A PROJECT REPORT ON

Detection Of Phishing Website Using Machine Learning

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING
IN
(Information Technology)

BY

Faheem Shaikh	17
Pooja Garg	14
Kunal Rokde	16
Omkar Shelke	19

Under The Guidance of
Prof. A. N. Varade

Sinhgad Institutes

DEPARTMENT OF INFORMATION TECHNOLOGY

Sinhgad Academy of Engineering
Kondhwa,Pune, Maharashtra,India

2021-2022

Sinhgad Institutes

CERTIFICATE

This is to certify that the project phase-I report entitled

Detection Of Phishing Website Using Machine Learning

Submitted by

Faheem Shaikh	17
Pooja Garg	14
Kunal Rokde	16
Omkar Shelke	19

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. A. N. Varade** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of Bachelor of Engineering (Information Technology)

Prof. A. N. Varade
Guide

Dr. S. S. Kulkarni
HOD

Dr. K.P.Patil
Principal.

Abstract

Phishing is regularly an ordinary assault on persons via making them reveal their total one-of-a-kind info the use of counterfeit web sites. The purpose of phishing records process tool URLs is to pinch the personal information like consumer name, passwords and online banking transactions. Phishers(attackers) uses the websites that rectangular diploma visually and semantically the photo of these actual web-sites. As the era continues to grow, phishing techniques started to progress quickly and this may be prevented by way of exercise anti-phishing mechanisms to find phishing. Machine planning to apprehend can be a powerful device that is regularly used in the direction of phishing attacks. The proposed system has also surveyed the capabilities used for the detection and detection techniques by the use of Machine learning.

Keywords:-Phishing, Phishing Websites, Detection, Machine Learning.

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on **' Detection Of Phishing Website Using Machine Learning'**

I would like to take this opportunity to thank my internal guide **Prof. A. N. Varade** Department of Information Technology, Sinhgad Academy of Engineering, for his unconditional guidance. I am really grateful to them for their kind support. Their valuable suggestions were very helpful

An erudite teacher, a magnificent person and a strict disciplinarian, we consider ourselves fortunate to have worked under her supervision. We are thankful to Honorable Principal, Sinhgad Academy of Engineering **Dr.K.P.Patil** for the support.

We are highly grateful to **Dr. S. S. Kulkarni** Head of Department, Information Technology, Sinhgad Academy of Engineering, for indispensable support, suggestions.

We admit thanks to project coordinator, Department of Information Technology, Sinhgad Academy of Engineering for giving us such an opportunity to carry on such mind stimulating and innovative Project.

Faheem Shaikh 17

Pooja Garg 14

Kunal Rokde 16

Omkar Shelke 19

(B E. Information Technology)

List of Figures

List of Tables

INDEX

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Phishing imitates the characteristics and alternatives of emails and makes it appear similar due to the fact the original one. It seems nearly like that of the legitimate supply. The consumer thinks that this e-mail has come back from a real employer or a corporation. This makes the consumer to forcefully visit the phishing internet site thru the hyperlinks given inside the phishing email. These phishing web sites region unit created to mock the seams of an ingenious website. The phishers force person to inventory up the non-public info via giving baleful messages or validate account messages etc. so that they inventory up the preferred data which might be utilized by them to misuse it. They devise things such as the user isn't always left with the other choice but to go to their spoofed web site. Phishing is the most hazardous criminal physical activities in the cyber region. Since the maximum of the customers logs on to get admission to the services supplied with the aid of government and financial establishments, there has been a significant boom in phishing attacks for the beyond few years. Phishers commenced to earn cash and that they try this as a thriving business. Phishing may be law-breaking, the explanation behind the phishers doing this crime is that it is terribly trustworthy to try to do this, it doesn't value something and it effective.

The phishing will truly get entry to the e-mail identity of somebody it's terribly sincere to are looking for out the email identification currently every day and you will send an email to every person is freely offered throughout the globe. These attacker's vicinity terribly much less price and electricity to urge valuable know-how quick and truly. The phishing frauds effects malware infections, statistics loss, fraud, etc. information at some stage in which those cyber criminals have an interest is that the crucial data of a user similar to the password, OTP, credit/ debit card numbers CVV, sensitive know-how associated with business, medical understanding, confidential information, etc commonly these criminals conjointly acquire data which may provide them directly get admission to do the social media account their emails. A lot of software /ways and algorithms area unit used for phishing detection. These area unit used at academic and industrial organization levels. A phishing address and also the

parallel online page have several capabilities which could be one-of-a-kind from the address that allows us to take associate degree instance to cover the initial domain decision the phishing assaulter will sense terribly protracted and confusing name of the domain. This is often terribly effortlessly visible.

1.2 OBJECTIVES OF PROJECT REPORT

- To conduct a comparative assessment between various classification data mining algorithms techniques, and various feature selection scenarios.
- To develop multi-classifier integration model by combining clustering and more than one classification technique to enhance detection and protecting phishing websites.
- To determine the best classification algorithm for phishing detection.

1.3 ORGANIZATION OF PROJECT REPORT

- Chapter 2 Deals with the Project Related Work i.e Literature Survey.
- Chapter 3 Giving an overall view of the techniques used in the system
- Chapter 4 Deals with System Design.
- Chapter 5 Project Plan
- Chapter 6 Implementation Part
- Conclusion and at last references

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

- A. Lakshmanarao and P.Surya Prabhakara Rao, “Phishing website detection using novel machine learning fusion approach ”, IEEE 2021

  various machine learning algorithms logistic regression, decision tree classifier, random forest classifier, AdaBoost, gradient boosting classifier for the phishing detection

- Jitendra Kumar and A. Santhanavijayan , “Phishing Website Classification and Detection Using Machine Learning ”, International Conference on Computer Communication and Informatics, 2020

  In this paper logistic regression , Gaussian Naïve Bayes and Random Forest were been proposed

- Mehmet Korkmaz and Ozgur Koray Sahingoz, “Detection of Phishing Websites by Using Machine Learning-Based URL Analysis”, IEEE 2020

  Proposed a machine learning-based phishing detection system by using eight different algorithms to analyze the URLs, and three different datasets to compare the results with other works.

- Charu Singh , “Phishing Website Detection Based on Machine Learning: A Survey”,IEEE 2020

  Review was made With the huge number of phishing emails or messages received every day, companies or individuals are not able to detect all of them, where different reviews were given for detection of phishing attack, by using machine learning.

- Vaibhav Patil and Pritesh Thakkar , “Detection and Prevention of Phishing Websites using Machine Learning Approach”, IEEE 2018

Proposed three approaches for detecting phishing websites. First is by analyzing various features of URL , second is by checking legitimacy of website by knowing where the website is being hosted and who are managing it, the third approach uses visual appearance based analysis for checking genuineness of website.

- T. Peng, I. Harris, and Y. Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning," Proc. - 12th IEEE Int. Conf. Semant. Comput. ICSC 2018, vol. 2018–Janua, pp. 300–301, 2018.

