SQL Injection Attack Detection using Naive Bayes Classifier

Submitted By

Yash N. Variya

20MCEI14

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2022

SQL Injection Attack Detection using Naive Bayes Classifier

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Yash N. Variya

(20MCEI14)

Guided By

Dr Zunnun Narmawala

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2022

Certificate

This is to certify that the major project entitled **"SQL Injection Attack Detection using Naive Bayes Classifier"** submitted by **Yash N. Variya (20MCEI14)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Major Project Part-II, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Dr Zunnun Narmawala
Internal Guide & Associate Professor
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Sharada Valiveti
Professor & PG Cordinator
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Rajesh N Patel
Director
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Madhuri Bhavsar
Professor & Head
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Statement of Originality

I, **Yash N. Variya**, **20MCEI14**, give undertaking that the Major Project entitled **"SQL Injection Attack Detection using Naive Bayes Classifier"** submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

_____

Signature of Student

Date:

Place:

Endorsed by

Dr. Zunnun Narmawala

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Zunnun Narmawala**, Associate Professor, Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr Madhuri Bhavsar**, Hon'ble Head of Computer Science And Engineering Department, Institute of Technology, Nirma University, Ahmedabad for her kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr Rajesh N Patel,**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Science and Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Yash N. Variya

20MCEI14

Abstract

The current world relies heavily on web applications. As a result, providing security to every web application is a huge challenge. In most cases, the information is already in the database on the back-end of the web application. The number of online platform hacks is growing daily as everything becomes digital. Hackers frequently target online database applications. One of the most common types of attacks is SQL injection. A malicious code is injected into the SQL query of the user by the attacker. As a result, they get access to the database and they can change the information. The internet is the most reliable and commonly utilized channel for communication and business activities in today's modern world. Users load massive amounts of data onto the web every day through numerous means, and user input might be malicious. As a result, web application security getting more crucial. Because they are so easily available, they are vulnerable to a variety of flaws that, if ignored, can result in harm. Attackers use these weaknesses to get unauthorized access through a variety of illegal activities. The machine learning concept with the Support Vector Machine (SVM) algorithm was introduced to overcome the above-mentioned attacks. It is used to detect and prevent SQL injection queries. Attacks on the internet are increasing in number and severity on a regular basis. The massive amount of data available on the internet encourages hackers to attempt innovative attacks. The Structured Query Language is a most dangerous attack that targets web applications. Several studies work had been carried out to mitigate this assault both by stopping it from an early Level or detecting it whilst it happens. We present an overview of SQL Injection attacks in this paper, as well as a classification of the recently presented detection and prevention solutions. More and more persons are using computers in their daily lives in this planet. As a result, more data is stored. The core aspects of computer backups are recovery and storage. Unfortunately, data loss occurs for a variety of reasons, including accident deletion, software or hardware failure, and cybercrime-related actions. Every organization should have a solid security plan in place because a breach of security can cost a lot of money and risk embarrassing the company in the eyes of customers or clients.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

In this chapter, The importance, overview, and quality assurance of SQL injection attacks are discussed. The chapter goes over why SQL injection is so crucial in today's world and how machine learning is used in SQL injection attacks.

1.1 Introduction and Problem Summary

The SQL injection is always at the top of the OWASP (Open Web Application Security Project) top 10 lists. As a result, SQL injection detection and prevention are extremely important. The research's major goal is to create a machine learning-based classifier that can detect SQL injection attacks. Moreover, in recent years deep learning-based approach is improved bug prediction, bug localization, and identifying code.According to a survey conducted in 2018, 953 thousand daily web attacks were banned, compared to 611 thousand daily blocked attacks the previous year. According to OWASP, SQL injection is the most dangerous type of attack. It compromises the main security offerings: confidentiality, authentication, authorization, and integrity. SQL injection attacks are defined as the injection of malicious SQL queries into a web application's input field in order to gain access to the web application's database. Most websites nowadays utilize a back-end database to store user data or any other information. Interaction Forms and emails are frequently used to communicate with these users in order to retrieve information that they have chosen. Hackers attempt to exploit this functionality by injecting malicious code. Injecting harmful code into these user inputs, which will eventually be utilised to construct SQL queries, is incorrect. The effectiveness of the SQL injection attacks can be related to user inputs. As a result, it can have a harmful effect, such as the harmful of

the databases or the collecting of sensitive and confidential information Clients' personal information is kept private via the web application. Not only at educational institutions, but also to avoid indicator attacks, a lot of research has been done in recent years. Researchers have suggested the following preventive actions. Web vulnerabilities can put personal information and other sensitive data at risk. Resources of great value When a user tries to send a request to a web server, he does so via the HTTP protocol. Forms created in Hypertext Markup Language (HTML), Uniform Resource Positions (URLs), or other formats fields in which information can be entered Users can employ SQL injection with the unfiltered form. This is due to the fact that the data from the agreement form gets processed without being reviewed.The SQL Hall of Fame Search examines current trends in SQL injection attacks and data triggers. In the world of big data, securing your back-end database from SQL injection attacks is a subject-based challenge. By using SQL injection, the attackers can change their own data with users' personal data. By this attacker can have direct access to the database server. In this paper, We discuss the present SQL injection methods, types, and goals, as well as the machine learning method and neural network model for detecting SQL injection attacks.

