Neuro Fuzzy Approach for Anti Phishing on Fog Networks

A project submitted to

ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY

TIRUNELVELI

in partial fulfilment of the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Submitted by

M. KATHIJA APSANA

Reg. No.: 21SPCS02

Under the Supervision and Guidance of

Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc(CS)., M.Phil., B.Ed.,

PG DEPARTMENT OF COMPUTER SCIENCE (SSC)

St. Mary's College (Autonomous), Thoothukudi - 628001

APRIL 2023

CERTIFICATE

This is to certify that this project work entitled **"Neuro Fuzzy Approach for Anti Phishing on Fog Networks"** is submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli, in partial fulfilment for the award of the degree of Master of Science in Computer Science for the work done during the year 2022-2023 by **M. KATHIJA APSANA** (Reg. No. 21SPCS02)

Signature of the Guide

Signature of the Co-ordinator

Signature of the Director

Director
Self Supporting Courses
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Principal

Principal
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Examiner

DECLARATION

I do hereby declare that the project entitled **"Neuro Fuzzy Approach for Anti Phishing on Fog Networks"** submitted for the degree of Master of Science in Computer Science in my original work carried out under the guidance of **Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., M.Phil., B.Ed**, Assistant Professor, PG Department of Computer Science(SSC),St.Mary's College(Autonomous), Thoothukudi.

Station: Thoothukudi

Date: 10.04.2023

M. Kothija Apsana

Signature of the student

ACKNOWLEDGEMENT

I express my first and foremost thanks to God Almighty for his gracious help and shower of blessings for having rendered us the strength and support to finish our project successfully.

My sincere thanks to **Dr. Sr. A. S. J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., Principal, Dr. Sr. C. Shibana M.Sc., B.Ed., M.Phil., Ph.D., Secretary, Rev. Sr. S. Josephine Jeyarani, Director of SSC,** St. Mary's College (Autonomous), Thoothukudi, for giving permission to work on this project.

I convey my heartfelt thanks to **Ms. C. Nayanthra Mascarenhas M.Sc., M.Phil., SET.,** Assistant Professor and Coordinator, PG Department of Computer Science (SSC) for her support and counsel.

I express my hearty thanks to my guide **Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., M.Phil., B.Ed.,** Assistant professor, PG Department of Computer Science (SSC),for her valuable suggestions, gentle guidance, enthusiastic ideas, to carry out and complete my work entirely effortlessly.

I also express my boundless thanks **to Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D.,** Assistant Professor, PG Department of Computer Science (SSC), for her encouragement and support.

I am much indebted to **Dr. P. Johnson Durai Raj** Director of Postulate Info Tech for his untiring effort, Immense Knowledge and priceless contribution without which I couldn't have finished my work effectively on time.

I thank my family members especially my parents for their encouragement and support both morally and financially which helped me to finish the project successfully.

POSTULATE

01.04.2023

To

The Head of the Department
PG Department of Computer Science (SSC)
St. Mary's College(Autonomous)
Thoothukudi

Dear Sir/Madam,

Sub: Project Work Completion-Reg

On behalf of **POSTULATE**, we are pleasure to inform you that **Ms. M. Kathija Apsana, Reg no: 21SPCS02** studying **Master of Computer Science** Final year has been done the project work at our concern Postulate Info Tech Private Limited in the title "**Neuro Fuzzy Approach for Anti-Phishing Fog Networks**" during the period from December 2022 to March 2023.

During this period, the candidate was found diligent and sincere. We wish all the best for future endeavors.

POSTULATE INFO TECH PRIVATE LIMITED

DIRECTOR

ABSTRACT

ABSTRACT

Fog networks are an emerging technology that enable users to access resources from multiple distributed locations. As fog networks become more popular, the risk of malicious actors using them for phishing attacks has increased. To protect users from such attacks, it is necessary to develop an effective anti-phishing system. One such system is a neuro-fuzzy approach for anti-phishing on fog networks. The neural networks are used to analyze user behavior and detect suspicious activities such as visits to suspicious websites and attempts to access restricted resources. The fuzzy logic is then used to classify the suspicious activities as either benign or malicious. The system then takes action based on the classified activities. If the activity is classified as benign, the user is allowed to access the requested resources. If the activity is classified as malicious, the user is blocked from accessing the resources and is redirected to a warning page. In addition, the system is designed to be implemented in PHP, a popular scripting language for web development. This makes it easy for web developers to integrate the system into their existing applications and websites. Furthermore, the system is designed to be highly scalable and can be used to protect large numbers of users in a fog network. This makes it an attractive option for protecting users on fog networks.

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

Phishing is a type of social engineering attack that seeks to take advantage of a flaw in the system at the user's end. For example, a system may be technically secure enough to prevent password theft, but an unwitting user's password may be leaked if the attack er sends a falsified (phished) update password request. Phishing is analogous to fishing in the water, except that instead of catching a fish, attackers attempt to steal personal information from consumers. When a person visits a bogus website and inputs their login and password, the attacker obtains the victim's credentials, which can be used for harmful reasons. The term "phishing" comes from the word "fishing." Because phishing websites look so much like legitimate ones, online consumers can easily be duped into providing personal information. Phishers utilise a variety of strategies to deceive their victims when they create phishing sites, including email messages, instant chats, forum postings, phone calls, and social networking information. Phishing causes significant economic damage all around the world, and the number and complexity of phishing sites is continually increasing. The number of phishing attacks is increasing by 5% monthly, according to reports from the Anti-Phishing Working Group. However, the anti-phishing problem has not been well addressed at the network edge for the following reasons. Many research employ the blacklist/whitelist strategy. However, all methods require a manual/automated update procedure to maintain a list of phishing websites. Before launching, URL requests are validated against a local database or a cloud database.

Despite the fact that blacklist/whitelist approaches detect phishing sites quickly, administering the blacklist/whitelist database for both the local and cloud databases is inefficient due to the continually expanding number of phishing sites. However, the anti-phishing problem has not been well addressed at the network edge for the following reasons. For starters, mobile users are more likely than desktop users to check their emails and utilise online browsers. As a result, people are far more likely to visit phishing sites that have not yet been discovered or blocked by anti-phishing software or firewalls on their local networks or devices. Second, because mobile devices are continually "hungry" for energy and computational resources, anti-phishing software is frequently neglected or disabled. As a result, it's difficult for consumers to tell if an incoming link is real or not. Third, existing anti-phishing solutions are ineffective at detecting phishing attempts, and mobile users may be vulnerable to phishing attacks while performing routine. tasks According to the survey, mobile users are three times more likely than desktop users to submit their login credentials. As a result, phishing attempts against terminal users are a serious concern at the network's edge.

SYSTEM SPECIFICATION

SYSTEM SPECIFICATION

HARDWARE REQUIREMENT:

- **Processor:** 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10GHz 3.11 GHz
- **Ram:** 8 GB
- **Hard Disk:** 50GB To 90GB
- **System type:** 64-bit Operating System

SOFTWARE REQUIREMENTS:

- **Front End:** PHP,HTML,CSS
- **Back End:** MySql

SOFTWARE DESCRIPTION

PHP

- PHP (Hypertext Pre-processor) is a popular, open-source scripting language used for web development.
- It is a server-side scripting language, meaning it is executed on the server before the page is rendered in the browser.
- It is used to create dynamic webpages, which can interact with databases and perform tasks such as user authentication, and data manipulation.
- It is primarily used for server-side programming, but can be used to create desktop applications.
- It is an object-oriented language, meaning it has classes and objects to help organize and structure code.
- It is platform-independent, meaning it can run on all major operating systems.

- It is relatively easy to learn and use, compared to other scripting languages such as Java and C#.
- It is a popular language for web development, and is used by some of the most popular websites in the world.
- It is secure, and can be used to create secure web applications.
- It is actively maintained, and is constantly being updated to incorporate new features and fix bugs.

MYSQL

- MySQL is an open-source relational database management system (RDBMS).
- It is based on Structured Query Language (SQL) which is used to access, manage, and manipulate data stored in relational databases.
- It is very popular among web developers as it is fast and easy to use.
- It is available on multiple operating systems, including Linux, Windows, and Mac OS X.
- It provides features such as triggers, stored procedures, views, and user-defined functions.
- It supports a wide range of programming languages such as Java, PHP, Python, and Ruby.
- It is highly secure and can be used for both small and large-scale applications.

HTML

- HTML stands for Hypertext Markup Language.
- HTML is used to structure and present content on the World Wide Web.
- HTML is written in the form of tags, which are surrounded by angle brackets.
- HTML tags are used to create HTML documents and render their content.
- HTML documents can include images, links, and other objects.

CSS

- CSS stands for Cascading Style Sheets.
- CSS escribes how HTML elements are to be displayed on screen,paper, or in other media.
- CSS saves a lot of work. It can control the layout of multiple web pages all at once.
- External stylesheets are stored in CSS files.

SERVER SCRIPTING LANGUAGE

- PHP is a server scripting language, and a powerful tool for makingdynamic and interactive Web pages.
- PHP is a widely-used, free, and efficient alternative to competitorssuch as Microsoft's ASP.
- PHP 7 is the latest stable release.

PROJECT DESCRIPTION

PROJECT DESCRIPTION

Neuro-fuzzy systems are a type of artificial intelligence technology that combines the learning capability of neural networks with the logical capabilities of fuzzy logic to deliver a more accurate and efficient system for decision making. This technology has recently been applied in the field of anti-phishing on fog networks. The goal of this application is to detect and prevent the spread of malicious phishing attempts through a fog network. Fog networks are composed of a large number of geographically distributed nodes, each with their own computing power and communication capabilities. This allows for a distributed computing architecture, where data can be processed and stored across multiple nodes. However, this also makes fog networks vulnerable to malicious attacks such as phishing attempts. These attempts involve using deceptive emails or websites to trick users into revealing sensitive information such as passwords or credit card numbers. To combat these threats, neuro-fuzzy systems can be used to detect and prevent the spread of phishing attempts. Neuro-fuzzy systems are able to process large amounts of data quickly and accurately and can be used to detect anomalous patterns in the data. This data can then be used to identify potential phishing attempts and alert the user before any sensitive information is disclosed. The neuro-fuzzy approach to anti-phishing on fog networks can be implemented in a variety of programming languages, including PHP. In conclusion, neuro-fuzzy systems are a powerful tool for detecting and preventing the spread of malicious phishing attempts on fog networks. It is an effective way to identify anomalous patterns in the data and alert users before any sensitive information is disclosed.

DESCRIPTION OF MODULES

URL and derived features

These features are included in the address of the website. Phishers generally adopt following methods to attack:

1. Long URL: Suspicious or phishing domains are hidden under long URL
2. Providing IP instead of URL: The IP address are not well recognised by general public and hence phishing URLs can easily be spoofed.
3. Using shortened URLs: Shortened URLs always redirect and hence seem unsuspicious to naked eyes.
4. "@" symbol in URL: Web browsers ignore anything preceding "@" symbol, hence the phishing part can follow the "@" symbol.
5. URLs with "//": "//" is used to redirect the URL, hence can lead to landing up on a phishing site post redirections.

6. URL with "-": Legitimate websites rarely use "-". However, phishing websites use "-" in URLs to mimic the names of legitimate websites.

7. Number of subdomain: Legitimate websites generally use no or only one sub-domain, however, phishers generally redirect via multiple sub-domains.

8. Use of HTTPs security: Websites on HTTPS are generally secure and have a valid certificate issued by a trusted authority. However, phishing websites generally operate over unprotected HTTP layer or do not have a valid HTTPS certificate.

9. Period for which Domain has been registered: Legitimate websites usually operate over several years. Most phishing websites operate for a short period of time and do not have domain registered for more than one year.

10. Favicon: Website favicons are used to relate identity to URL, and generally load from the same domain. If it is being loaded from an external website then, it can be an attempt to spoof the identity of URL and a possible attack.

11. Ports: All websites running over HTTP use port 80 and running over HTTPS use port 443. The other ports should remain closed for security reasons. However, when we inspected some popular websites like www.google.com, www.linkedin.com, www.yahoo.com, etc. we found most of them have their FTP (21), SSH (22) and other non-standard ports open. Hence we decided to drop this feature.

12. Use of "https" in domain part: Phishers can use "https" in domain part to trick users into believing that the URL passes through secure "HTTPS" protocol.

Page's source code-based features

A common trick employed by phishers is to make the interface textually and graphically very similar to a legitimate webpage. However, these features are difficult to tap especially when the content is loaded dynamically. However, there are a number of distinguishing features between legitimate and phishing websites based on code structure of the webpage: Based on URLs embedded in webpage The URLs being accessed/accessible by the webpage generally carry a good amount of information about their nature. If the links belong to the website itself, it increases the credibility of the website. Few features identified on the basis of the embedded URLs are:

1. Embedded objects' URLs: Legitimate pages share their domains with the objects embedded in them. However, phishing websites usually load embedded objects from external resources to resemble them.
2. URL of Anchor tag: Anchor tag in HTML is used for hyperlinking. A legitimate website will never have a void source in anchor tag. However, phishers can use it to discard useless information for them and to redirect personal information to alternate sources.
3. Tags: Legitimate pages have the domain name for the page and domain name of URLs in it's <Meta>, <Script> and <Link> tags as same. However, they usually differs for suspicious websites.
4. Server Form Handler (SFH): Legitimate websites always take action on the content submitted via a form. However, if the form handler is void, or of a different domain than the actual website, then the chances of being phishing are higher.
5. Submitting information to a mail: Legitimate websites generally process the information submitted either on frontend itself or on some backend. However, a phisher might redirect the information to his personal mail.
6. Abnormal URL: The host name is usually included in the URL of all the objects on webpage. However, since this feature was already well explored by other factors above, we decided to drop it.

K-nearest Neighbours:

1. K-Nearest Neighbours (KNN) is a supervised machine learning algorithm used for classification and regression.

2. It is a non-parametric and lazy learning algorithm, meaning that it does not make any assumptions about the underlying data.

3. In KNN, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small).

4. The input consists of the k closest training examples in the feature space.

5. The algorithm works by calculating the distance between a query point and all the training data points. The distance can be of any type e.g. Euclidean, Manhattan or Chebyshev.

6. The k nearest neighbors are then determined by selecting the k points with the lowest distance to the query point.

7.The output is then determined by majority vote of the k neighbor.

DATABASE DESCRIPTION

DATABASE DESCRIPTION

REGISTRATION

NAME	TYPE	NULL	DEFAULT
ID	int(11)	NO	NONE
FIRSTNAME	varchar(20)	NO	NONE
LASTNAME	varchar(20)	NO	NONE
USERNAME	varchar(20)	NO	NONE
EMAIL	varchar(50)	NO	NONE
PASSWORD	varchar(250)	NO	NONE
CONTACT	double	NO	NONE
USER_TYPE	varchar(20)	NO	NONE
FORGOT_PASS -ID	varchar(32)	YES	NULL
CREATED	datetime	NO	NONE
MODIFIED	datetime	NO	NONE

USER_FEEDBACK

NAME	TYPE	NULL	DEFAULT
FID	int(11)	NO	NONE
RATE	varchar(10)	NO	NONE
NAME	varchar(20)	NO	NONE
EMAIL	varchar(30)	NO	NONE
COMMENT	text	NO	NONE

URLS

NAME	TYPE	NULL	DEFAULT
UID	int(11)	NO	NONE
URL	varchar(20)	NO	NONE
TYPE	int(11)	NO	NONE

SYSTEM STUDY

SYSTEM STUDY

EXISTING SYSTEM

The existing system for this project would involve creating a database of known phishing websites and malicious URLs, and then using it to detect and block any such URLs when accessing a website. This can be done by using a combination of automated tools like web crawlers and web filters, as well as manual processes like manual user verification. A neuro fuzzy approach for this project would involve using machine learning techniques to create a model that can detect and block malicious URLs. This model would use input from both the user and the website to create a decision tree that can accurately identify suspicious URLs. This model could be trained using data from known phishing websites and malicious URLs, as well as from user input and feedback. The model could then be used to detect and block any suspicious URLs on fog networks. Additionally, the model could be used to flag any suspicious activity on the network, such as unusual data transfers or network connections. This could help to prevent any malicious activities from occurring on the network.

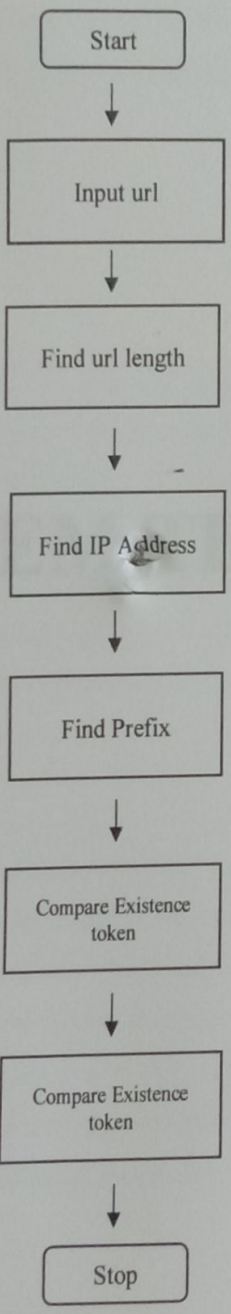
PROPOSED SYSTEM

A Neuro Fuzzy Approach for Anti-Phishing on Fog Network project is a novel system developed to detect and prevent malicious phishing attacks in fog networks. This system uses a combination of artificial intelligence and fuzzy logic to create an environment of self-learning algorithms that are capable of detecting and responding to phishing attempts in real-time. The system is divided into two components, namely, the detection and the prevention. The detection component is used to identify any suspicious phishing emails by using features like the sender's identity, the content, and the domain of the email. The prevention component is used to protect the users from any malicious phishing attempts by blocking the emails and preventing them from being sent to the intended recipient. The system is implemented in PHP and is built on an open-source framework to ensure scalability and reliability. This system will greatly improve the security of fog networks and protect users from malicious attacks.

SYSTEM DESIGN

SYSTEM DESIGN

ATA FLOW DIAGRAM



SYSTEM TESTING

SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTING

LOAD TESTING

Load testing is a type of performance testing that simulates a real-world load on any software, application, or website. Without it, your application could fail miserably in real-world conditions. That's why we build tools like retrace to help you monitor application performance and fix bugs before your code ever gets to production. Load testing examines how the system behaves during normal and high loads and determines if a system, piece of software, or computing device can handle high loads given a high demand of end-users. This tool is typically applied when a software development project nears completion.

UNIT TESTING

A program represents the logical elements of a system. For a program to run satisfactorily, it must compile and test data correctly and tie in properly with other programs. Achieving an error free program is the responsibility of the programmer. Program testing checks for two types of errors: syntax and logical. Syntax error is a program statement that violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are common syntax errors. These errors are shown through error message generated by the computer. For Logic errors the programmer must examine the output carefully.

FUNCTIONAL TESTING

Functional testing of an application is used to prove the application delivers correct results, using enough inputs to give an adequate level of confidence that will work correctly for all sets of inputs. The functional testing will need to prove that the application works for each client type and that personalization function work correctly. When a program is tested, the actual output is compared with the expected output. When there is a discrepancy the sequence of instructions must be traced to determine the problem. The process is facilitated by breaking the program into self-contained portions, each of which can be checked at certain key points. The idea is to compare program values against desk-calculated values to isolate the problems.