  Provides an approach to detecting phishing email attacks using an analysis of linguistic communication and machine learning. It is accustomed search the text's syntax to detect malicious intent. A natural language processing (NLP) technique is employed in conjunction with a predicate to decode each sentence and identifies the semantic jobs of words within the sentence. Computer supervised learning is employed to come up with the blacklist of malicious pairs.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 INTRODUCTION

This software requirement specification (SRS) report expresses complete description about proposed System. This document includes all the functions and specifications with their explanations to solve related problems.

3.1.1 Problem Statement

Phishing is an online crime that tries to trick unsuspected users to expose their sensitive (and valuable) personal information, for example, usernames, passwords, financial account details, personal addresses, SSN, and social relationships, to the miscreant, often for malicious reasons. Phishing is normally perpetrated by disguising as a trustworthy entity in internet communication which is achieved by combining both social engineering and technical tricks.

3.1.2 Project Scope

The main purpose of the proposed system is to detect the phishing websites who are trying to get access to sensitive data or by creating the fake website and trying to get access of user's personal credentials. We are using machine learning algorithm to safeguard the sensitive data and detect the phishing website who are trying to gain access on sensitive data.

3.1.3 Project Perspective

The proposed system will provide good and easy graphical user interface to both new as well as experienced user of the computer.The web based application is been developed specifically for detecting various pishing websites.

3.1.4 User Classes and Characteristics

Basic knowledge of using computers is adequate to use this application.
Knowledge of how to use a mouse or keyboard and internet browser is necessary.
The user interface will be friendly enough to guide the user.

3.1.5 Assumptions and Dependencies

- Assumptions:

  [1] The product must have an interface which is simple enough to understand.

  [2] All the software such as python, etc are installed and running on the computers

- Dependencies:

  [1] All necessary software's are available for implementing and use of the system. proposed system would be designed, developed and implemented based on the software requirements specifications document. users should have basic knowledge of computer and we also assure that the users will be given software training documentation and reference material.

  [2] Well Trained dataset

3.2 FUNCTIONAL REQUIREMENT

3.2.1 System Feature 1(Functional Requirement)

Functional requirement describes features, functioning, and usage of a product/system or software from the perspective of the product and its user. Functional requirements are also called as functional specifications were synonym for specification is design. Provide User friendly Interface and Interactive as per standards.

3.3 NON-FUNCTIONAL REQUIREMENT

3.3.1 Performance Requirements

- High Speed :- System should process requested task in parallel for various action to give quick response. Then system must wait for process completion.
- Accuracy :- System should correctly execute process, display the result accurately. System output should be in user required format.

3.3.2 Safety Requirements:

The data safety must be ensured by arranging for a secure and reliable transmission media. The source and destination information must be entered correctly to avoid any misuse or malfunctioning. Password generated by user is consisting of characters, special character number so that password is difficult to hack. So, that user account is safe.

3.3.3 Security Requirements

Secure access of confidential data (user's details). Information security means protecting information and information systems from unauthorized access, use, disclosure, disruption, modification or destruction.The terms information security, computer security and information assurance are frequently incorrectly used interchangeably. These fields are interrelated often and share the common goals of protecting the confidentiality, integrity and availability of information; however, there are some subtle differences between them.

3.3.4 Software Quality Attributes

[1] Runtime System Qualities: Runtime System Qualities can be measured as the system executes.

[2] Functionality: The ability of the system to do the work for which it was intended.

[3] Performance:The response time, utilization, and throughput behavior of the system. Not to be confused with human performance or system delivery time.

[4] Security:A measure of systems ability to resist unauthorized attempts at usage or behavior modification, while still providing service to legitimate users.

[5] Availability: (Reliability quality attributes falls under this category) the measure of time that the system is up and running correctly; the length of time between failures and the length of time needed to resume operation after a failure.

[6] Usability: The ease of use and of training the end users of the system. Sub qualities: learn ability, efficiency, affect, helpfulness, control.

[7] Interoperability: The ability of two or more systems to cooperate at runtime.

3.4 SYSTEM REQUIREMENT

3.4.1 Software Requirements(Platform Choice)

- Tools - Python IDE
- Programming Language - Python
- Software Version - Python 3.5

3.4.2 Hardware Requirements

- Processor - Pentium IV/Intel I3 core
- Speed - 1.1 GHz
- RAM - 512 MB (min)
- Hard Disk - 20GB
- Keyboard - Standard Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - LED Monitor

3.5 ANALYSIS MODELS: (SDLC MODEL TO BE APPLIED)

One of the basic notions of the software development process is SDLC models which stands for Software Development Life Cycle models. SDLC is a continuous process, which starts from the moment, when its made a decision to launch the project, and it ends at the moment of its full remove from the exploitation. There is no one single SDLC model. They are divided into main groups, each with its features and weaknesses. Evolving from the first and oldest waterfall SDLC model, their variety significantly expanded.

The SDLC models diversity is predetermined by the wide number of product types starting with a web application development to a complex medical software. And if you take one of the SDLC models mentioned below as the basis in any case, it should be adjusted to the features of the product, project, and company. The most used, popular and important SDLC models are given below:

[1] Waterfall Model

[2] Iterative Model

[3] Spiral Model

[4] V-shaped Model

[5] Agile Model

Waterfall Model

Waterfall is a cascade SDLC model, in which development process looks like the flow, moving step by step through the phases of analysis, projecting, realization, testing, implementation, and support. This SDLC model includes gradual execution of every stage completely. This process is strictly documented and predefined with features expected to every phase of this software development life cycle model.

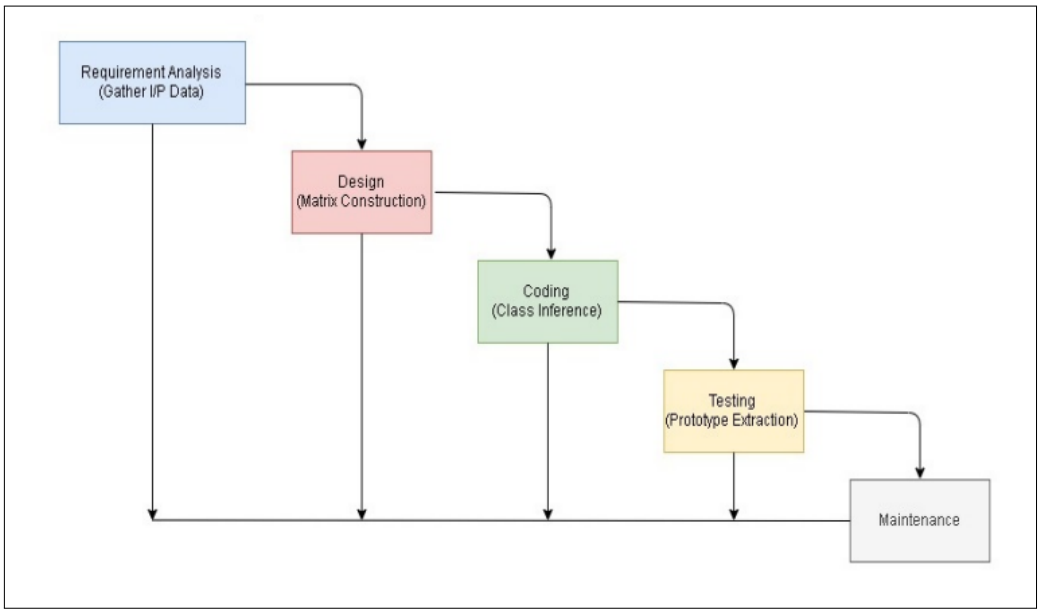


Figure 3.1: Waterfall Model

[1] **Planning and requirement analysis**

Each software development life cycle model starts with the analysis, in which the stakeholders of the process discuss the requirements for the final product. The goal of this stage is the detailed definition of the system requirements. Besides, it is needed to make sure that all the process participants have clearly understood the tasks and how every requirement is going to be implemented. Often, the discussion involves the QA specialists who can interfere the process with additions even during the development stage if it is necessary.