In today's environment, the most significant word in the IT sector is security. As a result, providing security is our main responsibility. According to our research, SQL injection attacks are extremely damaging in the security domain. We explore SQL injection attacks and how they operate in this research paper. For SQL injection attacks, we also define the research gap and problem statement. The Naive Bayes classifier method was used to improve the accuracy of the detection of SQL injection attacks. First, we trained our data set and compared it to several machine learning approaches, determining that the Naive Bayes method provided better accuracy.

Chapter 2

Literature Survey

In this chapter, I have included a summary of research papers which are related to SQL injection attacks detection, how the other researchers have done it, what kind of approach they have used, and the advantages and disadvantages of those approaches.

2.1 Summary of Related Researches

For detecting SQL injection attacks, a variety of studies have been done. We give a critical review of some recently published related works in this section.

In this section, An overview of SQL injection attacks is presented. So, in this article, we'll talk about SQL injection attack sources and types.[8]. Any application parameter that can be used in such a database query could be vulnerable to SQL injection. The authors suggested four possible sources for the SQL Injection Attack (SQLIA). User input, cookies, server variables, and stored injection are examples of these sources[7, 2].

Injection Through User Input

Forms are commonly used in web applications to gather data from users (such as signup, login, etc.) or to let users select the data to be received (such as search, adapted view, etc.). Hackers could inject malicious code into these forms with "text fields," allowing them to get indented data (such as secret info) or perform indented operations (manipulate a database, etc.). Login Name, Password, Address, Phone Number, Credit Card Number, and Search are all common fields[6].

Injection Through User Cookies

The cookie variable is used to attack users on the web application and website. Generally, an attacker needs to access the victim's account and for that, they may use cookies to achieve their goals. Cookies inside the first region aren't supposed to be treated as a person enters. On the other hand, cookies may contain data encoded in hexadecimal, hashes, serialization information, or plain data. For example, we can use commands to inject the cookies. The query will let the attacker use the provided password and it turns into an unauthorized entry right into a machine. As we know that we can use many HTTP intercepts before we sent this to the server. now attackers may add a malicious query into the cookie field. so, in this case, an attacker may use the HTTP GET/POST SQL injection to get the password from the website or web application[1].

Stored Injection

Attackers utilise stored injection (also known as second-order injection) to inject malicious data into a system, causing SQL injection attacks every time that input is used. A different code exists for second-order SQL injection. The attacker initially registers for the application as a valid user of the website, using a seeded username such "admin'- -." Following that, the attacker will try to change his password.[11]

Injection Through Server Variables

Network headers, HTTP metadata, and environmental variables are all part of the server variables collection. Such server parameters are usually used by web applications to monitor user data and identify browsing trends. If these variables are not verified before being stored in a database, attackers can take advantage of this vulnerability by inserting an SQL injection attacks straight into the server[9].

In this chapter, we cover the various types and forms of SQL injection attacks[14]. Table 2.1, 2.2 summarises the many types of SQL injection attacks.

2.2 Conventional Approach to Detection of SQL Injection Attacks

This section provided a list of research publications that cover SQL injection attack detection and prevention using machine learning. After studying a number of papers, I

prepared a comparison table in which we discuss each paper's research, the future scope and prevention of SQL injection attack detection and prevention using machine learning, and whether or not a vulnerability is used. The comparison table is mentioned in Table 2.3, 2.4 and 2.5[13, 4, 10, 12, 19, 1, 16, 17, 4].

- (Sangeeta Nagpure and Sonal Kurkure (2017)) They present the difference between vulnerability assessment and pen testing by using automatic techniques and manual techniques and also they provide some techniques through which we can identify a vulnerability with more accuracy[14].

- (R Sri Devi and M Mohan Kumar (2020)) They present an overview of the vulnerability assessment of web applications using the NIKTO tool. How attackers may attack the website and takeover the accessed website. For that query, they also provide remediation techniques to provide a better result for the website[7].

- (José Fonseca and Marco Vieira (2007)) The authors presented one framework in which they focus on the most critical vulnerability of OWASP top 10 which Is SQL injection and XSS (Cross-Site Scripting). Focus on the workflow of network and penetration testing. The proposed techniques for caber threats like SQL injection and XSS. Once completed manual testing on a website and compare results with different tools to get accuracy[8].

- (Yugansh Khera and Deepansh Kum (2019)) The authors presented the impact of the Vulnerability assessment tool on cyber security. They compare the result of manual testing and automated testing on the web application. The accuracy we got in automatic testing is higher than compared manual testing. Also, they explain why cyber threats are increasing rapidly in the cyber world [11].

