

Detection and Prevention of Sql Injection Using Machine Learning Based Methods

Brahmaniya Chetankumar Nareshkumar¹, Professor Nimesh Vaidya²

¹PG Scholar – Faculty of Engineering, Computer Engineering Department Swaminarayan University, India, Email id-brahmaniyachetankumar@gmail.com

²Assistant Professor & HOD - Faculty of Engineering, Computer Engineering Department Swaminarayan University, India, nimesh.vaidya001@gmail.com

Abstract:

Data is one type of most vital components or part of information systems. Machine learning which is mostly used and SQL injection is most commonly attacks on those days. Machine Learning is one type of programming language but it uses differently to compare with traditional programming. It uses to approach analysis is a great solution and also effective. Machine learning deals with structured and semi-structured data. Its included learning of new data and selfstudy to modify with existing and new data or dataset also learn the new things from the data. SQL injection attacks occurs when our web application having vulnerabilities or any loop holes in that also called bad things in programming which is easily attacker to do get login credentials to remotely control of whole under the attacker. SQL injection is one type of an injection attack that makes it possible to execute malicious SQL queries or statements. There are having identified issues are like that lack of open SQL injection datasets that its easily to get data from dataset but less. Limited work is reported for attacks on join SQL queries which are mainly used for processing but it's on necessary for all the data sets.

Keywords: SQL Injection, Machine Learning, Cyber Security, Vulnerability.

1. Introduction

SQL Injection is one of the most serious security threats on the internet today. In fact, SQL injection it is the technique of that allows an opposes to insert arbitrary SQL commands in the queries that web application or websites makes to its database. SQL injection works on the vulnerable webpages and application that use a backend database like MySQL, ORACLE and MSSQL, etc. [1]

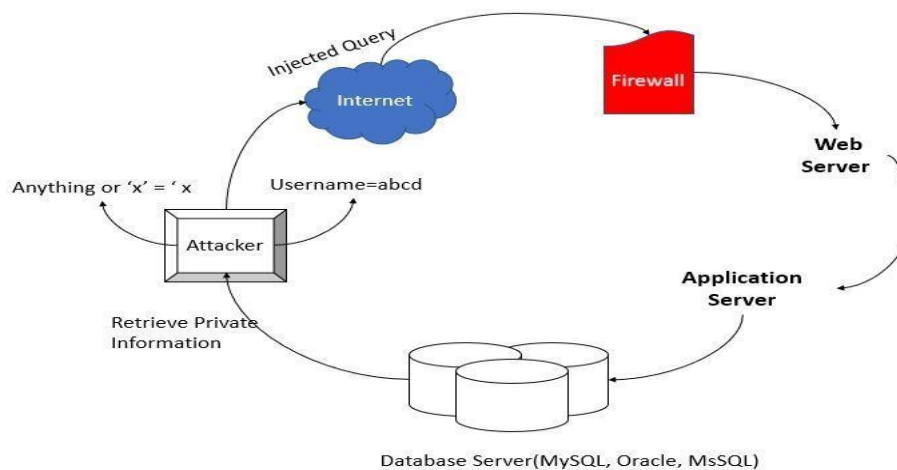


Figure 1. About SQL Injection

SQL Injection attack is occurred when our web application or web pages having an vulnerabilities or may be any loop holes in that after that attacker is easily done the inject the sql query to do get login or confidential data which is whole under the attacker. [2]

2. Background

2.1 Machine Learning

Machine learning is the one the most used technology in present day. Machine learning deals with structured and semistructured data. Its included learning of new data and self-study to modify with existing and new data or dataset and also learn the new things from the data. [1]

Machine Learning is one type of programming language but it uses differently to compare with traditional programming. It uses to approach analysis is a great solution and also effective. Machine learning deals with structured and semi structured data. Its included learning of new data and self-study to modify with existing and new data or dataset also learn the new things from the data.

