DETECTION OF PHISHING WEBSITES FROM URLS

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

Submitted By

MOHAMMED ABDUL RASHEED KHALID	**21UK1A6769**
MOHAMMAD SUBHAN	**21UK1A6768**
MODEYA RITHVIK KUMAR	**21UK1A6766**
SANGA RAKESH	**22UK5A6714**

Under the guidance of

Mr. E.MAHESH

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) – 506005

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING(DATA SCIENCE)
VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION
INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled "DETECTING OF PHISHING WEBSITES FROM URLS" is being submitted by MOHAMMED ABDUL RASHEED KHALID(21UK1A6769), MOHAMMAD SUBHAN(21UK1A6768), MODEYA RITHVIK KUMAR(21UK1A6766),SANGA RAKESH(22UK5A6714), in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science & Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023- 2024.

Project Guide	**HOD**
Mr. E. MAHESH	**Dr. K. Sharmila**
(Assistant Professor)	(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr.P.PRASAD RAO,** Principal, Vaagdevi EngineeringCollege for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **Dr.K.SHARMILA**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **E.MAHESH,** Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friendsfor their encouragement and outpouring their knowledge and experience throughoutthe thesis.

MOHAMMED ABDUL RASHEED KHALID	**(21UK1A6769)**
MOHAMMADSUBHAN	**(21UK1A6768)**
MODEYA RITHVIK KUMAR	**(21UK1A6766)**
SANGA RAKESH	**(22UK5A6714)**

ABSTRACT

Phishing websites pose a significant cybersecurity risk, as they trick users into accessing malicious content and divulging sensitive information. These fraudulent sites often appear identical to legitimate webpages in terms of their interface and URL. Existing strategies for detecting phishing websites, such as blacklists and heuristics, have limitations due to inefficient security technologies and the rapid evolution of phishing tactics. Enhanced detection mechanisms are needed to mitigate this threat effectively.

TABLE OF CONTENTS:-

1.INTRODUCTION

OVERVIEW

Detecting phishing websites from URLs is crucial in safeguarding online users from cyber threats. Our project focuses on developing an efficient system that utilizes machine learning and data analysis techniques to identify and classify URLs as either legitimate or malicious. Phishing websites, which mimic trusted domains to deceive users into divulging sensitive information, pose significant risks such as identity theft and financial fraud. By analyzing features like domain characteristics, presence of suspicious elements, and similarity to known phishing patterns, our system aims to accurately differentiate between safe and harmful URLs.

This initiative seeks to enhance cybersecurity measures by empowering users and organizations to detect potential threats proactively. By raising awareness and providing actionable insights into phishing attempts, our project aims to mitigate the impact of online fraud and protect individuals' digital identities. Understanding and implementing effective measures against phishing attacks are crucial steps in creating a safer digital environment for users worldwide.

PURPOSE

The purpose of detecting phishing websites from URLs serves several critical objectives aimed at enhancing cybersecurity and protecting online users. These purposes include:

1. **Informing and Educating Users:** Providing clear and accessible information to users about the risks associated with phishing websites. By raising awareness, individuals can make informed decisions and take proactive measures to safeguard their personal information.

2. **Enhancing Cybersecurity**: Strengthening cybersecurity measures by developing a system that can accurately detect and classify phishing URLs. This helps mitigate the risks of identity theft, financial fraud, and other cyber threats posed by malicious websites.

3. **Supporting Regulatory Efforts**: Supporting regulatory agencies and policymakers by providing insights into the prevalence and characteristics of phishing attempts. This information can inform the development of policies and regulations aimed at combating online fraud.

4. **Monitoring and Evaluation**: Similar to environmental monitoring, monitoring the prevalence and evolution of phishing techniques over time. By evaluating detection methods and their effectiveness, aiming to continuously improve the ability to protect users.

5. **Promoting Safe Online Practices**: Promoting safe online practices by empowering users and organizations with tools to detect and avoid phishing attacks. This contributes to creating a more secure digital environment and reducing the economic and personal impacts of cybercrime.

2.LITERATURE SURVEY

EXISTING PROBLEM

- The existing problem of phishing websites poses a complex global challenge affecting cybersecurity across personal, organizational, and societal levels. Phishing, a form of cyber attack, involves deceptive techniques to trick individuals into revealing sensitive information such as passwords, financial details, and personal data.
- Phishing attacks not only compromise individual privacy but also have significant implications for organizational security, leading to financial losses, reputational damage, and legal consequences.
- Moreover, phishing undermines trust in digital communications and can facilitate other forms of cybercrime such as identity theft and malware

distribution. Vulnerable populations, including less tech-savvy users and businesses with weaker cybersecurity measures, are particularly at risk.

- Inadequate user awareness, evolving phishing tactics, and the widespread use of online platforms further exacerbate the issue. Addressing these challenges requires robust technical solutions, user education, and collaborative efforts to mitigate the impacts of phishing attacks on individuals and organizations.

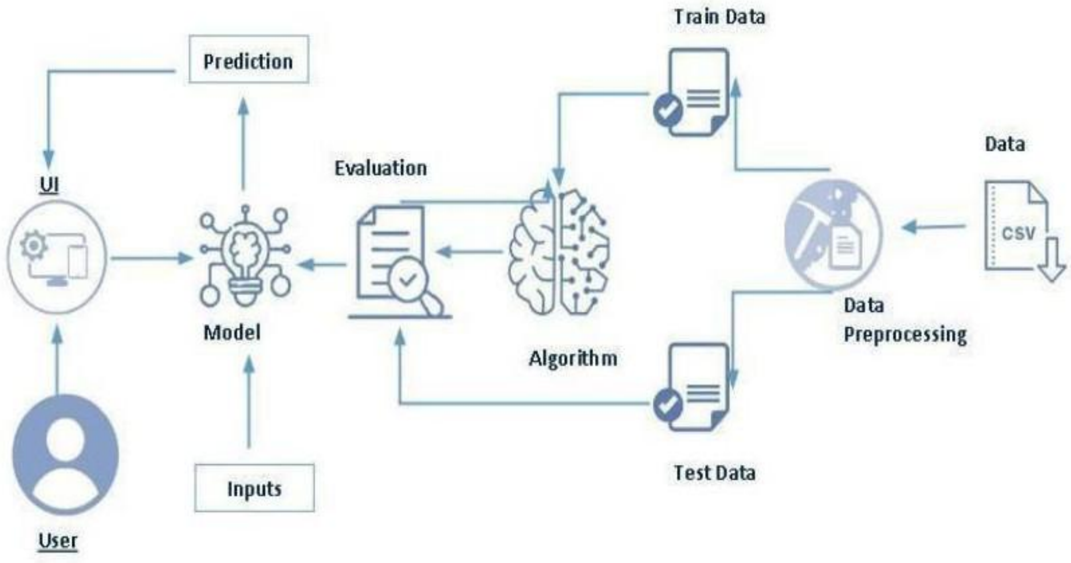
PROPOSED SOLUTION

- Our innovative proposed solution employs advanced machine learning models to detect and classify phishing URLs accurately and efficiently. By leveraging sophisticated algorithms and data analytics, we aim to enhance cybersecurity measures and protect users from falling victim to phishing attacks.
- Key components of our solution include:

1. Advanced Machine Learning Models: Utilizing state-of-the-art machine learning techniques to analyze URL features, detect patterns indicative of phishing, and classify URLs as safe or malicious in real-time.
2. Personalized Threat Information: Providing tailored information about phishing risks associated with detected URLs, including common tactics used by attackers and recommended precautions for users to safeguard their personalinformation.
3. Intuitive User Interfaces: Developing user-friendly interfaces that present clearand actionable insights about phishing threats. This ensures that users can easily understand the risks associated with URLs they encounter online and take appropriate preventive actions.

3.THEORITICAL ANALYSIS

BLOCK DIAGRAM



SOFTWARE DESIGNING

The following software components are essential for completing the project on detecting phishing websites from URLs:

- **Python Programming Environment**: Python will serve as the primary programming language for developing the detection system. It offers robust libraries and frameworks for machine learning, data analysis, and web development.

- **Google Colab**: Google Colab will be used as the development and execution environment for model training, data preprocessing, and analysis tasks. It provides a cloud-based Jupyter Notebook environment with access to necessary Python libraries and computational resources.

- **Dataset (CSV File):** A curated dataset in CSV format containing labeled URLs, distinguishing between legitimate and phishing websites. This dataset is crucial for training and evaluating machine learning models.

- **Data Preprocessing Tools**: Python libraries such as Pandas, NumPy, and Scikit-learn will be employed for data preprocessing tasks. This includes handling missing values, text normalization, and feature extraction from URLs.

- **Feature Engineering**: Utilizing Scikit-learn and custom Python scripts for feature engineering to extract relevant features from URLs. This process involves analyzing URL structures, domain characteristics, and other indicators of phishing behavior.

- **Machine Learning Models**: Employing machine learning frameworks like Scikit-learn, TensorFlow, or PyTorch for developing and training classification models. These models will learn to differentiate between benign and malicious URLs based on extracted features.

- **Model Evaluation**: Evaluating model performance using metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques and validation sets will ensure robust evaluation of model generalization and effectiveness in real-world scenarios.

- **User Interface (UI) Development**: Developing a user-friendly web interface using Flask, a Python web framework. The Flask application will allow users to input URLs for evaluation, visualize detection results, and receive actionable insights about identified phishing threats.

- **Integration and Deployment**: Integrating the trained machine learning models with the Flask-based UI and deploying the system on a web server or cloud platform for accessibility and scalability.