- (Keyur Patel (2019)) The author presents a vulnerability assessment on any specific target. That is how vulnerability assessment is a process of identifying the security loopholes or bugs in the computer system, network, or web application of an organization. In the assessment, we have to go through the organization with the necessary knowledge, understanding of infrastructure, and understanding of the threats to the environment. Also, explain that vulnerability assessment is unable to identify the logical attack vectors[15].

- (Anna L. Buczak (2015)) The author created a structure in which they discuss about difference between the machine learning approach and the data mining approach in cyber security threats. How machine learning techniques use different types of algorithms and

data mining uses a different algorithm. Main focus is on the detection of a target using machine learning and data mining[3].

- (Marco Barreno and Blaine Nelson (2015)) They explain the security of machine learning and that we use machine learning in every aspect of IT infrastructure. As we know that in past years machine learning are used in the cyber world to improve cyber threats. If we are using machine learning in the security domain so, is it safe or helpful or is there any impact is exiting or not. They explain each point of machine learning which is used in the cyber security domain[2].

- (Ovidiu Valea and Ciprian Opris¸ (2020)) The author gives an overview of the Metasploit tool framework. How Metasploit tool is used in vulnerability assessment and penetration testing. How much accuracy Metasploit will provide when we are using this tool to detect loopholes in any organization[10].

- (Muhammad Saidu Aliero and Imran Ghani (2015)) The authors give an overview of SQL injection attack detection using a tool. There is a difference between SQL injection detection by manual testing and injection detection by an automated tool. We can get higher accuracy by using a tool for SQL injection. SQL injection attacks are the most common attack in the security domain so, at that time, accuracy is important for SQL injection detection[11].

- (Anil Lamba (2014)) The author explains that in the current world scenario the use of computers is increasing day by day. For that reason, the system's complexity is increasing. As we all know that each organization's systems are connected to the internet. Currently, new and complex software is coming in the market. So, all these reasons may increase loopholes in systems[13].

- (Prashant S. Shinde and Shrikant B. Ardhapurkar(2016)) The authors present the importance of cyber security role when systems are hacked by an attacker. Because once attacker enters your system then they can do anything with your systems. So, at that time we realize why we have to secure our organization in a security area from the attackers using techniques of security[18].

2.3 Machine Learning Approach to Detection of SQL Injection Attacks

Various approaches for preventing SQL injection attacks have been suggested by researchers and many authors. According to the author's survey, the most popular techniques are static, dynamic, and hybrid (both static and dynamic)[9]. The dynamic analysis is identified as a more advanced technique that allows the system to identify the SQL injection in the valid queries. In static analysis, it will check whether it is correct generated SQL queries and examines the mismatch in the queries[3]. Both static and dynamic analyses are used in the hybrid technique. They start with a static approach and then move on to a dynamic approach. Machine learning is used in both dynamic and hybrid analysis techniques.

The author has proposed various solutions based on the techniques listed above. Machine learning algorithms are considered as being the most successful in detecting SQL injection attacks[18]. According to the survey, the machine learning approach is not only suitable for preventing identified attacks, but also for detecting unknown attacks. The machine learning method to SQL injection attacks is discussed in this section[20].discusses the results of many authors' surveys on SQL injection using machine learning.

The survey table of machine learning approaches is in Table 2.6, 2.7[5].

2.4 Research Gap

SQL injection is a very common and in-demand type of attack. SQL injection attacks were one of the top ten vulnerabilities identified by OWASP. According to a survey, nearly a thousand cyber attacks were blocked every day from 2018 until the present. According to Owasp's survey, SQL injection attacks are the most serious of the top ten vulnerabilities. SQL injection attacks are carried out on a variety of websites by different types of attackers. As a result, proper SQL injection remediation must be implemented. As I survey and research many papers related to machine learning, deep learning, and artificial neural network model. After that, we can state that new solutions to protect against SQL injection vulnerabilities are being developed every day. Many authors developed a technique based on an artificial neural network and created a model that included a URL generator, URL classifier, and NN model. But there is no accuracy in

detecting whether there is a malicious URL at input validation or not. And if malicious queries are detected at the input field, then, it can be SQL injection type attack or not.These two items have not yet been implemented. So we'll try to do something about it.

2.5 Problem Statement

SQL injection has become the most prevalent attack in recent years all around the world. As a result, many software and website development firms rely on cyber security firms. In today's world, security is very important. SQL injection attacks must be prevented and detected using various strategies, such as the machine learning approach to SQL injection detection. We can detect SQL injection manually, but the findings will be imperfect and inaccurate. To achieve accuracy, we must apply several machine learning methods. Many authors have developed a neural network (NN) based model for the detection and classification of SQL injection attacks in previous publications. I've proposed a neural network-based approach that includes a URL generator. We'll get the results of SQL injection attacks through that technique. Section IV explains the neural network model's operation and flowchart. According to the survey, they focused on generated URL and SQL injection attacks. Our focus, meanwhile, is on developing a proposed model that attempts to identify each produced URL as malicious or benign. Second, for each malicious URL, we can identify which type of SQL injection attack it is. We'll use a different type of data set for this.