[2] **Designing project architecture**

At the second phase of the software development life cycle, the developers are actually designing the architecture. All the different technical questions that may appear on this stage are discussed by all the stakeholders, including the customer. Also, here are defined the technologies used in the project, team load, limitations, time frames, and budget. The most appropriate project decisions are made according to the defined requirements.

[3] **Development and programming**

After the requirements approved, the process goes to the next stage actual development. Programmers start here with the source code writing while keeping

in mind previously defined requirements. The system administrators adjust the software environment, front-end programmers develop the user interface of the program and the logics for its interaction with the server. The programming by itself assumes four stages:-

- Algorithm development
- Source code writing
- Compilation
- Testing and debugging

[4] **Testing**

The testing phase includes the debugging process. All the code flaws missed during the development are detected here, documented, and passed back to the developers to fix. The testing process repeats until all the critical issues are removed and software work ow is stable.

[5] **Deployment**

When the program is finalized and has no critical issues it is time to launch it for the end users. After the new program version release, the tech support team joins. This department provides user feedback; consult and support users during the time of exploitation. Moreover, the update of selected components is included in this phase, to make sure, that the software is up-to-date and is invulnerable to a security breach.

CHAPTER 4

SYSTEM DESIGN

4.1 DATA FLOW DIAGRAM

4.1.1 Level 0 Data Flow Diagram

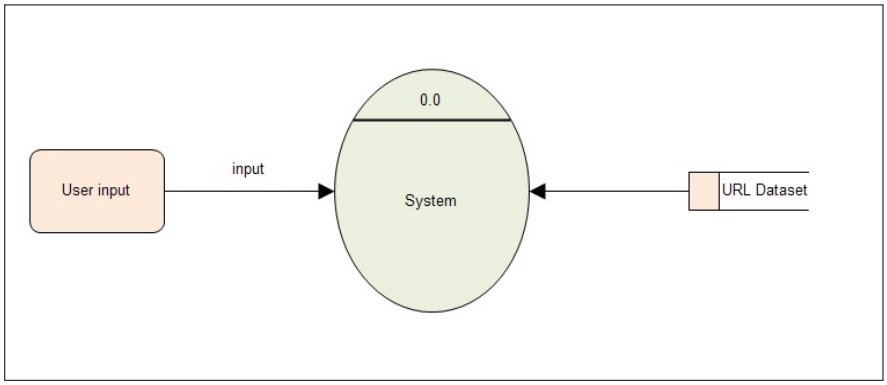


Figure 4.1: DFD Level 0

4.1.1.1 Level 0 Data Flow Diagram

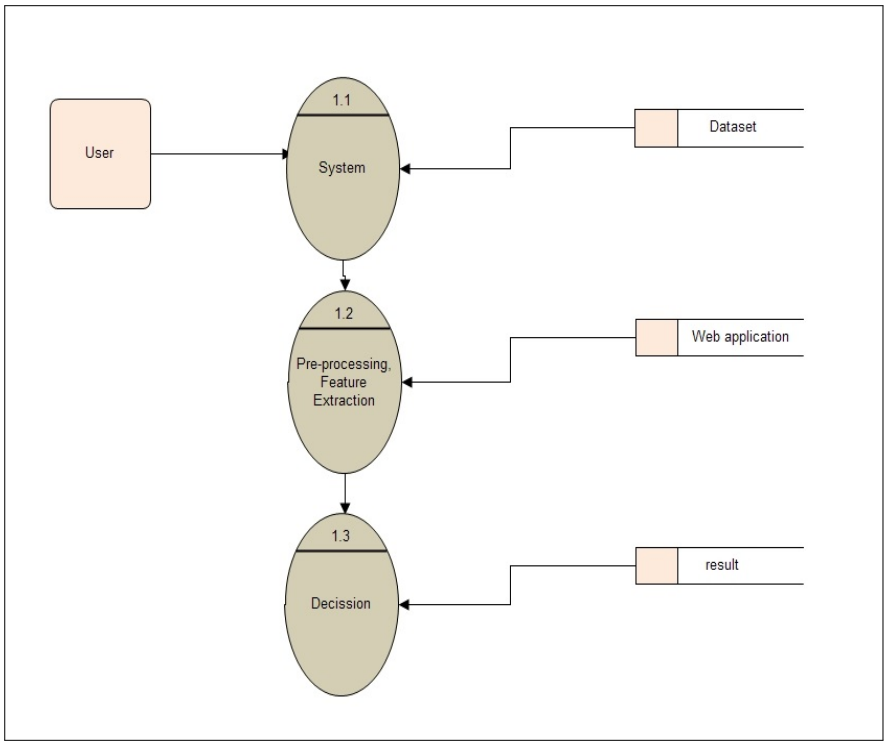


Figure 4.2: DFD Level 1

4.2 UML DIAGRAM

4.2.1 Use-cases

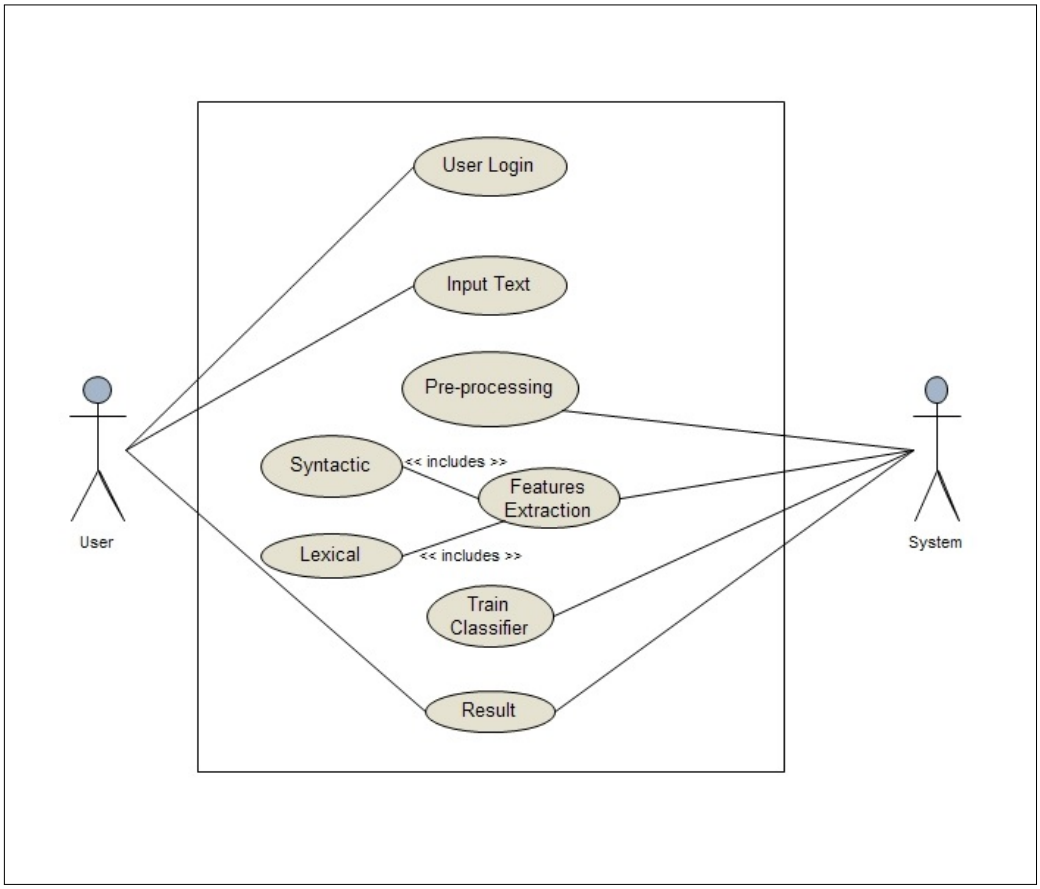


