**PROJECT REPORT**

**ON**

**PHISHING WEBSITE CHECKER SYSTEM**

**Submitted For Partial Fulfillment For Award Of**

**BACHELOR OF COMPUTER APPLICATIONS**

**BY**

**ABHISHEK VERMA (1604005)**

**UNDER THE GUIDANCE OF**

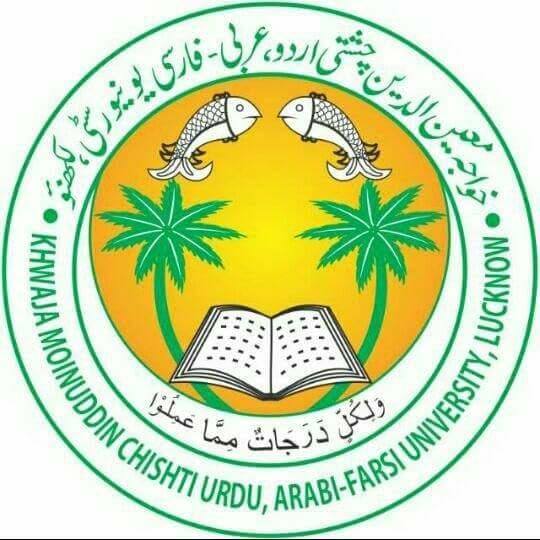
Dr. Mazhar KhaliqMr. Zaid Kamil

Internal Supervisor External Supervisor

(Head of Department Project Manager

Computer Science & (DIGIPODIUM)

Information Technology)

****

**Khwaja Moinuddin Chisti Urdu Arabi Farsi University , Lucknow**

**2018-2019**

**DEPARTMENT OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**

**CANDIDATE DECLARATION**

I, hereby certify that the work which is being presented I the project work entitled **“PHISHING WEBSITE CHECKER SYSTEM”** is the fulfillment of therequirement for award of the Bachelors in Computer Applications submitted in the Computer Science and Information Technology Department of KMCUAFU.

This is an authentic record of my work during 6 months period under the supervision of **Mr Zaid Kamil.**

The matter presented in the project work has not been submitted by me for the award of any Degree/Diploma of this or any other University.

Date : / /2019

Place : Lucknow (Signature of Candidate )

**TITLE OF THE PROJECT**

**PHISHING WEBSITE CHECKER**

**SYSTEM**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S NO.** | **DESCRIPTION**  Title of the project | **PAGE NO.** |
| 1 | Acknowledgement | 5 |
| 2 | Introduction | 6 |
| 3 | About the project | 7 |
| 4 | Problem statement | 8-11 |
| 5 | Objective and scope | 12 |
| 6 | Terms defined | 13 |
| 7 | Description and working | 14 |
| 8 | Theoretical/conceptual framework | 15 |
| 9 | Technology used | 16-24 |
| 10 | Python libraries used | 25-27 |
| 11 | Modules and files | 28 |
| 12 | Gantt chart | 29-30 |
| 13 | Problem statement | 31-32 |
| 14 | System analysis and design | 33-39 |
| 15 | UML diagram | 40 |
| 16 | Use case diagram | 41 |
| 17 | Lifecycle model: spiral | 42-43 |
| 18 | Snapshots | 44-47 |
| 19 | Coding | 48-69 |
| 20 | Testing | 70-84 |
| 21 | Implementation | 85 |
| 22 | Maintenance and support | 86 |
| 23 | Software requirements | 87 |
| 24 | Conclusion | 88 |
| 25 | Bibliography | 89 |

**ACKNOWLEDGEMENT**

I would like to express my thanks to the people who have helped me the most throughout my project. I am grateful to my teacher Dr. Mazhar Khaliq for nonstop support for the project.

A special thanks of mine goes to my Project Manager Mr. Zaid Bin Kamil who helped me in completing the project, where he exchanged their own interesting ideas, thoughts and made this possible to complete my project with all accurate information.

Last but not the least I want to thanks my friends who treasured me for my hard work and encouraged me and finally to God who made all the things possible for me till the end.

Thanking You

Abhishek Verma

**INTRODUCTION**

In the fast changing world , information technology and information management are going to play an important role. We are living in computer ages during past some years. The computer is gaining popularity. Computer revolution found its way into almost every aspect of human life and living. A computer is admirably suited to handle any information and hence is an information processor that is, it can receive data, perform some basic operations on that data and produces results according to a predetermined program.

**Problem**

Cybercriminals use phishing because it’s easy, cheap, and effective. Email addresses are easy to obtain, and emails are virtually free to send. With little effort and little cost, attackers can quickly gain access to valuable data. Those who fall for phishing scams may end up with malware infections (including [ransomeware](https://www.proofpoint.com/us/threat-reference/ransomware)), identity theft, and data loss.

Cybercriminals also use phishing attacks to gain direct access to email, social media, and other accounts — or to obtain permissions to modify and compromise connected systems, like point of sale terminals and order processing systems. Many of the biggest data breaches — like the headline-grabbing 2013 Target breach — start with a phishing email. Using a seemingly innocent email, cybercriminals can gain a small foothold and build on it.

**ABOUT THE PROJECT**

This project’s name is “Phishing Website Checker System” , this is being made for detecting phishing websites. It will check whether the url inputted by the user is a phishing’s website url or not.

Phishing is when cybercriminals send malicious emails designed to trick people into falling for a scam. The intent is often to get users to reveal financial information, system credentials, or other sensitive data.

The term “phishing” came about in the mid-1990s, when hackers began using fraudulent emails to “fish for” information from unsuspecting users. Since these early hackers were often referred to as “phreaks,” the term became known as “phishing,” with a “ph.” Phishing emails try to lure you in and get you to take the bait. And once you’re hooked, you’re in trouble.

Phishing is an example of social engineering: a collection of techniques scam artists use to manipulate human psychology. Social engineering techniques include forgery, misdirection, and lying, all of which can play a part in phishing attacks. On a basic level, phishing emails use social engineering to encourage you to act without thinking things through.

**PROBLEM STATEMENT**

**WHAT IS PHISHING?**

Phishing is when cybercriminals send malicious emails designed to trick people into falling for a scam. The intent is often to gets users to reveal financial information, system credentials or other sensitive data.

The term “phishing” came about in the mid 1990s when hackers began using fraudulent emails to “fish for” information from unsuspecting users. Since these early hackers were often referred to as “phreaks” the term became known as “phishing” with a “ph”. Phishing emails try to lure you in and get you to take the bait. And once you are hooked, you are in trouble.

Phishing is an example of social engineering: a collection of techniques scam artists use to manipulate human psychology. Social engineering techniques include forgery, misdirection, and lying , all of which can play a part in phishing attacks. On a basic level, phishing emails use social engineering to encourage you to act without thinking things through.

The types of phishing are:

1)**VISHING:** vishing refers to phishing done over phone calls. Since voice is used for this type of Phishing, it is called Vishing ---voice+ phishing.

Considering the ease and enormity of data available in social networks, it is no surprise that phishers communicate confidently over a call in the name of friends , relatives or any related brand, without raising any suspicion.

2)**SMISHING:**SMS phishing or smishing is one of the easiest type of attacks in phishing.

The user is targeted by using SMS alerts. In SMSishing , users may receive a fake DM or fake order detail with a cancellation link.

The link would actually be a fake webpage designed to gather personal details.

3)**SEARCH ENGINE PHISHING:** Search engine phishing is the type of phishing that refers to the creation of a fake webpage for targeting specific keywords and waiting for the searcher to land on the fake webpage.