INTEGRATION TESTING

Integration testing is known as the second level of the software testing process, following unit testing. Integration testing involves checking individual components or units of a software project to expose defects and problems to verify that they work together as designed. Integration testing ensures that every integrated module functions correctly.Integration testing uncovers interface errors.Testers can initiate integration testing once a module is completed and doesn't require waiting for another module to be done and ready for testing.Testers can detect bugs, defects, and security issues.Integration testing provides testers with a comprehensive analysis of the whole system, dramatically reducing the likelihood of severe connectivity issues

CODING

CODING

Admin.php

```
<?php
include('server.php');
if (!isLoggedIn()) {
        $_SESSION['msg'] = "You must log in first";
        header('location: login.php');
}

if (isset($_POST['logout'])) {
        session_destroy();
        unset($_SESSION['username']);
        header("location: login.php");
}
?>


<!DOCTYPE html>
<html>
<head>
        <meta charset="utf-8">
        <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-
scalable=no">
        <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
        <meta name="HandheldFriendly" content="true">

        <meta name="descripton" content="free website check">
        <meta name="keyword" content="phishing, check websites, fake websites">
        <meta name="author" content="Suzan Dhungana">
        <title>detecting phishhing website</title>
        <link rel="stylesheet" type="text/css" href="./css/bootstrap.min.css">
        <link rel="stylesheet" type="text/css" href="./js/bootstrap.min.js">
        <link rel="stylesheet" type="text/css" href="./js/jquery-3.3.1.min.js">
        <link rel="stylesheet" type="text/css" href="./css/styles.css">
        <link rel="stylesheet" type="text/css" href="./css/styles2.css">
</head>
<body id="show">
        <header>
                <div class="container">
                        <div id="logo">
                                <h1><span class="highlight">Phishing</span> <span
class="second">Detection</span></h1>
                        </div>
```

```
                <nav>
                    <ul>
                        <li class="current"><a href="admin.php">users</a></li>
                        <li><a href="feedbacks.php">feedbacks</a></li>
                        <li><a href="index.php">logout</a></li>
                    </ul>
                </nav>
            </div>
        </header>

        <section>
            <div class="container">
                <?php
                include("dbcon.php");
                $sql="select * from registration";
                $users=mysqli_query($db,$sql);
                ?>
                <div class="container middle">
                    <br>
                    <div id="formt" align="center">
                        <table cellpadding="1" cellspacing="1"       width="600"
border="1" id="registration" style="margin-bottom: 20px;">
                            <tr>
                                <th>firstname</th>
                                <th>lastname</th>
                                <th>username</th>
                                <th>email</th>
                                <th>contact_no</th>
                                <th>user type</th>
                                <tr>
                                    <?php
                                    while
($registration=mysqli_fetch_assoc($users)) {
                                        echo "<tr>";
                                        echo
"<td>".$registration['firstname']."</td>";
                                        echo
"<td>".$registration['lastname']."</td>";
                                        echo
"<td>".$registration['username']."</td>";
                                        echo
"<td>".$registration['email']."</td>";
                                        echo
"<td>".$registration['contact']."</td>";
                                        echo
"<td>".$registration['user_type']."</td>";
                                        echo "</tr>";
                                    }
                                    ?>
```

```
                                        </tr>
                                    </tr>
                                </table>
                            </div>
                        </div>

                    </div>
                </section>
</body>
</html>
```

dbcon.php

```
<?php

$db = mysqli_connect('localhost:3308', 'root', 'Postulate@1234','php_project_db');

?>
```

errors.php

```
<?php  if (count($errors) > 0) : ?>
  <div class="error">
        <?php foreach ($errors as $error) : ?>
          <p><?php echo $error ?></p>
        <?php endforeach ?>
  </div>
<?php  endif ?>
```

Feedback.php

```
<?php
include('server.php');
if (!isLoggedIn()) {
        $_SESSION['msg'] = "You must log in first";
        header('location: login.php');
}

if (isset($_POST['logout'])) {
        session_destroy();
        unset($_SESSION['username']);
        header("location: login.php");
}
?>
<!DOCTYPE html>
<html>
<head>
        <meta charset="utf-8">
        <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-
```

```
scalable=no">
        <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
        <meta name="HandheldFriendly" content="true">

        <meta name="descripton" content="free website check">
        <meta name="keyword" content="phishing, check websites, fake websites">
        <meta name="author" content="Suzan Dhungana">
        <title>detecting phishhing website</title>
        <link rel="stylesheet" type="text/css" href="./css/bootstrap.min.css">
        <link rel="stylesheet" type="text/css" href="./js/bootstrap.min.js">
        <link rel="stylesheet" type="text/css" href="./js/jquery-3.3.1.min.js">
        <link rel="stylesheet" type="text/css" href="./css/styles.css">
        <link rel="stylesheet" type="text/css" href="./css/styles2.css">
</head>
<body id="show">
        <header>
                <div class="container">
                        <div id="logo">
                                <h1><span class="highlight">Phishing</span> <span
class="second">Detection</span></h1>
                        </div>
                        <nav>
                                <ul>
                                        <li><a href="admin.php">users</a></li>
                                        <li class="current"><a href="feedbacks.php">feedbacks</a></li>
                                        <li><a href="index.php">logout</a></li>
                                </ul>
                        </nav>
                </div>
        </header>

        <section>
                <div class="container">
                        <?php
                        include("dbcon.php");
                        $sql="SELECT * FROM user_feedback";
                        $records=mysqli_query($db,$sql);
                        ?>
                        <div class="container middle">
                                <br>
                                <div id="formt" align="center">
                                        <table width="550" border="1" cellpadding="1" cellspacing="1"
id="registration" style="margin-bottom: 20px;">
                                                <tr>
                                                        <!-- <th>User_id</th> -->
                                                        <th>Rate</th>
                                                        <th>Name</th>
                                                        <th>email</th>
                                                        <th>Comment</th>
```

```
                                    </tr>
                                    <?php
                                    while($user_feedback= mysqli_fetch_assoc($records)){
                                        echo "<tr>";
    // echo "<td>".$user_feedback['user_id']."</td> " ;
                                        echo "<td>".$user_feedback['rate']."</td>";
                                        echo "<td>".$user_feedback['name']."</td>";
                                        echo "<td>".$user_feedback['email']."</td>";
                                        echo "<td>".$user_feedback['comment']."</td>";
                                        echo "</tr>";
                                    }
                                    ?>

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

                    </div>
                </section>
</body>
</html>
```

Forgetpass.php

```
<?php
session_start();
$sessData = !empty($_SESSION['sessData'])?$_SESSION['sessData']:'';
if(!empty($sessData['status']['msg'])){
    $statusMsg = $sessData['status']['msg'];
    $statusMsgType = $sessData['status']['type'];
    unset($_SESSION['sessData']['status']);
}
?>
<!DOCTYPE html>
<html>
<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-scalable=no">
    <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
    <meta name="HandheldFriendly" content="true">

    <meta name="descripton" content="free website check">
    <meta name="keyword" content="phishing, check websites, fake websites">
    <meta name="author" content="Suzan Dhungana">
    <title>detecting phishhing website</title>
    <link rel="stylesheet" type="text/css" href="./css/bootstrap.min.css">
    <link rel="stylesheet" type="text/css" href="./js/bootstrap.min.js">
```

```
        <link rel="stylesheet" type="text/css" href="./js/jquery-3.3.1.min.js">
        <link rel="stylesheet" type="text/css" href="./css/styles.css">
        <link rel="stylesheet" type="text/css" href="./css/styles2.css">
        <link rel="stylesheet" type="text/css" href="./css/fogpas.css">
</head>
<body>
        <header>
                <div class="container">
                        <div id="logo">
                                <h1><span class="highlight"
onclick="location='index.php'">Phishing</span> <span class="second"
onclick="location='index.php'">Detection</span></h1>
                        </div>
                        <nav>
                                <ul>
                                        <li><a href="index.php">Home</a></li>
                                        <li><a href="register.php">Register</a></li>
                                        <li class="current"><a href="login.php">Login</a></li>
                                </ul>
                        </nav>
                </div>
        </header>

        <section id="logins">
                <h2 style="color: blue">Enter the Email of Your Account to Reset New Password</h2>
                <?php echo !empty($statusMsg)?'<p class="'.$statusMsgType.'" style="background-
color: #051019;
                opacity: 0.6;">'.$statusMsg.'</p>':''; ?>
                <div class="container1">
                        <div class="regisFrm">
                                <form action="userAccount.php" method="post">
                                        <input type="email" name="email" placeholder="EMAIL"
required="">
                                        <div class="send-button">
                                                <input type="submit" name="forgotSubmit"
value="CONTINUE">
                                        </div>
                                </form>
                        </div>
                </div>
        </section>

        <section id="check">
                <div class="container">
                        <h1>Check URL</h1>
                        <form action="" method="POST">
                                <input type="text" placeholder="Paste URL..." name="url"
required="paste url first">
                                <button type="submit" name="submit"
```

```
class="button1"><span>CHECK</span></button><span id="result"><?php

// initializing variables
                        $username = "";
                        $email    = "";
                        $errors = array();

// connect to the database
                        include("dbcon.php");


                        if (isset($_POST['submit'])) {
                                $url = mysqli_real_escape_string($db, $_POST['url']);

                                if (count($errors) == 0) {
                                        $query = "SELECT * FROM urls WHERE url='$url'";
                                        $results = mysqli_query($db, $query);
                                        if (mysqli_num_rows($results) == 1) {
                                                $check_url = mysqli_fetch_assoc($results);
                                                if ($check_url['type'] == '1') {
                                                        echo "<font color='red'>THIS IS
PHISHING URL</font>";

                                                }else{
                                                        echo "<font color='green'>THIS IS NOT
PHISHING URL</font>";
                                                }
                                        }else {
                                                echo("Not Found, Please Login");
                                        }
                                }
                        }?>
                </span>
        </form>
    </div>
</section>
<footer>
    <p>Phishing Website Detection System, Copyright &copy, 2018<br></p>
    <!-- <p>Contact us:<br>Phone no: 9813225420<br>Email:
sujan.dhungana321@hotmail.com</p> -->
</footer>
</body>
</html>
```

Index.php

```
<?php include('server.php') ?>
<!DOCTYPE html>
<html>
<head>
```

```
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width,initial-scale=1,maximum-scale=1,user-
scalable=no">
    <meta http-equiv="X-UA-Compatible" content="IE=edge,chrome=1">
    <meta name="HandheldFriendly" content="true">

    <meta name="descripton" content="free website check">
    <meta name="keyword" content="phishing, check websites, fake websites">
    <meta name="author" content="Suzan Dhungana">
    <title>detecting phishhing website</title>
    <link rel="stylesheet" type="text/css" href="./css/bootstrap.min.css">
    <link rel="stylesheet" type="text/css" href="./js/bootstrap.min.js">
    <link rel="stylesheet" type="text/css" href="./js/jquery-3.3.1.min.js">
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap-
theme.min.css" >
    <link rel="stylesheet" type="text/css" href="./css/styles.css">
    <link rel="stylesheet" type="text/css" href="./css/styles2.css">
    <!-- <link rel="stylesheet" type="text/css" href="./css/form.css"> -->
    <!-- <script src="form.js"></script> -->
</head>
<body>
    <header>
        <div class="container">
            <div id="logo">
                <h1><span class="highlight"
onclick="location='index.php'">Phishing</span><span class="second"
onclick="location='index.php'">Detection</span></h1>
            </div>
            <nav>
                <ul>
                    <li class="current"><a href="index.php">Home</a></li>
                    <li><a href="register.php">Register</a></li>
                    <li><a href="login.php">Login</a></li>
                </ul>
            </nav>
        </div>
    </header>

    <section id="showcase">
        <div class="container">
            <h1>Do Not Make Mistake - Check Every Website</h1>
            <p>Stay Alert!!!, Be Safe From The Phishers</p>
        </div>
    </section>

    <section id="check">
        <div class="container">
            <h1>Check URL</h1>
            <form action="" method="POST">
```

```
                        <input type="text" placeholder="Paste URL...." name="url" required="">
                        <button type="submit" name="submit"
class="button1"><span>CHECK</span></button><span id="result"><?php
// initializing variables
                        $username = "";
                        $email    = "";
                        $errors = array();
// connect to the database
                        include("dbcon.php");
                        if (isset($_POST['submit'])) {
                                $url = mysqli_real_escape_string($db, $_POST['url']);
                                if (count($errors) == 0) {
                                        $query = "SELECT * FROM urls WHERE url='$url'";
                                        $results = mysqli_query($db, $query);
                                        if (mysqli_num_rows($results) == 1) {
                                                $check_url = mysqli_fetch_assoc($results);
                                                if ($check_url['type'] == '1') {
                                                        echo "<font color='red'>THIS IS
PHISHING URL</font>";
                                                }else{
                                                        echo "<font color='green'>THIS IS NOT
PHISHING URL</font>";
                                                }
                                        }else {
                                                echo("Not Found, Please Login");
                                        }
                                }
                        }?>

                </span>
        </form>
    </div>
</section>
<section class="slide">
    <div class="slideshow-container">
<div class="mySlides fade">
 <img src="css/img/Capture.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture2.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture3.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture4.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture5.png" style="width:100% height:100%">
```

```
</div>
<div class="mySlides fade">
 <img src="css/img/Capture6.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture7.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture8.png" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture9.jpg" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture10.jpg" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture11.jpg" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture12.jpg" style="width:100% height:100%">
</div>
<div class="mySlides fade">
 <img src="css/img/Capture13.png" style="width:100% height:100%">
</div>
</div>
<br>
<div style="text-align:center">
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
        <span class="dot"></span>
</div>
</section>
<section style="background-color: #F2F2F2">
        <div class="container">
                <div class="imagebg"></div>
                <div class="row " style="margin-top: 50px">
                        <div class="col-md-6 col-md-offset-3 form-container">
                                <h2>Feedback</h2>
```

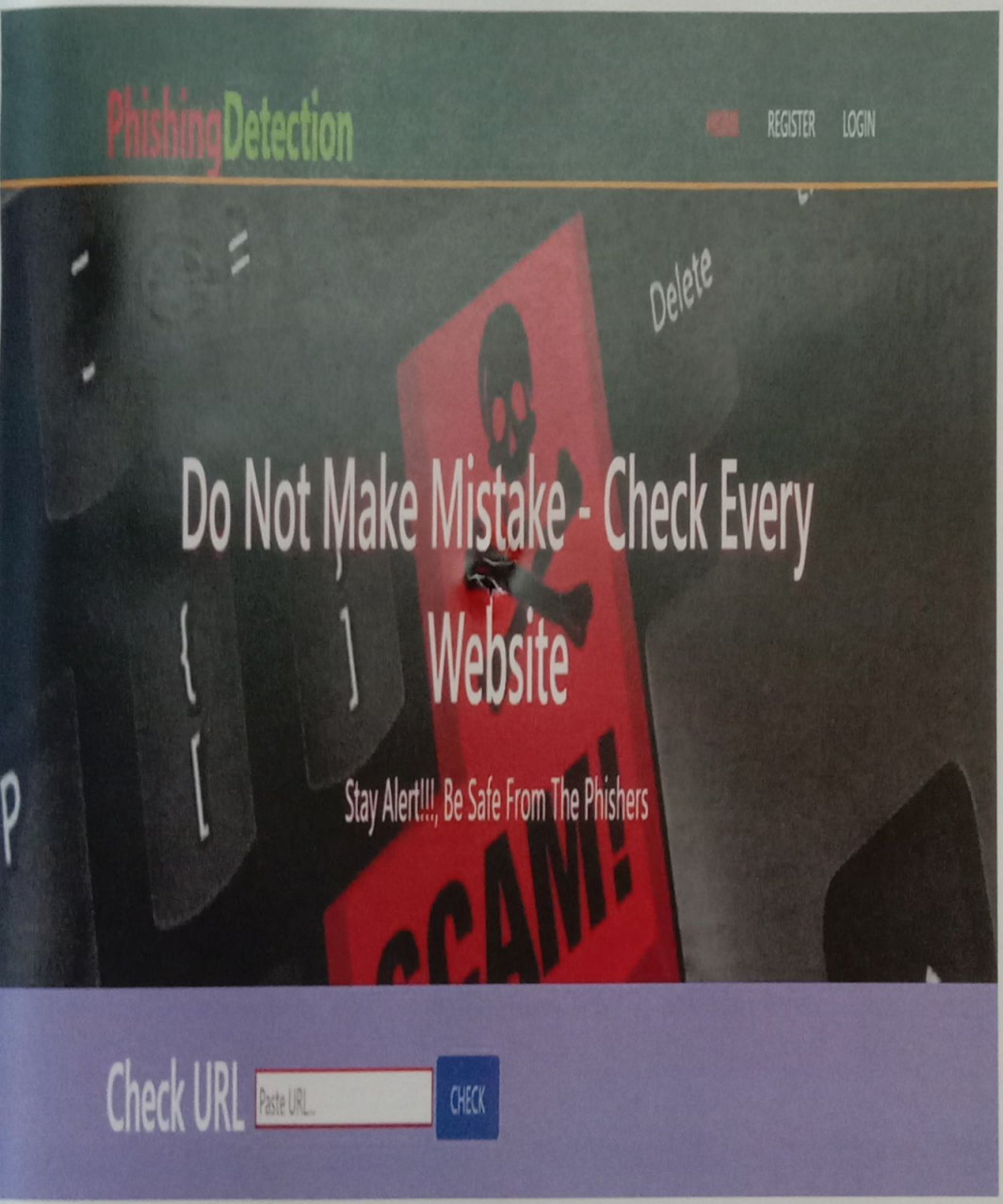
```
                    <p> Please provide your feedback below: </p>
                    <form role="form" method="post" id="reused_form">
                        <div class="row">
                            <div class="col-sm-12 form-group">
                                <label>How do you rate your overall
experience?</label>
                                <p>
                                    <label class="radio-inline">
                                        <input type="radio"
name="experience" id="radio_experience" value="bad" required="">
                                        Bad
                                    </label>
                                    <label class="radio-inline">
                                        <input type="radio"
name="experience" id="radio_experience" value="average" >
                                        Average
                                    </label>
                                    <label class="radio-inline">
                                        <input type="radio"
name="experience" id="radio_experience" value="good" >
                                        Good
                                    </label>
                                </p>
                            </div>
                        </div>
                        <div class="row">
                            <div class="col-sm-12 form-group">
                                <label for="comments"> Comments:</label>
                                <textarea class="form-control" type="textarea"
name="comments" id="comments" placeholder="Your Comments" maxlength="6000" rows="7"
required=""></textarea>
                            </div>
                        </div>
                        <div class="row">
                            <div class="col-sm-6 form-group">
                                <label for="name"> Your Name:</label>
                                <input type="text" class="form-control"
id="name" name="name" required>
                            </div>
                            <div class="col-sm-6 form-group">
                                <label for="email"> Email:</label>
                                <input type="email" class="form-control"
id="email" name="email" required>
                            </div>
                        </div>
                        <div class="row">
                            <div class="col-sm-12 form-group">
                                <button type="submit" class="btn btn-lg btn-
warning btn-block" name="send">Post </button>
```

```
                                    </div>
                                </div>
                            </form>
                        </div>
                    </div>
                </div>
            </section>
            <footer>
                <p>Phishing Website Detection System, Copyright &amp;copy, 2018<br></p>
                <!-- <p>Contact us:<br>Phone no: 9813225420<br>Email:
            sujan.dhungana321@hotmail.com</p> -->
            </footer>
            <script>
            var slideIndex = 0;
            showSlides();
            function showSlides() {
              var i;
              var slides = document.getElementsByClassName("mySlides");
              var dots = document.getElementsByClassName("dot");
              for (i = 0; i < slides.length; i++) {
                slides[i].style.display = "none";
              }
              slideIndex++;
              if (slideIndex > slides.length) {slideIndex = 1}
              for (i = 0; i < dots.length; i++) {
                dots[i].className = dots[i].className.replace(" active", "");
              }
              slides[slideIndex-1].style.display = "block";
              dots[slideIndex-1].className += " active";
              setTimeout(showSlides, 5000); // Change image every 2 seconds
            }
            </script>
            <script type="text/javascript"
            src="http://ajax.googleapis.com/ajax/libs/jquery/1.8.3/jquery.min.js"></script>
            <script type="text/javascript">
                // $(".reset").bind("click", function() {
                //         $("input[type=text], textarea").val("");
                // });
            </script>
            </body>
            </html>
```