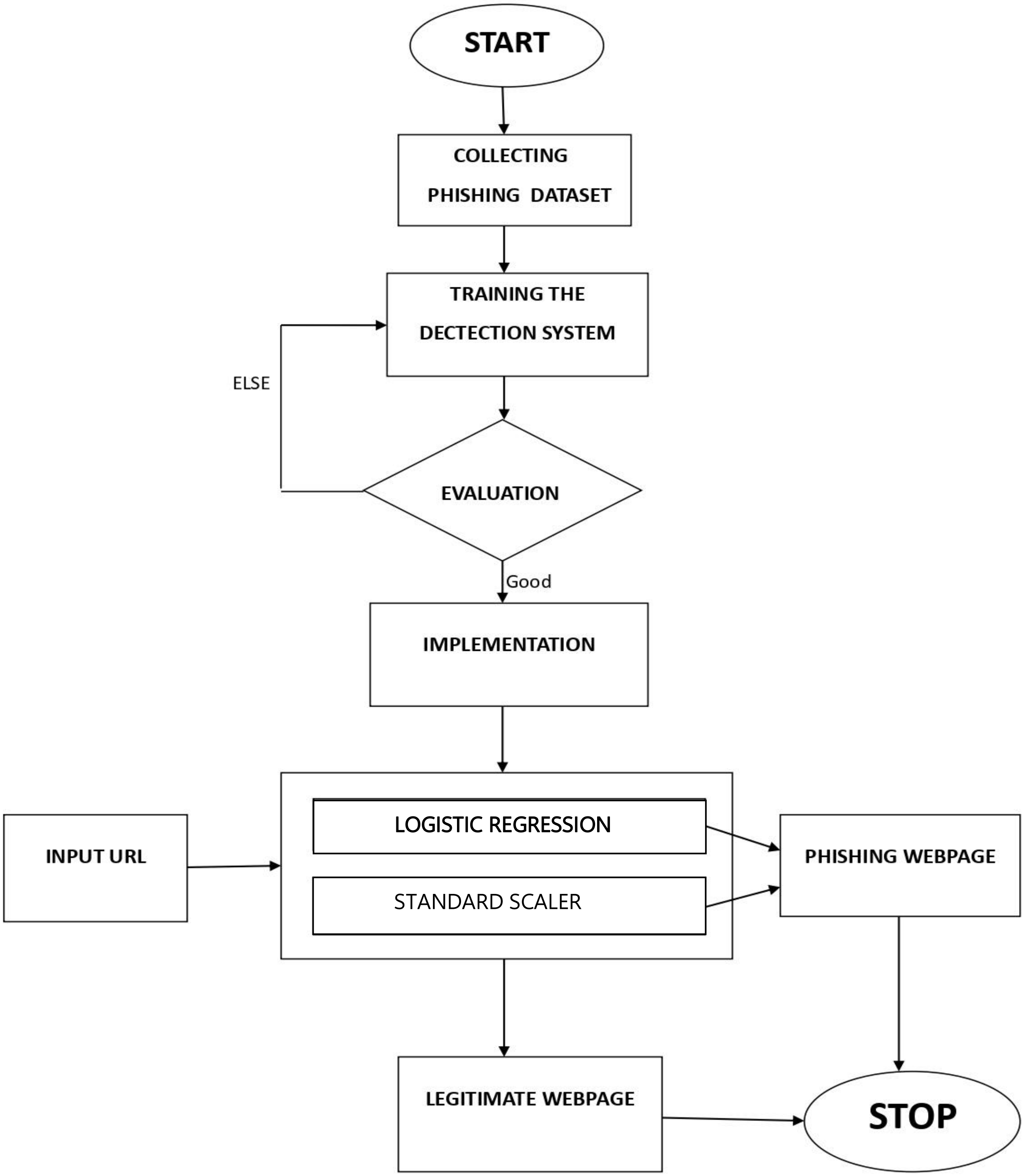
4. EXPERIMENTAL INVESTIGATION

In this project, we utilized a Phishing Website Dataset. This dataset is in CSV format and includes labeled data with the following columns:

1. URL: The web address of the website being evaluated.
2. Length_of_URL: Number of characters in the URL.
3. Length_of_Hostname: Number of characters in the hostname part of the URL.
4. IP_Address_Presence: Boolean indicating whether the URL contains an IP address instead of a hostname.
5. URL_Shortening_Service: Boolean indicating whether the URL uses a URL shortening service.
6. Having_At_Symbol: Boolean indicating whether the URL has the "@" symbol, common in phishing URLs.

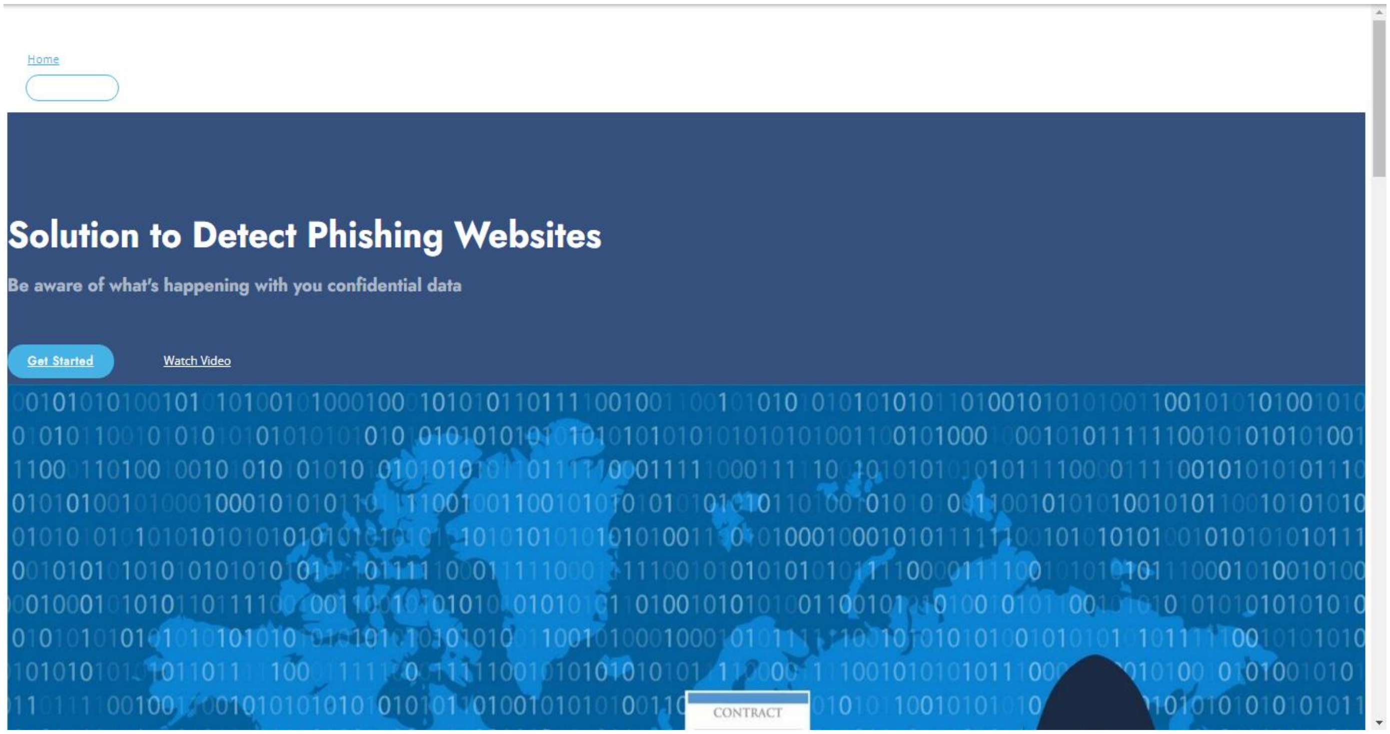
7. Redirection_Count: Number of redirections in the URL.
8. Prefix/Suffix: Boolean indicating whether the URL uses a prefix or suffix technique to deceive users.
9. Having_Sub_Domain: Boolean indicating whether the URL has a subdomain.
10. SSLfinal_State: State of the SSL certificate validity.
11. Domain_registeration_length: Length of time the domain is registered for.
12. Favicon: Boolean indicating presence of a favicon.
13. HTTP_Host: HTTP header field indicating the host name of the requested resource.
Request_URL: Domain of URL's request.

5. FLOW CHART


START
COLLECTING
PHISHING DATASET
TRAINING THE
DECTECTION SYSTEM
ELSE
EVALUATION
Good
IMPLEMENTATION
INPUT URL
LOGISTIC REGRESSION
STANDARD SCALER
PHISHING WEBPAGE
LEGITIMATE WEBPAGE
STOP