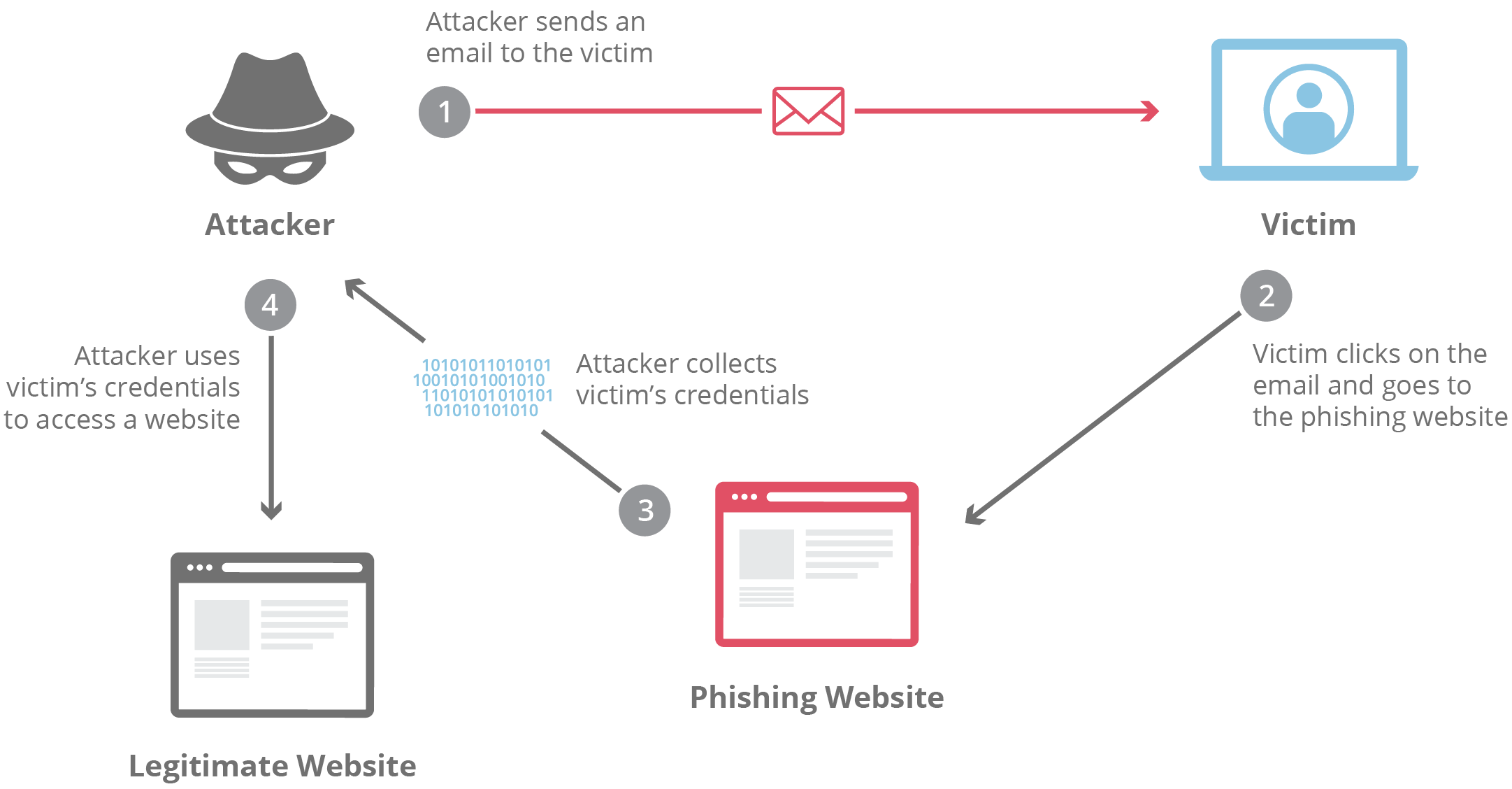
Once a searcher clicks on the page link, he/she will never recognize that he/she is hooked until it is too late.

4)**SPEAR PHISHING:** Unlike traditional phishing—which involves sending emails to millions unknown users –spear phishing is typically targeted in nature, and the emails are carefully designed to target a particular user.

These attacks have a greater risk because phishers do a complete social profile research about the user and their organization-throuugh their social media profile and company website.

5)**WHALING:** whaling is not very different from spear phishing, but the targeted groups becomes more specific and confined in this type of phishing attacks.

This techniques targets C-suite posts like CEO,CFO,COO—or any other senior management positions—who are considered to be the big players in the information change of any organization, commonly known as “whalers” in phishing terms.



**PHARMING**

Pharming is a scamming practice in which malicious code is installed on a personal computer or server , misdirecting users to fraudulent websites without their knowledge or consent. Pharming has been called “phishing with a lure”.

In phishing the perpetrator sends out legitimate looking emails, appearing to come from some of the web’s most popular sites, in an effort to obtain personal and financial information from individual recipients. But in pharming, large numbers of computer users can be victimized because it is not necessary to target individuals pne by one and no conscious action is required on the part of the victim. In one form of pharming attack, code sent in an e-mail modifies local host files on a personal computers. The host file converts URLs into the number strings that the computer uses to access websites. A computer with a compromised host file will go the fake websites even if a user types in the correct internet address or clicks on an affected network entry. Some spyware removal program can correct the corruption , but it frequently resurs unless the user changes browsing habitals.

A particularly ominous pharming tactic is known as domain name system poisoning(DNS poisoning), in which the domain name system table in a server is modified so that someone who thinks they are accessing legitimate websites are actually directed towards fraudulent ones. In this method of pharming individual personal computer host files need not be corrupted . Instead the problem occurs in the DNS server which handles thousands or millions of internet users request for URLs . Victims end up at the bogus site without any visible indicator of discrepancy. Spyware removal programs cannot deal with this type of pharming because nothing needed to be technically wrong with the end users computers.

Once personal information such as a credit card number, bank account number

With the advancements in technology phishing attacks are growing day by day, hackers or phishers are coming up with new techniques and tools to do the phishing attacks, phishing attack is an attack in which a phisher or attacker send an fake email or do social engineering to bring user on a webpage which looks similar to a popular website’s webpage but actually its not , it’s a fake webpage which is being created by the hacker to steal the information of user regarding to that website, whenever user enters their details related to that website , user gets no response but those details are being saved in the database of the hacker or phisher which he will use for his own benefit in future.

**OBJECTIVE OF THE SYSTEM**

Although today everyone is known to the internet and its benefits but most of them do not know about the harms of using it, they get attacked by different Trojans, viruses etc, their personal details got stolen by attacks like phishing, Pharming etc.

So the main objective of making this system is to reduce phishing attacks which is happening on the Internet , in this system we will ask user to input a URL which he thinks might be a part of phishing then the system will recognize and analyse that if that URL is real or fake , and after analysing and identifying the URL we will tell or show the user that if it is a fake URL or a legitimate URL.

1. The main objective is to differentiate between phishing and non phishing website.
2. It will help in reducing data leaks.
3. It will provide security to user.
4. Provide various details for determination of fake and legitimate URLs.

**SCOPE**

This project can be used:

1. By the users who are using internet on a daily basis.
2. By the users who wants security on Internet.
3. By anyone who regularly do money based transactions online.

**TERMS DEFINED**

**IP ADDRESS:**

An internet protocol address (IP Address) is a logical numeric address that is assigned to every single computer, printer , switch, router or any other device that is a part of TCP/IP network.

The IP address is the core component on which the networking architecture is built, no network exists without it. An IP address is a logical address that is used to uniquely identify every node in the network. Because IP addresses are logical , they can change. They are similar to addresses in a town or city because the IP address gives the network node an address so that it can communicate with other nodes or networks, just like mail is sent to friends and relatives.

**DOMAIN**

A domain in the context of networking refers to any groups of users , workstations, devices, printers, computers, and databases servers that share different types of data via network resources. These are also many types of subdomains.

A domain has a domain controller that governs all basic domain functions and manages network security, Thus a domain is used to manage all user functions, including username/password and shared system resource authentication and access . A domain is also used to assign specific resource privilages, such as user accounts.

**SUBDOMAIN:**

A subdomain is a domain that is a part of a larger domain under the domain name system hierarchy. It is used as an easy way to create a more memorable web address for specific or unique content with a website. For example, it could make it easier for users to remember and navigate to the picture gallery of a site by placing it in the address gallery.mysite.com, as opposed to mysite.comedia/gallery.

**DESCRIPTION AND WORKING**

In this system we will ask a user to input a URL which he wants to check, then we will perform some techniques to determine whether that URL is real or fake, the techniques which are going to be used are:

1)**FINDING IP ADDRESS:**  We will find the IP address of the URL inputted by the user, every URL or website has an unique IP address so if the IP is found then we can say that it may not be a phishing website.

2)**DOMAIN AND SUBDOMAIN:**  Every website need to be registered to get a domain , so we will check whether the URL entered by the user contains a domain or subdomain or not.

3)**TYPESQUATTING:** Typesquatting is a technique in which hacker redirect a user to a fake webpage to steal his details if he enters a wrong URL like he types goggle.com instead of typing google.com. In this section we will gather a list of popular domain names and websites names and match them with the URL inputted by the user.