Type of attack	Attacker's aim	Description	Description
Tautologies	bypassing authentication and extracting	Conditional statements Are formed in this sort of way That they are constantly real	Select * from emp¡*nfowhereempid* = *"or'*7 = 7*'*;
Logically Incorrect queries	To extract information about database and identify injectable patterns	Invalid queries are Accomplished main to error Messages which Constitute statistics Approximately information type or desk Call.	Aggregate functions applied on invalid data types or using 'having' and 'group by' clauses.
Union Query	Bypassing authentication and extracting data	By using the usage of operator 'union', malicious query Is joined with safe question.	Select * from user where user='ravi' union select * from admin where id='3142'– 'pass='2=2';
Stored procedure	Privilege escalation, executing remote commands, DoS	The usage of integrated techniques, Malicious movements are Performed.	Commands like DROPTABLE, SHUTDOWN are executed.
Piggy-backed queries	Data extraction and modification, DoS	Malicious question is Appended to valid Question. On execution of First query, 2nd also Gets performed.	Select * from user where name= 'ravi' and pass='1234'; drop table user;

Table 2.1: Types of SQL Injection

Type of attack	Attacker's aim	Description	Description
Alternate Encodings	To evade detection	Some databases have Filters which come across Characters like –, percentage, and so forth., As bad individual. As a way to Avoid detection, attacker Encode the question in Ascii or Unicode.	SELECT salary FROM users WHERE login=" AND pin=0; exec
Blind injection		Database schema is Guessed by using gathering Responses on basis of Real/false questions.	Attackers injects query to discover the vulnerabilities like select * from user where id='12' and pass='1=0'; to check if there is input validation or not.
Timing attacks		Information collection is Performed thru observing Reaction time taken in Answering questions	Keywords like wait for are inserted to delay execution if query is true etc.

Table 2.2: Types of SQL Injection

No	Paper title	Year	Analysis	Prevention	vulnerability
1	Vulnerability Assessment and Penetration Testing of Web Application	2017	In this paper, they discuss about vulnerability assessment and pen testing are two different vulnerability testing. And how manual and automatic work.	Organizations must plan an integrated guide and automated checking out a technique to grow accuracy in the identification of vulnerabilities in net packages.	XSS, SQL injection, CSRF
2	Testing for Security Weakness of Web Applications using Ethical Hacking	2020	In this paper, they discuss about how ethical hackers find the weakness of web applications using the Nikto tool.	Apprehend openness and flaws in networks and internet programs the use of penetration testing to shield the institutions from cyber threats.	XSS
3	Testing comparing web vulnerability scanning tools for SQL injection and XSS	2007	In this paper, they discuss about first testing of web applications and then compare this test with different tools.	Recognize openness and flaws in networks and web applications the use of penetration checking out to guard the institutions from cyber threats. For future work, we intend to apply this benchmark system to other internet applications to better Recognize the connection between software program faults and vulnerabilities.	XSS, SQL injection

Table 2.3: Comparison Table

No	Paper title	Year	Analysis	Prevention	vulnerability
4	Analysis and Impact of Vulnerability Assessment and Penetration Testing	2019	In this paper, they discuss about how VAPT tools are important in assessment and testing. And the impact of each tool that is used in the assessment.	India and other countries are Forcing virtual payments and the statistics are stored digitally. Due to lack In verbal exchange network increases more cyber-attacks.	-
5	A Survey on Vulnerability Assessment and Penetration Testing for Secure Communication	2019	Present day vulnerabilities, Determination of these vulnerabilities, the methodology used for determination, gear used to decide the vulnerabilities to secure the groups from cyber risk.	We will mitigate the Danger of assault from those vulnerabilities. Often updating protection rules and mechanism of the security model may also lower the possibilities of being exploited from the developed Vulnerabilities.	XSS, SQL injection
6	A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection	2015	In this paper, they discuss about machine learning and data mining for cyber security IDS.	We can provide a better solution in device studying for cyber safety in a specific segment.	Cyber Analytics
7	Towards Pen testing Automation Using the Metasploit Framework	2020	In this paper, they discuss about the overview of the Metasploit tool and how it works to check the vulnerability.	A framework that automates the pen checking out steps and may be used by pen testers to see if a gadget may be easily exploited or no longer.	SSRF, CSRF