Types of SQL Injection

2.1.1 Error Based SQL Injection

It is one of the most common types of SQLI Vulnerable and it's quite easy to understand. It relies on feeding unexpected commands or invalid input, typically through a user input to cause the database server to reply with an error that may collect the details about the users that attacker try to get the details.

2.1.2 Union Based SQL Injection

It's another type of SQLI, in that union operator extends the results returned by original query, enabling users to run two or more inject the statements if they have the same structure as original query.

2.1.3 Blind Injection

This type of injection attack does not show any error, that's why it's called "BLIND".

This attack is more difficult to exploit as compare to above two types of injections. It is more difficult to exploit as it returns information when the application is given SQL payloads which it does not know that it's true or false response from server or not.

It also has two types of Attack:

1] Boolean- Based: In this type of attack, Boolean query causes the application to give different response for a valid or invalid result in the database.

2] Time Based: In this type of blind SQLI it depends on the waiting for specific time or period before vulnerable application responds to an attacker's queries changes with a time delay value.

3. Literature Review

B. Gautam, J. Tripathi, Dr. Satwinder Singh [1] they proposed the "A Secure Coding Approach For Prevention of SQL Injection Attacks" objective is to using different approach to All types of SQL injection attack were compared and they get a resultant that all of them and they are able to prevent SQL Injection attacks either partially.[1]

K. Choubey, Prof. Priyanka Jain [2] "A Survey on SQL Injection Attacks and Methods to Detect and Prevent Them," this presents the secure coding approach that can be used by web developers and security professionals to secure their application against such type of attacks at the time of coding and also compared different types of prevention techniques and different approaches are tautology, Input-URL validation, Prepared Statements etc. used in this proposed paper.[2]

QI LI, WEISHI LI, JUNFENG WANG, AND MINGYU CHENG [3] "A SQL Injection Detection Method Based on Adaptive Deep Forest" the objective is adjusting the structure of the tree model and deal with multi-dimensional features to avoid overfitting problem effectively.[3]

XIN XIE, CHUNHUI REN, YUSHENG FU, JIE XU , AND JINHONG GUO [4] "SQL Injection Detection for Web Applications Based on Elastic-Pooling CNN" the objective this paper presents a method of SQL injection detection based on Elastic-Pooling CNN (EPCNN) and compares it with traditional detection methods. This method can output a fixed two-dimensional matrix without truncating data and effectively detects the SQL injection of web applications. Based on the irregular matching characteristics, it can identify new attacks and is harder to bypass. [4]

Sonakshi, Rakesh Kumar, Girdhar Gopal [5] "CASE STUDY OF SQL INJECTION ATTACKS" INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY in this paper it is to make programmers and researchers aware of SQL injection techniques which are still prevalent these days.

Garima Singh, Dev Kant, Unique Gangwar, Akhilesh Pratap Singh [6] "SQL INJECTION DETECTION AND CORRECTION USING MACHINE LEARNING TECHNIQUES" in this paper presents technique for the detection and correction of different type of SQL attacks and how to detect and correct using different techniques.

4. The Proposed Method

According to literature survey, we saw that there are various types of SQLIA performed in different web applications through their loop holes or also says VULNERABILITY in their respective application which is used by user.

Attacker attacks on them through vulnerability to sneak confidential and financial information of user for their own advantages.