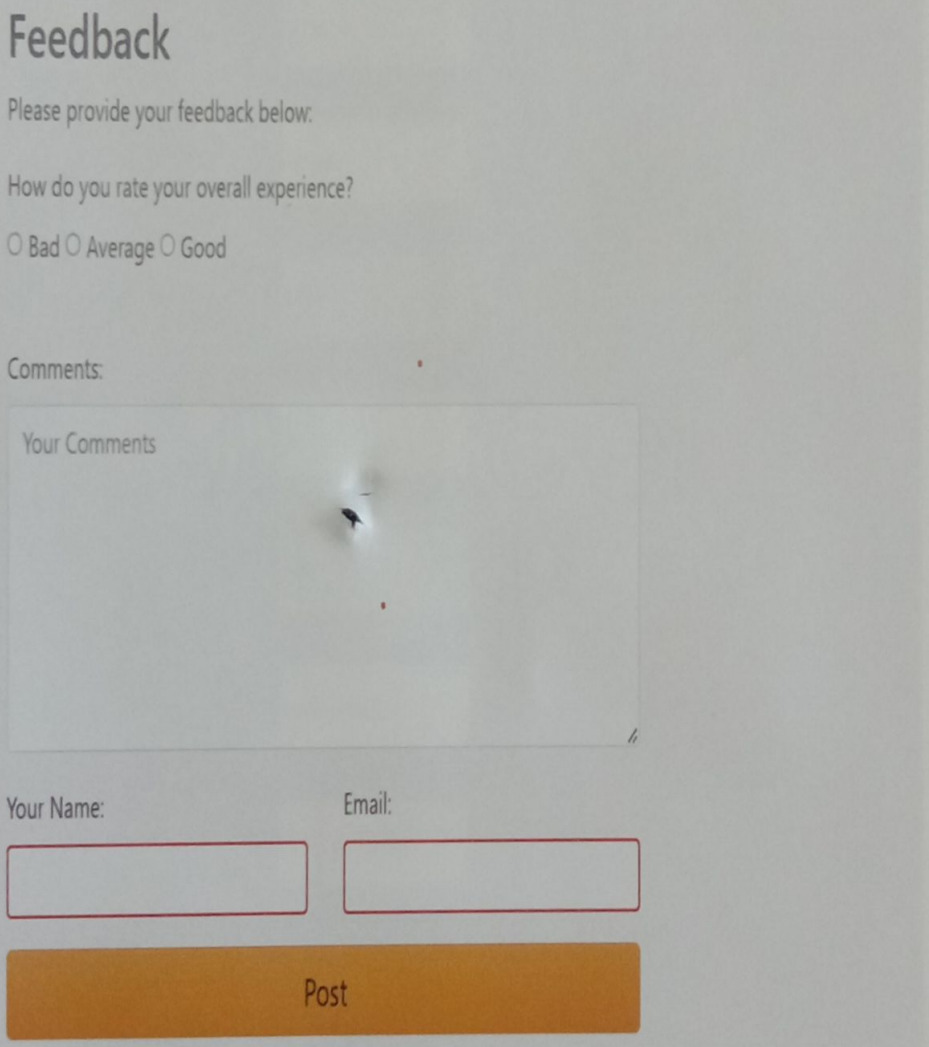
SCREENSHOTS

SCREENSHOTS

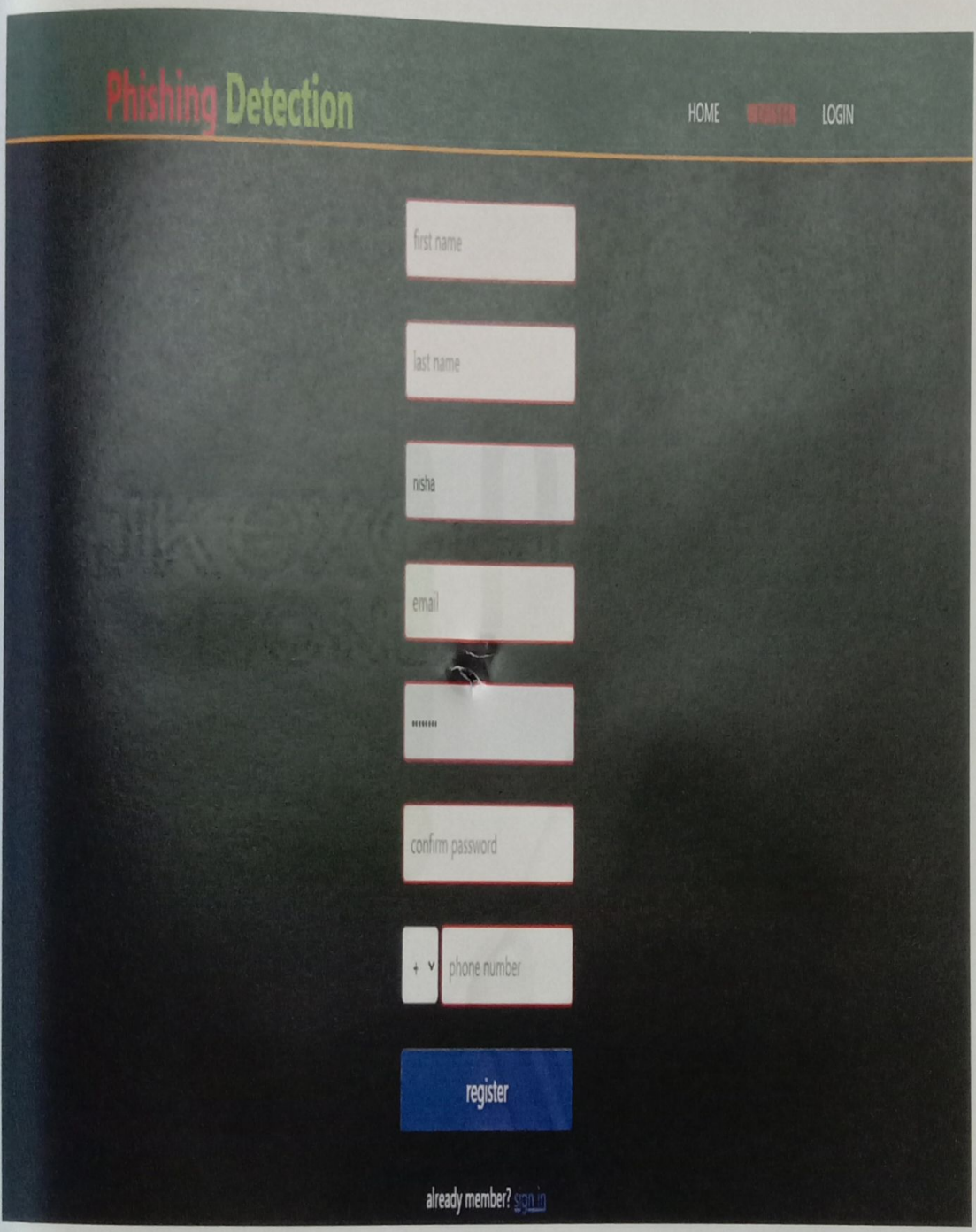
HOME PAGE



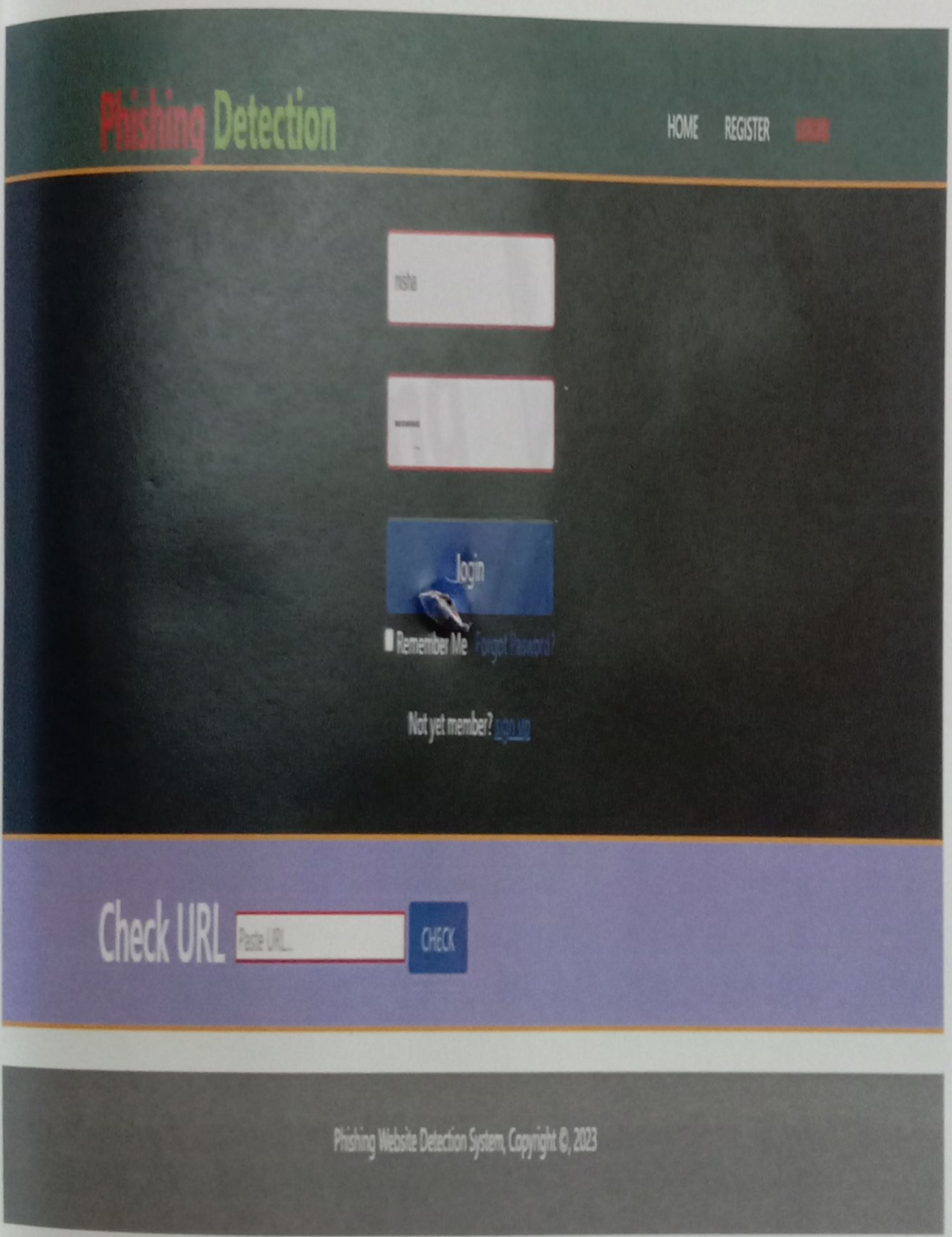
FEEBACK PAGE

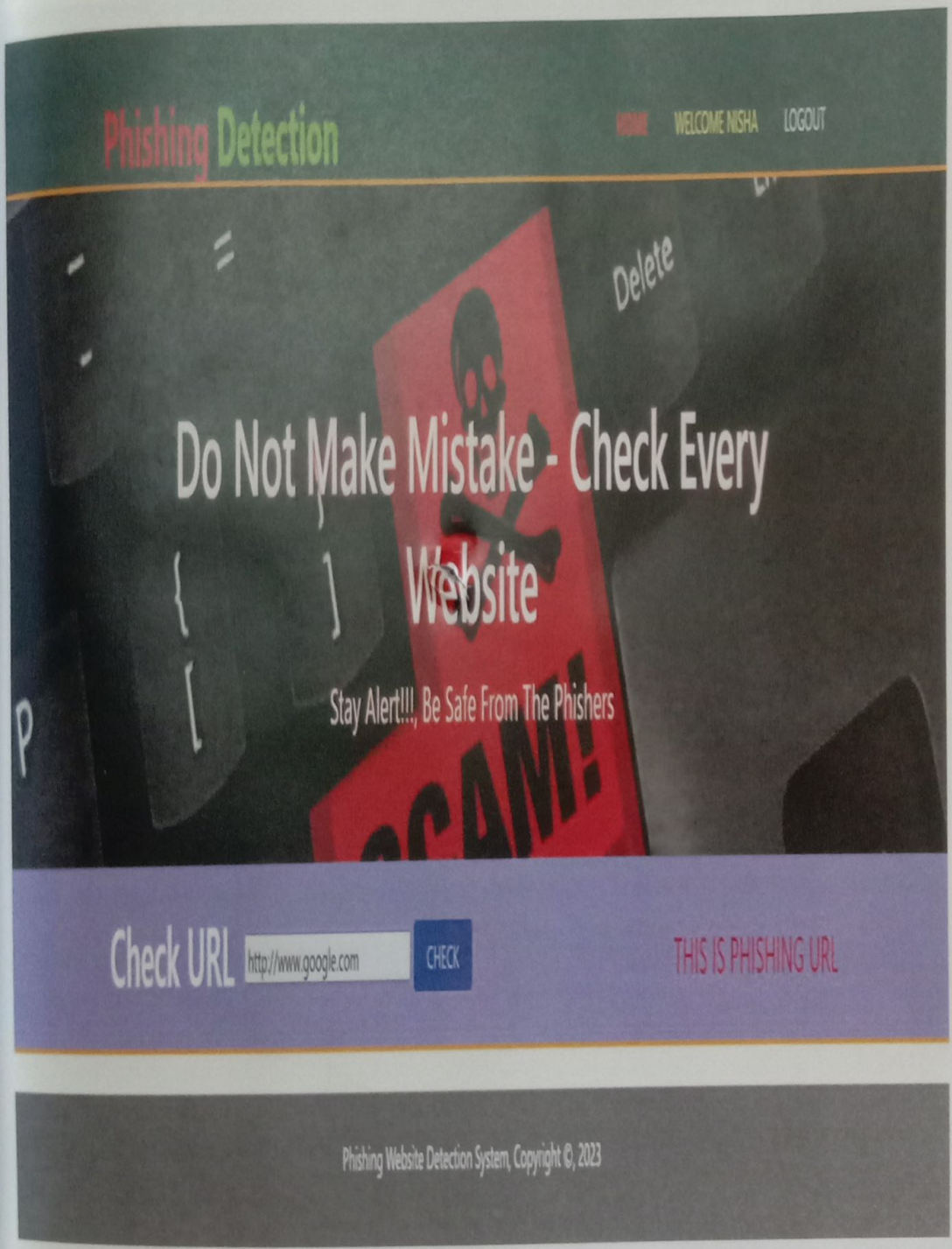


REGISTER PAGE

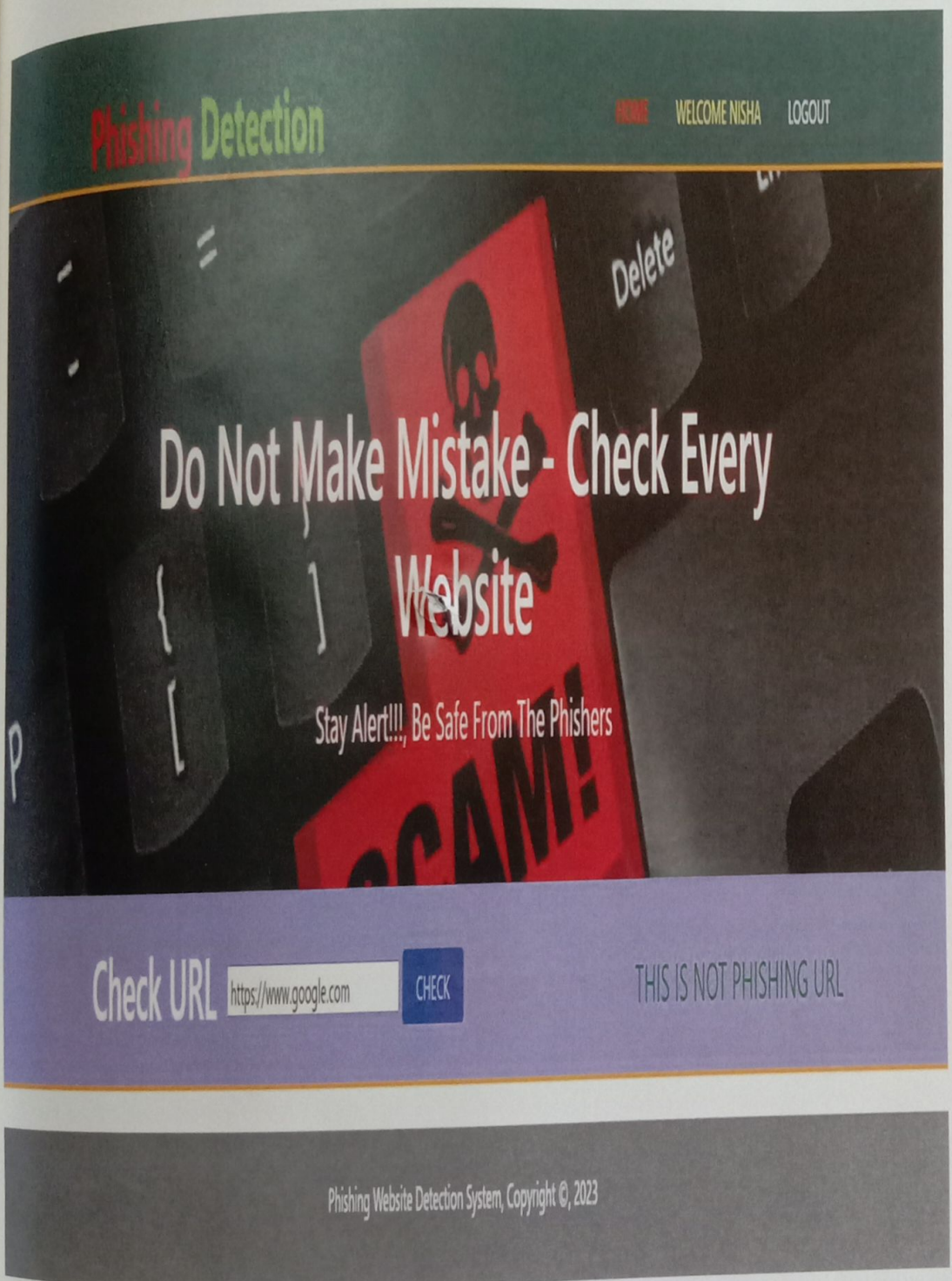


LOGIN PAGE


Phishing Detection
HOME
REGISTER
login
Remember Me
Not yet member?
Check URL
Paste URL...
CHECK
Phishing Website Detection System, Copyright ©, 2023


Phishing Detection
WELCOME NISHA
LOGOUT
Delete
Do Not Make Mistake - Check Every Website
Stay Alert!!!, Be Safe From The Phishers
SCAM!
Check URL
http://www.google.com
CHECK
THIS IS PHISHING URL
Phishing Website Detection System, Copyright ©, 2023

LEGIMATE WEBSITE



CONCLUSION

CONCLUSION

A Neuro-Fuzzy approach for anti-phishing on Fog Network using PHP can be used to provide strong security against malicious attacks. This technology uses a combination of artificial neural networks and fuzzy logic to identify potentially malicious activities and alert the user or system. The neural network can be used to classify incoming data, detect patterns, and identify suspicious activities. The fuzzy logic can be used to create rules that specify how the system should respond to different inputs. The use of this technology can help to reduce false positives and false negatives, allowing for more accurate detection and prevention of malicious activities. PHP can be used to create an interface for the Neuro-Fuzzy system, as well as to provide an automated response to malicious activities. This system can be used to create a secure and reliable fog network, allowing for secure communication and data transfer.

FUTURE ENHANCEMENT

FUTURE ENHANCEMENT

- Develop a system architecture based on Neuro-Fuzzy approach to detect malicious phishing activities on Fog Networks.
- Design a fuzzy logic system with input variables such as IP addresses, network traffic flow, suspicious URLs, user behavior, etc.
- Develop a real-time detection system to detect suspicious activities on the Fog Network.
- Develop a robust system to block or blacklist the malicious URLs and IP addresses.
- Design an alert system to notify the users of the malicious activities.
- Implement proper security measures to protect sensitive data from malicious activities.
- Design an interface in PHP to provide user-friendly access to the system.
- Test the system for accuracy and performance.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Hussain, A., & Kaur, H. (2020). A Neuro-Fuzzy Approach for Anti-Phishing in Fog Network. IEEE Access, 8, 39785-39794.

2. Hamed, M. A., & Dutta, M. (2018). A fuzzy based approach for anti-phishing in fog networks. In 2018 IEEE 15th Annual Consumer Communications & Networking Conference (CCNC) (pp. 537-541). IEEE.

3. Wang, W., & Jiang, D. (2017). Research on anti-phishing mechanism based on fuzzy cognitive map. In 2017 16th International Conference on Fuzzy Systems and Knowledge Discovery (pp. 738-742). IEEE.

4. Li, X., Li, Y., Li, Y., & Zhu, H. (2016). Research on anti-phishing system based on fuzzy cognitive map. In 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1686-1690). IEEE.

5. Zeng, C., & Yang, P. (2016). An intelligent anti-phishing system based on fuzzy cognitive map. International Journal of Automation and Computing, 13(2), 169-175.

6. Durga Prasad, S. S., & Singh, S. (2015). A fuzzified cognitive map based anti-phishing system. In 2015 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT) (pp. 1-4). IEEE.

7. Chen, C., & Chen, H. (2014). A fuzzy-based anti-phishing email detection system. In 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-5). IEEE.

8. Mokhtar, M., & El-Baz, A. (2013). A fuzzy-based anti-phishing system. International Journal of Computer Science and Security, 7(2), 158-170.

WEBSITE LINKS

- https://towardsdatascience.com/whataphish-detecting-phishing-websites-e5e1f14ef1a9
- https://nevonprojects.com
- https://github.com/S

IDENTIFICATION OF PCB DEFECT USING COMPUTER VISION

A project submitted to

ST. MARY'S COLLEGE (Autonomous), THOOTHUKUDI

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY,
TIRUNELVELI

In partial fulfilment of the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Submitted by

B. MAHALAKSHMI

Reg.No.: 21SPCS03

Under the Supervision and Guidance of

Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D.,

PG DEPARTMENT OF COMPUTER SCIENCE (SSC)

St. Mary's College (Autonomous), Thoothukudi - 628001

APRIL 2023

CERTIFICATE

This is to certify that this project work entitled **"IDENTIFICATION OF PCB DEFECT USING COMPUTER VISION"** is submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli, in partial fulfilment for the award of the degree of Master of Science in Computer Science for the work done during the year 2022-2023 by **B. MAHALAKSHMI** (Reg. No.: 21SPCS03).

Signature of the Guide

Signature of the Co-ordinator

Signature of the Director

Director
Self Supporting Courses
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Principal

Principal
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Examiner

DECLARATION

I do hereby declare that the project entitled **"IDENTIFICATION OF PCB DEFECT USING COMPUTER VISION"** submitted for the degree of Master of Science in Computer Science is my original work carried out under the guidance of **Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D.,** Assistant Professor, PG Department of Computer Science (SSC), St. Mary's College(Autonomous), Thoothukudi.

B. Mahalakshmi

Place: Thoothukudi **Signature of the student**

Date: 10-04-2023

ACKNOWLEDGEMENT

I express my first and foremost thanks to God Almighty for his gracious help and shower of blessings for having rendered us the strength and support to finish our project successfully.

My sincere thanks to Dr. Sr. A. S. J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., Principal, Dr. Sr. C. Shibana M.Sc., B.Ed., M.Phil., Ph.D., Secretary, Rev. Sr. S. Josephine Jeyarani, Director of SSC, St. Mary's College (Autonomous), Thoothukudi, for giving permission to work on this project.

I convey my heartfelt thanks to Ms. C. Nayanthra Mascarenhas M.Sc., M.Phil., SET., Assistant Professor and Coordinator, PG Computer Science (SSC), for her support and counsel.

I express my hearty thanks to my guide Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D., Assistant Professor, PG Department of Computer Science (SSC) for her valuable suggestions, gentle guidance, enthusiastic ideas, to carry out andcomplete my work entirely effortlessly.

I also express my boundless thanks to Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., M.Phil., B.Ed., Assistant professor, PG Department of Computer Science (SSC), for her encouragement and support.

I am much indebted to Dr. P. Johnson Durai Raj, Director, Postulate Info tech for his untiring effort, Immense Knowledge and priceless contribution without which I couldn't have finished my work effectively on time.

I thank my family members especially my parents for their encouragement and support both morally and financially which helped me to finish the project successfully.

POSTULATE

01.04.2023

To

The Head of the Department
PG Department of Computer Science (SSC)
St. Mary's College(Autonomous)
Thoothukudi

Dear Sir/Madam,

Sub: Project Work Completion-Reg

On behalf of **POSTULATE**, we are pleasure to inform you that **Ms.B.Mahalakshmi, Reg no: 21SPCS03** studying **Master of Computer Science** Final year has been done the project work at our concern Postulate Info Tech Private Limited in the title **"Identification of PCB Defect Using Computer Vision"** during the period from December 2022 to March 2023.

During this period, the candidate was found diligent and sincere. We wish all the best for future endeavors.

POSTULATE INFO TECH PRIVATE LIMITED

DIRECTOR

ABSTRACT

ABSTRACT

A printed circuit board is used to connect different electronic components mounted on it using pathway or tracks which is etched from copper sheet. In this project different approaches have been implemented on template images and the input test image to detect the defects on the PCB board. After that, compare the template image and input test image by using structural similarity index-based template matching technique. Thus, the common defects are pin hole, mouse bite, open circuit, short, spur and the spurious copper. After that localize the defect and create a boundary box for that defect. To get the details about the defects, extracting structural features based on the region properties such as area, width, height, and the shape features can be found out. Finally, the extracted feature is given to the BPN classifier is used to classify the defect and identify the PCB defects in the image.

TABLE OF CONTENTS

INTRODUCTION

1. INTRODUCTION

Computer vision is the field of artificial intelligence that enable computers and system to drive meaningful information from digital images, videos and other visual inputs and take action or make recommendation based on that information. The process involves image acquiring, screening, analyzing, identifying and extracting information. This extensive processing helps computer to understand any visual content and act on it accordingly. Printed circuit boards (PCBs) are the boards that are used as the base in most electronics – both as a physical support piece and as the wiring area for the surface-mounted and socketed components. PCBs are most commonly made out of fiberglass, composite epoxy, or another composite material.