4)**POPULARITY:** In this section we will check the popularity of the website , firstly we will download a dataset of TOP 1MILLION POPULAR WEBSITES BY ALEXA and we will match the URL with this dataset , if the URL is in this dataset the we will display the website is “popular” and if that is not present in this dataset we will show “not popular”.

**Theoretical / Conceptual Framework**

User input

Final Result

IP, Domain, Popular, Typesquatting.

Evaluation

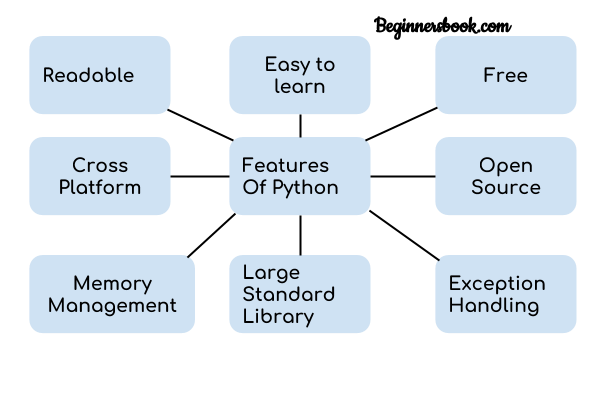
**TECHNOLOGY USED**

**PYTHON:**

Python is a high-level , interpreted scripting language developed in the late 1980s by Guido van Rossum at the National Research Institute for Mathematics and Computer Science in the Netherlands. The initial version was published at alt.sources newsgroup in 1991, and version 1.0 was released in 1994.

Python 2.0 was released in 2000, and the 2.x versions were the prevalent releases until December 2008. At that time, the development team made the decision to release version 3.0, which contain a few relatively small but significant changes that were not backward compatible with 2.x versions. Python 2 and 3 are very similar, and some features of Python 3 have been backported to Python 1. But in general, they remain not quite compatible.

**Features of python:**



1. **Readable:** Python is a very readable language.

2. **Easy to Learn:** Learning python is easy as this is a expressive and high level programming language, which means it is easy to understand the language and thus easy to learn.

3. **Cross platform:** Python is available and can run on various operating systems such as Mac, Windows, Linux, Unix etc. This makes it a cross platform and portable language.

4. **Open Source:** Python is a open source programming language.

5. **Large standard library:** Python comes with a large standard library that has some handy codes and functions which we can use while writing code in Python.

6. **Free:** Python is free to download and use. This means you can download it for free and use it in your application. See: [Open Source Python License](https://docs.python.org/3/license.html). Python is an example of a FLOSS (Free/Libre Open Source Software), which means you can freely distribute copies of this software, read its source code and modify it.

7. **Supports exception handling:** If you are new, you may wonder what is an exception? An exception is an event that can occur during program exception and can disrupt the normal flow of program. Python supports exception handling which means we can write less error prone code and can test various scenarios that can cause an exception later on.

8. **Advanced features:** Supports generators and list comprehensions. We will cover these features later.

9. **Automatic memory management:** Python supports automatic memory management which means the memory is cleared and freed automatically. You do not have to bother clearing the memory.

**What can you do with python:**

## You may be wondering what all are the applications of Python. There are so many applications of Python, here are some of the them.

## 1. Web development – Web framework like Django and Flask are based on Python. They help you write server side code which helps you manage database, write backend programming logic, mapping urls etc.

2. **Machine learning** – There are many machine learning applications written in Python. Machine learning is a way to write a logic so that a machine can learn and solve a particular problem on its own. For example, products recommendation in websites like Amazon, Flipkart, eBay etc. is a machine learning algorithm that recognises user’s interest. Face recognition and Voice recognition in your phone is another example of machine learning.

3**. Data Analysis** – Data analysis and data visualisation in form of charts can also be developed using Python.

4. **Scripting** – Scripting is writing small programs to automate simple tasks such as sending automated response emails etc. Such type of applications can also be written in Python programming language.

5. **Game development** – You can develop games using Python.

6. You can develop Embedded applications in Python.

7. **Desktop applications** – You can develop desktop application in Python using library like TKinter or QT.

**FLASK:**

Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries except for some basic standards libraries. Flask is a web framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some web page, a blog, a wiki or go as big as a web based application or a commercial website.

You can start small with app contained entirely in one file, then slowly scale up to multiple files and folders in a well-structured manner as your site becomes more and more complex.

**HOW FLASK HELP PYTHON INTERACT WITH HTML:**

1. Templating
2. User systems
3. Password hashing
4. Databases
5. Content management

…… and a whole lot more.

Jinja2 is a web template engine which combines a template with a certain data source to render the dynamic webpages.

To work with flask we need to have python 2.7 or above versions to be installed on our system.

**BOOTSTRAP:**

Bootstrap is a free and open source front-end web framework for faster and easier web development . It contains HTML and CSS-based design template for typography, forms, buttons, navigation and other interface components, as well as optional Javascript extensions. It concerns itself with front-end development only. Bootstrap gives you the ability to easily create responsive designs. Bootstrap’s responsive CSS adjusts to phones, tablets, and desktops. Bootstrap is compatible with all modern browsers(Chrome, Firefox, Internet Explorer, Edge, Safari and Opera).

Bootstrap includes components such as buttons, navbars, dropdown menus, alert boxes and more. In most cases, you can make use of a component simply by using the appropriate class name.

**ANACONDA:**

Anaconda is a free and open-source distribution of Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc), that aims to simplify package management and deployment . Package various are managed by the package management system conda. The Anaconda distribution is used by over 12 million users and includes more than 1400 popular data-science packages suitable for Windows, linux and MacOS.

Anaconda Navigator is a desktop graphical user interface(GUI) includes in anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a Local Anaconda Repository, install them in an environment, run the packages and update the. It is available for Windows, MacOS and Linux.

**HTML:**

HTML stands for Hyper Text Markup Language. It is used to design web pages using markup languages. HTML is the combination of Hypertext and Markup Language. Hypertext defines the link between the web pages. Markup language is used to define the text document within tag which defines the structure of web pages. This language is used to annotate(make notes for the computer) text so that a machine can understand it and manipulate text accordingly. Most of the Markup (e.g. HTML) languages are human readable. Language uses tags to define what manipulation has to be done on the text.

HTML is a markup language which is used by the browser to manipulate text, images and other content to display it in required format. HTML was created by Tim Berners-Lee in 1991. The first ever version of HTML was HTML 1.0 but the first standard version was HTML 2.0 which was published in 1999.

**Advantages:**

* HTML is used to build a websites.
* It is supported by all browsers.
* It can be integrated with other languages like CSS, JavaScript etc.

**Disadvantages:**

* HTML can create only static webpages so for dynamic web page other languages have to be used.
* Large amount of code has to be written to create a simple web page.
* Security feature is not good

**CSS:**

Cascading Style Sheet, fondly referred to as CSS, is a simple design language intended to simplify the process of making web pages presentable.

CSS handles the look and feel part of a web page. Using CSS , you can control the color of text , the styles of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images and colors are used, layout design, variations in display for different device and screen sizes as well as a variety of other effects.

CSS is easy to learn and understand but it provides powerful control over the presentation of an HTML document. Most commonly, CSS is combined with the markup languages HTML or XHTML.

## Advantages of CSS

* **CSS saves time** − You can write CSS once and then reuse same sheet in multiple HTML pages. You can define a style for each HTML element and apply it to as many Web pages as you want.
* **Pages load faster** − If you are using CSS, you do not need to write HTML tag attributes every time. Just write one CSS rule of a tag and apply it to all the occurrences of that tag. So less code means faster download times.
* **Easy maintenance** − To make a global change, simply change the style, and all elements in all the web pages will be updated automatically.
* **Superior styles to HTML** − CSS has a much wider array of attributes than HTML, so you can give a far better look to your HTML page in comparison to HTML attributes.
* **Multiple Device Compatibility** − Style sheets allow content to be optimized for more than one type of device. By using the same HTML document, different versions of a website can be presented for handheld devices such as PDAs and cell phones or for printing.
* **Global web standards** − Now HTML attributes are being deprecated and it is being recommended to use CSS. So its a good idea to start using CSS in all the HTML pages to make them compatible to future browsers.