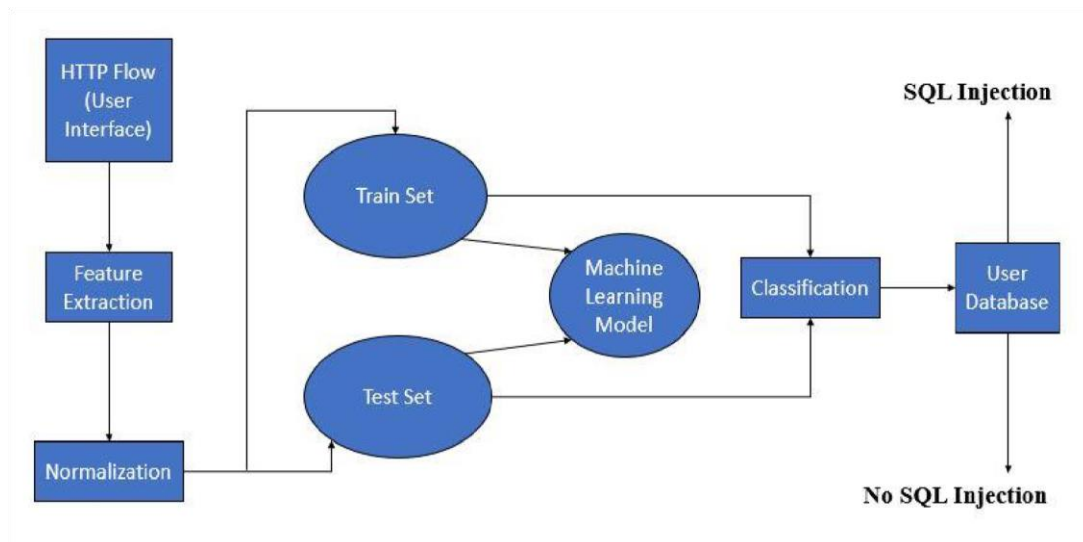


Figure 2. Flow of Work

Input Data Set:

The given input dataset contains SQL Injection sample of different queries that are collected from various sources of analysis.

Feature Selection:

The main objective of feature selection is to simplify our model that appropriate and most significant features are selected from the dataset that the different malware sample. In which the most appropriate features will result in reducing the dimensionality of the raw dataset. Such reduction will reflect in return to the overfitting or underfitting problem, and also reducing the model computation cost.

Classifier Evaluation:

In machine Learning, it is most required to build the computational models that are able to generalize the extracted feature. While training the model, a disturbed generalization is recognized by over-training. The most common way to avoid the over-training is to use an appropriate of data splitting. Classification is defined as the process of finding a model or mapping function which separate data into multiple classes.

Training Dataset or Trained Dataset:

The objective of train/test experiments is to realize some patterns of data. Those patterns are used by the learning models to make almost accurate predictions. Based on experimental results, learning models could be judged according to their efficiency. The word efficiency means the capability of a learning

model to generalize well to new data. Also, it was necessary to observe the performance of the different classifiers with the reduced features set.

1. Implementation

In this Section Jupyter (Anaconda) open-source platform is used to perform the experiment. The Experiment are done on SQL Injection dataset.

The Libraries used are pandas, numpy, matplotlib, re, nltk and the Machine Learning Models are Bagging, Adaboost and Gradient boosting, Random Forest and Decision tree classifier.

1.1 Dataset Description

Data Set Description The data is related with the SQL Injection in which different types of Queries and malicious code of which was in numeric data are include. It was 2 attributes and 4201 of queries.

Table No 1. Description of Dataset

Dataset Characteristics	Multivariate
Language:	Python
Domain:	SQL Injection
Attribute Characteristics:	Integer, Real
Number of Attributes:	2
Association:	Classification
Training and Testing:	60% and 40% ratio
Output:	Binary (0/1)

For the implementation, I'm using my own website to perform the SQL injection attacks on that and how we can protect from the loop holes of application. We can used different algorithm of machine learning methods are naive byes, gradient boosting, random forest etc. are used respectively.

Jupyter Notebook platform is used to perform experiments using python.

- Experiments are working on Datasets.
- Algorithms used

1. Naive Bayes Algorithm (Used)

Naive Bayes algorithm has already been implemented for detecting SQL Injections. Naive Bayes is a classification model in supervised learning that is based on Bayes Theorem.

2. Bagging Boosting Algorithm

Bagging is an ensemble learning approach that predicts a value of data by using multiple supervised learning models and then combining the results of all these individual learning models by a chosen technique.

3. Decision Tree classifier

Decision tree classifier is supervised learning technique which is used both classification and regression problems also but mostly used for classification problem.