This project aims to develop an automated system for the identification of defects in printed circuit boards (PCBs). The system will utilize an image processing and machine vision system to detect and classify the type of defect, and to identify the type of corrective action required to repair the defect. The system will be able to identify and classify defects such as cracks, delamination's, missing or misaligned components, solder joint fractures, and other types of defects. The system will also be able to distinguish between defects that require manual repair and those that can be fixed automatically. The system will be able to provide real-time feedback to production personnel, allowing for the prompt and accurate identification of defects and the implementation of corrective action.

The dataset consists of the image with 6 type of defects which are missing hole, mouse bite, open circuit, spur, short, spurious copper. The main aim is to detect the type of defects in the input PCB image. In other words, the PCB image is taken as input, and find structural similarity index between standard template image in order to identify the defected region, After, extraction of the defected region, extracted the feature descriptors from that defected region and these descriptors are given for training and testing process and for PCB defect type identification BPN model is used. Each defected PCB image has involved into SSIM operation with the template PCB board image.

Template matching is a technique for finding the most similar to the image patch in the source image, and it is often used for target recognition, tracking and detection. The most similarity must be based on a certain similarity criterion, that is, similarity measurement is required. The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between template PCB image and input test PCB image.

HOW DOES COMPUTER VISION WORK:

Computer Vision primarily relies on pattern recognition techniques to self-train and understand visual data. The wide availability of data and the willingness of companies to share them has made it possible for deep learning experts to use this data to make the process more accurate and fast.

While machine learning algorithms were previously used for computer vision applications, now deep learning methods have evolved as a better solution for this domain. For instance, machine learning techniques require a humongous amount of data and active human monitoring in the initial phase monitoring to ensure that the results are as accurate as possible. Deep learning on the other hand, relies on neural networks, and uses examples for problem solving. It self-learns by using labeled data to recognize common patterns in the examples

ORIGIN OF COMPUTER VISION:

Computer vision is not a new concept; in fact, it dates back to the 1960s. It started with an MIT project "Summer Vision Project" which analyzed scenes to identify objects. David Marr, the celebrated neuroscientist, laid down the building blocks of computer vision, taking a cue from the functions of the cerebellum, hippocampus, and cortex of human perception. He has been dubbed the father of computer vision since, and the field has evolved to include much more complicated functionalities.

APPLICATIONS OF COMPUTER VISION:

- **COMPUTER VISION FOR DEFECT DETECTION:**

This is, perhaps, the most common application of computer vision. Until now the detection of defects is carried out by trained people in selected batches, and total production control is usually not possible. With computer vision we can detect defects such as cracks in metals, paint defects, bad prints etc. in sizes smaller than 0.05mm. Much better than the human eye! These vision cameras need an algorithm that is the "intelligent brain" which is able to differentiate what is a defect and what is not. This algorithm is designed and trained specifically for each particular application through images.

- **COMPUTER VISION FOR METROLOGY:**

It is another of the most important applications. What has been done so far with complex laser metrology equipment or probes, now it can be measured using computer vision. The key on this topic is to make a good adjustment of the reference to be able to measure with the necessary precision, and above all, use the appropriate lighting for each type of material and work environment. Using artificial vision systems we can measure variable part sizes, straightness and parallelism.

- **COMPUTER VISION FOR INTRUDER DETECTION:**

Through hyper spectral cameras it is possible to differentiate between a fruit and a stone, which allows, especially in food, safer products for the consumer. Hyper spectral cameras are able to differentiate the type of material through the measurement they make of the wavelength. In this way, it is possible to differentiate a stone from a fruit, a plastic from a metal or other combinations while the material is different.

- **COMPUTER VISION FOR ASSEMBLY VERIFICATION:**

Every day more and more complex assemblies are made, with more parts or connections. Computer vision allows us to verify, step by step, that each piece is in its place, or at the end of the process, that the final assembly is correct. This application is very useful for the assembly of machinery, equipment, electronic boards or pre-assemblies with a lot of complexity. These systems significantly reduce cycle times of very complex operations and reoperation times.

- **COMPUTER VISION FOR SCREEN READER:**

Sometimes it is not possible to extract data from a display screen either because it is a closed supplier system or because that system is incompatible with the one installed. A solution to this problem is to install a computer vision camera to read the screen and extract the data that appears on it (temperatures, codes, tensions... any useful information that appears on the screen and you need it). To do this, we look for the interest regions in which the information is located, we use a character recognition algorithm (OCR) to extract it, and everything perfect!

- **COMPUTER VISION FOR CODE AND CHARACTER READER (OCR):**

Let's be honest, the designers are very nice people, but they tend to change typography to more complex letters (and prètty!) frequently. We go back to the character recognition algorithms to have a computer vision recognition system trained, so even you have a complex typography, the OCR will be able to read them. It is a system so robust capable of reading even handwritten letters. Proof of best designers!

- **COMPUTER VISION + ROBOTICS FOR BIN PICKING:**

And finally, one of the applications that we are most requested, combining collaborative robotics with computer vision to be able to perform bin picking of pieces placed disorderly. The pieces are messy and therefore we need to optimize the trajectories and detect the grip coordinates. The robot needs help (the computer vision) telling him what a piece is and where it is, so it can decide what is the best way to pick up.

PRINTED CIRCUIT BOARD

Printed circuit boards (PCBs) are the boards that are used as the base in most

electronics – both as a physical support piece and as the wiring area for the surface-mounted and soldered components. PCBs are most commonly made out of fiberglass, composite epoxy, or another composite material.

DEFECTED REGION LOCALIZATION:

In the localization step the identified defected regions are drawing with a bounding box around their extent. After localize the defects in the image, next the structure feature extraction can be done. Shape feature descriptors are used for extracting the features from localized defected image region. The final step is to classify defect by using BPN classifier. The extracted shape features are fed into the BPN model in order to train the network weights. Finally, the trained class detects the PCB defects name that may be a missing hole, mouse bite, open circuit, short, spur, spurious copper.

Python

Python is a high level interpreted language used for general purpose programming. It is widely used for scientific computing and can be used for a wide variety of general tasks from data mining to software development. Python is the main language used for this project.

This allows for quick prototyping and debugging, and makes it easier to use for non-programmers. It has a wide range of libraries and frameworks for developing applications, from web development to data science. Python is an object-oriented language, meaning that it is composed of objects and classes. This makes it easier to develop complex applications, since objects can be reused and modified. The language features an easy-to-learn syntax and a large standard library, which is the collection of functions and modules that come with the language. Python is also used for scripting and automation. This makes it possible to automate repetitive tasks, such as collecting data from websites and running simulations. The language also provides powerful tools for data analysis and machine learning. Python is often used to develop artificial intelligence and deep learning applications. Python is an open source language, meaning that anyone can contribute to it and all of its source code is freely available. This makes it easy to learn and use, and encourages collaboration and development. Python is also platform independent, meaning that applications written in Python can be used on any operating system. Python is an excellent choice for developing applications and automating tasks. It is easy to learn and use, and provides a wide range of libraries and frameworks for developing applications. It is also an open source language, meaning that anyone can contribute to its development. Finally, it is platform independent, making it possible to use applications written in Python on any operating system. Python is a high level interpreted language used for general purpose programming. It is widely

used for scientific computing and can be used for a wide variety of general tasks from data mining to software development. Python is the main language used for this project.

Numpy

NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open source extension module for Python, which provides fast precompiled functions for mathematical and numerical routines. Furthermore, NumPy enriches the programming language Python with powerful data structures for efficient computation of multi-dimensional arrays and matrices. The implementation is even aiming at huge matrices and arrays. Besides that the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays. It is the fundamental package for scientific computing with Python.

It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier Transform, and random number capabilities.

Matplotlib

Matplotlib is a Python plotting library that allows programmers to create a wide variety of graphs and visualizations with ease of use. The great feature about Matplotlib is that it integrates very well with Jupyter Notebook and creating visualizations is simplified. Matplotlib also works very well with pandas and numpy. Matplotlib's high-level plotting API provides a convenient way to create and customize plots. For example, the pyplot module allows users to quickly create line graphs, histograms, and scatter plots. The library is highly customizable, with numerous settings and parameters that can be adjusted to create a customized plot.

For example, users can set the figure size, add labels and annotations, set the font size and style, and adjust the colors used in the plot. Matplotlib also supports a variety of backends, allowing the user to select which backend to use in order to generate a plot. Matplotlib is used in a wide variety of applications, ranging from data analysis to scientific visualization. It is often used in conjunction with other libraries such as NumPy, SciPy, and pandas to create powerful visualizations. It is also commonly used in machine learning applications to visualize the accuracy of a model's predictions. Overall, Matplotlib is an incredibly powerful and versatile plotting library for Python. It is easy to use and highly customizable, making it a great tool for creating professional plots and charts. Matplotlib is a Python plotting library that allows programmers to create a wide variety of graphs and visualizations with ease of use. The great feature about Matplotlib is that it integrates very well with Jupyter Notebook and creating visualizations is simplified. Matplotlib also works very well

with pandas and numpy.

OpenCV

OpenCV (Open Source Computer Vision Library) is an open source computer visio n and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 14 million. The library is used extensively in companies, research groups and by governmental bodies. It has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available.

A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. 7 OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

Tensor Flow

TensorFlow is an open-source deep learning library by Google. It was originally developed by Google's engineers who were working on Google Brain and has been used for research on machine learning and deep learning. TensorFlow at it is core is about computations of multidimensional arrays called tensors but what makes TensorFlow great is its ability to be flexible to deploy computations on different devices such as CPU's and GPU's.

Keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Keras contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with

image and text data easier. The code is hosted on Git Hub, and community support forums include the Git Hub issues page, and a Slack channel. Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of distributed training of deep learning models on clusters of Graphics Processing Units (GPU).

Tkinter

Tkinter is a python based graphical user interface (GUI) library. It is the most commonly used library for the development of GUI applications. Tkinter is a standard package in python and is very easy to learn. It provides a powerful object-oriented interface to the Tk GUI toolkit. Tkinter is used to create a variety of GUI applications such as games, dialog boxes, and more.

Tkinter provides a wide range of widgets. These widgets can be used to create graphical user interfaces (GUIs) for applications. Widgets are the basic building blocks of a GUI application. They provide the user with a way to interact with the application. The most commonly used widgets are buttons, labels, entry fields, radio buttons, check boxes, scrollbars, and spin boxes.

- Tkinter also provides a selection of geometry managers which are used to arrange widgets in a window.
- These geometry managers can be used to create different layouts for your application.
- The most commonly used geometry managers are pack, grid, and place.
- Tkinter also provides a variety of events which can be used to create interactive applications.
- Events are used to detect user interaction with widgets.
- For example, a button click event is used to detect when a user clicks on a button.
- Tkinter also provides a set of tools for creating graphical images.
- These tools can be used to create simple images or complex animations.

Objectives

1. To develop a cost-effective and reliable solution for automated identification of PCB defects.
2. To identify the various types of defects that can occur in PCBs and develop a detection algorithm for each of them.
3. To create a model that is capable of accurately identifying PCB defects from images.
4. To develop a user-friendly interface for the detection system.
5. To validate the accuracy of the detection system with a real-world dataset.
6. To optimize the system for maximum speed and accuracy.
7. To ensure that the system is reliable and robust enough to be used in a manufacturing setting.

SYSTEM SPECIFICATION

2. SYSTEM SPECIFICATION

HARDWARE REQUIREMENTS:

- Processor : 11th Gen Intel(R) Core(TM) i3-1115G4 @ 3.00GHz 3.00 GHz
- RAM : 8 GB
- Hard Disk : 40GB
- System Type : 64-bit Operating System

SOFTWARE REQUIREMENTS:

- Operating system : Windows 10 pro
- Front-end : Python, tkinter
- Back-end : TensorFlow

PROJECT DESCRIPTION

3. PROJECT DESCRIPTION

This project aims to develop a computer vision-based system for the automated identification of Printed Circuit Board (PCB) defects. The system will be able to detect and classify various defects in a PCB such as pin hole, mouse bite, open circuit, short, spur and the spurious copper. The specialized software will be used to detect and classify the defects. The system will then output the defect information to the user. The ultimate goal of this project is to provide a fast and accurate way to identify and classify PCB defects, thereby improving the quality and reliability of PCBs.

The PCB Defect Identification Project is a research and development project aimed at developing a reliable and efficient automated system to detect and identify defects in printed circuit boards (PCBs). The system is expected to reduce the time, cost and complexity of manual inspection of PCBs, and enable the production of reliable, high-quality products. The project will involve designing, developing and testing the system, which is expected to include software algorithms to detect and identify defects, a computer vision system and an embedded system for controlling the system. The system will be tested in a laboratory environment and then be deployed on production lines for actual use. The project will involve collaboration with industrial partners to ensure the system meets their requirements. The project team will also be responsible for providing technical support and training to the users of the system. The objective of the project is to develop and deploy a reliable, automated system to detect and identify defects in printed circuit boards. The system is expected to reduce the costs and complexity of manual inspection, and enable the production of high-quality products.

The 6 types of PCB defects used in this project are missing hole, mouse bite, open circuit, short, spur, spurious copper.

1. **MISSING HOLE:**

A missing hole defect is one where the solder mask is undefined or the setting for a longer board are carried over to a smaller board.

2. **MOUSE BITE:**

They are used when the board is going to be panelized, routed, and populated with a pick and place machine. There is no standardized symbol for mouse bites. In fact, mouse bites may be directly added to the PCB using hole/via primitives and not feature on the schematics at all

3. **OPEN CIRCUIT:**

Depending on your source, open circuits constitute about a third of PCB defects, especially in the form of open solder joints. A number of issues can cause open circuits on your board, varying

from materials to processing to handling.

4. SHORT:

Shorting occurs when the solder does not separate from two or more leads before the solder solidifies. Increasing the flux solids or quantity is one way of decreasing shorting. A reduction of the lead length and the pad size will reduce the amount of solder being held on the base of the board. Alternative pads were increased in length on the exit side of the wave. This made the actual separation distance between the adjacent terminations larger and decreased the shorting.

5. SPUR:

Spur defect is one in which it has a small spike in a circuit board. It is caused when two conductive paths on the board are connected inadvertently, either through a short circuit or a solder bridge. The conductive paths can be traces, vias, or other components. This defect can cause a range of problems, from signal interference to complete circuit failure. It can also result in increased power consumption and increased heat generation.

6. SPURIOUS COPPER:

Spurious copper is one in which it has a line in between the board in the PCB defected. This could be due to excessive heat, insufficient solder, or poor soldering technique. It could also be caused by a defect in the copper trace itself, such as an open circuit or a break. In addition, it could be caused by improper cleaning of the board before soldering or corrosion of the copper.

SYSTEM STUDY

4. SYSTEM STUDY

EXISTING SYSTEM

The existing system for identification of PCB defect project is based on automated visual inspection (AVI) techniques. AVI is a computer-controlled process that uses a camera and specialized software to detect defects or anomalies in printed circuit boards (PCBs). AVI inspects the entire board for defects such as lifted or missing components, poor solder joints, incorrect part placement, and incorrect solder paste deposition. The system can also detect cracks in the solder mask, which can lead to circuit failure. AVI is used to quickly identify and classify defects, so that they can be fixed in the production line. In addition, AVI can provide real-time feedback to the production team, allowing them to make adjustments to their process to reduce the likelihood of further defects.

PROPOSED SYSTEM

The proposed consists of 5 modules:

- Dataset description
- Structural Similarity Based template matching
- Defect localization
- Structural feature extraction
- BPN classifier

1. DATASET DESCRIPTION:

We used PCB Defect dataset from kaggle online community for data scientist and machine learning practitioners. This dataset contains total of 1386 images with 6 kinds of defects such as missing hole, mouse bite, open circuit, short, spur and spurious copper. This dataset includes 115 images of each PCB defect type. This is a public synthetic PCB defect dataset which is used for detection, classification and registration tasks.

2. STRUCTURAL SIMILARITY INDEX BASED TEMPLATE MATCHING:

Template matching is a technique for finding the most similar to the image patch in the source image, and it is often used for target recognition, tracking and detection. The most similarity must be based on a certain similarity criterion, that is, similarity measurement is required. The structural similarity index measure (SSIM) is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos. SSIM is used for measuring the similarity between two images. The SSIM index is a full reference metric; in

words, the measurement or prediction of image quality is based on an initial uncompressed distortion-free image as reference.

DEFECTED REGION LOCALIZATION:

Object localization refers to identifying the location of one or more objects in an image and drawing abounding box around their extent. Object detection combines these two tasks and localizes and classifies one or more objects in an image.

- Object Localization: Locate the presence of objects in an image and indicate their location with a bounding box.
- Input: An image with one or more PCB defect.
- Output: One or more bounding boxes (e.g. defined by a point, width, and height).

SINGLE OBJECT LOCALIZATION:

- Image classification: Algorithms produce a list of PCB defect categories present in the image.
- Single-object localization: Algorithms produce a list of PCB defect categories present in the image, along with an axis-aligned bounding box indicating the position and scale of one instance of each PCB defect category.
- Object detection: Algorithms produce a list of PCB defect categories present in the image along with an axis-aligned bounding box indicating the position and scale of every instance of each PCB defect category.

STRUCTURAL FEATURE EXTRACTION:

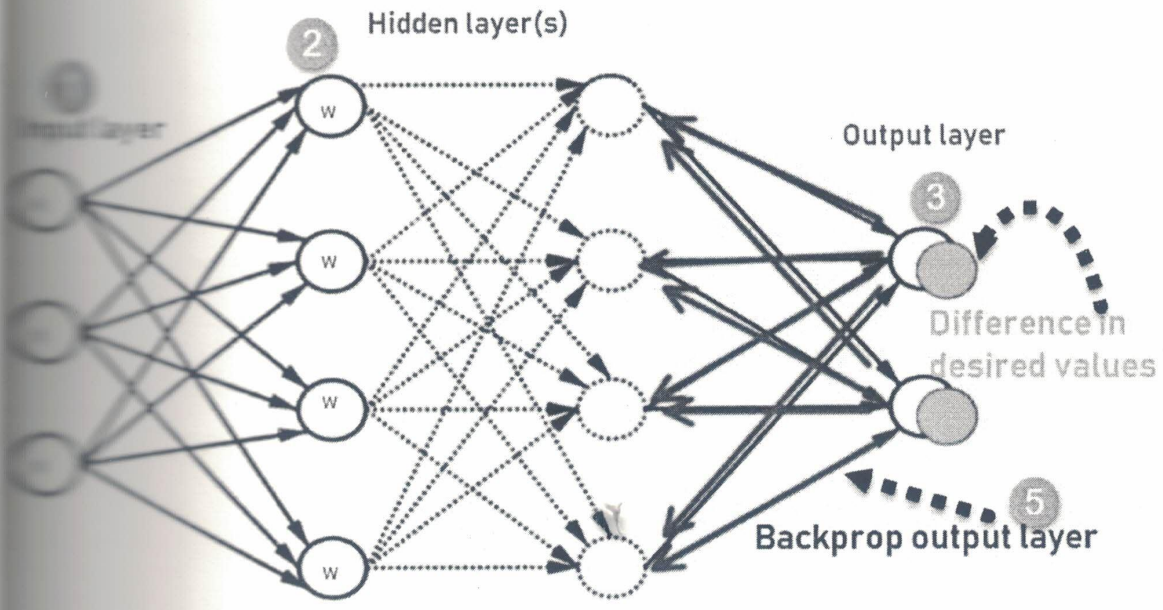
Feature extraction can be considered as finding a set of features that define the shape of the underlying character as precisely and uniquely as possible. The term feature selection refers to methods that select the best subset of the input feature set. There are two feature extraction methods Structural method and Statistical method. In this work we used Structural Method. Structural method identifies structural features of a character. Structural features are based on topological and geometric properties of the character. The structural features used in this work are area, width, height and the shape feature. The shape features are convex hull, eccentricity, perimeter Crompton, equivalent diameter, perimeter, filled area, solidity etc. This type of representation may also encode some knowledge about the structure of the object or may provide some knowledge as to what sort of components make up that object

BPN TRAINING:

Back-propagation is the essence of neural net training. It is the practice of fine-tuning the weights

of a neural net based on the error rate (i.e. loss) obtained in the previous epoch (i.e. iteration). Proper tuning of the weights ensures lower error rates, making the model reliable by increasing its generalization. A neural network is a group of connected it I/O units where each connection has a weight associated with its computer programs. Back propagation is a short form for "backward propagation of errors." It is a standard method of training artificial neural networks.

Back propagation algorithm in machine learning is fast, simple and easy to program. A feedforward BPN network is an artificial neural network. In this network, it has all its extracted features on the input layer, hidden layer and output layer holds all the types of PCB defect and shows the output as identified PCB defect.



BPN Classifier

PCB DEFECT TYPE IDENTIFICATION:

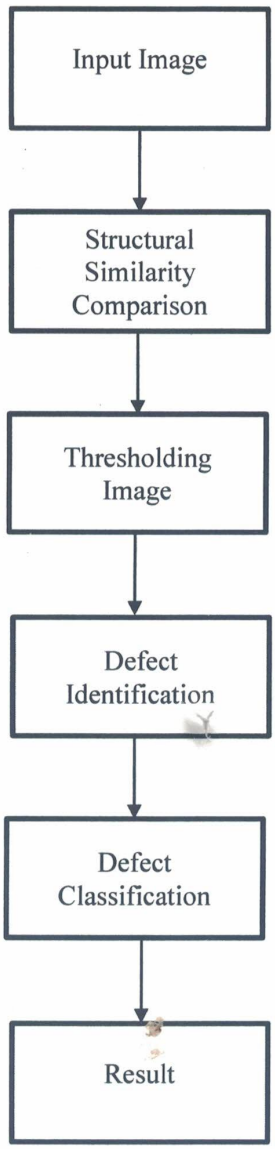
While training a model, we generated the template for all kinds of the PCB defect form the PCB defect dataset and they are load into the model while testing. Then, the model is fed with the input images of PCB for testing. Then, the model finds the similarity between two different images and it measures or predicts the image quality using the structural similarity index based on the template matching process. Then, it draws the bounding boxes over the defected location of the PCB. Then, it extracts the structural features like plating voids, insufficient copper to edge clearance, bad soldering, slivers, missing solder mask between pads, acid traps, poorly manufactured components, starved thermals, pin holes, line spacing, line width, specs, voids, etc.