**WHY CSS?**

* **CSS saves time :**You can write CSS once and reuse same sheet in multiple HTML pages.
* **Easy Maintainence :**To make a global change simply change the style, and all elements in all the webpages will be updated automatically.
* **Search Engines :**CSS is considered as clean coding technique, which means search engines won’t have to struggle to “read” its content.
* **Superior styles to HTML :**CSS has a much wider array of attributes than HTML, so you can give a far better look to your HTML page in comparison to HTML attributes.
* **Offline Browsing :**CSS can store web applications locally with the help of offline cache. Using of this we can view offline websites.

**PYTHON LIBRARIES USED**

**PANDAS:**

Pandas is an open source python library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data- an Econometrics from Multidimensional data.

Prior to Pandas, python was majorly used for data mugging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data- load, prepare, manipulate, economics, Statistics, analytics, etc.

Key Features of Pandas:

* Fast and efficient dataframe object with default and customized indexing.
* Tools for loading data into in-memory data objects from different file formats
* Data alignment and integrated handling of missing data.
* Reshaping and pivoting of datasets.
* Label based slicing, indexing and subsetting of large data sets.
* Columns from a data structure can be deleted or inserted.
* Group by data for aggregation and transformations.
* High performance merging and joining of data.
* Time series functionality

**FLASK:**

Flask is a web-framework. This means flask provides you with tools, libraries and technologies that allow you to build a web application. The web application can be some web pages, a blog, a wiki or go as big as a web based calendar application or a commercial website.

Flask is apart of the categories of the micro framework. Micro framework are normally framework with little to no dependencies to external libraries. This has pros and cons. Pros would be Framework is light and Con is that sometime you have to do more work by yourself.

How flask helps python interact with HTML:

* 1. Templating
  2. User systems
  3. Password hashing
  4. Databases
  5. Content management

….and a whole lot more.

**General terms :**

1)**WSGI**: web server gateway interface has been adopted as a standard for python web application development. WSGI is a specification for a universal interface.

2) **JINJA2**: Jinja2 is a popular templating engine for Python. A web templating system combines a template with a certain data source to render dynamic web pages.

**URL\_FOR:**

The url\_for function is very useful for dynamically building a URL for a specific function. The function accepts the name of a function as first argument, and one or more keyword arguments, each corresponding to the variable part of URL.

**REQUESTS:**

Requests is an Apache2 Licensed HTTP library, written in Python. It is designed to be used by humans to interact with the language. This means you don’t have to manually add query strings to URLs or form-encode your POST data. Don’t worry if that made no sense to you. It will in due time.

Requests will allow you to send HTTP/1.1 requests using python. With it , you can add content like headers , form data, multipart files, and parameters via simple python libraries. It also allows you to access the response data of Python in the same way.

**SOCKET:**

Socket library is used to find the ip address of the url which is being passed in the function.

**MODULES AND FILES**

**PYTHON FILES**

1. app.py

import pandas, flask, render\_template, url\_for, request, session

def home():

def analyse():

def response():

def about():

1. result.py

def calculate\_result():

1. predict.py

import socket, pandas

def find\_ip():

def popular():

def domain():

def typo():

def check():

**HTML FILES:**

**1)** About.html

2) Home.html

3) Analyse.html

4) Home.html

5) Response.html

**GANTT CHART**

A Gantt chart is a graphical depiction of project schedule. A Gantt chart is type of bar chart that shows the start and finish dates of several elements of a project that include resources, milestones, tasks and dependencies. Henry Gantt, an American mechanical engineer, designer the Gantt chart in 1917.

The Gantt chart is the most widely used chart in project management, these charts are useful in planning a project and defining the sequence of tasks that require completion. In most instances , the chart displays as a horizontal bar chart. Horizontal bars of different lengths represent the project timeline that includes task sequences, duration , and start and end dates for each task. The horizontal bar also shows how much of a task require completion. The horizontal bar length is proportional to the time necessary for a tasks completion. In addition, the project tasks are visible on the vertical bar.

A Gantt chart aids in scheduling, managing, and monitoring specific tasks and resources in a project . The chart shows the project timeline that includes scheduled and completed work over a period. The Gantt chart aids project managers in communicating project status or plans helps ensure the project remains on track.

Having a perfect chart does not guarantee a successful project execution on deadline, it is also crucial to compare the projected progress to the real progress and then update the chart accordingly.



**PROBLEM FORMULATION**

**INTRODUCTION**

Problem formulation or problem starting is the starting point of the software development activity. The objective of this statement is to answer. Exactly what must the system do ? The software project is initiated by the client’s need. In the beginning, these needs are on the minds of various people in the client’s organisation . The analyst has to identify the requirements by talking to the people and understanding to their needs. It goes without saying that an accurate and through understanding of software requirement are essentials to the success of software development effort. All further development like system analysis. System design and coding will depends on how accurate and well understood the requirements are poorly analyzed and specified software will disappoint the user and will bring brief to the developer. No matter how well designed and well appearances are often deceiving. Chances of misinterpretation are very high , ambiguity is probable and communication gap between customer and developer is bound to bring confusions. Requirements understanding begin with a clear and concise heading stating in sentence the task to be performed. Then requirements are describe in a technical manner precise statement.

**FEASIBILITY STUDY:**

All projects are feasible given unlimited and infinite time! Unfortunately , the development of computer based system is more likely to be plagued by a scarcity of resources and difficult delivery dates. It is both necessary and prudent to evaluate the feasibility of the project at the earliest possible time. Months or years of effort, Money loss and unfold Professional embarrassment can be averted l few better understand the project at its study time.

**TECHNICAL FEASIBILITY:**

Internet is available all through the world these days. The hardware required is not a costly affair. Every student who wants to study on the website may use his/her personal computers or tablets etc. to access the website.

**ECONOMICAL FEASIBILITY:**

**The project is not a costly product for the customer, as the database used in the product is MySQL which is open source. only one or two technical are required to host the website. The administrator himself can upload the files (study material) which does not requires extra training.**

**The JSP is used to design the front end of the product which does not require any license.**

**OPERATIONAL FEASIBILITY:**

The website would have a good GUI so no training would be required to access the product. The student registering for study and certification should only know English language, basic knowledge of computer and knowledge of internet browsing and surfing.

**SYSTEM ANALYSIS AND DESIGN**

**REQUIREMENT ANALYSIS:**

Requirement Analysis is the process of defining the expectations of the users for an application that is to be built or modified. Requirements analysis involves all the tasks that are conducted to identify the needs of different stakeholders . Therefore requirements analysis means to analyse, document, validate and manage software or system requirements. High-quality requirements are documented, actionable, measurable, testable, traceable. Helps to identify business opportunities, and are defined to a facilitate system design.

**PROCESS**

1) **Eliciting requirements**: The process of gathering requirements by communicating with the customers is known as eliciting requirements.

2) **Analysing requirements:**This step helps to determine the quality of the requirements . It involves identifying whether the requirements are unclear, incomplete, ambiguous, and contradictory. These issues resolved before moving to the next step.

3) **Requirement modeling**: In requirements modeling, the requirements are usually documented in different formats such as use cases, user stories, natural language documents, or process specification.

4) **Review and retrospective**: This step is conducted to reflect on the previous iterations of requirements gathering in a bid to make improvements in the process going forward.

**PLANNING**

### Planning the system *requires the user to define what the problem is*. The planning may also include how the user would like to solve the problem. Defining the scope of the problem is also important in this stage as well. Defining the scope helps to prevent the project from [scope creep](https://en.wikipedia.org/wiki/Scope_creep). Once the problem is determined, and one or more solutions have been selected, planning to implement the solution begins. Multiple scenarios may be enacted to determine the best course of action for implementing the system.

Course of action should be well documented and take into consideration a schedule showing anticipated start and completion times of activities (milestones) leading to the objectives, knowing expenditures required to achieve objectives, scheduling regular status reviews (are we on course?), anticipating any organizational restructuring to accommodate the objectives, anticipating and planning for mitigation of risks that may hinder achievements, implementing policies and procedures for decision making, and defining a standard level of performance.

**CHARACTERISTICS OF PLANNING:**

1) **First and foremost managerial function**: Planning provides the base for other functions of the management, i.e, organizing, staffing, directing, and controlling, as they are performed within the periphery of the plans mode.

2) **Goal oriented**: It focuses on defining the goals of the organization, identifying alternative courses of action and deciding the appropriate action plan, which is to be undertaken for reaching the goals.

3) **Pervasive**: It is pervasive in the sense that it is present in all the segments and is requires at all the levels of the organization. Although the scope of planning varies at different levels and departments.