SQLIAs on Local server or on local websites

Now, shown the attacks are performed on my web site using the injected query on the perform the attack for sneak the basic information of users.

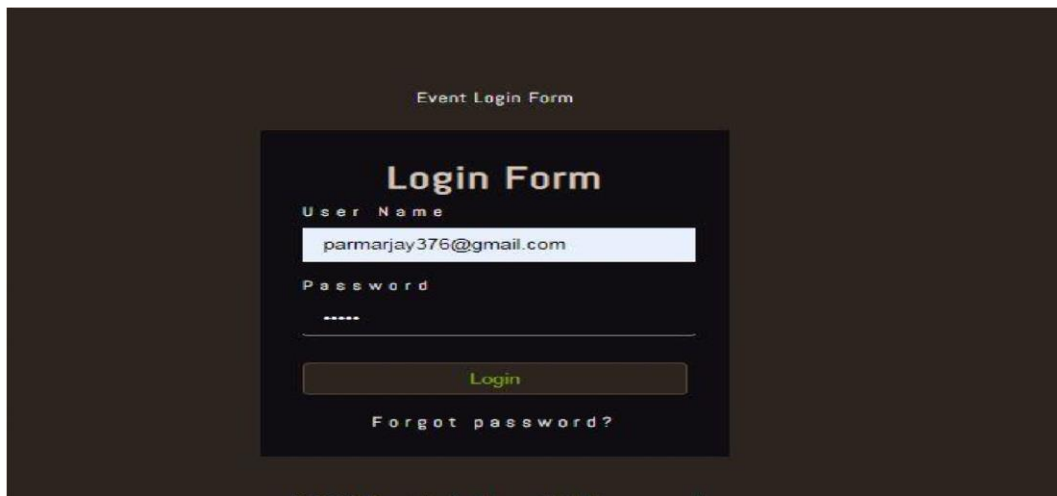
In below figure shown that is my website which is title of collage event management under collage of event details which is only held in my collage.



Figure 5.1

For first case: checked for valid user interacts with login credentials of above page website using authenticated username and passwords.

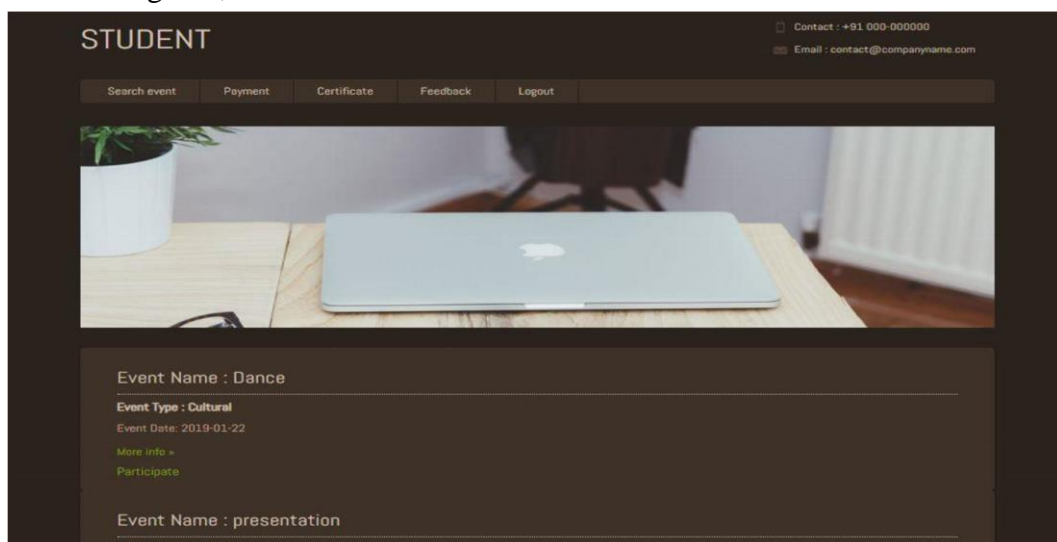
For it has been demonstrated using XAMP server and PHP as scripting language. It is shown in below fig 5.2.,



The image shows a web page titled "Event Login Form". It features a central dark box with the title "Login Form" in white. Below the title, there are two input fields: "User Name" with the value "parmarjay376@gmail.com" and "Password" with masked characters "*****". A green "Login" button is positioned below the password field. At the bottom of the box, there is a link that says "Forgot password?".