Figure 4.3: UseCase Diagram

4.2.2 Activity Diagram:

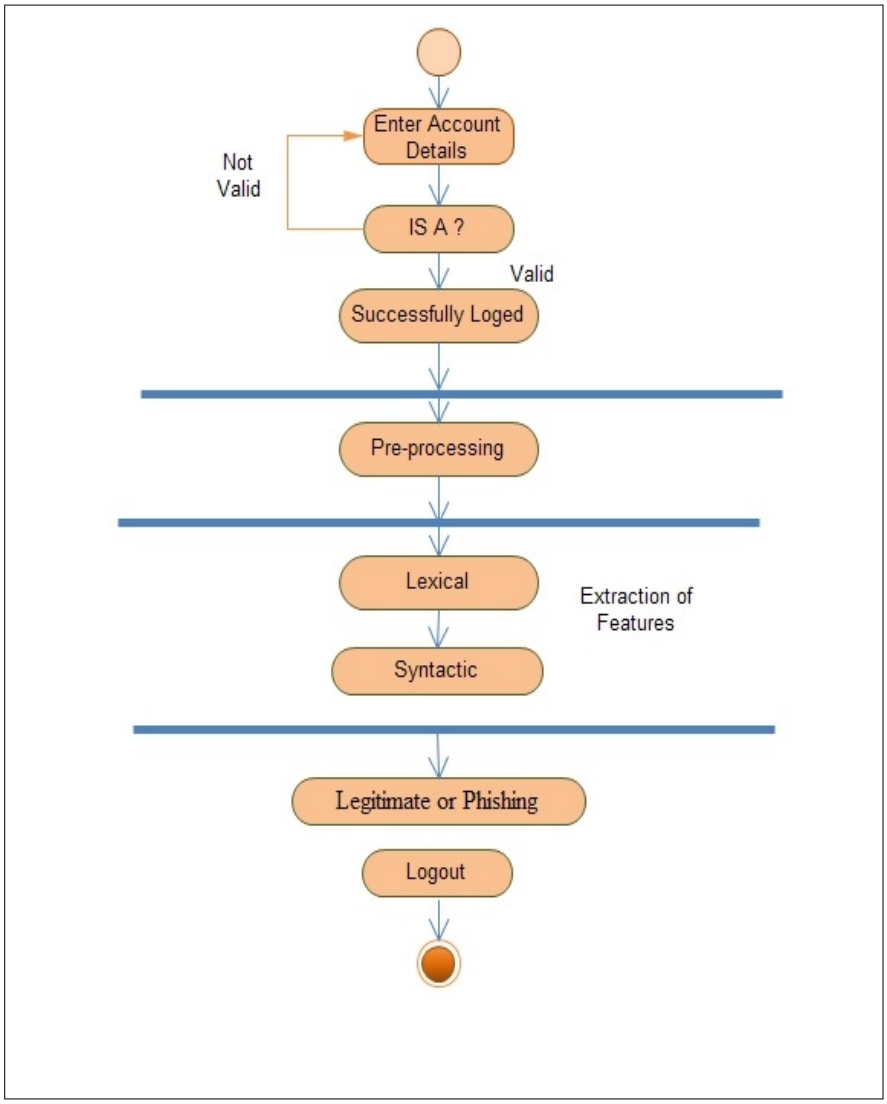


Figure 4.4: Activity Diagram

4.2.3 Class Diagram:

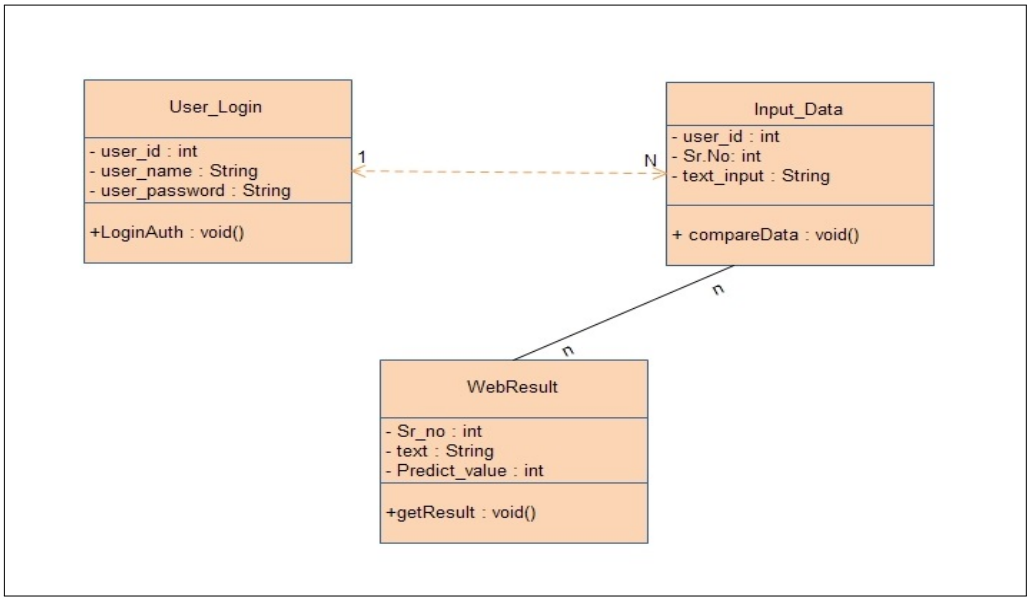


Figure 4.5: Class Diagram

4.2.4 Sequence Diagram:

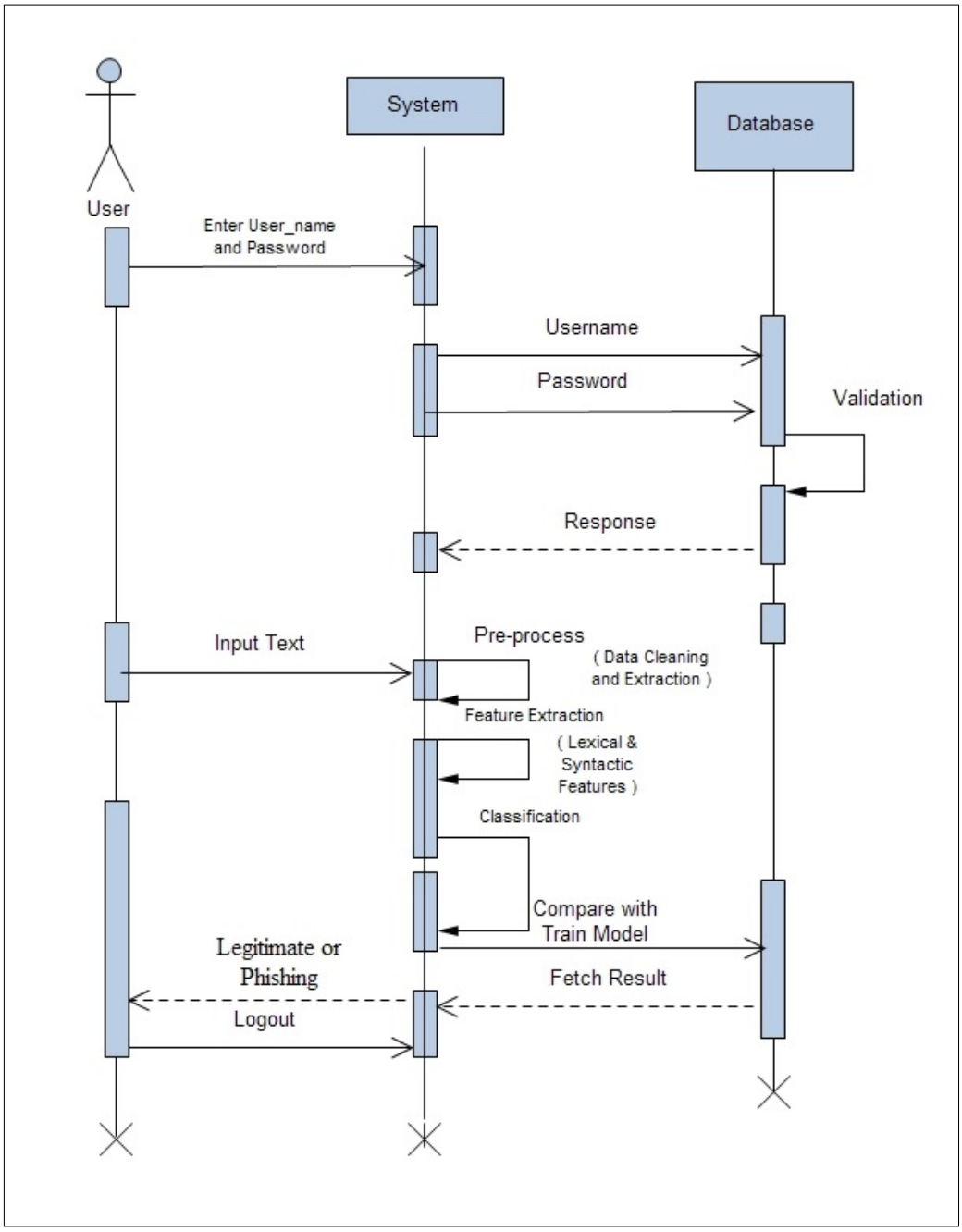


Figure 4.6: Sequence Diagram

4.2.5 Component Diagram:

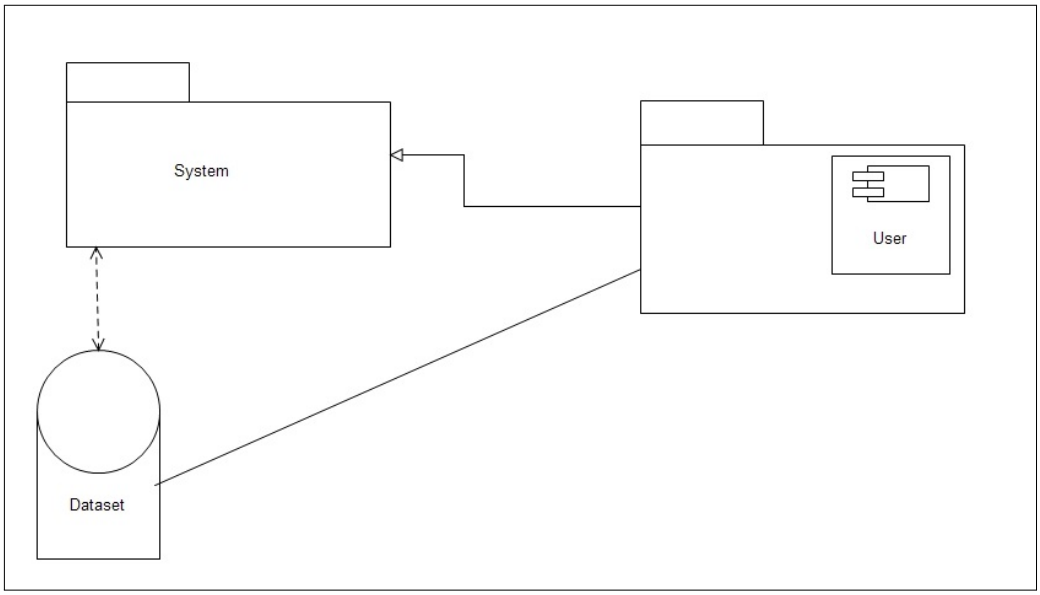


Figure 4.7: Component Diagram

4.2.6 Deployment Diagram:

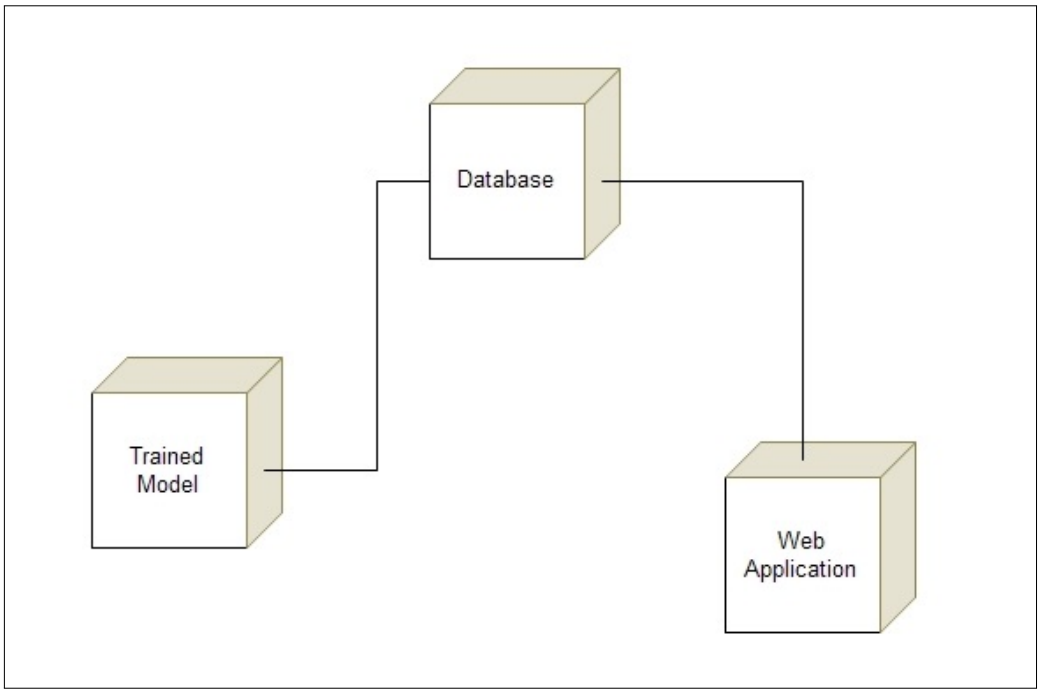


Figure 4.8: Deployment Diagram

CHAPTER 5

PROJECT PLAN

5.1 PROJECT ESTIMATES

5.1.1 Effort Estimate Table:

Task	Effort weeks	Deliverables	Milestones
Analysis of existing systems & compare with proposed one	4 weeks		
Literature survey	1 weeks		
Designing & planning	2 weeks		
System flow	1 weeks		
Designing modules & itś deliverables	2 week	Modules: design document	
Implementation	7 weeks	Primary system	
Testing	4 weeks	Test Reports	Formal
Documentation	2 weeks	Complete project report	Formal

Table 5.1: Effort Estimate Table

5.1.2 Project Description:

Phase	Task	Description
Phase 1	Analysis	Analyse the information given in the IEEE paper.
Phase 2	Literature survey	Collect raw data and elaborate on literature surveys.
Phase 3	Design	Assign the module and design the process flow control.
Phase 4	Implementation	Implement the code for all the modules and integrate all the modules.
Phase 5	Testing	Test the code and overall process weather the process works properly.
Phase 6	Documentation	Prepare the document for this project with conclusion and future enhancement.

Table 5.2: Project Scheduling

5.1.3 Estimation of KLOC:

The number of lines required for implementation of various modules can be estimated as follows:

Sr.No.	Modules	KLOC
1	Graphical User Interface	0.20
2	Back-end Algorithm Implementation	1.2
3	Front-Side Coding	1.2
4	Back-end Connectivity	0.6

Thus the total number of lines required is approximately 2.60 KLOC.

D = (Total KLOC / KLOC in a day)/30

= (3.6/0.025)/30

= 4.8

5.2 RISK MANAGEMENT

5.2.1 Overview of Risk Mitigation, Monitoring, Management