4) **Continuous process**: Plans are made for a specific term, say for a month , quarter, year and so on. Once that period is over, new plans are drawn, considering organization’s present and framed, executed and followed by another plan.

5) **Intellectual** **Process**:  It is a mental exercise at it involves the application of mind, to think, forecast, imagine intelligently and innovate etc.

6) **Futuristic**: In the process of planning we take a sneak peek of future. It encompasses looking into future, to analyse and predict it, so that the organisation can face the future challenges effectively.

7) **Decision** **making**:  Decisions are made regarding the choice of alternative courses of action that can be undertaken to reach the goal. The alternative chosen should be best among all, with least number of negative and highest number of positive outcomes.

**IMPORTANCE OF PLANNING:**

* It helps managers to **improve future performance**, by establishing objectives and selecting a course of action, for the benefit of the organisation.
* It **minimises risk and uncertainty**, by looking ahead into future.
* It **facilitates coordination of activities**. Thus, reduces overlapping among activities and eliminates unproductive work.
* It states in advance, what should be done in future, so it provides **direction for action.**
* It uncovers and identifies **future opportunities and threats**.
* It **sets out standards for controlling**. It compares actual performance with the standard performance and efforts are made to correct the same.

. By planning process, an organisation not only gets the insights of future, but it also helps the organisation to shape its future. Effective planning involves simplicity of the plan, i.e. the **plan should be clearly stated and easy to understand**, because if the plan is too much complicated it will create chaos among the members of the organisation. Further, the **plan should fulfil all the requirements of the organisation**.

**ANALYSIS**

The analysis phase involves gathering requirements for the system. At this stage, business needs are studied with the intention of making business processes more efficient. The system analysis phase focuses on what the system will do in an effort that views all stakeholders, as viable sources of information. In the analysis phase, a significant amount of time is spent talking with stakeholders and reviewing the stakeholder’s input. Common stakeholders for IT projects are:

* Architecture office
* Testing & certification office
* Records management team
* Application support group

Once stakeholders have been recognized, the gathering and analysis of the requirements can begin.Requirement gathering must be related to business needs or opportunities. Requirement analysis involves capturing requirements and analyzing requirements. Capturing requirements is communicating with stakeholders to agree on what the requirements are. Analyzing requirements is using standard tools to produce a baseline of the requirements. Once the stakeholders concur on the requirements, the baseline is created and becomes the formal requirement source.

***Analysis Stages***

1. **Study and analyse the current system:**  
   System analysts collect facts from existing users.  
   Existing systems limitations and details.
2. **Define new system objectives:**Define and Prioritise Users Requirements.  
   The analyst approaches the users to find out what they want from the system.  
   They use different data gathering techniques such as interviews, observations, and surveys.  
   This is an attempt to understand all aspects of the current system and eventually indicate how things may be improved by a new system.

**In short…**

The analyst understands the nature of the information and the functions of the software which is required for the system. The analyst makes a brief survey of the requirements and tries to analyze the performance of the system which is to be developed. He also makes sure that he gets enough information and resources for building the appropriate system.

**DESIGN**

The design phase is concerned with the physical construction of the system. Included are the design or configuration of the network (hardware, operating system, programming, etc.), design of user interfaces (forms, reports, etc.), design of system interfaces (for communication with other systems), and security issues. It is important that the proposed design be tested for performance, and to ensure that it meets the requirements outlined during the analysis phase. In other words, the main objective of this phase is to transform the previously defined requirements into a complete and detailed set of specifications which will be used during the next phase. Some of the activities that need to take place during the design phase are:

* Design the application
* Design and integrate the network
* Design and integrate the database
* Create a contingency plan
* Start a Maintenance, Training and Operations plan
* Review the design
* Articulate the business processes and procedures
* Establish a transition strategy
* Deliver the System Design Document
* Review final design

Design is being done in 2 phases:

1. Preliminary design phase
2. Structured or detailed design phase

**PRELIMINARY DESIGN PHASE:**

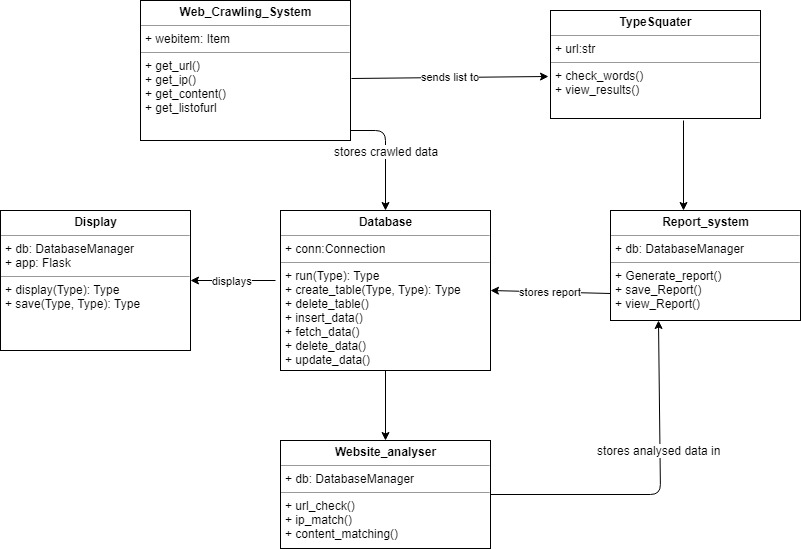
In this phase, the features of the new system to be designed are specified. The cost of implementing these features and the benefits to be derived from the system so developed are estimated. If the project is still considered to be feasible, then we move to the detailed design phase.

In the project “ Phishing website checker system “ initially the design were made to develop an interactive GUI for the end user to work on it easily. And making the interface simple enough so that the user can easily understand its functioning.

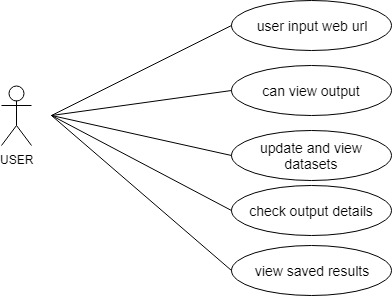
**STRUCTURED OR DETAILED DESIGN:**

In the detailed design phase, the in depth designing of the system takes place and the serious development of the system is done covering the user requirements. At this stage, the design of the system becomes more structured . Structured design is a blueprint of a computer system solution to a given problem having the same components and inter-relationships among the same components as the original problem. Input, output, databases, forms, coding and processing specifications are drawn up in details.

**UML DIAGRAM**

****

**USE CASE DIAGRAM**

****

**LIFECYCLE MODEL: SPIRAL**

#### 

#### Spiral Model

[The Spiral Model](https://www.softwaretestinghelp.com/spiral-model-what-is-sdlc-spiral-model/) includes iterative and prototype approach.

Spiral model phases are followed in the iterations. The loops in the model represent the phase of the SDLC process i.e. the innermost loop is of requirement gathering & analysis which follows the Planning, Risk analysis, development, and evaluation. Next loop is Designing followed by Implementation & then testing.

**Spiral Model has four phases:**

* Planning
* Risk Analysis
* Engineering
* Evaluation

**(i) Planning:**

Planning phase includes requirement gathering wherein all the required information is gathered from the customer and is documented. Software requirement specification document is created for the next phase.

**(ii) Risk Analysis:**

In this phase, the best solution is selected for the risks involved and analysis is done by building the prototype.

**For Example**, the risk involved in accessing the data from a remote database can be that the data access rate might be too slow. The risk can be resolved by building a prototype of the data access subsystem.

**(iii) Engineering:**

Once the risk analysis is done, coding and testing are done.

**(iv) Evaluation:**

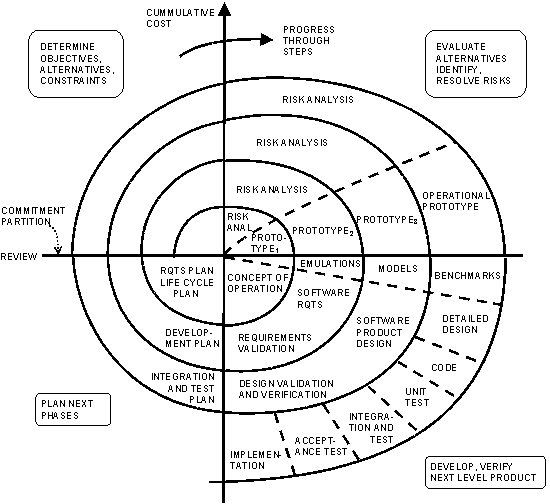
Customer evaluates the developed system and plans for the next iteration.