Figure 5.2

If they both username and password are corrected than it is authorized user is right person or valid person and shown in below fig 5.3.,

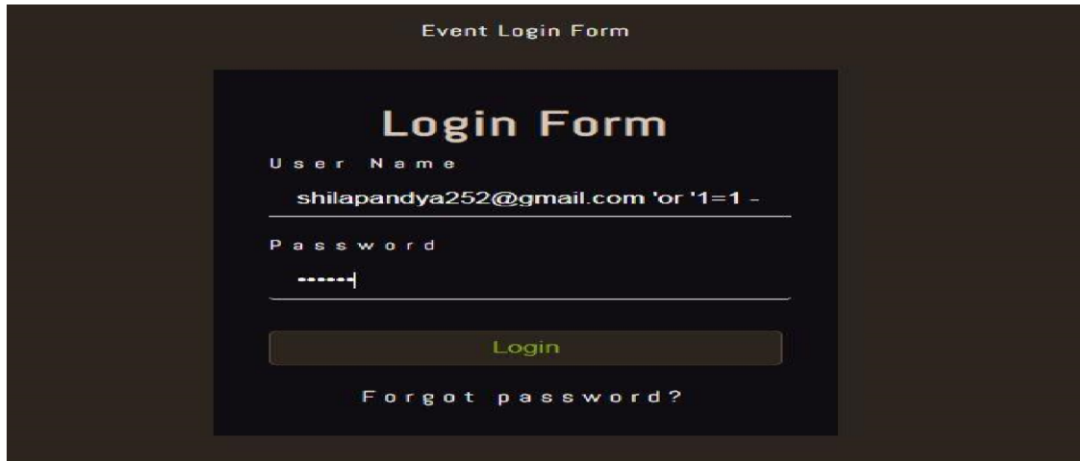


The image shows a web page titled "STUDENT". At the top right, there is contact information: "Contact : +91 000-000000" and "Email : contact@companyname.com". Below this is a navigation bar with buttons: "Search event", "Payment", "Certificate", "Feedback", and "Logout". The main content area features a large image of a laptop on a desk. Below the image, there is a section for "Event Name : Dance" with details: "Event Type : Cultural" and "Event Date: 2019-01-22". There are links for "More info >" and "Participate". Below this, another section shows "Event Name : presentation".

Figure 5.3

In Second case: we can again check for login credential but without knowing the password using injected query in password via 'or' 1=1' for bypass the password of login page and directly entry in the website.

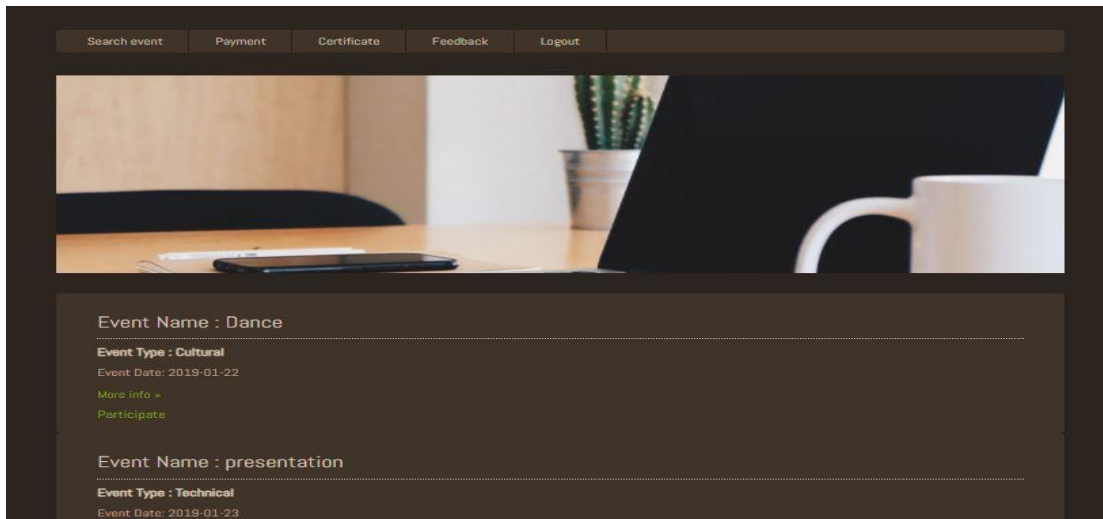
In below fig 5.6 shown that detect the attack on my website through the login credentials of username without password.