Risk management organizational role

Each member of the organization will undertake risk management. The development team will consistently be monitoring their progress and project status as to identify present and future risks as quickly and accurately as possible. With this said, the members who are not directly involved with the implementation of the product will also need to keep their eyes open for any possible risks that the development team did not spot. The responsibility of risk management falls on each member of the organization, while William Lord maintains this document.

Business Impact Risk

- Amount and quality of documentation that must be produced and delivered to customer the customer will be supplied with a complete online help file and users manual for Game Forge. Coincidentally, the customer will have access

to all development documents for Game Forge, as the customer will also be grading the project.

- Governmental constraints in the construction of the product none known.
- Costs associated with late delivery Late delivery will prevent the customer from issuing a letter of acceptance for the product, which will result in an incomplete grade for the course for all members of the organization.
- Costs associated with a defective product Unknown at this time.

Customer Related Risks

- Have you worked with the customer in the past? Yes, All team members have completed at least one project for the customer, though none of them have been to the magnitude of the current project.
- Does the customer have a solid idea of what is required? Yes, the customer has access to both the System Requirements Specification, and the Software Requirements Specification.
- Will the customer agree to spend time in formal requirements gathering meetings to identify project scope? Unknown. While the customer will likely participate if asked, the inquiry has not yet been made.

Process Risks