Table 2.4: Comparison Table

No	Paper title	Year	Analysis	Prevention	vulnerability
8	The security of machine learning	2015	In this paper, they discuss about Machine learning We can offer the greater higher answer in system studying for cyber safety in a different section. Potential to unexpectedly evolve to changing and complex situations has helped it turn out to be an essential tool for pc safety	We've presented a framework for articulating a complete view of various instructions of attacks on the device getting to know structures in phrases of 3 unbiased dimensions and an Hostile getting to know the sport.	–
9	A Component-Based SQL Injection Vulnerability Detection Tool	2015	In this paper, they discuss how we can detect SQL injection using a tool and how we can prevent it.	In this paper, they proposed aspect-based totally for SQLiv detection device for the cause of improving issue reusability, rapid integration, and preservation at less rate.	SQL injection
10	Cyber Attack Prevention using VAPT Tools	2014	In this paper they focused only on how we can Reduce vulnerabilities of internet utility using different tools.	Via this paper in the future, we can broaden new VAPT techniques and gear. Obligatory VAPT trying out can stop cyber-assault cases and provide fortify device safety.	–

Table 2.5: Comparison Table

Paper Title	Year	Author Name	Aim
Machine-Learning-Driven Evolutionary Approach for Testing Web Application Firewalls	2018	Minhas J. and Kumar R.	They used static and dynamic techniques. The main objective of their studies is to time-saving for incoming quires and detect them. And static and dynamic both reduce the possibility of false detection of SQL injection queries.
A Machine Learning-Based Technique for False Data Injection Attacks Detection in Industrial IoT	2020	Minhas J. and Kumar R.	They present a method for detecting the malicious queries at input field based on the combination of two classifiers Na¨ıve Bayes and RBC control mechanism. They used the tokenization method which means the work is to split the query into significant elements called tokens.
Research on SQL injection detection technology based on SVM	2018	Kamtuo K and Soomlek C	They created one framework to grab the SQL commands from the dataset and then this command sends to machine learning model for the prediction of the SQL injection attacks. The main approach is to prevent the SQL injection attack on illegal or logically incorrect queries, union queries on the server-side by applying machine learning. This author do not consider client-side actions.
Machine Learning for SQL Injection Prevention on Server-Side Scripting	2016	Kumar et al.	They propose a novel runtime technique. Through this, they can prevent SQL query injection attacks based on static and dynamic analysis. This technique depends on securing the value of the SQL query attribute of the web pages. At that time this technique matches them with query.
SQL Injection Detection Using Machine Learning.	2019	Uwagbole S and Lu Fan W	They proposed a classification system based on machine learning to detect and prevention of SQL injection attacks. This proposes a system to test the dataset by checking the token-based phase. this classification system through a support vector machine (SVM) algorithm blocks malicious web requests from entering the target back-end database

Table 2.6: Machine Learning Approach

Paper Title	Year	Author Name	Aim
IRJET- Detection of SQL Injection using Machine Learning: A Survey	2019	Huang and colleagues	They propose the Blackbox technique for testing at input field for SQL injection vulnerability. The tool can be identified all points in which applications are used in the input field for injecting SQL quires. And it also monitors the application and how machine learning is used in an application.
Detection of SQL injection based on artificial neural network	2020	Xiang Fu et al.	They proposed a design which is a static analysis framework. The main aim of this framework is to identify SQLIA vulnerabilities at compile time. The framework statically monitors the MSIL (Microsoft Symbolic intermediate language) byte code of an ASP.NET Web application, using symbolic execution. and this framework also can analyze the source code and also find delicate vulnerabilities that cannot be discovered by the Blackbox scanner.

Table 2.7: Machine Learning Approach

Chapter 3

Proposed Methodology

As explained previously, many authors deploy neural network-based algorithms to analyse SQL injection attack results. This section explains how the neural network-based model works, and we discoverer how to use the data set to gain the best accuracy.

3.1 Proposed Neural Network-based Model

This section describes the proposed neural network-based model and the processes for detecting SQL injections. The proposed model's process is shown in Fig. 3.1. Artificial neural networks (ANNs) and simulated neural networks (SNNs) are the heart of deep learning. Human brain designed the entire structure and name. Node layers are what comprise a neural network. A neural network has consisted of node layers. This, contains an input layer, one or more hidden layers, and an output layer. Each node is connecting with another node and it has an associated weight and gateway(entry). If the output of the node is based on a threshold value then that node is activated and sends data to the next layer of the network otherwise no data is passed along to the next layer of the network. As we can see from the below figure that there are mainly three blocks which are URL generator, NN model and output[15].

The URL Generator

In URL Generator there are two components which are "Benign URLs" and "Malicious URLs". The real URL addresses that exist in the world and don't have SQL injection attacks signature(s). These type of URL has been captured from the internet(2016). The Google search engine (Google, 2016) has been in work to find the URL addresses which are benign but have SQL injection attack signature(s). The "Malicious URLs"