6. RESULT

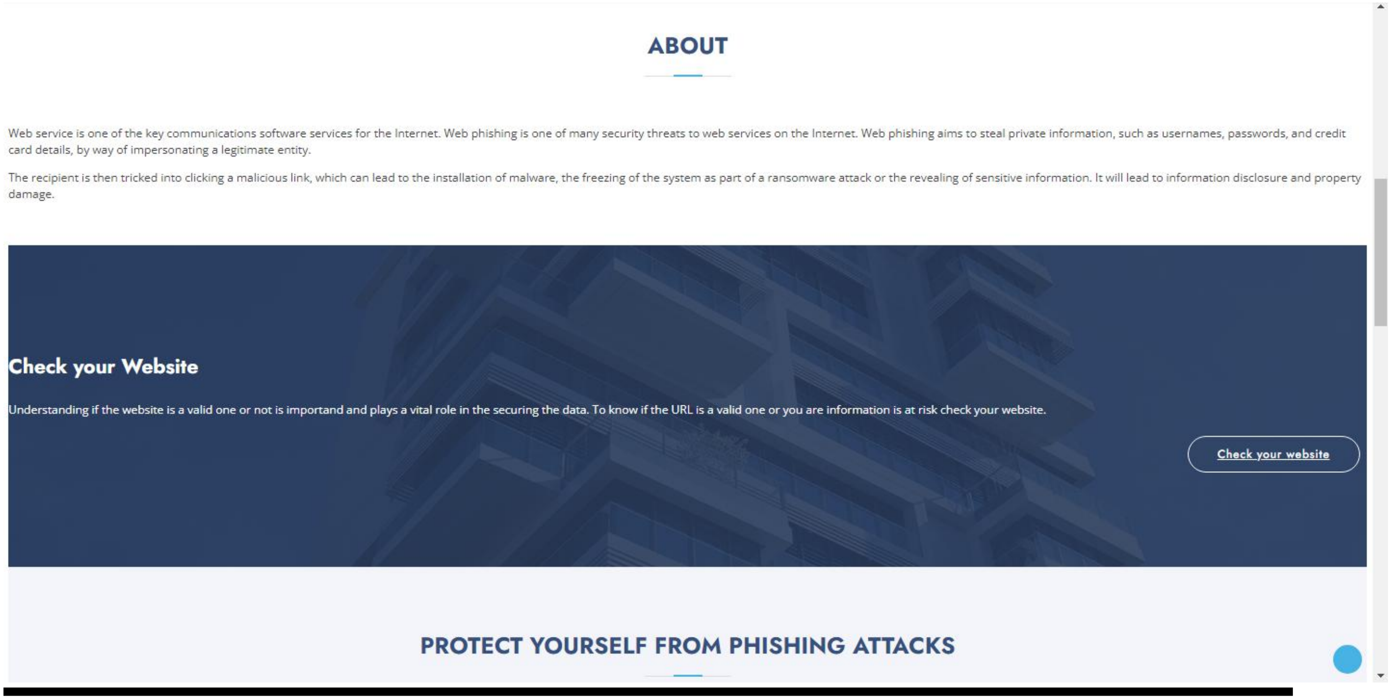
INDEX



Home

Solution to Detect Phishing Websites

Be aware of what's happening with you confidential data



ABOUT

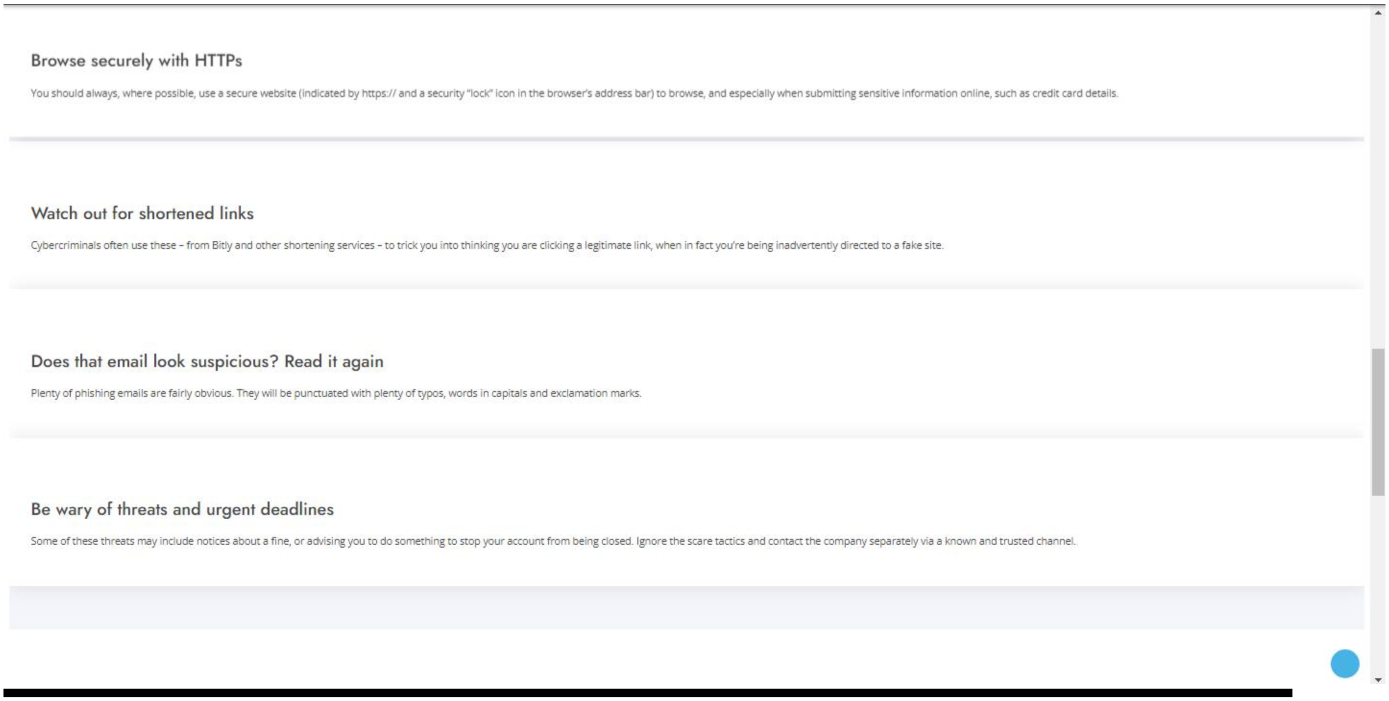
Web service is one of the key communications software services for the Internet. Web phishing is one of many security threats to web services on the Internet. Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity.

The recipient is then tricked into clicking a malicious link, which can lead to the installation of malware, the freezing of the system as part of a ransomware attack or the revealing of sensitive information. It will lead to information disclosure and property damage.

Check your Website

Understanding if the website is a valid one or not is importand and plays a vital role in the securing the data. To know if the URL is a valid one or you are information is at risk check your website.

PROTECT YOURSELF FROM PHISHING ATTACKS



Browse securely with HTTPs

You should always, where possible, use a secure website (indicated by https:// and a security "lock" icon in the browser's address bar) to browse, and especially when submitting sensitive information online, such as credit card details.

Watch out for shortened links

Cybercriminals often use these – from Bitly and other shortening services – to trick you into thinking you are clicking a legitimate link, when in fact you're being inadvertently directed to a fake site.

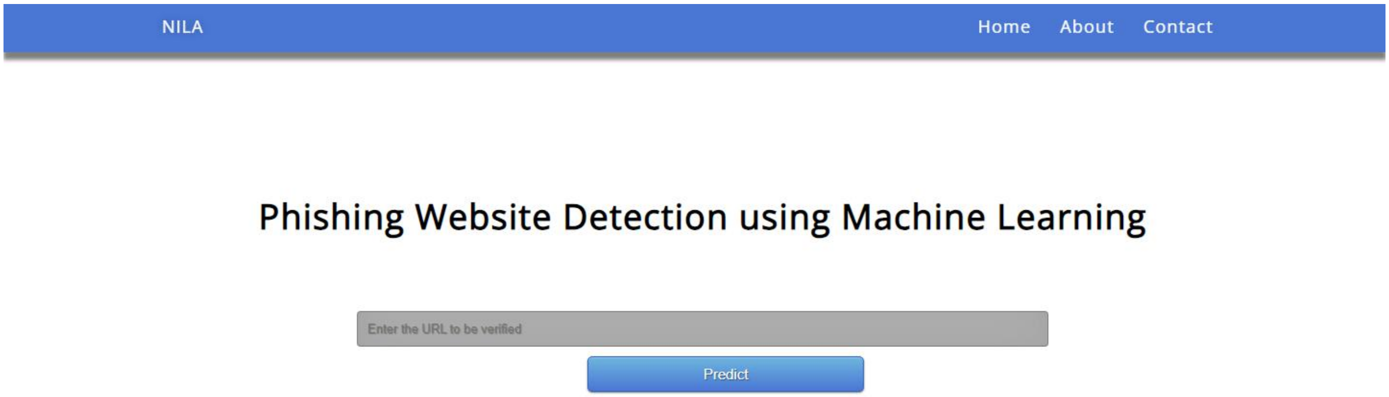
Does that email look suspicious? Read it again

Plenty of phishing emails are fairly obvious. They will be punctuated with plenty of typos, words in capitals and exclamation marks.

Be wary of threats and urgent deadlines

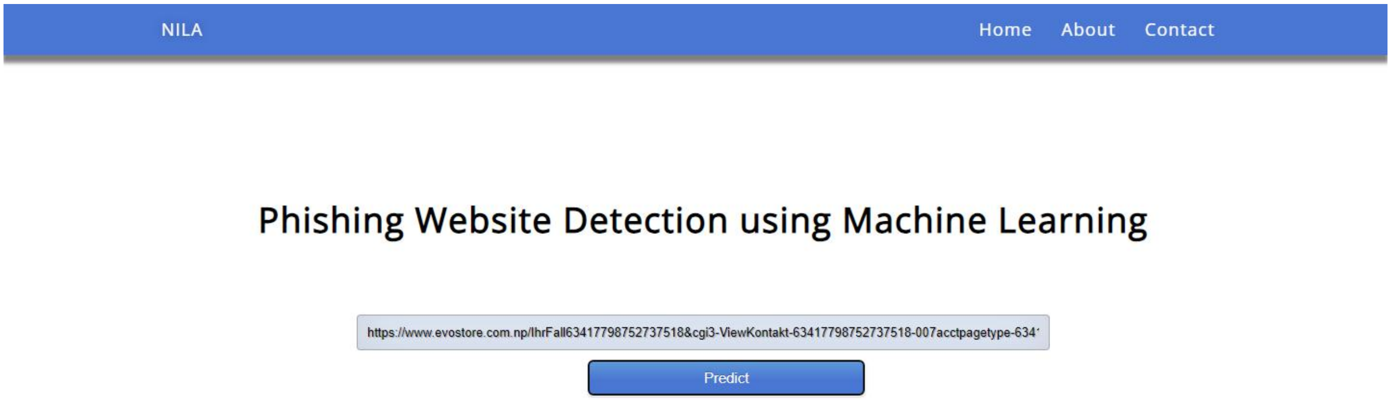
Some of these threats may include notices about a fine, or advising you to do something to stop your account from being closed. Ignore the scare tactics and contact the company separately via a known and trusted channel.