- Does senior management support a written policy statement that emphasizes the importance of a standard process for software development? N/A. PA Software does not have a senior management. It should be noted that the structured method has been adopted. At the completion of the project, it will be determined if the software method is acceptable as a standard process, or if changes need to be implemented.
- Has your organization developed a written description of the software process to be used on this project? Yes.

- Are staff members willing to use the software process? Yes. The software process was agreed upon before development work began.
- Is the software process used for other products? N/A. PA Software has no other projects currently.

Technical Issues

- Are facilitated application specification techniques used to aid in communication between the customer and the developer? The development team will hold frequent meetings directly with the customer. No formal meetings are held (all informal). During these meetings the software is discussed and notes are taken for future review.
- Are specific methods used for software analysis? Special methods will be used to analyze the software progress and quality. These are a series of tests and reviews to ensure the software is up to speed. For more information, see the Software Quality Assurance and Software Configuration Management documents.
- Do you use a specific method for data and architectural design? Data and architectural design will be mostly object oriented. This allows for a higher degree data encapsulation and modularity of code.

Technology Risk

- Is the technology to be built new to your organization? No
- Does the software interface with new or unproven hardware? No
- Is a specialized user interface demanded by the product requirements? Yes.

Development Environment Risks Is a software project management tool available? No. No software tools are to be used. Due to the existing deadline, the development team felt it would be more productive to begin implementing the project than trying to learn new software tools. After the completion of the project software tools may be implemented for future projects.