The image shows a web interface for an "Event Login Form". The form is centered on a dark background. It has a title "Login Form" in a large, bold, white font. Below the title, there are two input fields: "User Name" and "Password". The "User Name" field contains the text "shilapandya252@gmail.com 'or '1=1 -". The "Password" field contains several asterisks. Below the input fields is a green "Login" button. At the bottom of the form, there is a link that says "Forgot password?".

Figure 5.4

After pressing the “login” button, the attack is successful and the attacker is redirected the next page like shown in fig 5.7.,



The image shows a web interface for an "Event Details" page. The page has a dark background. At the top, there is a navigation bar with links: "Search event", "Payment", "Certificate", "Feedback", and "Logout". Below the navigation bar is a large image of a desk with a laptop, a smartphone, and a mug. Below the image, there are two event details sections. The first section is for an event named "Dance", with the type "Cultural" and the date "2019-01-22". It has links for "More info" and "Participate". The second section is for an event named "presentation", with the type "Technical" and the date "2019-01-23".

Figure 5.5

6. Experimental Results

The Experimental result is a Machine Learning.

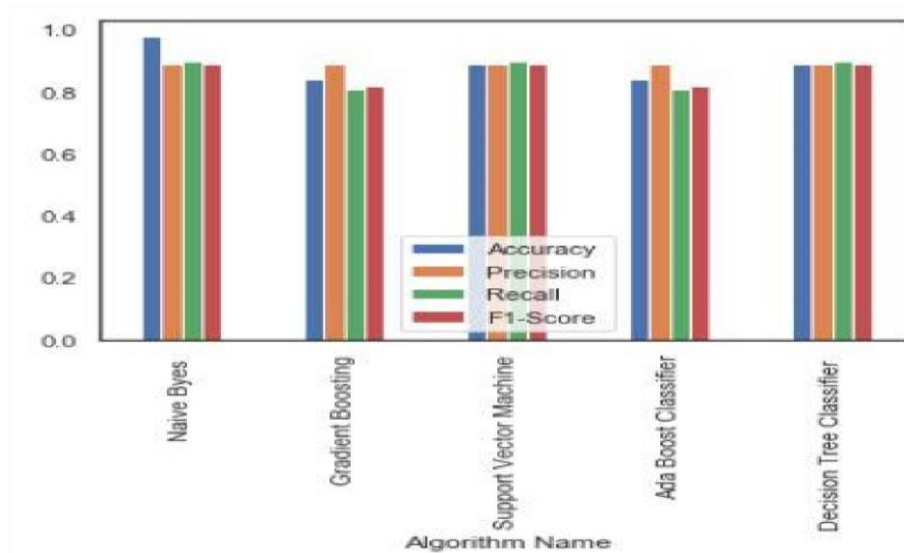


Figure 3. Comparison of Accuracy, Precision, Recall and f1_score

On above figure is shown that Comparison of Accuracy, Precision, Recall and F1- Score of different methods are used of machine learning with different percentage of all with 0.98, 0.89, 0.90 and 0.89.