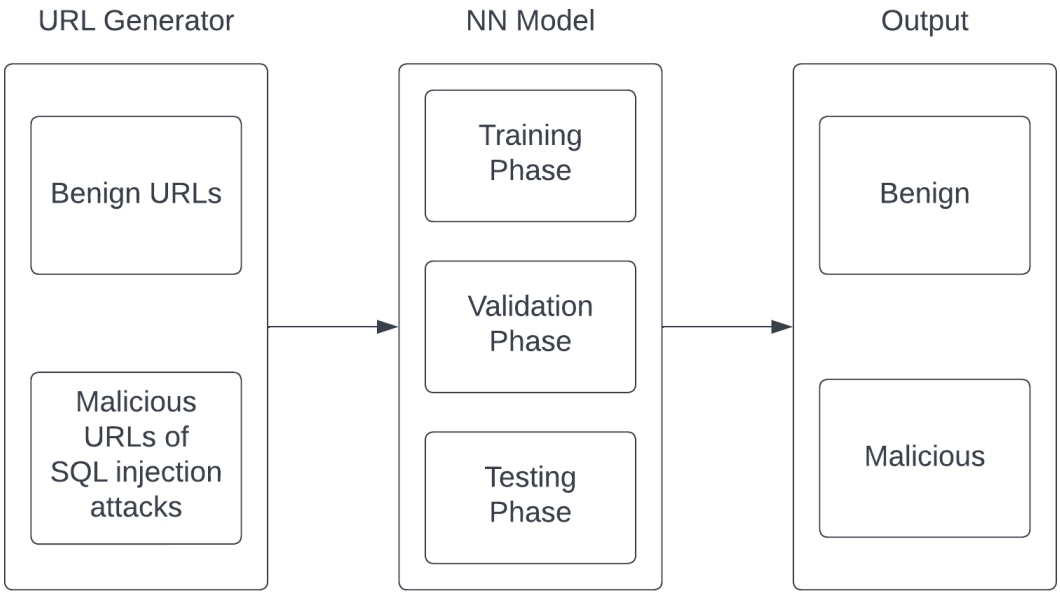

Figure 3.1: **Components of the proposed neural network-based model**

include the cruel and harmful URL addresses. It also includes SQL injection attack signature(s). Using PHP scripting language these URLs have been generated by adding the SQL injection attack signature(s) to the most popular and famous URL addresses in the world (Internet, 2016).

The URL Classifier

The URL classifier does two things: it checks existing URLs to see if they are benign or malicious, and it detects different types of SQL injection attacks for malicious and destructive URLs. To look at it another way, the URL Classifier deals directly with the URL addresses generated by the URL Generator.

The Neural Network (NN) Model

The neural network model deals with URLs that are identified as benign or malicious, as discussed in the URL generator and URL classifier sections. For malicious URLs, the neural network model detects the SQL injection attack type. The URL classifier sends this information to the neural network. Following that, it considers three phases: training, validating, and testing, with distribution rates of 70 percentage, 15 percentage, and 15 percentage, respectively. There are x and y inputs in a neural network model, with n hidden layers connecting them. Back-propagation is an abbreviation for backward

propagation algorithm, and it is a common NN-based algorithm used by the author. The algorithm operates on the basis of a set of inputs and outputs.

Chapter 4

Experimental Setup and Result Analysis

In this chapter, we discuss which methods I used for SQL injection attacks detection and which Classifier is used for detection. How can this be implemented and what are the things to be done in which way the proposed work is going to be performed?

4.1 Naive Bayes classifier Approach

We'll use data-set for SQL injection detection, as discussed in the previous section.I have an one data set for detecting SQL injection here. The Naive Bayes classifier was used. We used the Naive Bayes approach to train our data set in this case. Binary and multi-class classification are referred to as Naive Bayes. Variables are compared to numerical variables when category input is used. The Naive Bayes algorithm performed well. In predictions and forecasting data based on historical results, Naive Bayes is used. A supervised learning method is Naive Bayes. In other terms, it is a collection of supervised learning algorithms. The 'Bayes' Theorem is the basis for the Naive Bayes classification algorithm. We deployed Naive Bayes using the multinomial approach in this scenario. In Naive Bayes, the Multinomial technique refers to a Bayesian learning strategy common in Natural Language Processing (NLP). In Naive Bayes mainly three-technique which is Gaussin, Multinomial, and Bernoulli. considering this data set we used a multinomial approach for best accuracy.

First of all, we'll evaluate how the data set will train. Payload.txt and labels.txt are the two training files that we used. The payload.txt file is used as train data and the

labels.txt file is used as train labels. In this data set, we used three test files. owasp,burp-suite, and fuzzdb are these three test files that have labels and payload. Once you call this as a payload and labels you will get the final result of this data set. Once we have the results, we can evaluate which of the three test files has the highest accuracy.

Here we discuss our results and how they relate to various scenarios. As previously said, we train our data set using a multinomial approach. For training files we have payload and labels and for the test files, we have burp-suite, fuzzdb, and owasp. We have tried three scenarios. In the first phase, we have taken burp-suite payload.txt and burp-lables.txt as a training data set and burp suite (payload.txt labels.txt), owasp (payload.txt labels.txt) and fuzzdb (payload.txt labels.txt) as test files.We put it through three rounds of testing. In the second phase, we use the same scenario as the first, with the one change we use fuzzdb-payload.txt and fuzzdb-labels.txt as training files.We've also run it through three rounds of testing. In the last phase, we use owasp-payload.txt and owasp-labels.txt as training files. Below, I've created three tables based on the accuracy of the trained data set which are represented in figures 4.1, 4.2, and 4.3, respectively.

