecurity and Privacy. 2. 764-777. 10.3390/jcp2040039.

[6]W. Zhang, Y. Li, X. Li, M. Shao, Y. Mi, H. Zhang, G. Zhi, "Deep Neural Network-Based SQL Injection Detection Method", Security and Communication Networks, vol. 2022, Article ID 4836289, 9 pages, 2022. https://doi.org/10.1155/2022/4836289

[7]S. O. Uwagbole, W. J. Buchanan and L. Fan, "Applied Machine Learning predictive analytics to SQL Injection Attack detection and prevention," 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), Lisbon, Portugal, 2017, pp. 1087-1090,doi:10.23919/INM.2017.7987433.

[8]N. Gandhi, J. Patel, R. Sisodiya, N. Doshi and S. Mishra, "A CNN-BiLSTM based Approach for Detection of SQL Injection Attacks," 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE),Dubai,UnitedArabEmirates

[9]M. H. Ali AL-Maliki; Mahdi Nsaif Jasim. "Review ofSQL injection attacks: Detection, to enhance the security of the website from client-side attacks". International Journal of Nonlinear Analysis and Applications, 13, 1, 2022, 3773-3782. doi: 10.22075/ijnaa.2022.6152 Sharma, V., Kumar, S. (2023). Comparative Study of Machine Learning Algorithms for Prediction of SQL Injections. In: Shukla, P.K., Singh, K.P., TripathComputer Vision and Robotics. Algorithms for Intelligent Systems. Springer, Singapore.

[10]P. Roy, R. Kumar and P. Rani, "SQL Injection Attack Detection by Machine Learning Classifier," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India,

[11]Falor, A., Hirani, M., Vedant, H., Mehta, P., Krishnan, D. (2022). A Deep Learning Approach for Detection of SQL Injection Attacks Using Convolutional Neural Networks. In: Gupta, D., Polkowski, Z., Khanna, A., Bhattacharyya, S., Castillo, O. (eds) Proceedings of Data Analytics and Management . Lecture Notes on Data Engineering and Communications Technologies, vol 91.Springer, Singapore. https://doi.org/10.1007/978-981-16-6285-0_24

[12]D. Tripathy, R. Gohil and T. Halabi, "Detecting SQL Injection Attacks in Cloud SaaS using Machine Learning," 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Baltimore, MD, USA, 2020, pp. 145-150, doi: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00035.

[13]Hubskyi, O., Babenko, T., Myrutenko, L., Oksiiuk, O. (2021). Detection of SQL Injection Attack Using Neural Networks. In: Shkarlet, S., Morozov, A., Palagin, A. (eds) Mathematical Modeling and Simulation of Systems (MODS'2020). MODS 2020. Advances in Intelligent Systems and Computing, vol 1265. Springer, Cham. https://doi.org/10.1007/978-3-030-58124- 4_27

[14]P. Tang, W. Qiu, Z. Huang, H. Lian, G. Liu, Detection of SQL injection based on artificial neural network, Knowledge-Based Systems, Volume 190, 2020, 105528, https://doi.org/10.1016/j.knosys.2020.105528.