PREDICTIONS



Phishing Website Detection using Machine Learning

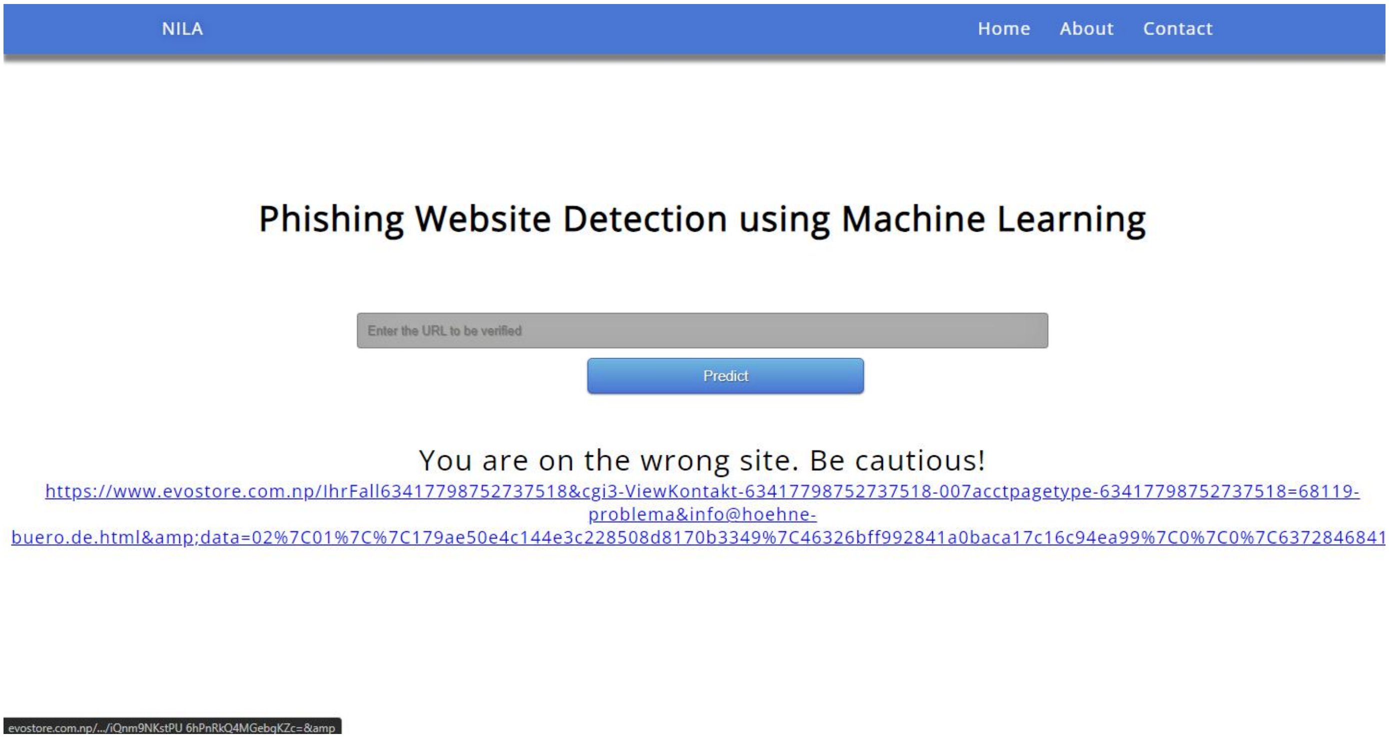
Enter the URL to be verified



Phishing Website Detection using Machine Learning

https://www.evostore.com.np/IhrFall634177987527375l8&cgi3-ViewKontakt-634177987527375l8-007acctpagetype-634'

RESULTS

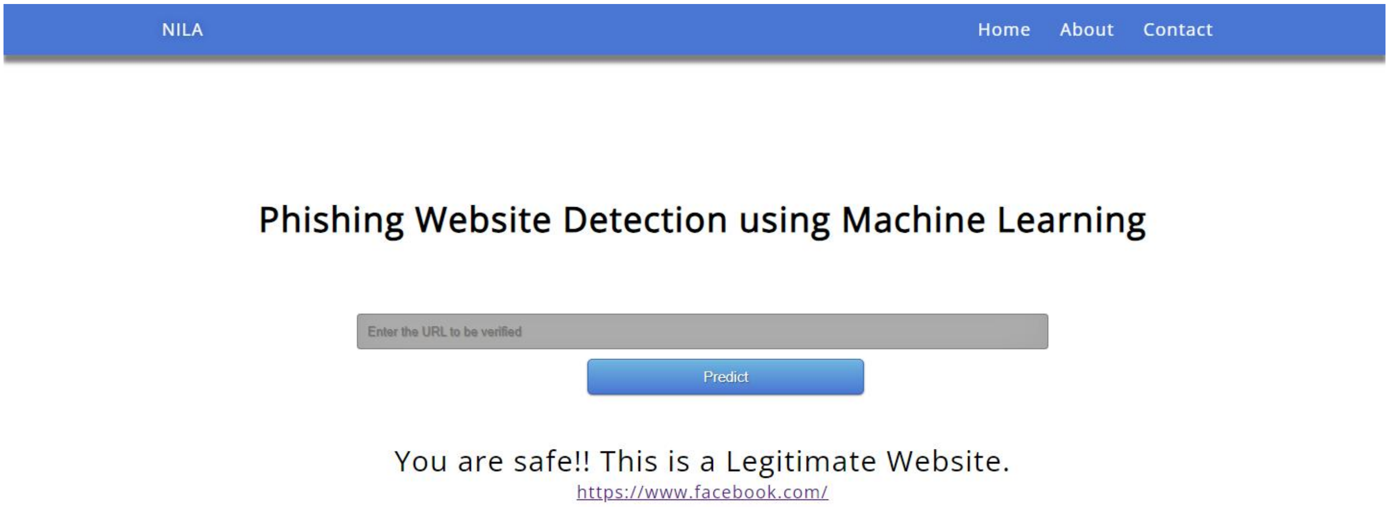


Phishing Website Detection using Machine Learning

You are on the wrong site. Be cautious!

https://www.evostore.com.np/IhrFall634177987527375l8&cgi3-ViewKontakt-634177987527375l8-007acctpagetype-634177987527375l8=68119-problema&info@hoehne-buero.de.html&amp;data=02%7C01%7C%7C179ae50e4c144e3c228508d8170b3349%7C46326bff992841a0baca17c16c94ea99%7C0%7C0%7C6372846841

evostore.com.np/.../iQnm9NKstPU 6hPnRkQ4MGebqKZc=&amp



Phishing Website Detection using Machine Learning

You are safe!! This is a Legitimate Website.

https://www.facebook.com/

7. ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

1. Enhanced Public Health: Provides timely information on phishing threats, promoting proactive security measures.
2. Proactive Decision-Making: Enables informed decisions regarding online activities.
3. Improved Security Awareness: Increases awareness of phishing techniques and cybersecurity risks.
4. Community Engagement: Engages users in cybersecurity best practices and feedback.
5. Customized Security Advice: Tailors recommendations based on phishing threat levels.

DISADVANTAGES:

1. Data Reliance: Relies on accurate and current phishing data for precise predictions.
2. Resource-Intensive: Requires substantial resources for data processing, modeling, and system maintenance.
3. Technical Expertise: Users may need technical knowledge to interpret phishing threat information.
4. Model Accuracy: The system's effectiveness depends on the quality of the predictive model.
5. Privacy Concerns: Handling user data raises privacy issues and requires secure practices.

8. APPLICATIONS

1. Enhanced Security: Empowers individuals to make informed decisions, reducing the risk of falling victim to phishing attacks.
2. Cybersecurity Monitoring: Monitors phishing trends and threats, supporting proactive cybersecurity measures.
3. Regulatory Compliance: Assists organizations in complying with cybersecurity regulations and standards.
4. Public Awareness: Raises awareness about phishing risks, influencing user behavior and promoting safer online practices.

9. CONCLUSION

- In conclusion, the proposed phishing detection system provides a comprehensive solution to address the critical challenges of cybersecurity monitoring, public awareness, and proactive user engagement. By integrating advanced predictive models with user-friendly interfaces, the system enables individuals and organizations to protect sensitive information, mitigate risks, and enhance overall cybersecurity resilience.

- This project represents a significant advancement in cybersecurity technology, aiming to safeguard users from phishing threats and foster a secure digital environment. With ongoing research and innovation, the system has the potential to make a positive impact on cybersecurity practices globally.

10. FUTURE SCOPE

Future Scope of the Phishing Detection System:

1. Global Implementation: Extend the system's deployment to a wider global audience, addressing phishing threats on a global scale.
2. Advanced Technology Integration: Integrate emerging technologies like AI and machine learning for more accurate and real-time phishing detection.
3. Enhanced User Education: Develop educational resources to empower users with knowledge on identifying and avoiding phishing attacks.
4. Industry Collaboration: Collaborate with cybersecurity firms and organizations to share threat intelligence and improve system efficacy.