If there are problems in feature extraction, then it will be back propagated through the network. And again, extracts the features and strengthen the classifier. Finally, the model detects the PCB defect and its kind as missing hole, mouse bite, open circuit, short, spur and spurious copper.

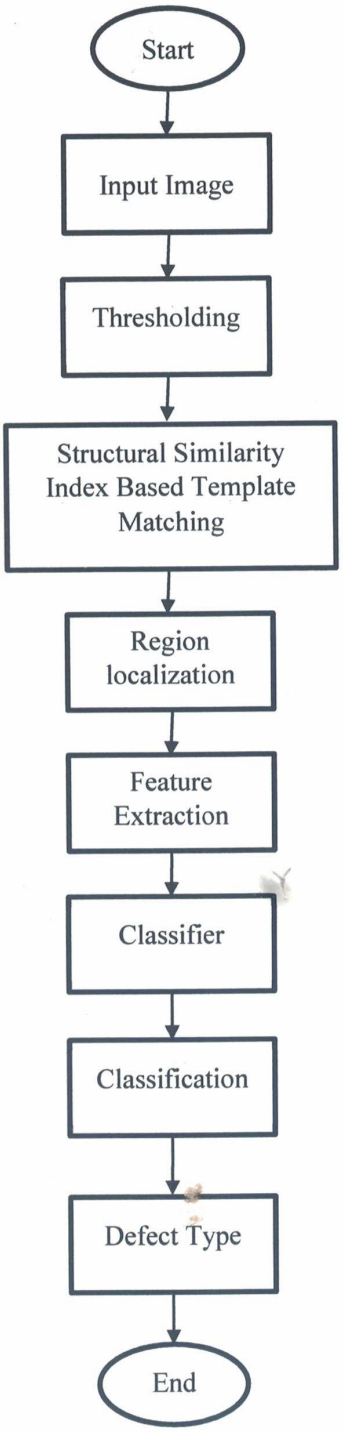
SYSTEM DESIGN

5. SYSTEM DESIGN

FLOW CHART



DATA FLOW DIAGRAM



SYSTEM TESTING

6. SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTING

REGRESSION TESTING

Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features. Regression Testing is nothing but a full or partial selection of already executed test cases that are re-executed to ensure existing functionalities work fine. This testing is done to ensure that new code changes do not have side effects on the existing functionalities. It ensures that the old code still works once the latest code changes are done.

LOAD TESTING

Load testing is a type of performance testing that simulates a real-world load on any software, application, or website. Without it, your application could fail miserably in real-world conditions. That's why we build tools like retrace to help you monitor application performance and fix bugs before your code ever gets to production. Load testing examines how the system behaves during normal and high loads and determines if a system, piece of software, or computing device can handle high loads given a high demand of end-users. This tool is typically applied when a software development project nears completion.

FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items: ¬

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.

- Systems Procedures: interfacing systems or procedures must be invoked.

[illegible] and preparation of functional tests is focused on requirements, key functions, or [illegible] cases. In addition, systematic coverage pertaining to identify Business process flows; [illegible] predefined processes, and successive processes must be considered for testing.

[illegible]ERY TESTING

[illegible] Testing is software testing technique which verifies software's ability to recover from [illegible] like software/hardware crashes, network failures etc. The purpose of Recovery Testing is [illegible] whether software operations can be continued after disaster or integrity loss. Recovery [illegible] involves reverting back software to the point where integrity was known and reprocessing [illegible] to the failure point.

[illegible]GRATION TESTING

[illegible] related to migration is also reviewed by the testing team who is performing migration [illegible]. Testing team verifies if the document is clear and there is no complexity to perform the [illegible] related to migration. Any kind of documentation error whether it is related to the [illegible], database, server and operating system, can be reported by the testing team. Time taken [illegible] the system completely known as downtime is recorded properly and it is very useful [illegible] the application is launched in production. Once the migration activities are completed as per [illegible] documentation, a basic check of application functionality is done and basic database [illegible] are done to ensure that database is properly connected to the application.

CODING

7. CODING

```
import cv2
import numpy as np
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.pyplot as plt
from skimage import data
import easygui
from skimage.feature import match_template
import os, numpy, PIL
from PIL import Image
from tkinter import filedialog as fd
from skimage.metrics import structural_similarity as compare_ssim
import argparse
import imutils
from tkinter.messagebox import showinfo
import tkinter as tk
import tensorflow as tf
model = tf.keras.models.load_model('best_model.h5')
color=[(0,50,255),(250,0,0),(255,255,255),(0,0,0),(255,20,147),(154,205,50)]
color_text=[(0,50,255),(250,0,0),(255,255,255),(0,0,0),(255,20,147),(154,205,50)]
def classifyinput(img_rgb):
    #img_rgb = cv2.imread(img_rgb,0)
    img = cv2.cvtColor(img_rgb, cv2.COLOR_BGR2GRAY)
    #img=cv2.resize(img_rgb,(25,25))
    img=np.resize(img,(25,25))
    #print(img)
    #print(img.shape)
    img=img.reshape(1,25,25,1)
    ch=['missing_hole',
    'mouse_bite',
    'Open_circuit',
    'Short',
    'Spur',
```

```
'spurious_copper']
p1=model.predict(img)
p2=list(p1[0])
p3=max(p2)
p4=p2.index(p3)
p5=l[p4]
return p5,p4
=np.unravel_index(np.argmax(result), result.shape)
=i[:,:,-1]

    title='alert',
    message=f"ready to selct original image"

original_image_path=fd.askopenfilename()
= cv2.imread(original_image_path)
=.shape
=cv2.resize(imageA,(350,350))

    title='alert',
    message=f"ready to selct Second image image"

=fd.askopenfilename()
= cv2.imread(edited)
=cv2.resize(imageB,(350,350))
convert the images to grayscale
= cv2.cvtColor(imageA, cv2.COLOR_BGR2GRAY)
= cv2.cvtColor(imageB, cv2.COLOR_BGR2GRAY)
compute the Structural Similarity Index (SSIM) between the two
images, ensuring that the difference image is returned
, diff) = compare_ssim(grayA, grayB, full=True)
=(diff * 255).astype("uint8")
print("SSIM: {}".format(score))
threshold the difference image, followed by finding contours to
obtain the regions of the two input images that differ
```

```
thresh = cv2.threshold(diff, 0, 255,
    cv2.THRESH_BINARY_INV | cv2.THRESH_OTSU)[1]
cnts = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL,
    cv2.CHAIN_APPROX_SIMPLE)
cnts = imutils.grab_contours(cnts)
# loop over the contours
i=0
resultf=[]
for c in cnts:
    # compute the bounding box of the contour and then draw the
    # bounding box on both input images to represent where the two
    # images differ
    (x, y, w, h) = cv2.boundingRect(c)
    print("width",w)
    print("Height",h)
    imgc=imageB
    #cv2.rectangle(imageB, (x, y,z), (x + w, y + h), (0, 0, 255), 2)
    cropped_image = imageA[y:y+h, x:x+w]
    #if((w>25 and h>25):
    res,index=classifyinput(cropped_image)
    print(res)
    resultf.append(res)
    print("total no of defeats",i+1)
    font = cv2.FONT_HERSHEY_SIMPLEX
    cv2.rectangle(imgc, (x, y), (x + w, y + h), color[index], 2)
    cv2.putText(imgc, f'{res}', (x,y), font, 0.5,color_text[index] , 1, cv2.LINE_AA)
    cv2.imwrite(f"plot/{res}_{i}.jpg",imgc)
    print("n")
    i+=1
print(resultf)
rf=set(resultf)
print(rf)
# show the output images
cv2.imshow("original", imageA)
```

```
cv2.resize(imageB,(500,500))
cv2.imshow("modified", imageB)
cv2.imshow("Diff", diff)
cv2.imshow("Thresh", thresh)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

```
import numpy as np
import matplotlib.pyplot as plt
import numpy as np
import matplotlib.pyplot as plt
from skimage import data
import easygui
from skimage.feature import match_template
import os, numpy, PIL
from PIL import Image
from tkinter import filedialog as fd
from skimage.measure import compare_ssim
from skimage.metrics import structural_similarity as compare_ssim
!pip install scikit-learn==0.16.1
import argparse
import imutils
from skimage.measure import label,regionprops, regionprops_table
from skimage import data, filters, measure, morphology
from skimage import measure
import pandas as pd
from tkinter.messagebox import showinfo
import tkinter as tk
from tkinter import *
import tensorflow as tf
win = Tk()
win.withdraw()
model = tf.keras.models.load_model('best_model_bpn.h5')
def load_image(img):
    img=cv2.imread("output/3/00041009_5.jpg",0)
    threshold = filters.threshold_otsu(img)
    mask = img > threshold
    mask = morphology.remove_small_objects(mask, 50)
    mask = morphology.remove_small_holes(mask, 50)
```

```
= measure.label(mask)
_img = label(labels)
= regionprops(label_img)

ps in regions:

ppend(props.filled_area)
ppend(props.convex_area)
ppend(props.eccentricity)
ppend(props.equivalent_diameter)
ppend(props.extent)
ppend(props.perimeter)
ppend(props.perimeter_crofton)
ppend(props.solidity)
ppend(props.area)
print("Filled Area:",l[0])
print("Convex Area:",l[1])
print("Eccentricity:",l[2])
print("Equivalent Diameter:",l[3])
print("Extent:",l[4])
print("Perimeter:",l[5])
print("Perimeter Crofton:",l[6])
print("Solidity:",l[7])
print("Area:",l[8])
return l

=[(0,100,255),(255,255,0),(255,0,255),(0,0,255),(205,100,0),(0,255,0)]
_text=[(0,100,255),(255,255,0),(255,0,255),(0,0,255),(205,100,0),(0,255,0)]
ture_classifier(img_rgb):

= cv2.cvtColor(img_rgb, cv2.COLOR_BGR2GRAY)
=cv2.resize(img_rgb,(30,30))
_image(img_rgb)
= None:
```

```
dfg=np.array(dt)
dfg=dfg.reshape(1,9)
X.append(dfg)
xd=np.array(X)
xd=xd.reshape(xd.shape[0],9)
dt=pd.DataFrame(xd)
cl=['Upset',
'Blast',
'Mousebite',
'Spur',
'Copper',
'Pin-hole']
p1=model.predict(dt)
p2=list(p1[0])
p3=max(p2)
p4=p2.index(p3)
p5=cl[p4]
return p5,p4
return "None",7
ind = np.unravel_index(np.argmax(result), result.shape)
ind=ind[::-1]
messagebox(
title='alert',
message=f"ready to selct original image"

root.update()
original_image_path=fd.askopenfilename()
imageA = cv2.imread(original_image_path)
s=imageA.shape
imageA=cv2.resize(imageA,(500,500))
messagebox(
title='alert',
message=f"ready to selct Second image image"
```

```
imageB = cv2.imread(edited)
imageB=cv2.resize(imageB,(500,500))
# convert the images to grayscale
grayA = cv2.cvtColor(imageA, cv2.COLOR_BGR2GRAY)
grayB = cv2.cvtColor(imageB, cv2.COLOR_BGR2GRAY)
# compute the Structural Similarity Index (SSIM) between the two
# images, ensuring that the difference image is returned
(score, diff) = compare_ssim(grayA, grayB, full=True)
diff = (diff * 255).astype("uint8")
print("SSIM: {}".format(score))
# threshold the difference image, followed by finding contours to
# obtain the regions of the two input images that differ
thresh = cv2.threshold(diff, 150, 255,
  cv2.THRESH_BINARY_INV | cv2.THRESH_OTSU)[1]
cnts = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL,
  cv2.CHAIN_APPROX_SIMPLE)
cnts = imutils.grab_contours(cnts)
# loop over the contours
i=0
img=imageB.copy()
res_all=[]
for c in cnts:
  # compute the bounding box of the contour and then draw the
  # bounding box on both input images to represent where the two
  # images differ
  (x, y, w, h) = cv2.boundingRect(c)
  #cv2.rectangle(imageB, (x, y,z), (x + w, y + h), (0, 0, 255), 2)
  cropped_image = grayB[y:y+h, x:x+w]
  res.index=Bpn_classifier(cropped_image)
```

```
    res_all.append(res)
    font = cv2.FONT_HERSHEY_SIMPLEX
    if index !=7:
        print("Defect Number:",i+1)
        print("width:",w)
        print("Height:",h)
        print("Area Of Defect:",w*h)
        print("Defect:",res)
        cv2.rectangle(imgc, (x, y), (x + w, y + h), color[index], 2)
        cv2.putText(imgc, f'{i+1}.{res}', (x,y), font, 0.5,color_text[index] , 1, cv2.LINE_AA)
    print("\n")
    i+=1
cv2.imwrite(f"plot/result.jpg",imgc)
# show the output images
#print("Total Defect:",res_all)
print("Total Number Of Defect:",len(res_all))
print("\n")
print("list of Defects:")
k=1
for i in res_all:
    print(f"{k}.{i}")
    k=k+1
cv2.imshow("original", imageA)
cv2.imshow("Defected Location", imgc)
#imageB=cv2.resize(imageB,(500,500))
cv2.imshow("modified", imageB)
cv2.imshow("Diff", diff)
cv2.imshow("Thresh", thresh)
cv2.waitKey(0)#press q for close all window
cv2.destroyAllWindows()
```

template_creation.py

```
import os
import easygui
import cv2
import numpy as np
from skimage.metrics import structural_similarity as compare_ssim
# from skimage.measure import  compare_ssim
#import argparse
import imutils
from tkinter import filedialog as fd
import tkinter as tk
from tkinter import ttk
from tkinter import filedialog as fd
from tkinter.messagebox import showinfo
import tkinter as tk
## Root window
#root = tk.Tk()
#root.title('Display a Text File')
#root.resizable(False, False)
#root.geometry('550x250')
ori_path=os.listdir('PCB_DATASET/images')
N=len(os.listdir('PCB_DATASET/images'))
for main_loop in ori_path:
    original_image_path=f"PCB_DATASET/PCB_USED/01.jpg"
    imageA= cv2.imread(original_image_path)
    j="S"
    p=1
    l=0
    list1=os.listdir(f"PCB_DATASET/images/{main_loop}")
    h=len(list1)
    k=0
    while l<=10:

        edit_img_path=f"PCB_DATASET/images/{main_loop}/{list1[k]}"
```

```
    imageB = cv2.imread(edit_img_path)
    grayA = cv2.cvtColor(imageB, cv2.COLOR_BGR2GRAY)
    grayB = cv2.cvtColor(imageA, cv2.COLOR_BGR2GRAY)

    # compute the Structural Similarity Index (SSIM) between the two
    # images, ensuring that the difference image is returned
    (score, diff) = compare_ssim(grayA, grayB, full=True)
    diff = (diff * 255).astype("uint8")
    # print("SSIM: {}".format(score))
    thresh = cv2.threshold(diff, 0, 255,
    cv2.THRESH_BINARY_INV | cv2.THRESH_OTSU)[1]
    cnts = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL,
        cv2.CHAIN_APPROX_SIMPLE)
    cnts = imutils.grab_contours(cnts)
    for c in cnts:
        # compute the bounding box of the contour and then draw the
        # bounding box on both input images to represent where the two
        # images differ
        (x, y, w, h) = cv2.boundingRect(c)

        #cv2.rectangle(imageA, (x, y), (x + w, y + h), (0, 0, 255), 2)
        #cv2.rectangle(imageB, (x, y), (x + w, y + h), (0, 0, 255), 2)
        cropped_image = imageB[y:y+h, x:x+w]
        if (w>25 and h>25):
            cropped_image=cv2.resize(cropped_image,(25,25))
            cv2.imwrite(f"output/{main_loop}/{main_loop}_{p}.jpg",cropped_image)
        p+=1
    k+=1
    l+=1
# j=input(f"Do You Want another one in{main_loop} currently{p}?")
```

train_model.py

```
import numpy as np
import cv2
import os
from sklearn.model_selection import train_test_split
import tensorflow as tf
from tensorflow.keras import datasets, layers, models
ori_path=os.listdir('output/')
N=len(os.listdir('output'))
X=[]
Y=[]
cl=[]

for i in range(N):
    cl.append(ori_path[i])
    path1=f"output/{ori_path[i]}"
    list1=os.listdir(path1)
    k=0
    for j in list1:
        path2=f"{path1}/{j}"
        image=cv2.imread(path2,0)
        X.append(image)
        Y.append(i)
        k=k+1
        if k==30:
            break
X=np.array(X)
Y=np.array(Y)
np.save("X",X)
np.save("Y",Y)

Y=tf.keras.utils.to_categorical(Y)
X=X.reshape(X.shape[0],X.shape[1],X.shape[2],1)
xtrain,xtest,ytrain,ytest=train_test_split(X,Y,test_size=0.03,shuffle=True,random_state=True)
```

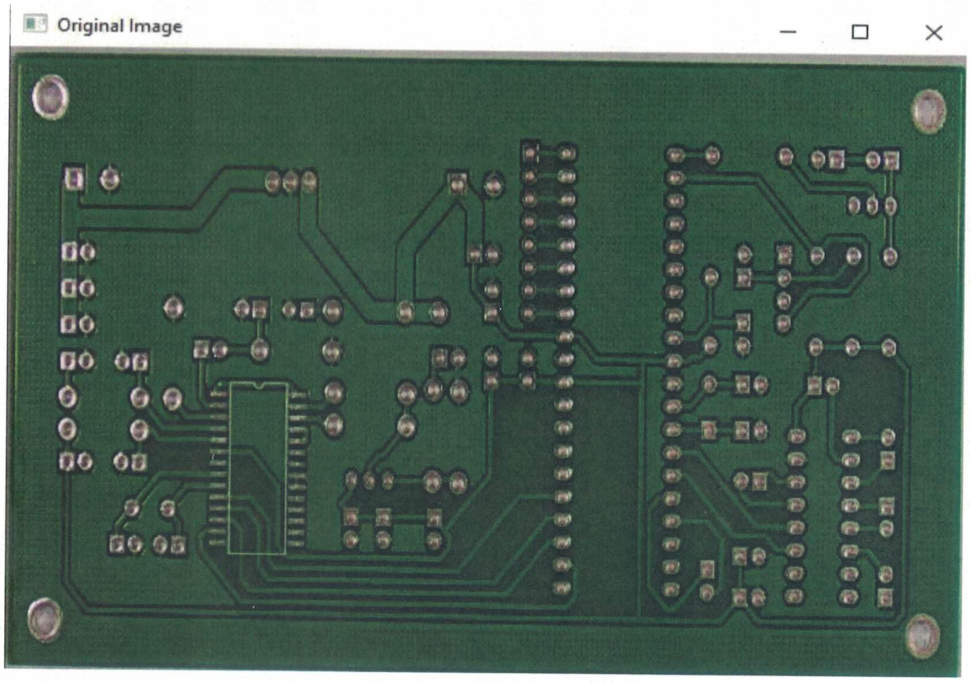
```
model = models.Sequential()
model.add(layers.Conv2D(64, (3, 3), activation='relu', input_shape=(25, 25, 1),padding="same"))
model.add(layers.MaxPooling2D((2, 2),padding='same'))
model.add(layers.Dropout(0.33))
model.add(layers.Conv2D(64, (3,3),activation='relu',padding='same'))
model.add(layers.MaxPooling2D((2, 2),padding='same'))
model.add(layers.Dropout(0.33))
model.add(layers.Conv2D(128,(3,3),activation='relu',padding='same'))
model.add(layers.MaxPooling2D((2, 2),padding='same'))
model.add(layers.Dropout(0.35))
#model.add(Conv2D(64, (3,3),activation='relu',padding='same'))
#model.add(MaxPooling2D((2, 2),padding='same'))
model.add(layers.Conv2D(32,(3,3),activation='relu',padding='same'))
model.add(layers.Dropout(0.35))
model.add(layers.Flatten())
model.add(layers.Dense(64, activation='relu'))
# model.add(layers.Dense(1012, activation='relu'))
model.add(layers.Dense(6,activation="softmax"))

model.compile(optimizer="Adam",
        loss='categorical_crossentropy',
        metrics=['accuracy'])

my_callbacks = [
  tf.keras.callbacks.EarlyStopping(monitor = 'val_accuracy',
                min_delta = 0,
                patience = 60,
                verbose = 1,
                restore_best_weights = True),
  tf.keras.callbacks.ModelCheckpoint(filepath='best_model2.h5',monitor='val_accuracy',
                mode='max',
                save_best_only=True,
                verbose=1),
]
```