5.3 TIME LINE CHART

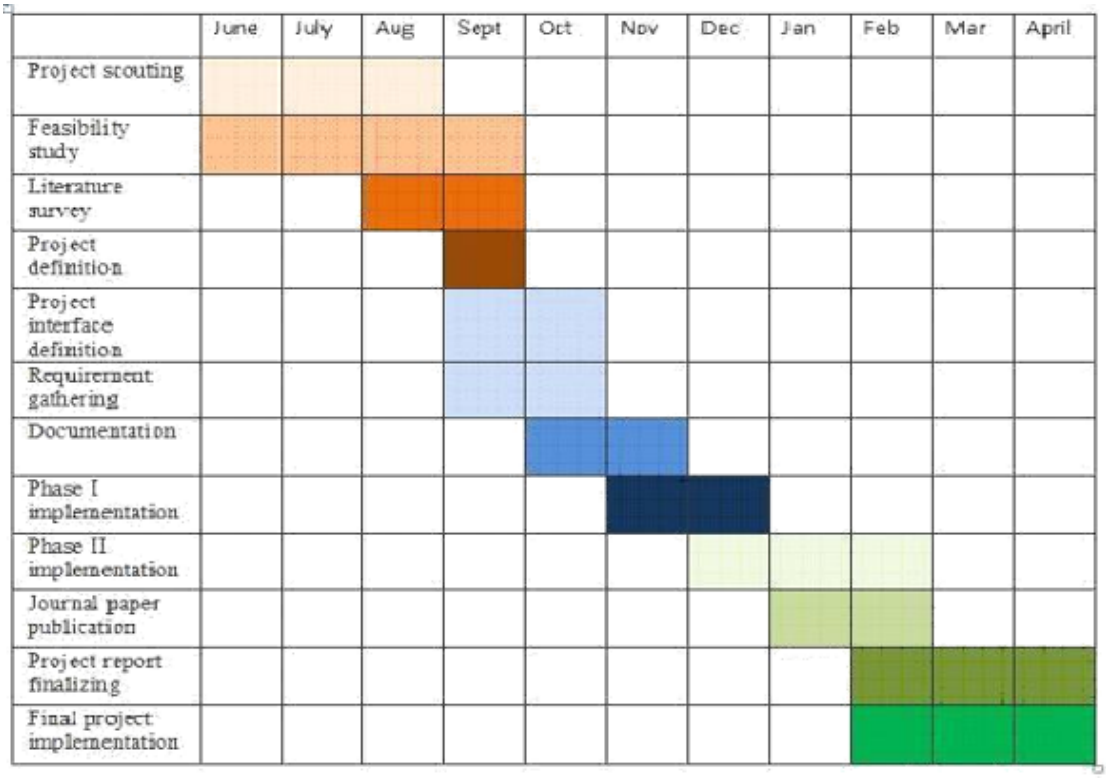


Figure 5.1: Time Line Chart

CHAPTER 6

IMPLEMENTATION

6.1 SYSTEM ARCHITECTURE

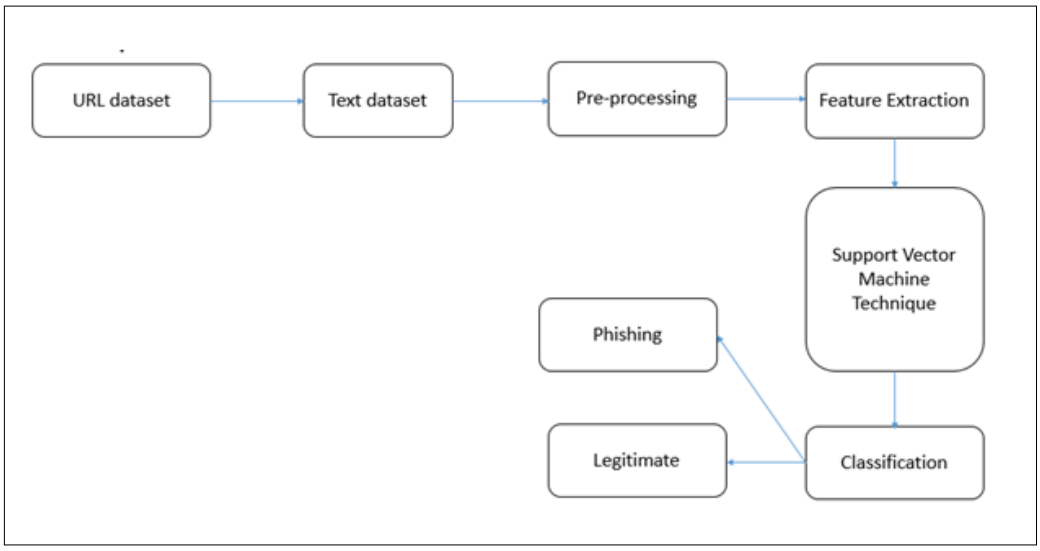


Figure 6.1: Block Diagram

The dataset of phishing and legitimate URLs is provided within the application which is then pre-processed so that the facts are within the working format for analysis. The functions have round various traits of phishing websites that have used to distinguish them from legitimate ones. Each category has its very own traits of phishing attributes and values are defined. The specified traits are extracted for every URL and legitimate stages of inputs are identified. These values are then assigned to every phishing internet site risk. The phishing properties esteems are spoken to with double no 0 and 1 which appears the characteristic is present or not.

6.2 ALGORITHM

6.2.1 SVM

SVM goes through following steps:-

- Import the dataset
- Explore the data to figure out what they look like
- Pre-process the data

- Split the data into attributes and labels
- Divide the data into training and testing sets
- Train the SVM algorithm
- Make some predictions
- Evaluate the results of the algorithm

6.2.2 KNN

K-Nearest Neighbors (KNN) KNN is used to solve the classification model problems. K-nearest neighbor or K-NN algorithm basically creates an imaginary boundary to classify the data. When new data points come in, the algorithm will try to predict that to the nearest of the boundary line. Therefore, larger k value means smother curves of separation resulting in less complex models. Whereas, smaller k value tends to overfit the data and resulting in complex models.

Example of finding accuracy using KNN :-

Import necessary modules

from sklearn.neighbors import KNeighborsClassifier

from sklearn.model_selection import train_test_split

from sklearn.datasets import load_shaming

import numpy as np

import matplotlib.pyplot as plt

shamingData = load_shaming()

Create feature and target arrays

X = shamingData.data

y = shamingData.target

Split into training and test set

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state=42)

neighbors = np.arange(1, 9)

train_accuracy = np.empty(len(neighbors))

```
test_accuracy = np.empty(len(neighbors))
Loop over K values
for i, k in enumerate(neighbors):
knn = KNeighborsClassifier(n_neighbors=k)
knn.fit(X_train, y_train)
Compute training and test data accuracy
train_accuracy[i] = knn.score(X_train, y_train)
test_accuracy[i] = knn.score(X_test, y_test)
Generate plot
plt.plot(neighbors, test_accuracy, label = 'Testing dataset Accuracy')
plt.plot(neighbors, train_accuracy, label = 'Training dataset Accuracy')
plt.legend()
plt.xlabel('n_neighbors')
plt.ylabel('Accuracy')
plt.show()
```