Model building :

11. APPENDIX

1) Dataset
2) Google colab and VS code Application Building
   1. HTML file (Index file, final file)
   2. CSS file
   3. Model in pickle format
   4. app.py and inputScript.py

SOURCE CODE:

INDEX.HTML

```
<!DOCTYPE html>
<html lang="en">

<head>
  <meta charset="utf-8">
  <meta content="width=device-width, initial-scale=1.0" name="viewport">

  <title>Phishing Website Detection</title>
  <meta content="" name="description">
  <meta content="" name="keywords">
        <link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet' type='text/css'>
        <link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet' type='text/css'>
```

```
	<link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet' type='text/css'>
	<link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300' rel='stylesheet'
type='text/css'>
	<link rel="stylesheet" href="{{ url_for('static', filename='css/final1.css') }}">
	<!--link rel="stylesheet" href="G:\Gayatri Files\Smartbridge\Nidhi\Phishing
Website\static\css\style1.css"-->



  <!-- Google Fonts -->
  <link
href="https://fonts.googleapis.com/css?family=Open+Sans:300,300i,400,400i,600,600i,700,700i|Jost:300,3
00i,400,400i,500,500i,600,600i,700,700i|Poppins:300,300i,400,400i,500,500i,600,600i,700,700i"
rel="stylesheet">


	<!--link href="{{ url_for('static', filename='/vendor/bootstrap/css/bootstrap.min.css') }}"
rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/icofont/icofont.min.css') }}" rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/boxicons/css/boxicons.min.css') }}"
rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/remixicon/remixicon.css') }}" rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/venobox/venobox.css') }}" rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/owl.carousel/assets/owl.carousel.min.css') }}"
rel="stylesheet">
	<link href="{{ url_for('static', filename='/vendor/aos/aos.css') }}" rel="stylesheet"-->

  <!-- Template Main CSS File -->
  <link href="assets/css/style.css" rel="stylesheet">
  <!--link href="{{ url_for('static', filename='css/style.css') }}" rel="stylesheet"-->


</head>

<body>

  <!-- ======= Header ======= -->
  <header id="header" class="fixed-top ">
    <div class="container d-flex align-items-center">

      <h1 class="logo mr-auto"><a href="index.html">NiLa</a></h1>
      <!-- Uncomment below if you prefer to use an image logo -->
      <!-- <a href="index.html" class="logo mr-auto"><img src="assets/img/logo.png" alt="" class="img-
fluid"></a>-->

      <nav class="nav-menu d-none d-lg-block">
    <ul>
     <li class="active"><a href="index.html">Home</a></li>
     <li><a href="#about">About</a></li>
     <li><a href="#contact">Contact</a></li>

    </ul>
   </nav><!-- .nav-menu -->

   <a href="http://localhost:5000/predict" class="get-started-btn scrollto">Get Started</a>

  </div>
 </header><!-- End Header -->

 <!-- ======= Hero Section ======= -->
 <section id="hero" class="d-flex align-items-center">

  <div class="container">
   <div class="row">
    <div class="col-lg-6 d-flex flex-column justify-content-center pt-4 pt-lg-0 order-2 order-lg-1" data-
aos="fade-up" data-aos-delay="200">
     <h1>Solution to Detect Phishing Websites     </h1>
     <h2>Be aware of what's happening with you confidential data</h2>
     <div class="d-lg-flex">
      <a href="http://localhost:5000/predict" class="btn-get-started scrollto">Get Started</a>
      <a href="https://www.youtube.com/watch?v=jDDaplaOz7Q" class="venobox btn-watch-video"
data-vbtype="video" data-autoplay="true"> Watch Video <i class="icofont-play-alt-2"></i></a>
     </div>
    </div>
    <div class="col-lg-6 order-1 order-lg-2 hero-img" data-aos="zoom-in" data-aos-delay="200">
     <img src="https://www.technologyvisionaries.com/wp-content/uploads/2020/01/bigstock-Data-
Phishing-Hacker-Attack-t-319270852-scaled.jpg" class="img-fluid animated" alt="">
    </div>
   </div>
  </div>

 </section><!-- End Hero -->

 <main id="main">



  <!-- ======= About Us Section ======= -->
  <section id="about" class="about">
   <div class="container" data-aos="fade-up">

    <div class="section-title">
     <h2>About</h2>
```

```
        </div>

        <div class="row content">
          <div class="col-lg-6">
            <p>
              Web service is one of the key communications software services for the Internet. Web phishing is
one of many security threats to web services on the Internet.  Web phishing aims to steal private
information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate
entity.
            </p>

          </div>
          <div class="col-lg-6 pt-4 pt-lg-0">
            <p>
             The recipient is then tricked into clicking a malicious link, which can lead to the installation of
malware, the freezing of the system as part of a ransomware attack or the revealing of sensitive information.
It will lead to information disclosure and property damage.
            </p>

          </div>
        </div>

      </div>
    </section><!-- End About Us Section -->

            <!-- ======= Cta Section ======= -->
    <section id="cta" class="cta">
      <div class="container" data-aos="zoom-in">

        <div class="row">
          <div class="col-lg-9 text-center text-lg-left">
            <h3>Check your Website</h3>
            <p>Understanding if the website is a valid one or not is importand and plays a vital role in the
securing the data. To know if the URL is a valid one or you are information is at risk check your website.
</p>
          </div>
          <div class="col-lg-3 cta-btn-container text-center">
            <a class="cta-btn align-middle" href="http://localhost:5000/predict">Check your website</a>
          </div>
        </div>

      </div>
    </section><!-- End Cta Section -->


    <!-- ======= Services Section ======= -->
    <section id="services" class="services section-bg">
      <div class="container" data-aos="fade-up">

        <div class="section-title">
          <h2>Protect yourself from Phishing Attacks</h2>
          <p>As a report from the Anti-Phishing Working Group (APWG) revealed earlier this year, there has
been a notable rise in the number phishing attacks. It’s a widespread problem, posing a huge risk to
individuals and organizations</p>
                          <p>Follow the tips below and stay better protected against phishing attacks.</p>
        </div>

        <div class="row">
                        <div class="col-xl-3 col-md-6 d-flex align-items-stretch mt-4 mt-xl-0" data-aos="zoom-in"
data-aos-delay="400">
            <div class="icon-box">
              <div class="icon"><i class="bx bx-layer"></i></div>
              <h4>Browse securely with HTTPs</h4>
              <p>You should always, where possible, use a secure website (indicated by https:// and a security
“lock” icon in the browser’s address bar) to browse, and especially when submitting sensitive information
online, such as credit card details.</p>
            </div>
          </div>
          <div class="col-xl-3 col-md-6 d-flex align-items-stretch" data-aos="zoom-in" data-aos-delay="100">
            <div class="icon-box">

              <div class="icon"><i class="bx bxl-dribbble"></i></div>
              <h4>Watch out for shortened links</h4>
              <p>Cybercriminals often use these – from Bitly and other shortening services – to trick you into
thinking you are clicking a legitimate link, when in fact you’re being inadvertently directed to a fake
site.</p>
            </div>
          </div>

          <div class="col-xl-3 col-md-6 d-flex align-items-stretch mt-4 mt-md-0" data-aos="zoom-in" data-
aos-delay="200">
            <div class="icon-box">
              <div class="icon"><i class="bx bx-file"></i></div>
              <h4>Does that email look suspicious? Read it again</h4>
              <p>Plenty of phishing emails are fairly obvious. They will be punctuated with plenty of typos,
words in capitals and exclamation marks.</p>
            </div>
          </div>

          <div class="col-xl-3 col-md-6 d-flex align-items-stretch mt-4 mt-xl-0" data-aos="zoom-in" data-aos-
delay="300">
            <div class="icon-box">
```

```
              <div class="icon"><i class="bx bx-tachometer"></i></div>
              <h4>Be wary of threats and urgent deadlines</h4>
              <p>Some of these threats may include notices about a fine, or advising you to do something to
stop your account from being closed. Ignore the scare tactics and contact the company separately via a
known and trusted channel.</p>
            </div>
          </div>




        </div>

      </div>
    </section><!-- End Services Section -->





    <!-- ======= Contact Section ======= -->
    <section id="contact" class="contact">
      <div class="container" data-aos="fade-up">

        <div class="section-title">
          <h2>Contact</h2>
          <p>Get in contect with us to build many more amazing projects.</p>
        </div>

        <div class="row">

          <div class="col-lg-9 mt-5 mt-lg-0 d-flex align-items-stretch">
            <div class="info">
              <div class="address">
                <i class="icofont-google-map"></i>
                <h4>Location:</h4>
                <p>Nacharam Main Road,Hyderabad</p>
              </div>

              <div class="email">
                <i class="icofont-envelope"></i>
                <h4>Email:</h4>
                <p>info@thesmartbridge.com</p>
              </div>
```

```
       <div class="phone">
        <i class="icofont-phone"></i>
        <h4>Call:</h4>
        <p>9122223335</p>
       </div>

                   <iframe width="100%" height="370" frameborder="0" allowfullscreen=""
style="border:0"
src="https://www.google.com/maps/embed?pb=!1m14!1m8!1m3!1d15227.078752666794!2d78.545269!3d
17.422837!3m2!1i1024!2i768!4f13.1!3m3!1m2!1s0x3bcb995da7380f6b%3A0x608ed5e4b34e713b!2sNar
mada+Arcade%2C+Nacharam+Mallapur+Rd%2C+Snehapuri+Colony%2C+Tarnaka%2C+Hyderabad%2C
+Telangana+500076%2C+India!5e0!3m2!1sen!2sin!4v1452762479907" frameborder="0" style="border:0;
width: 100%; height: 290px;" allowfullscreen></iframe>

      </div>

     </div>

     <div class="col-lg-7 mt-5 mt-lg-0 d-flex align-items-stretch">


     </div>

    </div>

   </div>
  </section><!-- End Contact Section -->

 </main><!-- End #main -->

 <!-- ======= Footer ======= -->
 <footer id="footer">



  <div class="footer-top">
   <div class="container">
    <div class="row">


    </div>
   </div>
  </div>

  <div class="container footer-bottom clearfix">
 </footer><!-- End Footer -->

 <a href="#" class="back-to-top"><i class="ri-arrow-up-line"></i></a>
 <div id="preloader"></div>

 <!-- Vendor JS Files -->
 <script src="assets/vendor/jquery/jquery.min.js"></script>
 <script src="assets/vendor/bootstrap/js/bootstrap.bundle.min.js"></script>
 <script src="assets/vendor/jquery.easing/jquery.easing.min.js"></script>
 <script src="assets/vendor/php-email-form/validate.js"></script>
 <script src="assets/vendor/waypoints/jquery.waypoints.min.js"></script>
 <script src="assets/vendor/isotope-layout/isotope.pkgd.min.js"></script>
 <script src="assets/vendor/venobox/venobox.min.js"></script>
 <script src="assets/vendor/owl.carousel/owl.carousel.min.js"></script>
 <script src="assets/vendor/aos/aos.js"></script>

 <!-- Template Main JS File -->
 <script src="assets/js/main.js"></script>


  <!--script src="{{ url_for('static',filename='vendor/jquery/jquery.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/bootstrap/js/bootstrap.bundle.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/jquery.easing/jquery.easing.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/php-email-form/validate.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/waypoints/jquery.waypoints.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/isotope-layout/isotope.pkgd.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/owl.carousel/owl.carousel.min.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/aos/aos.js') }}"></script>
 <script src="{{ url_for('static',filename='vendor/venobox/venobox.min.js') }}"></script>

 <script src="{{ url_for('static',filename='js/main.js') }}></script-->

</body>

</html>
```