**Advantages of Spiral Model:**

* Risk Analysis is done extensively using the prototype models.
* Any enhancement or change in the functionality can be done in the next iteration.

**Disadvantages of Spiral Model:**

* The spiral model is best suited for large projects only.
  + Cost can be high as it might take large no  of iterations which can lead to high time to reach the final product.

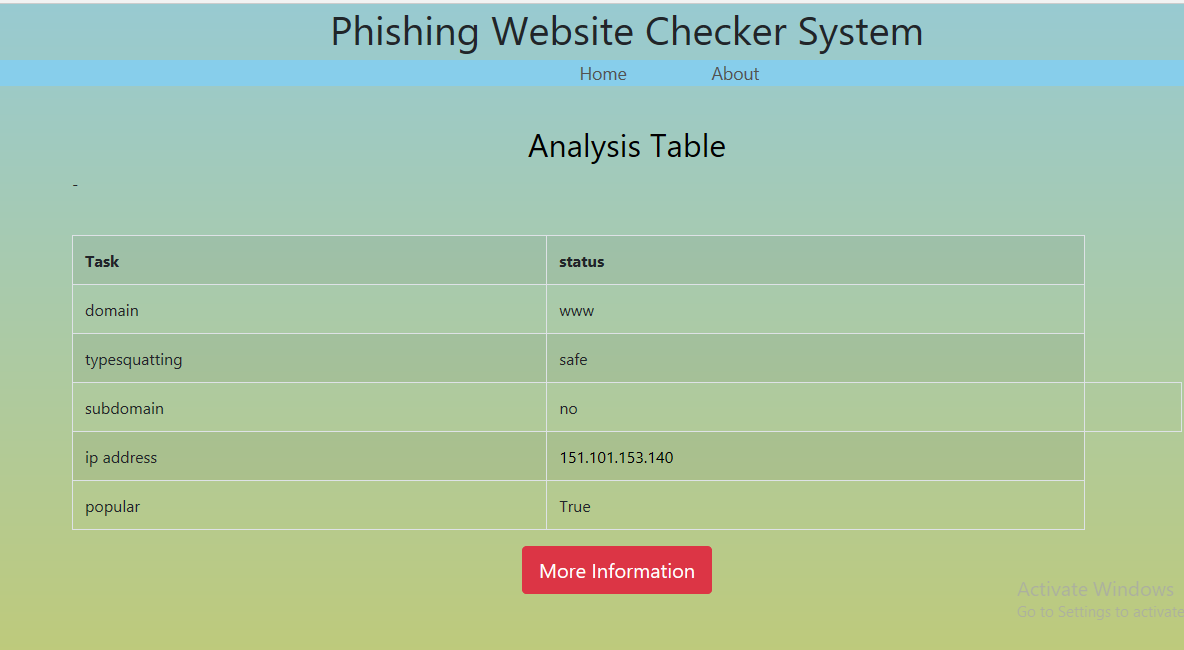


**SNAPSHOTS**

HOMEPAGE

****

ANALYSIS PAGE

****

RESULT PAGE

****

ABOUT PAGE

****

**CODING**

**app.py**

from flask import Flask, render\_template,url\_for, request, session

import predict as p

import pandas as pd

from result import calculate\_result

app=Flask(\_\_name\_\_)

app.secret\_key='qwesag21456de'

@app.route('/')

def home():

return render\_template('home.html')

@app.route('/about')

def about():

return render\_template('about.html')

@app.route('/response')

def response():

url=request.args.get('url')

if url:

prediction = calculate\_result(session)

return render\_template('response.html',url=url, pre=prediction)

return render\_template('response.html')

@app.route('/analyse', methods=['GET','POST'])

def analyse():

if request.method=='POST':

url=request.form.get('url')

if url:

ip=p.find\_ip(url)

session['ip']=str(ip)

pop=p.popular(url)

session['pop']=pop

sd=p.domain(url)

session['sd']=sd

squat=p.typo(url)

session['squat']=squat

sdom=p.check(url)

session['sdom']=sdom

if squat==1.0:

ts='safe'

else:

ts='unsafe'

return render\_template('analyse.html',ip=ip,pop=pop,sd=sd,url=url,sdom=sdom, ts=ts)

return render\_template('analyse.html')

if \_\_name\_\_ == "\_\_main\_\_":

app.run(debug=True)

**Predict.py**

import socket

import pandas as pd

df=pd.read\_csv("top-1m.csv",names=['sno','websites'])

df = pd.DataFrame(df.websites.str.split('.',1).tolist(),columns = ['name','domain'])

ls=df.name.values.tolist()

pop=False

def find\_ip(url):

try:

res=socket.gethostbyname(url)

return res

except Exception as e:

print(e)

return 'IP not Found'

def popular(url):

#df.drop\_duplicates(subset ="name",keep = False, inplace = True)

s=url.strip('www., .com,.co,.in, en.,.en')

if s in ls:

pop=True

return pop

else:

pop=False

return pop

def domain(url):

sub=url.split('.')

if 'www' in sub:

return 'www'

else :

return "no domain"

#def typo(url):

#a = set(map(str,url.split[1]))

#for url in ls:

#b=set(map(str,url.split('.')[1]))

#print(b)

# b = set(map(str,word2))

# c = a.intersection(b)

# return float(len(c)) / (len(a) + len(b) - len(c))

def typo(url):

lk=url.split('.')[1]

l=[i for i in ls if len(lk)==len(i)]

l=list(set(l))

lst=[]

a = set(map(str,url.split('.')[1]))

for k in l:

b=set(map(str,k))

c = a.intersection(b)

lst.append(float(len(c)) / (len(a) + len(b) - len(c)))

print(max(lst))

return max(lst)

def check(url):

lst=url.split('.')

if len(lst)<=3:

return "no"

else:

return "yes"

**Result.py**

def calculate\_result(result):

if result['squat']>=0.6 and result['ip'] and result['pop']:

return 'Not A Phishing Website'

else:

return 'Likely To Be A Phishing Website'

**Home.html**

{% extends "layout.html" %}

{% block css %}

<style>

\* {

margin: 0;

padding: 0;

}

p {

position: absolute;

left: 50%;

top: 65%;

transform: translateX(-50%) translateY(-50%);

text-align: center;

font-family: Pacifico, Calibri;

font-size: 1em;

}</style>

{% endblock %}

{% block content %}

<div class="jumbotron mt-5" style="background-color: rgba(245, 245, 245, 0.274)">

<h1 class="display-3 text-center">URL</h1>

<div class="col-md-8 mx-auto mt-2 text-center">

<form action='/analyse' method="POST">

<input type="text" class="form-control form-control-lg" placeholder="Enter a qualified Web URL" name="url">

<br>

<button class="btn btn-outline-danger btn-lg" type="submit">

Start Analysis

</button>

</form>

</div>

</div>

<div class="row justify-content-center">

<div class="col-md-8">

<p>

<b>This system is being made for detecting phishing</b>

<br>

<b>websites , so enter a url which you think is malicious.</b>

</p>

</div>

</div>

{% endblock %}

**Analyse.html**

{% extends "layout.html" %}

{% block content %}

<h2 align=center style="color:black;"><br>Analysis Table</h2>

<div class="row">

<div class="col-md-12">

<table class="table table-bordered table-striped" style="margin-top: 40px">

<tr>

<th>Task</th>

<th>status</th>-

</tr>

<tr>

<td>domain</td>

<td>{{sd}}</td>

</tr>

<tr>

<td>typesquatting</td>

<td>{{ts}}</td>

</tr>

<tr>

<td>subdomain</td>

<td>{{sdom}}<td>

</tr>

<tr>

<td>ip address</td>

<td style="color:black">{{ip}}</td>

</tr>

<tr>

<td>popular

</td>

<td>{{pop}}</td>

</tr>

</table>

<form action='/response' method="POST">

<!-- <button class="btn btn-danger btn-lg" type="submit" style="margin-left:450px;" >More Information</button> -->

<a class="btn btn-danger btn-lg" href="/response?url={{url}}" style="margin-left:450px;">More Information</a>

</form>

</div>

</div>

{% endblock %}

**Response.html**

{% extends "layout.html" %}

{% block content %}

{% block css %}

<style>

\* {

margin: 0;

padding: 0;