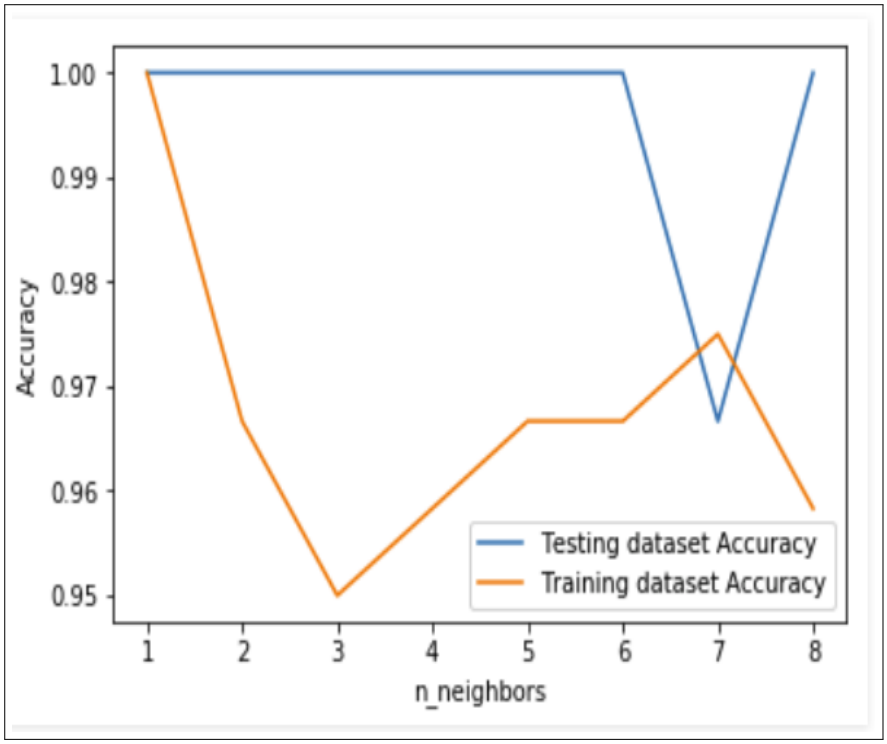


Figure 6.2: KNN Accuracy Result

6.3 TOOLS AND TECHNOLOGY USED

Python

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using signifficant whitespace. It provides constructs that enable clear programming on both small and large scales.Van Rossum led the language community until July 2018. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

Python features a comprehensive standard library, and is referred to as ”batteries included”. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open-source software and has a community-based development model. Python and CPython are managed by the non-profit Python Software Foundation.

Python is a general-purpose object-oriented programming language with high-level programming capabilities. It has become famous because of its apparent and easily understandable syntax, portability and easy to learn. Python is a programming language that includes features of C and Java. It provides the style of writing an elegant code like C, and for object-oriented programming, it offers classes and objects like Java.

- Python was developed in the late eighties, i.e., late 1980's by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands as a successor of ABC language capable of exception handling and interfacing.
- Python is derived from programming languages such as ABC, Modula 3, small talk, Algol-68. Van Rossum picked the name Python for the new language from a TV show, Monty Python's Flying Circus.
- Python page is a file with a .py extension that contains could be the combina-

tion of HTML Tags and Python scripts.

- In December 1989 the creator developed the 1st python interpreter as a hobby and then on 16 October 2000, Python 2.0 was released with many new features. On 3rd December 2008, Python 3.0 was released with more testing and includes new features.
- Python is an open source scripting language., which means that anyone can download it freely from www.python.org and use it to develop programs. Its source code can be accessed and modified as required in the project.Python is one of the official languages at Google.

Features of Python

[1] Easy to Learn and Use. Python is easy to learn and use.

[2] Expressive Language. Python language is more expressive means that it is more understandable and readable.

[3] Interpreted Language.

[4] Cross-platform Language.

[5] Free and Open Source.

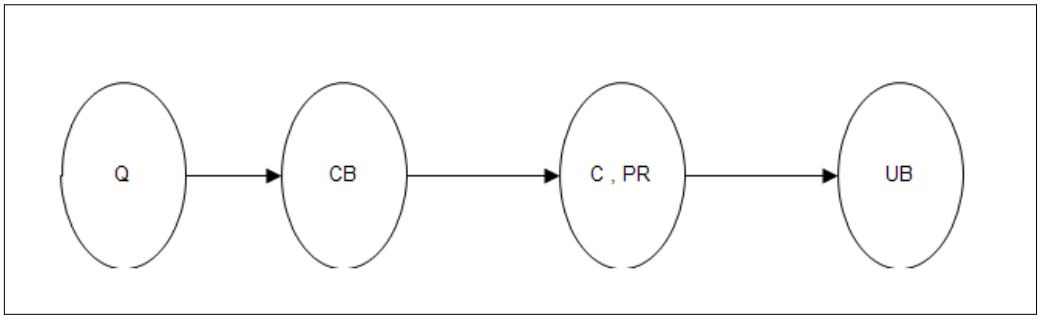
[6] Object-Oriented Language.

[7] Extensible.

[8] Large Standard Library.

6.4 MATHEMATICAL MODEL

A]



Where,

[1] Q = Input Image

[2] CB = Pre-processing

[3] C = Feature Extraction

[4] PR = Classification

[5] UB = Output

B] Set Theory

Let S be as system which allow users for detecting phishing sites using machine learning concepts. S = { In, P, Op, θ}

Identify Input In as

In ={ Q }

Where,

Q = Input Image

Identify Process P as

P = CB, C, PR

Where,

CB = Pre-processing

C = Feature Extraction

PR = Classification

Identify Output Op as

Op = { UB }

Where,

UB = Output

Failures: Huge database can lead to more time consumption to get the information. Hardware failure,Software failure.

Success: Search the required information from available in Datasets. User gets result very fast according to their needs.

Space Complexity: The space complexity depends on Presentation and visualization of discovered patterns. More the storage of data more is the space complexity.

Time Complexity: Check No. of patterns available in the datasets= n If (n¿1) then retrieving of information can be time consuming. So the time complexity of this algorithm is O(nn̂).

CHAPTER 7

CONCLUSION

Conclusion

The most important way to protect the user from phishing attack is the education awareness. Internet users must be aware of all security tips, which are given by experts. Every user should also be trained not to blindly follow the links to websites where they have to enter their sensitive information. It is essential to check the URL before entering the website. It has become a serious network security problem, facing financial loss of billions of dollars to both consumers and the e-commerce companies. And perhaps more eventually, phishing has made e-commerce distrusted and attractive to normal consumer.The most important way to protect the user from phishing attack is the education awareness. Internet users must be aware of all security tips which are given by experts. Every user should also be trained not to blindly follow the links to websites where they have to enter their sensitive information. It is essential to check the URL before entering the website in proposed system.

CHAPTER 8

REFERENCES

[1] A. Lakshmanarao and P.Surya Prabhakara Rao, “Phishing website detection using novel machine learning fusion approach ”, IEEE 2021

[2] Jitendra Kumar and A. Santhanavijayan , “Phishing Website Classification and Detection Using Machine Learning ”, International Conference on Computer Communication and Informatics, 2020

[3] Mehmet Korkmaz and Ozgur Koray Sahingoz, “Detection of Phishing Websites by Using Machine Learning-Based URL Analysis”, IEEE 2020 Charu Singh , “Phishing Website Detection Based on Machine Learning: A Survey”,IEEE 2020

[4] Vaibhav Patil and Pritesh Thakkar , “Detection and Prevention of Phishing Websites using Machine Learning Approach”, IEEE 2018

[5] T. Peng, I. Harris, and Y. Sawa, “Detecting Phishing Attacks Using Natural Language Processing and Machine Learning,” Proc. - 12th IEEE Int. Conf. Semant. Comput. ICSC 2018, vol. 2018–Janua, pp. 300–301, 2018.