```
odel = tf.keras.models.load_model('best_model.h5')
odel.fit(xtrain, ytrain, epochs=10,
            validation_data=(xtest, ytest),callbacks=my_callbacks)

del.fit(xtrain, ytrain, batch_size=64,epochs=100,validation_data=(xtest,
st),callbacks=my_callbacks)
odel.save("model.h5")
```

SCREENSHOTS

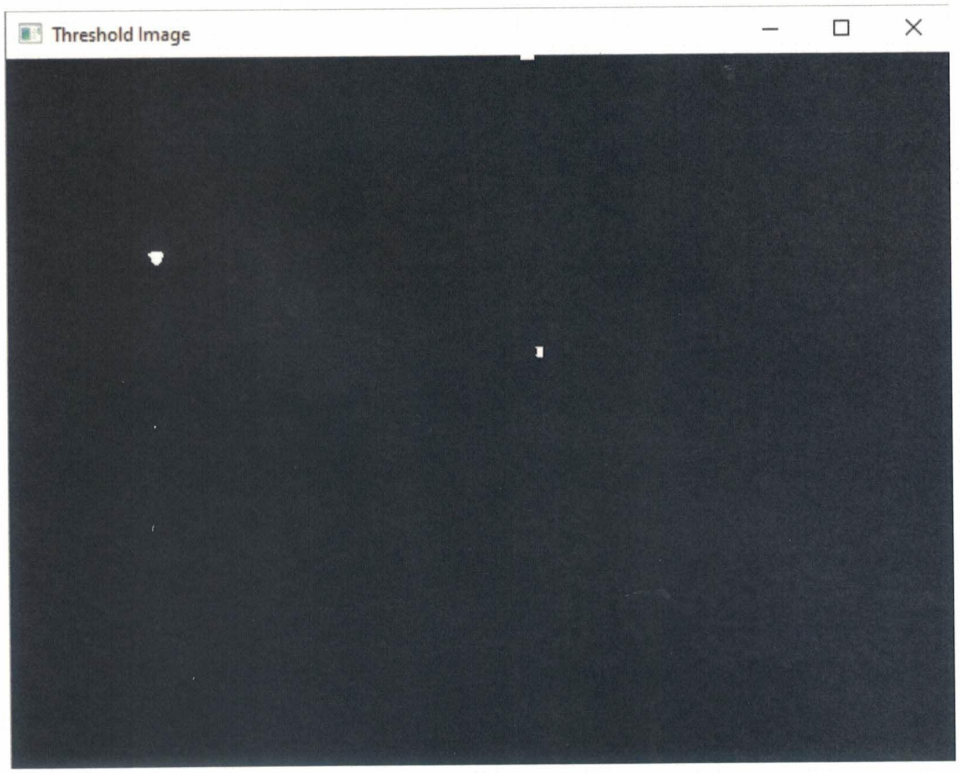
8. SCREENSHOTS

put Image

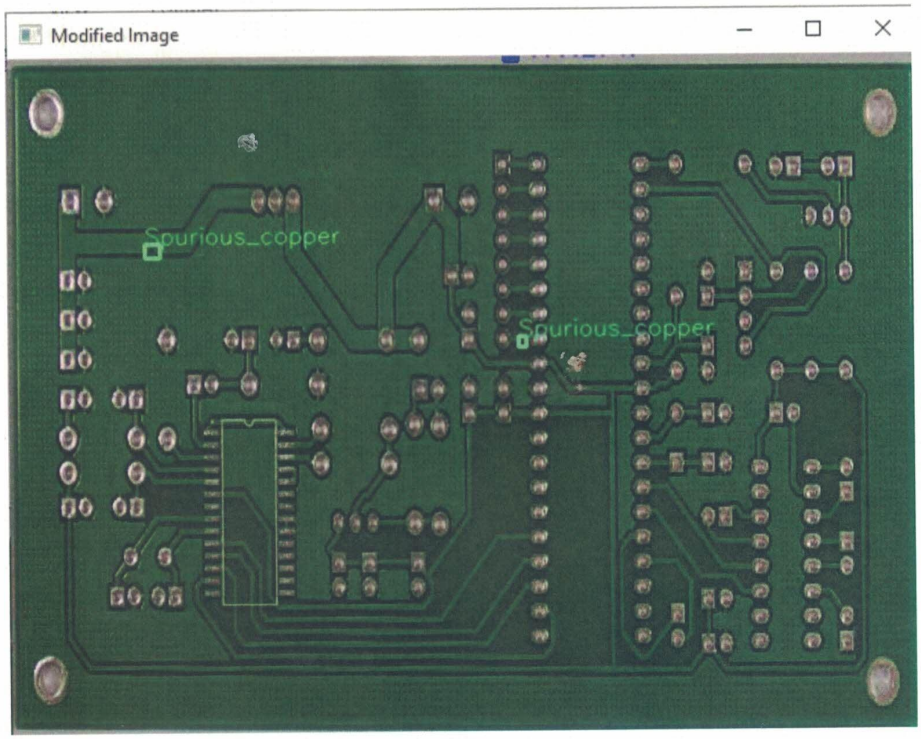


fference Image

Threshold Image



Modified Image



CONCLUSION

9. CONCLUSION

The conclusion of the project is that the template image and input test image is compared ing structural similarity index and the feature is extracted in this proposed work, this project signed a model with BPN network to predict the PCB defect. The model is built, trained and sted with the data in PCB dataset to achieve high efficiency, accuracy and very less computational ost and elapsed time. Therefore, the proposed method has contributed towards high or efficient CB defect detecting process.

FUTURE ENHANCEMENT

10. FUTURE ENHANCEMENT

Automated machine learning can be used to identify patterns in the data and make predictions about potential defects. This could be used to improve the accuracy and speed of defect detection.

Image recognition algorithms can be used to detect features and patterns in images that may indicate a potential defect.

Deep learning techniques can be used to analyze complex data sets to identify potential defects.

Robotic process automation can be used to automate tedious manual tasks, such as inspection and testing, to improve efficiency and accuracy.

IoT sensors can be used to monitor and measure the performance of PCBs and alert engineers to potential defects.

AI can be used to detect anomalies in PCBs and predict potential defects.

BIBLIOGRAPHY

BIBLIOGRAPHY

T.K. Ghose, P.K. Mallick, S.M. Abdin, "Machine vision based automatic defect detection and classification of printed circuit board", IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), pp. 2-7, 2014.

T.K. Ghose, S.M. Abdin, P.K. Mallick, "Detection and classification of defects of printed circuit boards using digital image processing techniques", International Journal of Image Processing, vol. 5, no. 1, pp. 1-14, 2011.

L.M. Pacheco, A.A. Cunha, E. M. L. Rocha, "Automatic PCB defect segmentation: A novel fuzzy approach", IEEE Transactions on Industrial Electronics, vol. 57, no. 10, pp. 3484-3493, 2010.

X. Yan, A.K. Jain, "Automatic detection of printed circuit board defects", IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics, vol. 32, no. 3, pp. 406-418, 2002.

H.Y. Hsieh, "Automatic detection of printed circuit board defects", IEEE Transactions on Industrial Electronics, vol. 51, no. 5, pp. 883-891, 2004.

Y.H. Lee, J.Y. Lee, "An automated system for inspection and classification of printed circuit board defects", IEEE Transactions on Industrial Electronics, vol. 55, no. 11, pp. 3965-3974, 2008.

Y.H. Lee, J.Y. Lee, "PCB defect recognition based on shape analysis and wavelet decomposition", IEEE Transactions on Industrial Electronics, vol. 56, no. 5, pp. 1583-1592, 2009.

K.R. Reddy, S.K. Mutha, "Automatic printed circuit board defect inspection using image processing", IEEE International Conference on Electronics, Communications and Computers (CONECCT), pp. 5-10, 2010.

https://medium.com/@Swathi_Nallapati/pcb-fault-detection-using-opencv-and-python-13f7ea7759c

https://towardsdatascience.com/building-an-end-to-end-deep-learning-defect-classifier-application-for-printed-circuit-board-pcb-6361b3a76232

https://blog.ldtalentwork.com/2021/03/07/how-to-detect-defects-on-a-printed-circuit-board-using-opencv/

https://docs.edgeimpulse.com/experts/prototype-and-concept-projects/identifying-pcb-defects

https://www.electroschematics.com/pcb-defects-detection-opencv/

https://github.com/Ixiaohuihuihui/Tiny-Defect-Detection-for-PCB

https://ieeexplore.ieee.org/document/8029117

ttps://www.techscience.com/cmc/v70n1/44420/pdf

ttps://iopscience.iop.org/article/10.1088/1742-6596/1827/1/012167/pdf

ttps://www.sciencedirect.com/science/article/pii/S0141933120309522

BITCOIN PRICE PREDICTION

A project submitted to

ST. MARY'S COLLEGE (Autonomous), THOOTHUKUDI

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY

TIRUNELVELI

In partial fulfillment of the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Submitted by

NISHA RANI. P

Reg. No. 21SPCS04

Under the Supervision and Guidance of

Ms. C. NAYANTHRA MASCARENHAS M.Sc., M. Phil., SET.,

PG DEPARTMENT OF COMPUTER SCIENCE (SSC)

St. Mary's College (Autonomous), Thoothukudi– 628001

APRIL 2023

CERTIFICATE

This is to certify that this project work entitled as **"BITCOIN PRICE PREDICTION"** is submitted to St. Mary's College(Autonomous), Thoothukudi affiliated to **Manonmaniam Sundaranar University, Tirunelveli**, in partial fulfillment for the award of the degree of Master of Science in **Computer Science** for the work done during the year 2022-2023 by **NISHA RANI. P** (Reg. No. 21SPCS04).

Signature of the Guide

Signature of the Co-ordinator

Signature of the Director

Director
Self Supporting Courses
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Principal

Principal
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Examiner

DECLARATION

I do hereby declare that the project entitled **"BITCOIN PRICE PREDICTION"** submitted for the degree of Master of Science in Computer Science in my original work carried out under the guidance **Ms. C. NAYANTHRA MASCARENHAS M.Sc., M.Phil., SET.**, Assistant Professor, PG Department of Computer Science (SSC), St. Mary's College (Autonomous), Thoothukudi.

Station: Thoothukudi

Date: 10.04.2023

P. Nithya

Signature of the Student

ACKNOWLEDGEMENT

I express my first and foremost thanks to God Almighty for his gracious help and shower of blessings for having rendered us the strength and support to finish our project successfully.

My sincere thanks to **Dr. Sr. A. S. J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., Principal, Rev. Sr. C.Shibana M.Sc.,B.Ed.,M.Phil.,Ph.D., Secretary, Sr. Josephine Jeyarani , Director of SSC,** St. Mary's College (Autonomous), Thoothukudi, for giving permission to work on this project.

I convey my heartfelt thanks to my guide **Ms. C. Nayanthra Mascarenhas M.Sc., M.Phil., SET.,** Assistant Professor and Coordinator, PG Department of Computer Science (SSC) for her support and counsel.

I express my hearty thanks to **Ms. A. Jenitta Jebamalar M.Sc. (IT), M.Sc. (CS), M.Phil., B.Ed., and Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D.,** Assistant Professor, PG Department of Computer Science (SSC) for their valuable suggestions and support.

I am much indebted to **DR . P. JOHNSONDURAIRAJ, DIRECTOR, POSTULATE** for his untiring effort, immense knowledge and priceless contribution without which I couldn't have finished my work effectively on time.

I thank my family members especially my parents for their encouragement and support both morally and financially which helped me to finish the project successfully.

POSTULATE

01.04.2023

To

The Head of the Department
PG Department of Computer Science (SSC)
St. Mary's College(Autonomous)
Thoothukudi

Dear Sir/Madam,

Sub: Project Work Completion-Reg

On behalf of **POSTULATE**, we are pleasure to inform you that **Ms.P.Nisha Rani, Reg no: 21SPCS04** studying **Master of Computer Science** Final year has been done the project work at our concern Postulate Info Tech Private Limited in the title "**Bitcoin Price Prediction**" during the period from December 2022 to March 2023.

During this period, the candidate was found diligent and sincere. We wish all the best for future endeavors.

POSTULATE INFO TECH PRIVATE LIMITED

DIRECTOR

TABLE OF CONTENT

ABSTRACT

ABSTRACT

Cryptocurrencies are a digital way of money in which all transactions are held electronically. It is a soft currency which doesn't exist in the form of hard notes physically. Here, we are emphasizing the difference of fiat currency which is decentralized that without any third-party intervention all virtual currency users can get the services. However, getting services of these cryptocurrencies impacts on international relations and trade, due to its high price volatility. There are several virtual currencies such as bitcoin, ripple, ethereum, ethereum classic, lite coin, etc. In our project, we especially focused on a popular cryptocurrency, i.e., bitcoin. From many types of virtual currencies, bitcoin has a great acceptance by different bodies such as investors, researchers, traders, and policy-makers. To the best of our knowledge, our target is to implement the efficient deep learning-based prediction models specifically long short-term memory (LSTM) to handle the price volatility of bitcoin and to obtain high accuracy.

INTRODUCTION

INTRODUCTION

1. INTRODUCTION

Cryptocurrency is a digital payment system that doesn't rely on banks to verify transactions. It's a peer-to-peer system that can enable anyone anywhere to send and receive payments. Instead of being physical money carried around and exchanged in the real world, cryptocurrency payments exist purely as digital entries to an online database describing specific transactions. When you transfer cryptocurrency funds, the transactions are recorded in a public ledger. Cryptocurrency is stored in digital wallets. Cryptocurrency received its name because it uses encryption to verify transactions. This means advanced coding is involved in storing and transmitting cryptocurrency data between wallets and to public ledgers. The aim of encryption is to provide security and safety.

Bitcoin is a form of digital money that is designed to be a secure, private, and anonymous way of conducting transactions over the internet. Bitcoin has become increasingly popular over the past few years, as its value has skyrocketed from a few cents to thousands of dollars. Along with its increasing popularity, the potential for bitcoin to be used as a form of investment has also grown. As a result, many people are now looking for ways to predict Bitcoin price changes in order to make a profit. One way to do this is by using deep learning techniques to analyze historical.

Bitcoin price data and build a model that can accurately predict future prices. The Python programming language is particularly well suited for this task due to its ability to work with large datasets and its easy-to-use syntax.

In this project, we will use Python to build a model that can predict Bitcoin price changes. The first step in this project is to gather the necessary data. We will use historical Bitcoin price data from various sources, including exchanges and online services.

Once the data is organized, we will use the Python scikit-learn library to build a model that can accurately predict Bitcoin price changes. We will use a variety of different machine learning algorithms, such as linear regression and decision trees, in order to create the most accurate model possible.

Finally, we will evaluate the performance of our model and make any necessary adjustments. Once we are satisfied with the performance of our model, we will be able to use it to predict future Bitcoin prices.

In conclusion, this project will use Python and machine learning techniques to build a model that can accurately predict future Bitcoin prices. By gathering historical data, organizing it into a usable format, and using machine learning algorithms to create the model, we will be able to make accurate predictions about future Bitcoin prices. Some of popular types of bitcoin are given below.

Ethereum:

Developed in 2015, Ethereum is a blockchain platform with its own cryptocurrency, called Ether (ETH) or Ethereum. It is the most popular cryptocurrency after Bitcoin.

Litecoin:

This currency is most similar to bitcoin but has moved more quickly to develop new innovations, including faster payments and processes to allow more transactions.

Ripple:

Ripple is a distributed ledger system that was founded in 2012. Ripple can be used to track different kinds of transactions, not just cryptocurrency. The company behind it has worked with various banks and financial institutions.

Binance Coin :

Binance Coin is native to Binance, the world's largest cryptocurrency exchange as of 2021. Transaction fees for this exchange are reduced for users who choose to pay in BNB. This has encouraged the adoption of Binance Coin, making it one of the largest cryptocoins in the market. To ensure its value remains stable, Binance destroys or "burns" a fixed percentage of the coins in circulation.

How to store cryptocurrency

Once you have purchased cryptocurrency, you need to store it safely to protect it from hacks or theft. Usually, cryptocurrency is stored in crypto wallets, which are physical devices or online software used to store the private keys to your cryptocurrencies securely. Some exchanges provide wallet services, making it easy

for you to store directly through the platform. However, not all exchanges or brokers automatically provide wallet services for you.

There are different wallet providers to choose from. The terms "hot wallet" and "cold wallet" are used:

- **Hot wallet storage:** "hot wallets" refer to crypto storage that uses online software to protect the private keys to your assets.
- **Cold wallet storage:** Unlike hot wallets, cold wallets (also known as hardware wallets) rely on offline electronic devices to securely store your private keys.

What can you buy with cryptocurrency?

When it was first launched, Bitcoin was intended to be a medium for daily transactions, making it possible to buy everything from a cup of coffee to a computer or even big-ticket items like real estate. That hasn't quite materialized and, while the number of institutions accepting cryptocurrencies is growing, large transactions involving it are rare. Even so, it is possible to buy a wide variety of products from e-commerce websites using crypto. Here are some examples:

Technology and e-commerce sites:

Several companies that sell tech products accept crypto on their websites, such as newegg.com, AT&T, and Microsoft. Overstock, an e-commerce platform, was the first sites to accept Bitcoin. Shopify, Rakuten, and Home Depot also accept it.

Cars:

Some car dealers – from mass-market brands to high-end luxury dealers – already accept cryptocurrency as payment.

Insurance:

In April 2021, Swiss insurer AXA announced that it had begun accepting Bitcoin as a mode of payment for all its lines of insurance except life insurance. Premier Shield Insurance, which sells home and auto insurance policies in the US, also accepts Bitcoin for premium payments.

SYSTEM REQUIREMENTS

2. SYSTEM REQUIREMENTS

2.1 HARDWARE SPECIFICATION:

COMPONENT	SPECIFICATION
Processor	AMD 3020e with Radeon Graphics 1.20GHz
Speed	1.1 GHZ
Ram	4.00 GB
Hard Disk	40 GB
Key Board	Standard Window Key

2.2 SOFTWARE SPECIFICATION:

COMPONENT	SPECIFICATON
Operating System	Windows 10
Tools	Spyder (python)
Front End	Html,css
Back End	python
Framework	flask

SOFTWARE DESCRIPTION

3. SOFTWARE DESCRIPTION

3.1 INTRODUCTION TO FRONT END

➢ **Python**

Python is a high level interpreted language used for general purpose programming. It is widely used for scientific computing and can be used for a wide variety of general tasks from data mining to software development. Python is the main language used for this project.

➢ **Anaconda**

Anaconda is a popular data science platform where you can create data science projects and machine learning. Libraries such as NumPy, Pandas, Matplotlib, TensorFlow and etc come with Anaconda and IDE's such as Jupyter Notebook, Spyder and etc.

➢ **Numpy**

NumPy is an acronym for "Numeric Python" or "Numerical Python". It is an open source extension module for Python, which provides fast precompiled functions for mathematical and numerical routines. Furthermore, NumPy enriches the programming language Python with powerful data structures for efficient computation of multi-dimensional arrays and matrices. The implementation is even aiming at huge matrices and arrays. Besides that the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays. It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier Transform, and random number capabilities.

- **Pandas**

Pandas is also a library in Python, like numpy is also used for data pre-processing and preparation. One of the main features about pandas is the DataFrame and Series data structure. These data structures are optimized and contain fancy indexing that allow a variety of features such as reshaping, slicing, merging, joining and etc to be available. Pandas and Numpy are extremely powerful when used together for manipulating data.

- **Matplotlib**

Matplotlib is a Python plotting library that allows programmers to create a wide variety of graphs and visualizations with ease of use. The great feature about Matplotlib is that it integrates very well with Jupyter Notebook and creating visualizations is simplified. Matplotlib also works very well with pandas and numpy.

- **OpenCV**

OpenCV (Open Source Computer Vision Library) is an open source computer visio n and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. OpenCV has more than 47 thousand people of user community and estimated number of downloads exceeding 14 million. The library is used extensively in companies, research groups and by governmental bodies. It

has C++, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS.

OpenCV leans mostly towards real-time vision applications and takes advantage of MMX and SSE instructions when available. A full-featured CUDA and OpenCL interfaces are being actively developed right now. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. 7 OpenCV is written natively in C++ and has a templated interface that works seamlessly with STL containers.

- **Tensor Flow**

TensorFlow is an open-source deep learning library by Google. It was originally developed by Google's engineers who were working on Google Brain and has been used for research on machine learning and deep learning. TensorFlow at it is core is about computations of multidimensional arrays called tensors but what makes TensorFlow great is its ability to be flexible to deploy computations on different devices such as CPU's and GPU's.

- **Keras**

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Keras contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on Git Hub, and community support forums include the Git Hub issues page, and a Slack channel. Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of distributed training of deep learning models on clusters of Graphics Processing Units (GPU).

- **HTML**

HTML (Hypertext Markup Language) is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page. The markup tells the Web browser how to display a Web page's words and images for the user. Each individual markup code is referred to as an element but many

people also refer to it as a tag.Some elements come in pairs that indicate when some display effect is to begin and when it is to end.

- **CSS**

A cascading style sheet (CSS) is a Web page derived from multiple sources with a defined order of precedence where the definitions of any style element conflict. The Cascading Style Sheet, level 1 recommendation from the World Wide Web Consortium, which is implemented in the latest versions of the Netscape and Microsoft Web browsers, specifies the possible style sheets or statements that may determine how a given element is presented in a Web page. CSS gives more control over the appearance of a Web page to the page creator than to the browser designer or the viewer. With CSS, the sources of style definition for a given document element are in this order of precedence.

- **Flask**

Flask is a micro-framework designed to create a web application in a short time. It only implements the core functionality giving developers the flexibility to add the feature as required during the implementation. It is a lightweight, WSGI application framework. This framework can either be used for pure backend as well as frontend if need be. The former provides the functionality of the interactive debugger, full request object, routing system for endpoints, HTTP utilities for handling entity tags, cache controls, dates, cookies etc. It also provides a threaded WSGI server for local development including the test client for simulating the HTTP requests. Werkzeug and Jinja are the two core libraries.

- **Jupyter Notebook IDE**

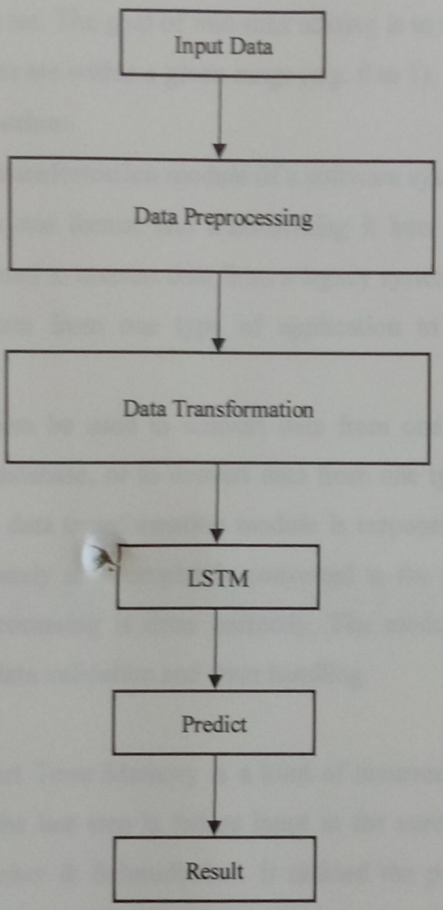
The Anaconda distribution comes with a variety of software that includes Jupyter Notebooks for scientific computing. Jupyter Notebooks is an open source software IDE that allows developers to create and share documents that contain live code and more.

SYSTEM DESIGN

AND DEVELOPMEN

4. SYSTEM DESIGN AND DEVELOPMENT

4.1 SYSTEM DESIGN



4.2 SYSTEM DEVELOPMENT

✓ **Data Preprocessing:**

Preprocessing simply refers to perform series of operations to transform or change data. It is transformation applied to our data before feeding it to algorithm. Data processing refers to perform operations on data to retrieve, transform, or change data, especially by computer.

Min-max scaling (also known as min-max normalization) is a data preprocessing technique used to transform numerical features in a dataset by rescaling them to a given range between 0 and 1. It is useful for data normalization and can help improve the accuracy of deep learning algorithms.

It is also used to standardize the range of independent variables or features of a data set. The goal of min-max scaling is to rescale the data set so that all the features are within a given range (e.g. 0 to 1).

✓ **Data Transformation:**

The data transformation module of a software system is responsible for taking data from one format and transforming it into another format. This module is often used to convert data from a legacy system into a newer format, or to convert data from one type of application to a different type of application.

It may also be used to convert data from one type of database to another type of database, or to convert data from one type of file to another type of file. The data transformation module is responsible for ensuring that the data is accurately and completely converted to the new format, and that any necessary processing is done correctly. The module may also include features such as data validation and error handling.

✓ **LSTM:**

Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is usedfor processing, predicting, and classifying on the basis of time-series data. Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that is specifically designed to handle sequential data, such as time series, speech, and text. LSTM networks are capable of learning long-term dependencies in sequential data, which makes them well suited for tasks such as language translation, speech recognition, and time series forecasting.

PROJECT DESCRIPTION

5. PROJECT DESCRIPTION

The project entiled"Bitcoin price prediction" has the following modules

- HOME PAGE
- RESULT PAGE

HOME PAGE:

In this page we need to enter the day, month and year after entering that we need enter number of future days to predicts to know future value of bitcoin.

RESULT PAGE:

In this page we can able see future value depends on the historical Price data, and also feed the model a look back value of 15 days, i.e.; it predicts on the basis of previous 15 days data.

SYSTEM STUDY

6. SYSTEM STUDY

6.1 EXISTING SYSTEM

The existing system used to make Bitcoin price predictions largely consists of analyzing market trends, news, and technical analysis. Market trend analysis involves looking at the historical data for Bitcoin prices to identify patterns and make predictions about future prices. News analysis involves looking at news stories related to Bitcoin and using that data to make predictions about future prices. Technical analysis involves using technical indicators such as moving averages and momentum oscillators to predict future prices.

6.2 DRAWBACKS OF EXISTING SYSTEM

1. Limited Data Sources:

A major drawback of existing Bitcoin price prediction models is their reliance on limited data sources. Many models use only historical price data and ignore other important factors, such as news, sentiment, macroeconomic events, and technical analysis.

2. Lack of Interoperability:

The existing models are not interoperable, meaning they cannot be used in different environments or with different software. This makes it difficult to compare the performance of different models and to properly evaluate their accuracy.

3. Difficulty in Obtaining Reliable Predictions:

Due to the lack of reliable data sources and limited interoperability, it can be difficult to obtain reliable predictions from existing models. This can lead to inaccurate predictions, resulting in poor decisions.

6.3 PROPOSED SYSTEM

1. **Data Collection and Pre-Processing:**

This component should include the collection of historical Bitcoin pricing data, as well as any other relevant data such as news headlines, economic data, and social media sentiment. This data should then be pre-processed to ensure it is in the correct format and contains no errors.

2. **Feature Engineering:**

This component should involve extracting useful features from the data collected in the previous step. This can include transforming the Bitcoin pricing data into a format that is more suitable for machine learning algorithms, as well as extracting features from other sources of data such as news headlines or economic data.

3. **Model Development:**

This component should involve developing and testing different machine learning models for the task of Bitcoin price prediction, such as neural networks, linear models, and support vector machines. The best performing model should be selected for deployment.

4. **Deployment:**

This component should involve deploying the model developed in the previous step so that it can be used to make predictions on incoming data. This can be done using a web service or an API.

5. **Evaluation:**

This component should involve evaluating the performance of the model on new data. This can include measuring the model's accuracy and evaluating its predictions against the actual Bitcoin prices.

6.4 ADVANTAGES OF PROPOSED SYSTEM

1. The system can provide an accurate prediction of the future price of bitcoin using historical data.

2. It can provide users with an understanding of the current market sentiment and help them make informed decisions.

3. The system can be used to identify entry and exit points for investments.

4. The system can be used to identify trends in the market that can be used to make more informed decisions.

5.The system can provide insight into the factors that affect the price of bitcoin, allowing users to make better investment decisions.

6. The system can be used to identify correlations between different market factors and the price of bitcoin.

7. The system can be used to identify potential investment opportunities.

SYSTEM TESTING AND IMPLEMENTATION

7.SYSTEM TESTING AND IMPLEMENTATION

7.1 SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

7.2 TYPES OF TESTING

1. REGRESSION TESTING :

Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features. Regression Testing is nothing but a full or partial selection of already executed test cases that are re-executed to ensure existing functionalities work fine. This testing is done to ensure that new code changes do not have side effects on the existing functionalities. It ensures that the old code still works once the latest code changes are done.

2. LOAD TESTING:

Load testing is a type of performance testing that simulates a real-world load on any software, application, or website. Without it, your application could fail miserably in real-world conditions. That's why we build tools like retrace to help you monit-or application performance and fix bugs before your code ever gets to production. Load testing examines how the system behaves during normal and high loads and determines if a system, piece of software, or computing device can handle high loads given a high demand of end-users. This tool is typically applied when a software development project nears completion.

3. FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is cantered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing.

4. RECOVERY TESTING:

Recovery Testing is software testing technique which verifies software's ability to recover from failures like software/hardware crashes, network failures etc. The purpose of Recovery Testing is to determine whether software operations can be continued after disaster or integrity loss. Recovery testing involves reverting back software to the point where integrity was known and reprocessing transactions to the failure point.

5. UNIT TESTING

A program represents the logical elements of a system. For a program to run satisfactorily, it must compile and test data correctly and tie in properly with other programs. Achieving an error free program is the responsibility of the programmer. Program testing checks for two types of errors: syntax and logical. Syntax error is a program statement that violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are

common syntax errors. These errors are shown through error message generated by the computer. For Logic errors the programmer must examine the output carefully.

7.3 SYSTEM IMPLEMENTATION

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system. Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly. Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which provides components and the operating procedures of the system.

CODE DESIGN

8.CODE DESIGN

Main.html