}

p{

position: absolute;

left: 50%;

top: 45%;

transform: translateX(-50%) translateY(-50%);

text-align: center;

font-family: Pacifico, Calibri;

font-size: 2em;

}

h2{

position: absolute;

left: 50%;

top: 25%;

transform: translateX(-50%) translateY(-50%);

text-align: center;

font-family: Pacifico, Calibri;

font-size:3em;

}

</style>

{% endblock %}

<h2 align=center style="color:blue;"><br>RESULT</h2>

<p align=center >

<br>We Just Analysed The Url Entered By You And After The Analysis Here Is The Result. <br>

<br>{{url}}<br>{{pre}}</br></p>

{% endblock %}

**Response.html**

{% extends "layout.html" %}

{% block content %}

{% block css %}

<style>

\* {

margin: 0;

padding: 0;

}

p{

position: absolute;

left: 50%;

top: 45%;

transform: translateX(-50%) translateY(-50%);

text-align: center;

font-family: Pacifico, Calibri;

font-size: 2em;

}

h2{

position: absolute;

left: 50%;

top: 25%;

transform: translateX(-50%) translateY(-50%);

text-align: center;

font-family: Pacifico, Calibri;

font-size:3em;

}

</style>

{% endblock %}

<h2 align=center>About</h2>

<p align=center> This Project Is to Detect Phishing Websites Using Machine Learning Techniques.</p>

{% endblock %}

**Layout.html**

<!doctype html>

<html>

<head>

<title>Phishing Website Checker System</title>

<!-- Required meta tags -->

<meta charset="utf-8">

<meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

<!-- Bootstrap CSS -->

<link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css" integrity="sha384-ggOyR0iXCbMQv3Xipma34MD+dH/1fQ784/j6cY/iJTQUOhcWr7x9JvoRxT2MZw1T" crossorigin="anonymous">

<link rel="stylesheet" type="text/css" href="{{url\_for('static',filename='main.css')}}">

<style type="text/css">

body {

height: 100vh;

background-image:linear-gradient(rgb(152, 202, 209), rgb(190, 202, 123));

background-size: cover;

background-repeat: no-repeat;

}

</style>

{% block css %}

{% endblock %}

</head>

<body>

<h1 style="color:black(213, 241, 212);" align=center> Phishing Website Checker System </h1>

<div class="nav" style="background-color:skyblue">

<ul class="nav-list">

<li class="nav-list\_\_item"><a class="nav-list\_\_link" href="/">Home</a></li>

<li class="nav-list\_\_item"><a class="nav-list\_\_link" href="/about">About</a>

</li>

</ul>

</div>

<div class='container'>

{% block content %}

{% endblock %}

</div>

<script src="https://code.jquery.com/jquery-3.3.1.slim.min.js" integrity="sha384-q8i/X+965DzO0rT7abK41JStQIAqVgRVzpbzo5smXKp4YfRvH+8abtTE1Pi6jizo" crossorigin="anonymous"></script>

<script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.14.7/umd/popper.min.js" integrity="sha384-UO2eT0CpHqdSJQ6hJty5KVphtPhzWj9WO1clHTMGa3JDZwrnQq4sF86dIHNDz0W1" crossorigin="anonymous">

</script>

<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/js/bootstrap.min.js" integrity="sha384-JjSmVgyd0p3pXB1rRibZUAYoIIy6OrQ6VrjIEaFf/nJGzIxFDsf4x0xIM+B07jRM" crossorigin="anonymous">

</script>

</body>

</html>

**TESTING:**

Software testing is a process of verifying and validation a software application program. The main aim of the system is to find out the bugs in the developed system before implementing it. It is an important phase of a successful system. After codifying the whole program of the system, a test is being performed on the system so developed. The output of the testing phase should match the expected results.

**Implementation** is the stage of a project during which theory is turned into practice. Themajor steps involved in this phase are:

**Acquisition** **and** **Installation** **of** **Software** **and** **Hardware**:

The hardware and the relevant software required for running the system must be made fully operational before implementation.

**Conversion:**

The conversion is also one of the most critical and expensive activities in the system development life cycle. The data from the old system needs to be converted to operate in the new format of the system. The database needs to be setup with security and recovery procedures fully define.

**User** **Training**: During this phase, all the programs of the system are loaded onto the user‘s computer. After loading the system, training of the user starts

**Main topics of such type of training are:**

How to execute the package

How to enter the data

How to process the data(processing details)

How to take out the reports

After the users are trained about the computerized system, working has to shift from manual to computerized working. The process is called ‗Changeover‘. Maintenance is necessary to eliminate errors in the system during its working life and to tune the system to any variation in its working environments. It has been seen that there are always some errors found in the systems that must be noted and corrected. It also means that review of the system must be taken from time to time.

The review of the system must be done for:

Knowing the full capabilities of the system.

Knowing the required changes or the additional requirements.

Studying the performance.

**Testing techniques and Strategies**

Software testing is a process of executing a program or application with the intent of finding the software bugs. It can also be stated as the process of validating and verifying that a software program or application or product:

Meets the business and technical requirements that guided its design and development.

Works as expected.

**Importance of Testing**

Software Testing is necessary because we all make mistakes. We need to check everything and anything we produce because things can always go wrong Since we assume that our work may have mistakes, hence we all need to check our own work. However some mistakes come from bad assumptions and blind spots, so we might make the same mistakes when we check our own work as we made when we did it. So we may not notice the flaws in what we have done.

Ideally, we should get someone else to check our work because another person is more likely to spot the flaws.

There are several reasons which clearly tells us as why Software Testing is important and what are the major things that we should consider while testing of any product or application.

Software testing is very important because of the following reasons:

1. Software testing is really required to point out the defects and errors that were made during the development phases.
2. It is essential since it makes sure of the Customer‘s reliability and their satisfaction in the application.
3. It is very important to ensure the Quality of the product. Quality product delivered to the customers helps in gaining their confidence.
4. Testing is necessary in order to provide the facilities to the customers like the delivery of high quality product or software application which requires lower maintenance cost and hence results into more accurate, consistent and reliable results.
5. Testing is required for an effective performance of software application.

**Purpose and Objectives**

Software Testing has different goals and objectives. The major objectives of Software testing are as follows:

* Finding defects which may get created by the programmer while developing the software.
* To gain confidence in and providing information about the level of quality.
* To make sure that the end result meets the business and user requirements.
* To ensure that it satisfies the BRS that is Business Requirement Specification and SRS that is System Requirement Specifications.
* To gain the confidence of the customers by providing them a quality product.

Software testing helps in finalizing the software application or product against business and user requirements. It is very important to have good test coverage in order to test the software application completely and make it sure that it‘s performing well and as per the specifications.

While determining the test coverage the test cases should be designed well with maximum possibilities of finding the errors or bugs. The test cases should be very effective. Higher the number of the defects reported the more effective are the test cases.

Once the delivery is made to the end users or the customers they should be able to operate it without any complaints. In order to make this happen the tester should know as how the customers are going to use this product and accordingly they should write down the test scenarios and design the test cases.

Software testing makes sure that the testing is being done properly and hence the system is ready for use. Good coverage means that the testing has been done to cover the various areas

like functionality of the application, compatibility of the application with the OS, hardware and different types of browsers, performance testing to test the performance of the application and load testing to make sure that the system is reliable and should not crash or there should not be any blocking issues. It also determines that the application can be deployed easily to the machine and without any resistance. Hence the application is easy to install, learn and use.

**Unit Testing and System Testing**

* Unit tests are basically written and executed by software developers to make sure that code meets its design and requirements and behaves as expected.
* The goal of unit testing is to segregate each part of the program and test that the individual parts are working correctly.
* It should handle the failures gracefully during the course of execution when any invalid input is given.
* A unit test provides a written contract that the piece of code must assure. Hence it has several benefits.

●Unit testing is basically done before integration as shown in the image below.

Method Used for unit testing:

White Box Testing method is used for executing the unit test.

* When Unit testing should be done?

Unit testing should be done before Integration testing.

* By whom unit testing should be done?

Unit testing should be done by the developers.

**Advantages of Unit testing:**

1. Issues are found at early stage. Since unit testing are carried out by developers where they test their individual code before the integration. Hence the issues can be found very early and can be resolved then and there without impacting the other piece of codes.
2. Unit testing helps in maintaining and changing the code. This is possible by making the codes less interdependent so that unit testing can be executed. Hence chances of impact of changes to any other code gets reduced.
3. Since the bugs are found early in unit testing hence it also helps in reducing the cost of bug fixes. Just imagine the cost of bug found during the later stages of development like during system testing or during acceptance testing.
4. Unit testing helps in simplifying the debugging process. If suppose a test fails then only latest changes made in code needs to be debugged.