FINAL.HTML

```
<!DOCTYPE html>
<html >
<!--From https://codepen.io/frytyler/pen/EGdtg-->
<head>
  <meta charset="UTF-8">
  <title>Prediction</title>
        <link href='https://fonts.googleapis.com/css?family=Pacifico' rel='stylesheet' type='text/css'>
        <link href='https://fonts.googleapis.com/css?family=Arimo' rel='stylesheet' type='text/css'>
        <link href='https://fonts.googleapis.com/css?family=Hind:300' rel='stylesheet' type='text/css'>
        <link href='https://fonts.googleapis.com/css?family=Open+Sans+Condensed:300' rel='stylesheet'
type='text/css'>
        <link rel="stylesheet" href="{{ url_for('static', filename='css/final1.css') }}">
        <!--link rel="stylesheet" href="G:\Gayatri Files\Smartbridge\Nidhi\Phishing
Website\static\css\style1.css"-->

<style>
.login{
top: 20%;
}
</style>
</head>

<body>
<div class="header">
<div>NILA</div>
        <ul>



                <li><a
href="file:///G:/Gayatri%20Files/Smartbridge/Nidhi/Phishing%20Website/index.html#contact">Contact</
a></li>
                <li><a
href="file:///G:/Gayatri%20Files/Smartbridge/Nidhi/Phishing%20Website/index.html#about">About</a>
</li>
                <li><a
href="file:///G:/Gayatri%20Files/Smartbridge/Nidhi/Phishing%20Website/index.html">Home</a></li>
        </ul>
</div>

<div class="main">
<h1>Phishing Website Detection using Machine Learning<h1>
</div>
<form action="{{ url_for('y_predict')}}"method="post">
        <input type="text" name="URL" placeholder="Enter the URL to be verified" required="required"
/>
```

```
                <button type="submit" class="btn btn-primary btn-block btn-large">Predict</button>
                </form>
                <br>
                <br>

                <div id='result',class='result' style='color:black;font-size:30px;'>{{ prediction_text }}</div>
                <a href=" {{ url }} "> {{ url }} </a>
</body>

</html>
```

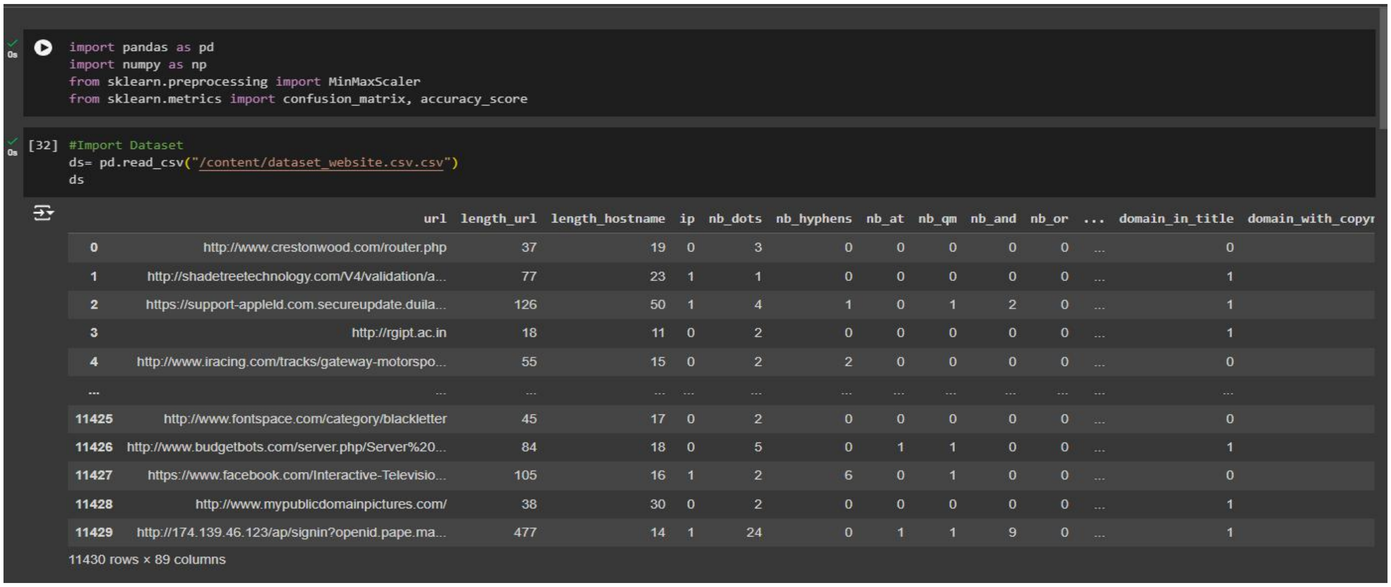
APP.PY

```
import numpy as np

from flask import Flask, request, jsonify, render_template

import pickle

import inputScript


# Load model

app = Flask(__name__)

model = pickle.load(open('Phishing_Website.pkl', 'rb'))


@app.route('/')

def helloworld():

    return render_template("index.html")


# Redirects to the page to give the user input URL.

@app.route('/predict')
```

```
def predict():
    return render_template('final.html')


# Fetches the URL given by the URL and passes to inputScript
@app.route('/y_predict', methods=['POST'])
def y_predict():
    '''
    For rendering results on HTML GUI
    '''
    url = request.form['URL']
    checkprediction = np.array(inputScript.main(url)).reshape(1, -1)
    prediction = model.predict(checkprediction)
    print(prediction)
    output = prediction[0]
    if output == 1:
        pred = "You are safe!! This is a Legitimate Website."
    else:
        pred = "You are on the wrong site. Be cautious!"
    return render_template('final.html', prediction_text='{}'.format(pred), url=url)


# Takes the input parameters fetched from the URL by inputScript and returns the predictions
@app.route('/predict_api', methods=['POST'])
def predict_api():
    '''
    For direct API calls through request
    '''
    data = request.get_json(force=True)
    prediction = model.y_predict([np.array(list(data.values()))])

    output = prediction[0]
    return jsonify(output)

if __name__ == "__main__":
        app.run(debug=True)
```

MODEL BUILDING

CODE SNIPPETS

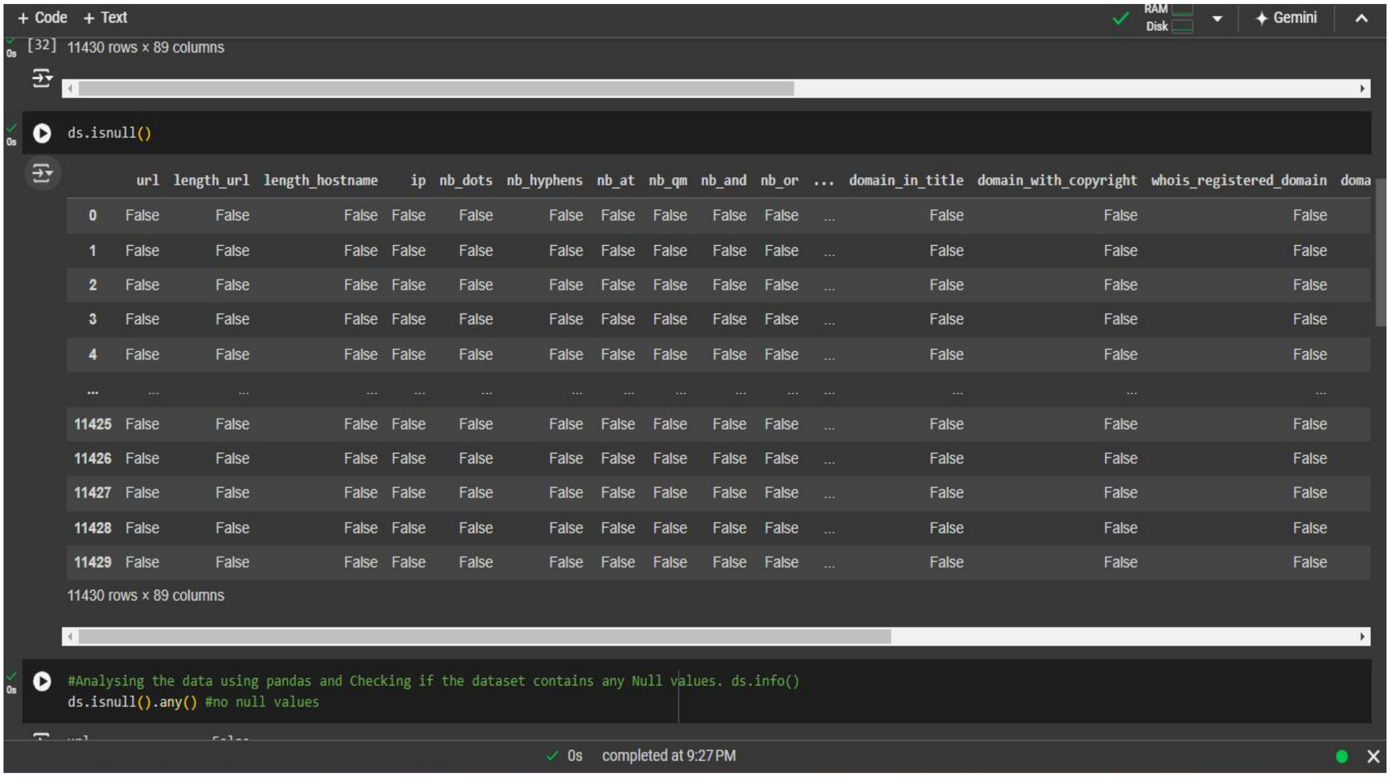


```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, accuracy_score
```

```python
#Import Dataset
ds= pd.read_csv("/content/dataset_website.csv.csv")
ds
```

	url	length_url	length_hostname	ip	nb_dots	nb_hyphens	nb_at	nb_qm	nb_and	nb_or	...	domain_in_title	domain_with_copyr
0	http://www.crestonwood.com/router.php	37	19	0	3	0	0	0	0	0	...	0	
1	http://shadetreetechnology.com/V4/validation/a...	77	23	1	1	0	0	0	0	0	...	1	
2	https://support-appleld.com.secureupdate.duila...	126	50	1	4	1	0	1	2	0	...	1	
3	http://rgipt.ac.in	18	11	0	2	0	0	0	0	0	...	1	
4	http://www.iracing.com/tracks/gateway-motorspo...	55	15	0	2	2	0	0	0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
11425	http://www.fontspace.com/category/blackletter	45	17	0	2	0	0	0	0	0	...	0	
11426	http://www.budgetbots.com/server.php/Server%20...	84	18	0	5	0	1	1	0	0	...	1	
11427	https://www.facebook.com/Interactive-Televisio...	105	16	1	2	6	0	1	0	0	...	0	
11428	http://www.mypublicdomainpictures.com/	38	30	0	2	0	0	0	0	0	...	1	
11429	http://174.139.46.123/ap/signin?openid.pape.ma...	477	14	1	24	0	1	1	9	0	...	1	