```
<!DOCTYPE html>

<html lang="en">

<head>

  <meta charset="UTF-8">

  <meta http-equiv="X-UA-Compatible" content="IE=edge">

  <meta name="viewport" content="width=device-width, initial-scale=1.0">

  <title>Bitcoin Price Prediction</title>

<linkrel="stylesheet"href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/css/boot
strap.min.css"integrity="sha384-Gn5384xqQ1aoWXA+058RXPxPg6fy4IWvTNh0
E263XmFcJlSAwiGgFAW/dAiS6JXm" crossorigin="anonymous">

</head>

<body>

  <div class="p-1 mx-auto mt-1" style="width: 38%; background-color: rgb(223, 223,
223);">

    <div class="card border-dark">

      <img class="card-img-top"

        src="https://images.unsplash.com/photo-1590283603385-
17ffb3a7f29f?ixlib=rb-
1.2.1&ixid=MnwxMjA3fDB8MHxwaG90by1wYWdlfHx8fGVufDB8fHx8&auto=fo
rmat&fit=crop&w=870&q=80"

        alt="Card image cap">

      <form action="{{ url_for('predict') }}" method="POST">

        <div class="card-header bg-transparent border-dark font-italic">
```

```
        </div>
        <div class="card-body">
          <h1 class="title is-1 mb-3">Bitcoin Price Prediction</h1>
          <div class="input-group mb-3">
            <div class="input-group-prepend">
              <label class="input-group-text" for="date">Date - First day of
prediction:</label>
            </div>
            <input id="date" type="date" name="date" min="2015-10-31"
max="2023-12-31" required
              class="form-control">
          </div>
          <div class="input-group mb-8">
            <div class="input-group-prepend">
              <label class="input-group-text" for="days">Number of future
                days to predict for:</label>
            </div>
            <select class="custom-select" name="n" id="days">
              <option value="1">1</option>
              <option value="2">2</option>
              <option value="3">3</option>
              <option value="4">4</option>
              <option value="5">5</option>
              <option value="6">6</option>
              <option value="7">7</option>
```

```
                <option value="8">8</option>
                <option value="9">9</option>
                <option value="10">10</option>
                <option value="11">11</option>
                <option value="12">12</option>
                <option value="13">13</option>
                <option value="14">14</option>
                <option value="15">15</option>
              </select>
            </div>
          </div>
          <div class="card-footer text-center bg-transparent border-dark">
            <button type="submit" class="btn btn-info btn-lg" style="width: 35%;">Predict</button>
            <button type="reset" class="btn btn-danger btn-lg" style="width: 35%;">Reset</button>
          </div>

        </form>
      </div>
    </div>
</body>
</html>
```

Result.html

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

<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Prediction Results</title>
  <linkrel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0/
  css/bootstrap.min.css"integrity="sha384-Gn5384xqQ1aoWXA
  +058RXPxPg6fy4IWvTNh0E263XmFcJlSAwiGgFAW/dAiS6JXm"
crossorigin="anonymous">
</head>
<body>
  <h1 class="title is-1 mb-4 p-4 bg-dark text-white">Results for Prediction</h1>
  <div class="ml-5">
    <p class="font-weight-bold h3 mb-4">
      Next {{n}} days Prediction (from {{date[8:]}}/{{date[5:7]}}/{{date[0:4]}}) :
    </p>
    <ul class="list-group list-group-flush">
      {%for i in range(0,len)%}
      <li class="list-group-item">Day {{i+1}}: $ {{res[i][0]}}</li>
      {%endfor%}
```

```
    </td>
  </div>
  <div>
  </div>
  <div>
    <!-- <img src=" {{ plot_url }}"> -->
  </div>
</body>
</html>
```

BitcoinPredictionApp.py

```
#!/usr/bin/env python
# coding: utf-8

# In[1]:

# importing essential libraries
from flask import Flask, render_template, request, url_for
from keras.models import model_from_json
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler

# In[2]:

# load json and create model
json_file = open('model.json', 'r')

loaded_model_json = json_file.read()
```

```
json_file.close()
model = model_from_json(loaded_model_json)
# load weights into new model
model.load_weights("NewBitcoinModel.h5")
print("Loaded model from disk")

model.compile(loss='mean_squared_error', optimizer='adam')
# In[3]:
import pickle as pkl
import pandas as pd
# with open("dfe.pkl", "rb") as f:
#     file = pkl.load(f)

# df = pd.DataFrame(file)
# df.to_csv(r'dsAll.csv')

df = pd.read_csv('ds.csv')
df=df.set_index('Date')
df.index = pd.to_datetime(df.index)
#df = pd.read_pickle('./dfe.pkl')



# In[4]:

app= Flask(__name__)

@app.route('/')
def home():
    return render_template('home.html')


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

def predict():
```

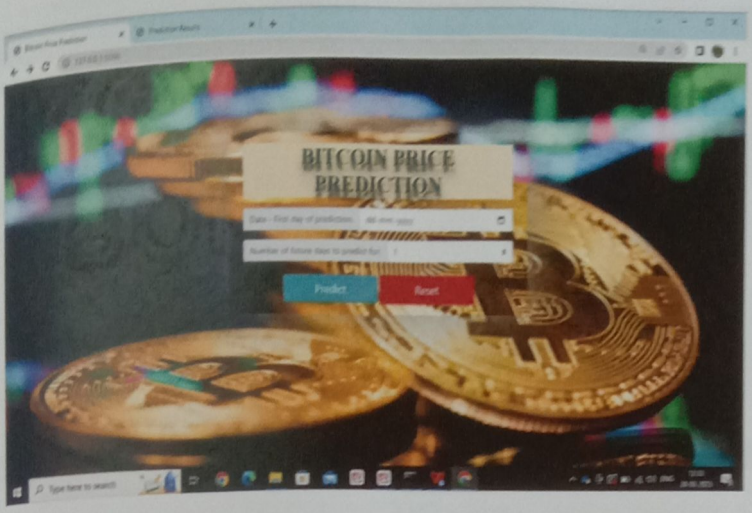
```
if request.method== 'POST':

    date = request.form['date']
    print(date)
    n = int(request.form['n'])
    print(n)

    loc = df.index.get_loc(date)
    print(loc)
    prev_data = df.iloc[loc:loc+15].Price.astype(float)
    #orig_data = df.iloc[loc:loc+n].Price.astype(float)

    min_max_scaler = MinMaxScaler(feature_range=(0, 1))
    ds = min_max_scaler.fit_transform(prev_data.values.reshape(-1, 1))
    ds = ds.reshape(1,15,1)

    look_back=15
    x_input = ds[len(ds)-look_back:].reshape(1,-1)

    temp_input=list(x_input)
    temp_input=temp_input[0].tolist()

    lst_output=[]
    i=0
    #n=15   # next number of days for which we are predicting
    while(i<n):

      if(len(temp_input)>look_back):
        x_input=np.array(temp_input[1:])
        x_input=x_input.reshape(1,-1)
        x_input = x_input.reshape((1, look_back, 1))
        yhat = model.predict(x_input, verbose=0)

temp_input.extend(yhat[0].tolist())
```

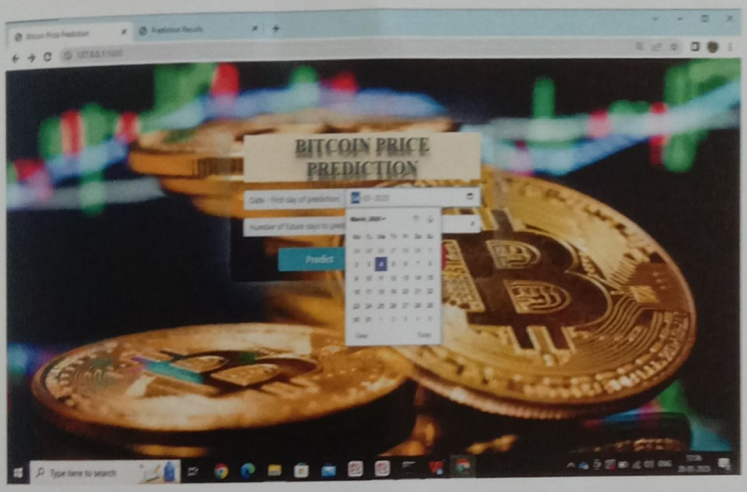
```
            temp_input=temp_input[1:]
            lst_output.extend(yhat.tolist())
            i=i+1
        else:
            x_input = x_input.reshape((1, look_back,1))
            yhat = model.predict(x_input, verbose=0)
            temp_input.extend(yhat[0].tolist())
            lst_output.extend(yhat.tolist())
            i=i+1

    res = min_max_scaler.inverse_transform(lst_output)

    #day_new=np.arange(1, look_back+1)
    #day_pred=np.arange(look_back+1, look_back+n+1)

    #trainScore = math.sqrt(mean_squared_error(orig_data, res))

    return render_template('result.html', res=res, len=len(res), n=n, date=date)
if __name__ == "__main__":
    app.run()
```

FORM DESIGN

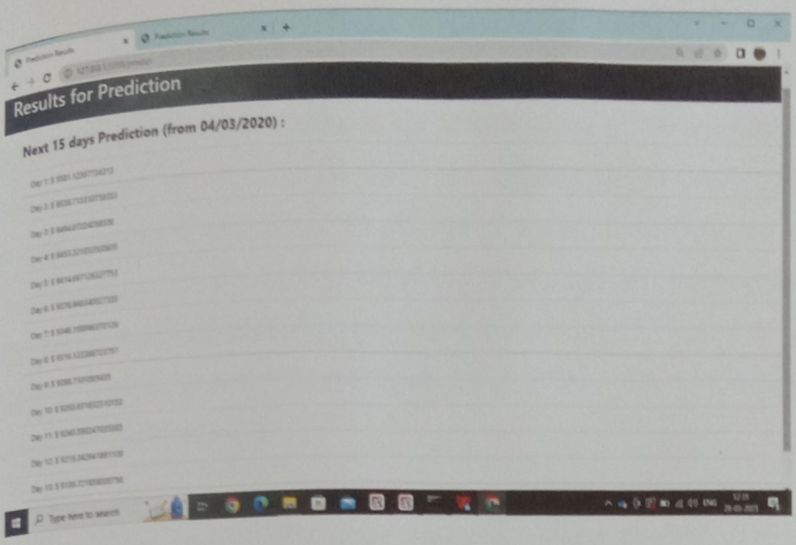
9. FORM DESIGN

9.1 Home page





9.2 Result page



CONCLUSION

CONCLUSION

CONCLUSION

The Bitcoin price prediction project concluded that predicting the future price of Bitcoin is a difficult endeavor due to the volatile nature of the cryptocurrency market. The project identified several factors that influence Bitcoin prices, such as news events, economic conditions, and investor sentiment. While there are many models and methods that attempt to predict Bitcoin prices, none of them are foolproof and they all come with potential risks. As a result, investors should be cautious when investing in Bitcoin and should understand the potential risks associated with investing in the digital currency.

FUTURE ENHANCEMENT

The outcome of this project provides a bitcoin price prediction, In future we use to predict the prices of other cryptocurrencies like litecoin, ether, ripple etc., The web application which is the final software product of this project can provide numerous other applications on single site like sentiment analysis of bitcoin, tutorial to do trading using bitcoin crypto currency.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Rizwan, M., Narejo, S. and Javed, M., 2019, December. Bitcoin price prediction using deep learning algorithm. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)* (pp. 1-7). IEEE.
2. Awoke, T., Rout, M., Mohanty, L. and Satapathy, S.C., 2020. Bitcoin price prediction and analysis using deep learning models. In *Communication Software and Networks: Proceedings of INDIA 2019* (pp. 631-640). Singapore: Springer Singapore.
3. Andi, H.K., 2021. An accurate bitcoin price prediction using logistic regression with LSTM machine learning model. Journal of Soft Computing Paradigm, 3(3), pp.205-217.
4. Rizwan, M., Narejo, S. and Javed, M., 2019, December. Bitcoin price prediction using deep learning algorithm. In 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS) (pp. 1-7). IEEE.
5. https://in.investing.com/crypto/bitcoin/historical-data
6. https://towardsdatascience.com/cryptocurrency-price-prediction-using-lstms-tensorflow-for-hackers-part-iii-264fcdbccd3f
7. https://www.geeksforgeeks.org/bitcoin-price-prediction-using-machine-learning-in-python/
8. https://en.wikipedia.org/wiki/Cryptocurrency

Credit Card Fraud Detection Framework for E-Commerce Sites

A project submitted to

ST. MARY'S COLLEGE (AUTONOMOUS), THOOTHUKUDI.

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY,

TIRUNELVELI

in partial fulfilment of the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Submitted by

U. PARAMESWARI BHARATHI

Reg. No.: 21SPCS05

Under the Supervision and Guidance of

Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc(CS)., MPhil., B.Ed.,

PG DEPARTMENT OF COMPUTER SCIENCE(SSC)

St. Mary's College (Autonomous), Thoothukudi-628001

April 2023

CERTIFICATE

This is to certify that this project work entitled **"CREDIT CARD FRAUD DETECTION FRAMEWORK FOR E-COMMERCE SITES"** is submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli, in partial fulfilment for the award of the degree of Master of Science in Computer Science for the work done during the year 2022-2023 by **U.PARAMESWARI BHARATHI** (Reg. No.: 21SPCS05).

Signature of the Guide

Signature of the Co-ordinator

Signature of the Director
Director
Self Supporting Courses
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Principal
Principal
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Examiner

DECLARATION

I do hereby declare that the project entitled " **CREDIT CARD FRAUD DETECTION FRAMEWORK FOR E-COMMERCE SITES"** submitted for the degree of Master of Science in Computer Science in my original work carried out under the guidance of **Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., MPhil., B.Ed**, Assistant Professor, PG Department of Computer Science(SSC),St. Mary's College(Autonomous), Thoothukudi.

Station: Thoothukudi

Date: 10.04.2023

Signature of the Student

ACKNOWLEDGEMENT

I express my first and foremost thanks to God Almighty for his gracious help and shower of blessings for having rendered us the strength and support to finish our project successfully.

My sincere thanks to Dr. Sr. A. S. J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., Principal, Dr. Sr. C. Shibana M.Sc.,B.Ed., M.Phil., Ph.D., Secretary, Rev. Sr. S. Josephine Jeyarani Director of SSC, St. Mary's College (Autonomous), Thoothukudi, for giving permission to work on this project.

I convey my heartfelt thanks to Ms. C. Nayanthra Mascarenhas M.Sc., M.Phil., SET., Assistant Professor and Coordinator, PG Computer Science (SSC), for her support and counsel.

I express my hearty thanks to my guide Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., M.Phil., B.Ed., Assistant Professor, PG Department of Computer Science (SSC),for her valuable suggestions, gentle guidance, enthusiastic ideas to carry out andcomplete my work entirely effortlessly.

I also express my boundless thanks to Dr. A. Vithya Vijayalakshmi., MCA., M.Phil., Ph.D., Assistant Professor, PG Department of Computer Science (SSC), for her encouragement and support.

I am much indebted to Dr.P.Johnson Durairaj Director, Postulates for his untiring effort, Immense Knowledge and priceless contribution without which I couldn't have finished my work effectively on time.

I thank my family members especially my parents for their encouragement and support both morally and financially which helped me to finish the project successfully.