Using Burp Suite Tarinng data set				
Testing Data set	Test 1	Test 2	Test 3	Accuracy
Burp suite	0.998	0.998	0.999	0.998
OWAPS	0.994	0.998	0.999	0.997
FUZZDB	0.996	0.998	0.999	0.997

Figure 4.1: Using Burp Suite

Using Fuzzdb Training Data Set				
Testing Data set	Test 1	Test 2	Test 3	Accuracy
Burp suite	0.998	0.998	0.999	0.998
OWAPS	0.994	0.998	0.999	0.997
FUZZDB	0.996	0.998	0.999	0.997

Figure 4.2: Using Fuzzdb

Using OWASP Training Data Set				
Testing Data set	Test 1	Test 2	Test 3	Accuracy
Burp suite	0.998	0.998	0.999	0.998
OWAPS	0.996	0.998	0.999	0.997
FUZZDB	0.996	0.998	0.999	0.997

Figure 4.3: Using OWASP

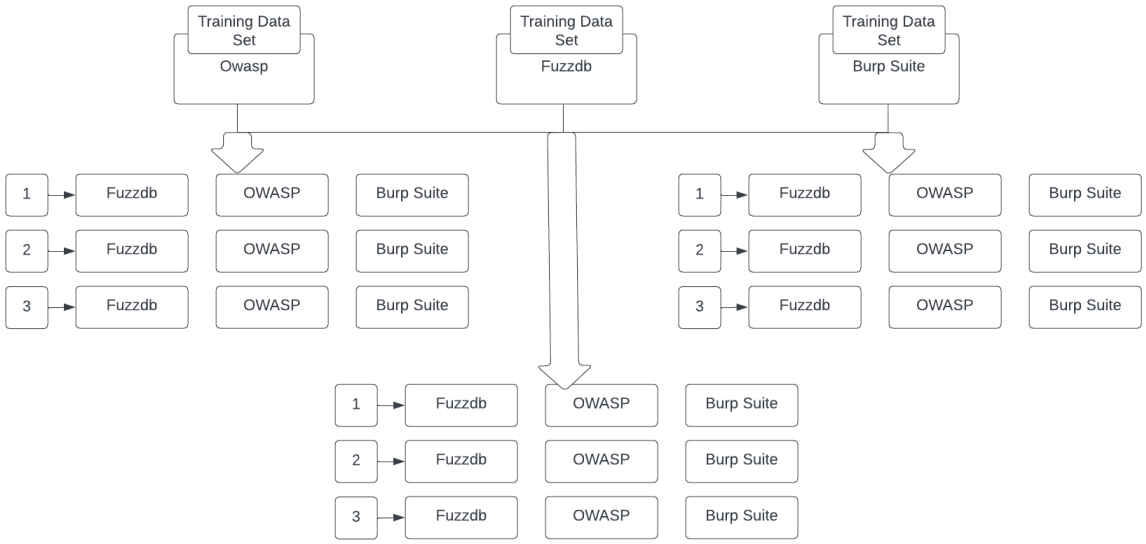

Figure 4.4: **Flow Chart**

In addition, i constructed a flow chart of our training data set, as shown in Figure 4.4. Each training data set was mentioned in the flowchart, along with their respective testing files.

4.2 Neural Network classifier Approach

Above we used a neural network-based model, In generally, a neural network is used to convert data from one form to the desired output, which is usually in another form. A computational learning system is a neural network. There are input layers, output layers, and processing layers in a neural network. Different types of layers are used in neural networks to analyse and learn data. A neural network is typically used to solve complicated problems.

We previously mentioned the neural network-based model, so now we'll compare our

naive Bayes result to the neural network-based SQL injection detection. We used one data file in a neural network-based data set, with two parameters: one is the unique value, and the other is the label. We have a 3951 value as a payload in unique values, and 0 and 1 in the label.As an outcome, Label values 1 and 0 indicate SQL injection and normal data, respectively. We used two ways in this data set using the neural network, and both methods produced accurate results. The first method is using Logistic Regression and the second is using a simple Neural Network.We have accuracy with a confusion matrix in the simple neural network. Using Logistic Regression, we get 0.928 accuracy after trained our data set. Using a simple Neural Network, we get 0.977 accuracy, 0.929 accuracy, and 1.0 recall value.

	Using simple Neural Network	Using Logistic Regression
Accuracy	0.977	0.928
Precision	0.929	0
Recall	1.0	0

Figure 4.5: Using Neural Network and Logistic Regression

The above figure 4.5 shows the accuracy of simple neural networks and logistic regression. As a result, we can compare the results of our naive Bayes method with the results of the neural network method for SQL injection detection. We may say that the Naive Bayes approach detects SQL injection attacks with the greatest accuracy. Another reason to use the Naive Bayes approach is that we used the multinominal methodology in the Naive Bayes method. This hasn't been used in the detection of SQL injection attacks in the past. As a conclusion of our survey, research, and implementation, we concluded that the Naive Bayes approach is currently the best at detecting SQL injection attacks. Finally, the accuracy of several machine learning approaches is shown in the table below. As a result, we may conclude that the Naive Bayes approach is the best at detecting SQL injection attacks.