**System testing**

●In system testing the behavior of whole system/product is tested as defined by the scope of the development project or product.

●It may include tests based on risks and/or requirement specifications, business process, use cases, or other high level descriptions of system

behavior, interactions with the operating systems, and system resources

●System testing is most often the final test to verify that the system to be delivered meets the specification and its purpose.

System testing is carried out by specialist‘s testers or independent testers.

●System testing should investigate both functional and non-functional requirements of the testing

TYPES OF TESTING

#### 1) Alpha Testing

It is the most common type of testing used in the Software industry. The objective of this testing is to identify all possible issues or defects before releasing it into the market or to the user.

Alpha testing is carried out at the end of the software development phase but before the Beta Testing. Still, minor design changes may be made as a result of such testing. [Alpha testing](https://www.softwaretestinghelp.com/what-is-alpha-testing-beta-testing/) is conducted at the developer’s site. In-house virtual user environment can be created for this type of testing.

#### 2) Acceptance Testing

An [acceptance test](https://www.softwaretestinghelp.com/what-is-acceptance-testing/) is performed by the client and verifies whether the end to end the flow of the system is as per the business requirements or not and if it is as per the needs of the end user. Client accepts the software only when all the features and functionalities work as expected.

It is the last phase of the testing, after which the software goes into production. This is also called User Acceptance Testing (UAT).

#### 3) Ad-hoc Testing

The name itself suggests that this testing is performed on [an ad-hoc](https://www.softwaretestinghelp.com/ad-hoc-testing/) basis i.e. with no reference to the test case and also without any plan or documentation in place for such type of testing. The objective of this testing is to find the defects and break the application by executing any flow of the application or any random functionality.

Ad-hoc testing is an informal way of finding defects and can be performed by anyone in the project. It is difficult to identify defects without a test case but sometimes it is possible that defects found during ad-hoc testing might not have been identified using existing test cases.

#### 4) Accessibility Testing

The aim of [accessibility testing](https://www.softwaretestinghelp.com/what-is-web-accessibility-testing/) is to determine whether the software or application is accessible for disabled people or not. Here disability means deaf, color blind, mentally disabled, blind, old age and other disabled groups. Various checks are performed such as font size for visually disabled, color and contrast for color blindness etc.

#### 5) Beta Testing

[Beta Testing](https://www.softwaretestinghelp.com/beta-testing/) is a formal type of software testing which is carried out by the customer. It is performed in **the Real Environment**before releasing the product to the market for the actual end users.

Beta testing is carried out to ensure that there are no major failures in the software or product and it satisfies the business requirements from an end-user perspective. Beta testing is successful when the customer accepts the software.

Usually, this testing is typically done by end-users or others. It is the final testing done before releasing an application for commercial purpose. Usually, the Beta version of the software or product released is limited to a certain number of users in a specific area.

So end user actually uses the software and shares the feedback to the company. Company then takes necessary action before releasing the software to the worldwide.

#### 6) Back-end Testing

Whenever an input or data is entered on front-end application, it stores in the database and the testing of such database is known as Database Testing or Backend testing. There are different databases like SQL Server, MySQL, and Oracle etc. Database testing involves testing of table structure, schema, stored procedure, data structure and so on.

In back-end testing GUI is not involved, testers are directly connected to the database with proper access and testers can easily verify data by running a few queries on the database. There can be issues identified like data loss, deadlock, data corruption etc during this back-end testing and these issues are critical to fixing before the system goes live into the production environment

#### 7) Browser Compatibility Testing

It is a subtype of Compatibility Testing (which is explained below) and is performed by the testing team.

[Browser Compatibility Testing](https://www.softwaretestinghelp.com/how-is-cross-browser-testing-performed/) is performed for web applications and it ensures that the software can run with the combination of different browser and operating system. This type of testing also validates whether web application runs on all versions of all browsers or not.

#### 8) Black Box Testing

Internal system design is not considered in this type of testing. Tests are based on the requirements and functionality.

Detailed information about the advantages, disadvantages, and [types of Black box testing](https://www.softwaretestinghelp.com/black-box-testing/) can be seen here.

#### 9) Branch Testing

It is a type of white box testing and is carried out during unit testing. Branch Testing, the name itself suggests that the code is tested thoroughly by traversing at every branch.

#### 10) Compatibility Testing

It is a testing type in which it validates how software behaves and runs in a different environment, web servers, hardware, and network environment. [Compatibility testing](https://www.softwaretestinghelp.com/software-compatibility-testing/) ensures that software can run on a different configuration, different database, different browsers, and their versions. Compatibility testing is performed by the testing team.

#### 11) Component Testing

It is mostly performed by developers after the completion of unit testing. [Component Testing](https://www.softwaretestinghelp.com/what-is-component-testing-or-module-testing/) involves testing of multiple functionalities as a single code and its objective is to identify if any defect exists after connecting those multiple functionalities with each other.

#### 12) End-to-End Testing

Similar to system testing, [End-to-end testing](https://www.softwaretestinghelp.com/what-is-end-to-end-testing/) involves testing of a complete application environment in a situation that mimics real-world use, such as interacting with a database, using network communications, or interacting with other hardware, applications, or systems if appropriate

**Importance of Testing:**

Software Testing is necessary because we all make mistakes. We need to check everything and anything we produce because things can always go wrong.

Since we assume that our work may have mistakes, hence we all need to check our own work. However some mistakes come from bad assumptions and blind spots, so we might make the same mistakes when we check our own work

**IMPLEMENTATION**

Initiating a project first requires the documenting of needs or requirements. Clear objectives should be developed from this study with reasons for selecting the objectives. Deliverables then need to be documented along with the project scope. Scope can be refined during this initialization process. Assumptions and constraints should also be documented. All stakeholders should be involved in this process. This information will become the projects charter and the basis for initiating the project. The project then follows the PLAN-DO CHECK-ACT cycle (as defined by Shewhart and modified by Deming, in the ASQ Handbook, pages 13-14, American Society for Quality, 1999). The results of each cycle will be linked to the next as input. This process should increase the likelihood of deliverable acceptance.

In order to achieve deliverable of acceptance and meeting of objectives, the new system being built must be tested. Aligned with this, the end users must be fully trained so the company will benefit from the new system. There are five activities that must be performed during the implementation phase: [[6]](https://en.wikibooks.org/wiki/Systems_Analysis_and_Design/Introduction#cite_note-6)

* Construct software components
* Verify and test
* Convert Data
* Training end users and document the system.

**MAINTENANCE AND SUPPORT**

Maintenance and support covers all activities that are required once the system is in place. Activities include, but are not limited to:

* Phone support for users
* Physical onsite user support
* Resolving any issues that may arise with the new system
* Providing support materials/tools for users

The amount of support required may be determined based on the system. If it is a large system involving many different departments, maintenance and support may be needed for a longer time. If is a smaller system, maintenance and support may only be needed for a short time.

**SOFTWARE REQUIREMENTS**

**Software Interface**

Client

* Interpreter : Python 3.6 or higher
* Libraries : Numpy, Scipy, Pandas, jupyter, sklearn, etc.,

Developer

* Operating system : Windows 7 or above or Linux
* IDE : Spyder or Pycharm
* Database : SQLite
* Interpreter : Python with Anaconda package
* Documentation tools : MS-word, MS-PowerPoint

**Hardware Interface**

Client

* + Processor : 1 GHz or above
  + RAM : 1 GB or higher

Developer

* + Processor : 2 GHz or above
  + RAM : 4 GB or higher

**CONCLUSION**

* In this project we tried to build a system that will help users to differentiate between genuine an fake website.

* This project aims to build a system which will help in reducing data theft and various information disclosure by reducing the phishing attacks.
* Hackers are becoming smart day by day , they are using new and different types of techniques to do hacking and steal our data so in future more functionalities will be added in this project like new techniques and methodolgy to detect the phishing websites and webpages.

**Bibliography**

**Online Sources:**

* https://stackoverflow.com
* https://www.wikipedia.org
* https://github.com
* https://www.tutorialspoint.com

**Books:**

* Python Programming Fundamentals – A Beginner‘s Handbook
* Learning Python Application Development –Ninad Sathaye
* Software Engineering: A Practitioner‘s Approach –Pressman