11430 rows × 89 columns



[32] 11430 rows × 89 columns

```
ds.isnull()
```

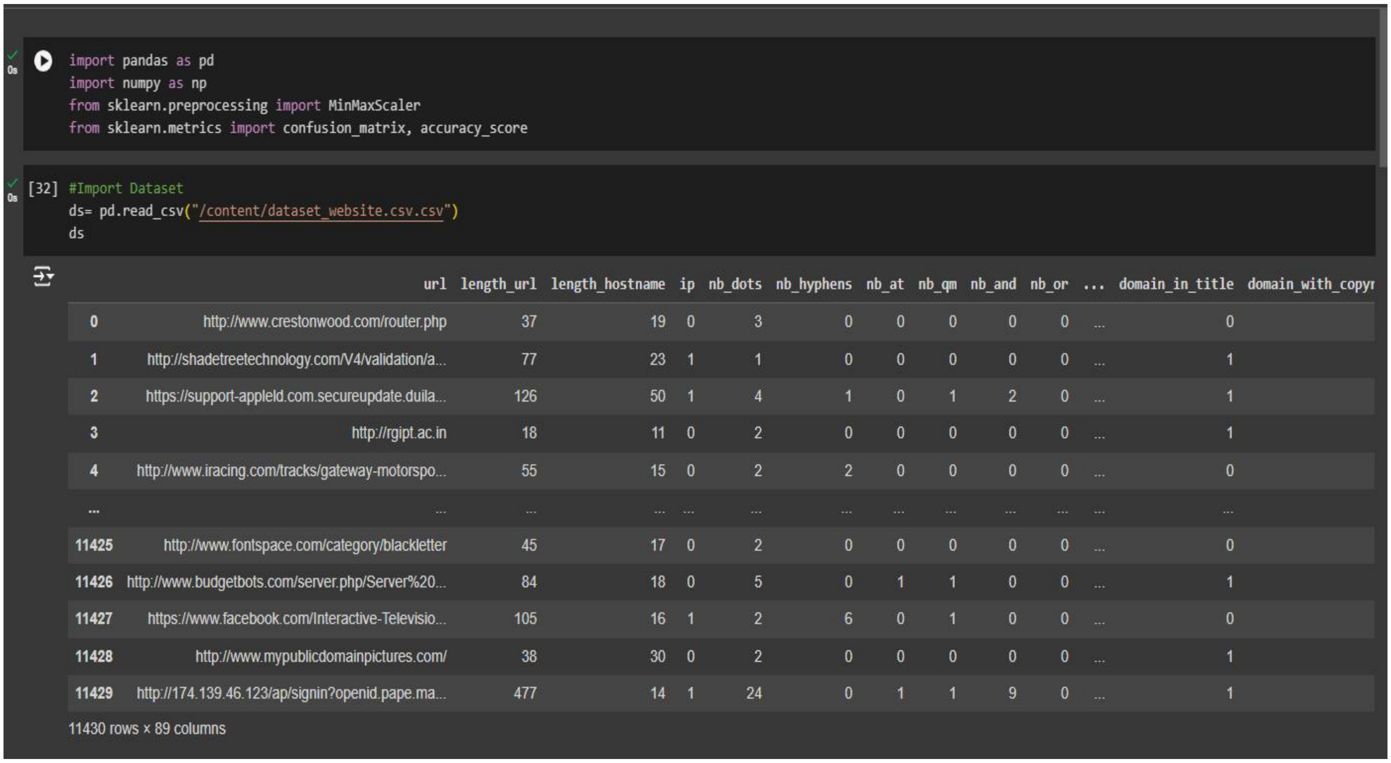
	url	length_url	length_hostname	ip	nb_dots	nb_hyphens	nb_at	nb_qm	nb_and	nb_or	...	domain_in_title	domain_with_copyright	whois_registered_domain	doma
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
11425	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
11426	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
11427	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
11428	False	False	False	False	False	False	False	False	False	False	...	False	False	False	
11429	False	False	False	False	False	False	False	False	False	False	...	False	False	False	

11430 rows × 89 columns

```
#Analysing the data using pandas and Checking if the dataset contains any Null values. ds.info()
ds.isnull().any() #no null values
```

```
url                False
```

✓ 0s completed at 9:27 PM

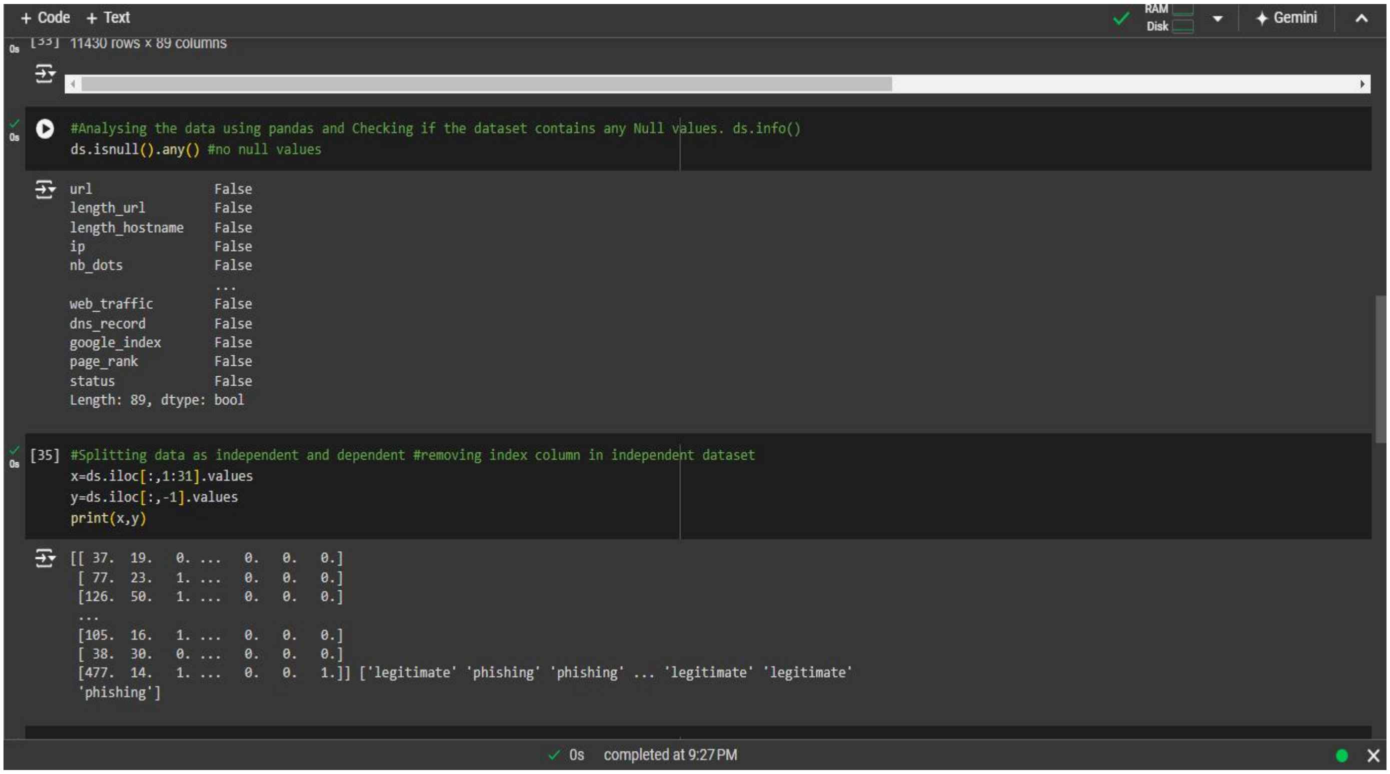


```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, accuracy_score
```

```
[32] #Import Dataset
     ds= pd.read_csv("/content/dataset_website.csv.csv")
     ds
```

	url	length_url	length_hostname	ip	nb_dots	nb_hyphens	nb_at	nb_qm	nb_and	nb_or	...	domain_in_title	domain_with_copyr
0	http://www.crestonwood.com/router.php	37	19	0	3	0	0	0	0	0	...	0	
1	http://shadetreetechnology.com/V4/validation/a...	77	23	1	1	0	0	0	0	0	...	1	
2	https://support-appleld.com.secureupdate.duila...	126	50	1	4	1	0	1	2	0	...	1	
3	http://rgipt.ac.in	18	11	0	2	0	0	0	0	0	...	1	
4	http://www.iracing.com/tracks/gateway-motorspo...	55	15	0	2	2	0	0	0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
11425	http://www.fontspace.com/category/blackletter	45	17	0	2	0	0	0	0	0	...	0	
11426	http://www.budgetbots.com/server.php/Server%20...	84	18	0	5	0	1	1	0	0	...	1	
11427	https://www.facebook.com/Interactive-Televisio...	105	16	1	2	6	0	1	0	0	...	0	
11428	http://www.mypublicdomainpictures.com/	38	30	0	2	0	0	0	0	0	...	1	
11429	http://174.139.46.123/ap/signin?openid.pape.ma...	477	14	1	24	0	1	1	9	0	...	1	