For Prevention steps for protect from SQLIA:

1. Prevention of attacks via URL's

- To prevent these types of attacks, the developers should always make use of HTTP Post method instead of HTTP GET method because of when information sent from a form with this get method it is visible to everyone.
- This is one point which is easily exploited by the attacker to inject the malicious queries to extract sensitive DB information.

2. Keep the page name other than admin name

- This is the best option for attacker tries to get DB of table name of users' information that's why we can change by default name of "admin" with another name.
- Because of that attacker firstly try to locate admin page of any website, by entering the complete URL and then affixed "/admin".
- Its easiest way to enter any websites through this way that's why we can keep changes the name to admin page.

3. Detection of malicious symbols

- Certain checks should apply to detect some known malicious word or symbols or keywords are such that “or”, “=”, “union “, “insert “, etc.
- These are responsible for an attack to take place.

7. Conclusion

In this paper, we describe that how SQL Injection is making an intensive impact on our computer systems, the internet, and data. SQL Injection itself is an enormous field and it is still a challenging dilemma in today's world. We used SQL Injection along with machine learning to detect attacks on data. We observe that our result shows great improvisation and it is more accurate as compared to conventional methods. As future work, we assume that we can have more features as an add-on in the system to detect attacks as well as prevent them as earliest as possible.

8. Acknowledgement

I would like to express my gratitude and Dedicated To “MY MOM AND DAD” who always picked me up on time. Their inspiration and constant support are of paramount importance in all aspect of my life. And also, I am highly indebted & would like to express my special thanks to Prof. Mahesh Panchal for his kind co-operation and encouragement which help me in completion of my work.

References

1. B. Gautam, J.Tripathi, and Dr. Satwinder Singh “ A Secure Coding Approach For Prevention of SQL Injection Attacks” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018), pp. 98749880.
2. K. Choubey , Prof. Priyanka Jain ,“A Survey on SQL Injection Attacks and Methods to Detect and Prevent Them,” International Journal of Innovative Research in Science, Engineering and Technology, Vol. 8, Issue 3, March 2019.
3. QI LI, WEISHI LI, JUNFENG WANG, AND MINGYU CHENG, “A SQL Injection Detection Method Based on Adaptive Deep Forest” IEEE Access ISSN: 2169-3536, Vol 7, October 17, 2019.
4. XIN XIE, CHUNHUI REN, YUSHENG FU, JIE XU, AND JINHONG GUO, “SQL Injection Detection for Web Applications Based on Elastic-Pooling CNN” IEEE ACCESS 2947527, Vol 7,October 30, 2019.
5. Sonakshi, Rakesh Kumar, Girdhar Gopal “CASE STUDY OF SQL INJECTION ATTACKS” INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, ISSN: 2277-9655, July, 2016.
6. <https://www.youtube.com/watch?v=2k6tB4fnUuM>
7. <https://dzone.com/articles/what-is-the-sql-injection-vulnerability-amp-how-to>
8. Subhranil Som, Sapna Sinha, Ritu Kataria “Study on SQL Injection Attacks: Mode, Detection and Prevention” International Journal of Engineering Applied Sciences and Technology, 2016 Vol. 1, Issue 8, ISSN no. 2455-2143.
9. Mussab Hasan, Zayed Balbahaith, Mohammed Tarique” Detection of SQL Injection Attacks: A machine Learning Approach” International Conference on Electrical Nd Applications, 2019.



10. Garima Singh, Dev Kant, Unique Gangwar, Akhilesh Pratap Singh “SQL Injection Detection and correction using
11. Machine Learning Techniques” Advances in Intelligent Systems and Computing, Volume 1, 2015.
12. Tareek Pattewar, Hitesh Patil, Harshada Patil, Neha Patil, Muskan Taneja, Tushar Wadile “Detection of SQL Injection Using Machine Learning: A Survey” International Research Journal of Engineering and Technology, Volume:06, Issue:11, November 2019.