The neural network-based model and the Naive Bayes model were discussed previously. As a result, we may say that the Naive Bayes model is more accurate than the Neural Network model. For this reason, we detailed the accuracy of the Naive Bayes-based data set above. When dealing with large or small data sets, Naive Bayes is commonly used.

As a result, we may conclude that the Burp Suite module can provide more accuracy.

Name	Accuracy
Naive Bayes	0.998
Simple Neural Network	0.977
Logistic Regression	0.928
Random Forest	0.991
Support Vector Machines	0.817
Convolutional Neural Network (CNN)	0.975

Figure 4.6: Final Result

Chapter 5

Conclusion

In this research, we have tried to enhance the detection of SQL injection attacks. the proposed system is based on a neural network model and we cover URL generator and URL classifier in that model. we have seen a detailed description of Naive Bayes algorithm. The detailed proposed model is also discussed in this paper. The testing result of the algorithm is listed and their accuracy of them is checked. a flowchart is also designed to see the working of the data set. we get higher accuracy in naive Bayes model compared to the neural network model so for future research, we may propose a better model to get accuracy in the neural network-based model.

Bibliography

[1] Muhammmad Saidu Aliero and Imran Ghani. A component based sql injection vulnerability detection tool. In *2015 9th Malaysian Software Engineering Conference (MySEC)*, pages 224–229, 2015.

[2] Marco Barreno, Blaine Nelson, Anthony D Joseph, and J Doug Tygar. The security of machine learning. *Machine Learning*, 81(2):121–148, 2010.

[3] Anna L. Buczak and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18:1153–1176, 2016.

[4] R Sri Devi and M Mohan Kumar. Testing for security weakness of web applications using ethical hacking. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, pages 354–361. IEEE, 2020.

[5] Jose Fonseca, Marco Vieira, and Henrique Madeira. Testing and comparing web vulnerability scanning tools for sql injection and xss attacks. In *13th Pacific Rim international symposium on dependable computing (PRDC 2007)*, pages 365–372. IEEE, 2007.

[6] Jai Narayan Goel and B.M. Mehtre. Vulnerability assessment penetration testing as a cyber defence technology. *Procedia Computer Science*, 57:710–715, 2015. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[7] Musaab Hasan, Zayed Balbahaith, and Mohammed Tarique. Detection of sql injection attacks: A machine learning approach. In *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pages 1–6, 2019.

[8] Anamika Joshi and V. Geetha. Sql injection detection using machine learning. *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 1111–1115, 2014.

[9] Krit Kamtuo and Chitsutha Soomlek. Machine learning for sql injection prevention on server-side scripting. *2016 International Computer Science and Engineering Conference (ICSEC)*, pages 1–6, 2016.

[10] Yugansh Khera, Deepansh Kumar, Nidhi Garg, et al. Analysis and impact of vulnerability assessment and penetration testing. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 525–530. IEEE, 2019.

[11] Anil Lamba. Cyber attack prevention using vapt tools (vulnerability assessment & penetration testing). *Cikitusi Journal for Multidisciplinary Research*, 1(2), 2014.

[12] Ao Luo, Wei Huang, and Wenqing Fan. A cnn-based approach to the detection of sql injection attacks. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)*, pages 320–324. IEEE, 2019.

[13] Sangeeta Nagpure and Sonal Kurkure. Vulnerability assessment and penetration testing of web application. In *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, pages 1–6, 2017.

[14] Keyur Patel. A survey on vulnerability assessment amp; penetration testing for secure communication. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 320–325, 2019.

[15] Naghmeh Moradpoor Sheykhkanloo. A learning-based neural network model for the detection and classification of sql injection attacks. *International Journal of Cyber Warfare and Terrorism (IJCWT)*, 7(2):16–41, 2017.

[16] Prashant S. Shinde and Shrikant B. Ardhapurkar. Cyber security analysis using vulnerability assessment and penetration testing. In *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, pages 1–5, 2016.

[17] Robin Sommer and Vern Paxson. Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy*, pages 305–316, 2010.

[18] Peng Tang, Weidong Qiu, Zheng Huang, Huijuan Lian, and Guozhen Liu. Detection of sql injection based on artificial neural network. *Knowledge-Based Systems*, 190:105528, 2020.

[19] Ovidiu Valea and Ciprian Oprişa. Towards pentesting automation using the metasploit framework. In *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 171–178. IEEE, 2020.

[20] Kevin Zhang. A machine learning based approach to identify sql injection vulnerabilities. In *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, pages 1286–1288. IEEE, 2019.