11430 rows × 89 columns



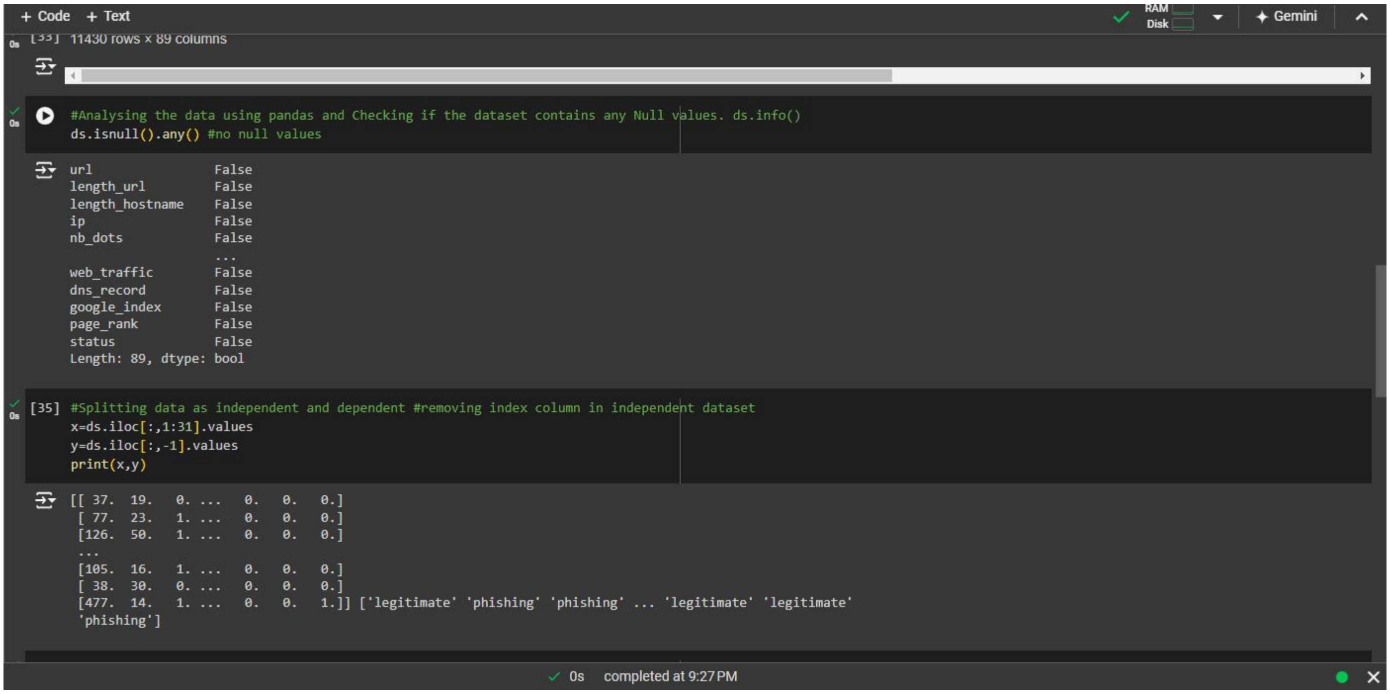
```
[33] 11430 rows × 89 columns
```

```
#Analysing the data using pandas and Checking if the dataset contains any Null values. ds.info()
ds.isnull().any() #no null values
```

```
url                False
length_url         False
length_hostname    False
ip                 False
nb_dots            False
                   ...
web_traffic        False
dns_record         False
google_index       False
page_rank          False
status             False
Length: 89, dtype: bool
```

```
[35] #Splitting data as independent and dependent #removing index column in independent dataset
x=ds.iloc[:,1:31].values
y=ds.iloc[:,-1].values
print(x,y)
```

```
[[ 37.  19.   0. ...   0.   0.   0.]
 [ 77.  23.   1. ...   0.   0.   0.]
 [126.  50.   1. ...   0.   0.   0.]
 ...
 [105.  16.   1. ...   0.   0.   0.]
 [ 38.  30.   0. ...   0.   0.   0.]
 [477.  14.   1. ...   0.   0.   1.]] ['legitimate' 'phishing' 'phishing' ... 'legitimate' 'legitimate'
 'phishing']
```



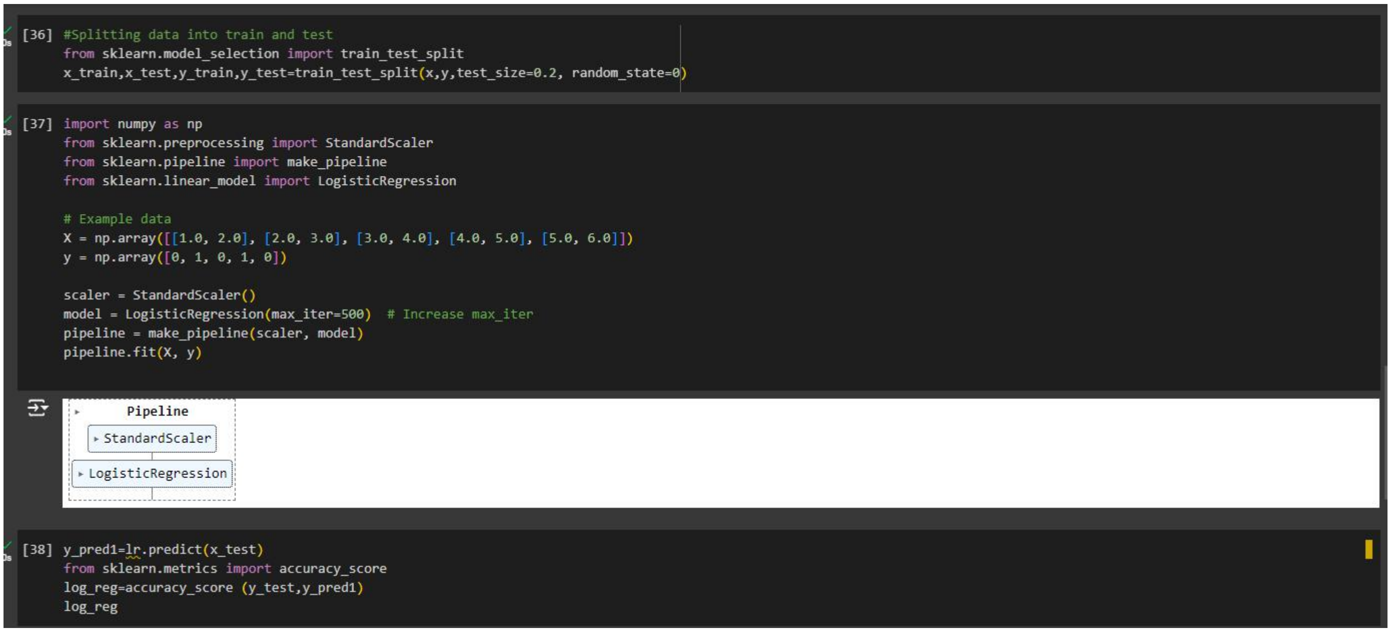
```
[33] 11430 rows × 89 columns
```

```
#Analysing the data using pandas and Checking if the dataset contains any Null values. ds.info()
ds.isnull().any() #no null values
```

```
url                False
length_url         False
length_hostname    False
ip                 False
nb_dots            False
                   ...
web_traffic        False
dns_record         False
google_index       False
page_rank          False
status             False
Length: 89, dtype: bool
```

```
[35] #Splitting data as independent and dependent #removing index column in independent dataset
x=ds.iloc[:,1:31].values
y=ds.iloc[:,-1].values
print(x,y)
```

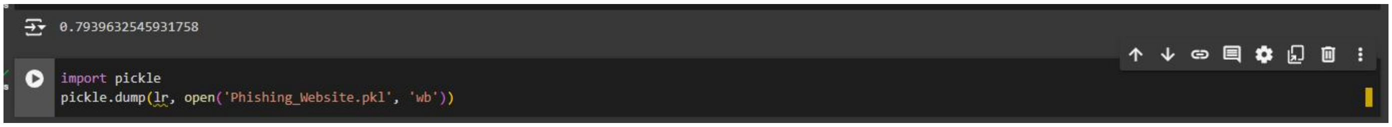
```
[[ 37.  19.   0. ...   0.   0.   0.]
 [ 77.  23.   1. ...   0.   0.   0.]
 [126.  50.   1. ...   0.   0.   0.]
 ...
 [105.  16.   1. ...   0.   0.   0.]
 [ 38.  30.   0. ...   0.   0.   0.]
 [477.  14.   1. ...   0.   0.   1.]] ['legitimate' 'phishing' 'phishing' ... 'legitimate' 'legitimate'
 'phishing']
```



```
[36] #Splitting data into train and test
     from sklearn.model_selection import train_test_split
     x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2, random_state=0)
```

```
[37] import numpy as np
     from sklearn.preprocessing import StandardScaler
     from sklearn.pipeline import make_pipeline
     from sklearn.linear_model import LogisticRegression

     # Example data
     X = np.array([[1.0, 2.0], [2.0, 3.0], [3.0, 4.0], [4.0, 5.0], [5.0, 6.0]])
     y = np.array([0, 1, 0, 1, 0])

     scaler = StandardScaler()
     model = LogisticRegression(max_iter=500)  # Increase max_iter
     pipeline = make_pipeline(scaler, model)
     pipeline.fit(X, y)
```

```
[38] y_pred1=lr.predict(x_test)
     from sklearn.metrics import accuracy_score
     log_reg=accuracy_score (y_test,y_pred1)
     log_reg
```



```
0.7939632545931758
```

```
import pickle
pickle.dump(lr, open('Phishing_Website.pkl', 'wb'))
```