POSTULATE

01.04.2023

To

The Head of the Department
PG Department of Computer Science (SSC)
St. Mary's College(Autonomous)
Thoothukudi

Dear Sir/Madam,

Sub: Project Work Completion-Reg

On behalf of **POSTULATE**, we are pleasure to inform you that **Ms. U. Parameswari Bharathi, Reg No: 21SPCS05** studying **Master of Computer Science** Final year has been done the project work at our concern Postulate Info Tech Private Limited in the title "**Credit Card Fraud Detection for E-Commerce Sites**" during the period from December 2022 to March 2023.

During this period, the candidate was found diligent and sincere. We wish all the best for future endeavors.

POSTULATE INFO TECH PRIVATE LIMITED

DIRECTOR

ABSTRACT

ABSTRACT

Credit Card Fraud can be defined as a case where a person uses someone else's credit card for personal reasons while the owner and the card-issuing authorities are unaware of the fact that the card is being used. Due to the rise and acceleration of E-Commerce, there has been a tremendous use of credit cards for online shopping which led to High amount of frauds related to credit cards. In the era of digitalization, the need to identify credit card frauds is necessary. Fraud detection involves monitoring and analyzing the behaviour of various users to estimate detect or avoid undesirable behaviour. To identify credit card fraud detection effectively, we need to understand the various technologies, algorithms and types involved in detecting credit card frauds. The algorithm can differentiate transactions which are fraudulent or not. Find fraud, they need to passed dataset and knowledge of the fraudulent transaction. They analyze the dataset and classify all transactions. Fraud detection involves monitoring the activities of populations of users to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting. Machine learning algorithms are employed to analyses all the authorized transactions and report the suspicious ones. These reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent.

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

From the moment the payment systems came to existence, there have always been people who will find new ways to access someone's finances illegally. This has become a major problem in the modern era, as all transactions can easily be completed online by only entering your credit card information. Even in the 2010s, many American retail website users were the victims of online transaction fraud right before two-step verification was used for shopping online. Organizations, consumers, banks, and merchants are put at risk when a data breach leads to monetary theft and ultimately the loss of customers' loyalty along with the company's reputation. Unauthorized card operations hit an astonishing amount of 16.7 million victims in 2017. Additionally, as reported by the Federal Trade Commission (FTC), the number of credit card fraud claims in 2017 was 40% higher than the previous year's number. There were around 13,000 reported cases in California and 8,000 in Florida, which are the largest states per capita for such type of crime. The amount of money at stake will exceed approximately $30 billion by 2020. Here are some credit card fraud statistics.

Machine Learning-based Fraud Detection:

- Detecting fraud automatically
- Real-time streaming
- Less time needed for verification methods
- Identifying hidden correlations in data

Conventional Fraud Detection:

- The rules of making a decision on determining schemes should be set manually.
- Takes an enormous amount of time
- Multiple verification methods are needed; thus, inconvenient for the user
- Finds only obvious fraud activities

Unsupervised:

Unsupervised Machine Learning methods use unlabeled data to find patterns and dependencies in the credit card fraud detection dataset, making it possible to group data samples by similarities without manual labeling.

PCA (Principal Component Analysis) enables the execution of an exploratory data analysis to reveal the inner structure of the data and explain its variations. PCA is one of the most popular techniques for Anomaly Detection.

LOF (Local Outlier Factor) is the score factor that helps understand how high the chance is for a certain data sample to be an outlier (anomaly). This is another of the most popular Anomaly Detection methods.

One-class SVM (Support Vector Machine) is a classification algorithm that helps to identify outliers in data. This algorithm allows one to deal with imbalanced data-related issues such as Fraud Detection.

Isolation Forest (IF) is an Anomaly Detection method from the Decision Trees family. The main idea of IF, which differentiates it from other popular outlier detection algorithms, is that it precisely detects anomalies instead of profiling the positive data points

Supervised:

Supervised ML methods use labelled data samples, so the system will then predict these labels in future unseen before data. Among supervised ML fraud identification methods, we define Decision Trees, Random Forest, KNN, and Naive Bayes.

K-Nearest Neighbors is a Classification algorithm that counts similarities based on the distance in multi-dimensional space. The data point, therefore, will be assigned the class that the nearest neighbors have.

XG Boost (Extreme Gradient Boosting) and Light GBM (Gradient Boosting Machine) are a single type of gradient-boosted Decision Trees algorithm, which was created for speed as well as maximizing the efficiency of computing time and memory resources. This algorithm is a blending technique where new models are added to fix the errors caused by existing models.

"Fraud detection is a set of activities that are taken to prevent money or property from being obtained through false preferences."

Fraud can be committed in different ways and in many industries. The majority of detection methods combine a variety of fraud detection datasets to form a connected overview of both valid and non-valid payment data to make a decision. This decision must consider IP address, geolocation, device identification, "BIN" data, global latitude/longitude, historic transaction patterns, and the actual transaction information. In practice, this means that merchants and issuers deploy analytically based responses that use internal and external data to apply a set of business rules or analytical algorithms to detect fraud.

SYSTEM SPECIFICATION

SYSTEM SPECIFICATION

HARDWARE SPECIFICATION :

- Processors : AMD Athlon Silver 3050U with Radeon Graphics 2.30 GHz
- Operating systems : Windows 11

SOFTWARE SPECIFICATION:

- Server Side : Python 3.7.4(64-bit) or (32-bit)
- Client Side : HTML, CSS, Bootstrap
- IDE : Flask 1.1.1
- Back end : MySQL 5.
- Server : WampServer 2i

SOFTWARE DESCRIPTION:

✓ PYTHON:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages. Python is a MUST for students when they are working in Web Development Domain.

✓ PANDAS:

Pandas is mainly used for data analysis and associated manipulation of tabular data in Data frames. Pandas allows importing data from various file formats such as comma-separated values, JSON, Parquet, SQL database tables or queries, and Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features. The development of pandas introduced into Python many comparable features of working with Data frames that were established in the R programming language. The panda's library is built upon another library NumPy, which is oriented to efficiently working with arrays instead of the features of working on Data frames.

✓ NUMPY:

NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object

✓ **MATPLOTLIB:**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

✓ **SEABORN:**

Seaborn is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures, Visualization is the central part of Seaborn which helps in exploration and understanding of data.

✓ **MySQL:**

MySQL tutorial provides basic and advanced concepts of MySQL. Our MySQL tutorial is designed for beginners and professionals. MySQL is a relational database management system based on the Structured Query Language, which is the popular language for accessing and managing the records in the database. MySQL is open-source and free software under the GNU license. It is supported by Oracle Company.MySQL database that provides for how to manage database and to manipulate data with the help of various SQL queries. These queries are: insert records, update records, delete records, select records, create tables, drop tables, etc. There are also given MySQL interview questions to help you better understand the MySQL database.

✓ **WAMPSERVER:**

WampServer is a Windows web development environment. It allows you to create web applications with Apache2, PHP and a MySQL database. Alongside, PhpMyAdmin allows you to manage easily your database.

✓ **BOOTSTRAP:**

Bootstrap is a free and open-source tool collection for creating responsive websites and web applications. It is the most popular HTML, CSS, and JavaScript framework for developing responsive, mobile-first websites.

✓ **FLASK:**

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 pages, a blog, a wiki or go as big as a web-based calendar application or a commercial website.Flask is often referred to as a micro framework. It aims to keep the core of an application simple yet extensible. Flask does not have built-in abstraction layer for database handling, nor does it have formed a validation support. Instead, Flask supports the extensions to add such functionality.

PROJECT DESCRIPTION

PROJECT DESCRIPTION

My project aims to build a credit card fraud detection model, which tells us if the transaction made by the card is fraud or not. So basically we will use the transaction and their labels as fraud or non-fraud to detect if the new transaction made by the customer is fraud or not.To prevent customers from being charged for the items they did not purchase, It is important for credit card companies to recognize fraudulent credit card transactions.

Implementation is the stage in the project where the theoretical design is turned into a working system. The implementation phase constructs, installs and operates the new system. The most crucial stage in achieving a new successful system is that it will work efficiently and effectively. There are several activities involved while implementing a new project. All projects go through a life cycle beginning with defining how the new software package will be used in our organization (requirements) through the end point of the project – a successful and effective implementation. Our activities matrix has been organized around six generic implementation life.

1.E-Commerce Raw and Transaction Dataset Annotation

1.1.Training Dataset

The dataset is a real dataset of e-commerce transactions and the objective was to detect anomalous e-commerce transactions.

1.2.Testing Dataset

Current purchase details of the e-commerce website customer.

2. Data Pre-processing

In this module selected data is formatted, cleaned and sampled. The data pre-processing steps includes following: a. Formatting: The data which is been selected may not be in a suitable format. The data may be in a file format and we may like it in relational database or vice versa. b. Cleaning: Removal or fixing of missing data is called as cleaning. The dataset may contain record which may be incomplete or it may have null values. Such records need to remove. c. Sampling: As number of frauds in dataset is less than overall transaction, class distribution is unbalanced in credit card transaction. Hence sampling method is used to solve this issue.

3. Clustering

The pre-processed data is clustered using clustering algorithms like K-Nearest Neighbour (KNN) to cluster relevant data in database.KNN algorithm provides a simplest way to classify the given dataset using k no. of clusters. We define k centroids, one for each cluster. In this algorithm we choose k no. of clusters (k fixed), find the centroids and distance between the objects to centroids, then we group them based o minimum distance loop is generated and iterated until there are no more changes to be done to change the centroids position after every loop. The main aim of this algorithm is to minimize the objective function.

4. Feature Selection

Feature selection has been proven to be effective and efficient for machine learning problems. The objectives of feature selection include building simpler and more comprehensible models, improving data mining performance such as predictive accuracy and comprehensibility. Also it includes preparing to remove redundancy and irrelevancy for understandable data. Feature selection can be divided into wrapper and filter method. The wrapper method relies on the predictive performance of a predefined learning algorithm to evaluate the selected features. It repeats the searching step and evaluating criteria until desired learning performance is obtained. The drawback of wrapper method is that the search space could be vast and it is relatively more expensive than other methods. Filter method is independent of any learning algorithms and relies on certain characteristics of data to assess the importance of features. Features are scored based on the scores according to the evaluation criteria, and the lowest scored features are removed. For this reason, we applied filter-based feature selection algorithms for feature selection method, which is the fastest and also suitable for practical use. Feature selection based on filter method can be categorized into ranker and the subset selector.

In the proposed research, we selected eight subset feature selection algorithms and six ranked feature selection algorithms to select features among existing features. Also, we assigned the score to evaluate features based on the frequency. The ranker algorithms are calculated by weighting the higher ranks. The results of two feature selection algorithms are combined to prioritize the features by selecting features which exceed the parameter in frequency and ranking.

The features important for fraud discoveries include:

- Customer's identity (email addresses, credit card numbers, etc.)
- The past order details
- Their preferred payment methods,
- The locations they have used for the transactions

- Their network (emails, phone numbers, and payment details entered with the online account).

5. DNN Classification

A fast-moving risk requires a fast-adapting approach. In this case, two parallel running deep learning models provide the mechanisms to identify fraud both within standalone transactional and demographic data, as well as across time. The two-model approach improves performance, better identifies threats, and reduces false positives that create unnecessary noise:

- Data Analysis: The first model consumes transactional and demographic information, as is from the dataset with a focus on real-time analysis.
- Activity Analysis: To improve accuracy beyond what is available with standalone data the second model analyses historical credit card activity with the appropriate feature to derive in-depth insights.

Deep learning architecture is composed of multi-level hidden layers of nonlinear processing units, where each neuron may send data to a connected neuron within hidden layers. A weighted combination of all input signals is aggregated and then an output signal transmitted by the connected neuron. Deep learning with default parameters using H2O library automatically handles missing values, data standardisation, load balancing, cross-validation, checkpointing and gridsearch. Deep Learning models are built through assessing different representations of raw data with exhibited high performance on complex data such as images, text and speech. There are many hyper parameters in deep learning which are used to tune the models.

6. Prediction

Current purchase details and feed the system historical data, along with markers indicating that the fraudulent and legitimate transactions. In the real-time detection phase, the feature vector is extracted from streaming transaction to decide whether it is fraudulent or not.

7. Blocking

It takes about a millisecond to estimate the probability of fraud. The results are sent to a cardholder's bank where the decision on whether to approve or decline a transaction is made.

8. Performance Analysis

The performance of the proposed classifier is evaluated in terms of 4 classification metrics relevant to credit card fraud detection—fraud detection rate, false alarm rate, balanced classification rate, and Matthews correlation coefficient. The other common metrics like accuracy and error rate are known to be bias metrics in the case of imbalance and hence we did not consider them.

DATABASE DESCRIPTION

DATABASE DESCRIPTION

ADMIN:

Field	Type	Null	Description
username	Varchar(20)	NO	User Name
password	Varchar(20)	NO	User Password

CUSTOMER:

Field	Type	Null	Description
Id	Int(11)	No	User id
Name	Varchar(20)	No	User Name
Card	Varchar(20)	No	User card
Mobile	Bigint(30)	No	User Mobile number
E-mail	Varchar(20)	No	User Email id

REGISTER:

Field	Type	Null	Description
Name	Varchar(20)	No	User Name
Mobile	Bigint(20)	No	User Mobile
E-mail	Varchar(50)	No	User Email
Address	Varchar(40)	No	User Address
Uname	Varchar(20)	No	User Username

PRODUCT:

Field	Type	Null	Description
Product	Varchar(50)	No	Product Name
Price	Double	No	Product Price
Details	Varchar(200)	No	Product Details

PURCHASE:

Field	Type	Null	Description
Id	Int(11)	No	Product id
Uname	Varchar(20)	No	User Name
Amount	Varchar(20)	No	Product Amount
Date	Varchar(20)	No	Date

SYSTEM STUDY

SYSTEM STUDY

EXISTING SYSTEM

The rule-based approach. Fraudulent activities in finance can be detected by looking at on-surface and evident signals. Unusually, large transactions or the ones that happen in atypical locations obviously deserve additional verification. Purely rule-based systems entail using algorithms that perform several fraud detection scenarios, manually written by fraud analysts. Today, legacy systems apply about 300 different rules on average to approve a transaction. That's why rule-based systems remain too straightforward. They require adding/adjusting scenarios manually and can hardly detect implicit correlations. On top of that, rule-based systems often use legacy software that can hardly process the real-time data streams that are critical for the digital space.The existing systems are carried out by considering machine learning algorithms like Support Vector Machine, Naïve Bayes, K-Nearest Neighbor and so on and some of them used random dataset. Very few have used artificial neural network for credit card fraud detection.

PROPOSED SYSTEM

DL-based fraud detection. However, there are also subtle and hidden events in user behaviour that may not be evident, but still signal possible fraud. Deep learning allows for creating algorithms that process large datasets with many variables and help find these hidden correlations between user behaviour and the likelihood of fraudulent actions. Another strength of deep learning system compared to rule-based ones is faster data processing and less manual work. For example, smart algorithms fit well with behavior analytics for helping reduce the number of verification steps.

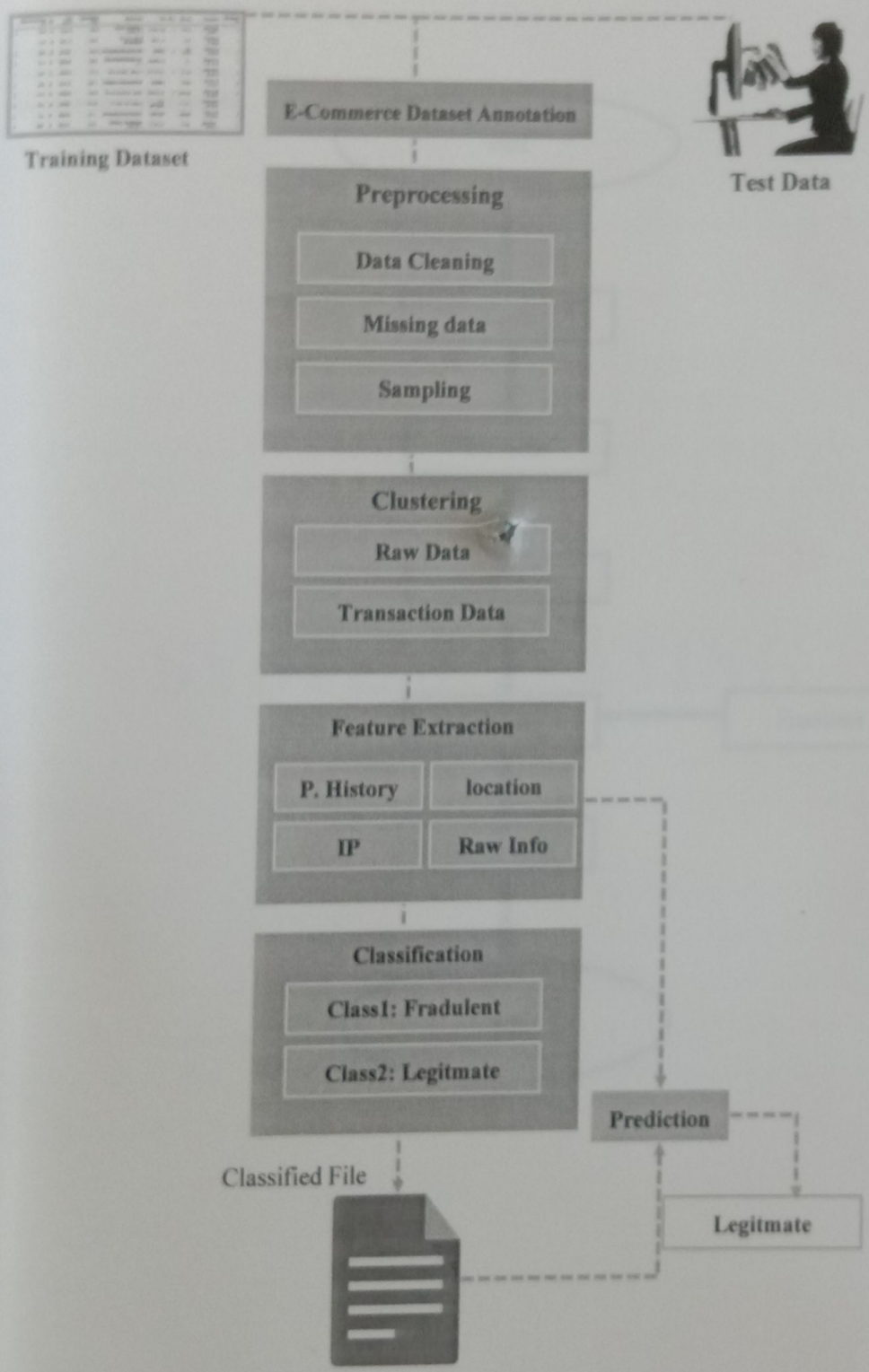
- **Deep Neural Networks**

It is an excellent complement to other techniques and improves with exposure to data. The neural network is a part of cognitive computing technology where the machine mimics how the human brain works and how it observes patterns.

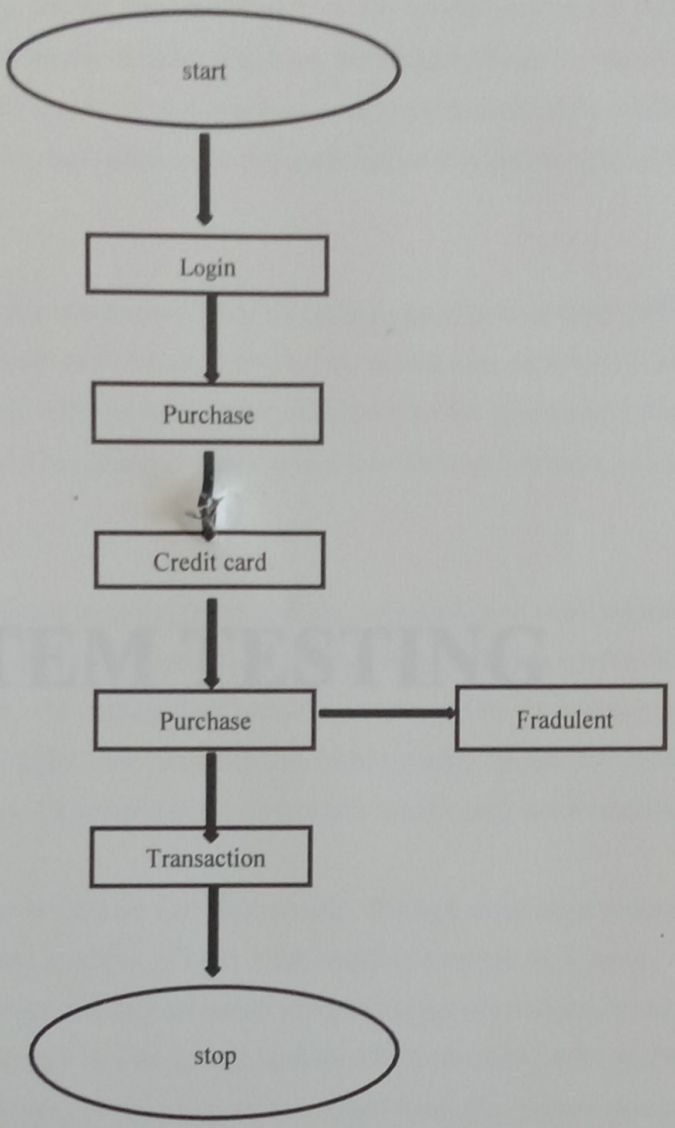
The neural networks are completely adaptive; able to learn from patterns of legitimate behaviour. These can adapt to the change in the behaviour of normal transactions and identify patterns of fraud transactions. The process of the neural networks is extremely fast and can make decisions in real time.

SYSTEM DESIGN

SYSTEM DESIGN



DATA FLOW DIAGRAM



SYSTEM TESTING

SYSTEM TESTING

In this phase of methodology, testing was carried out on the several application modules. Different kind of testing was done on the modules which are described in the following sections. Generally, tests were done against functional and non-functional requirements of the application following the test cases. Testing the application again and again helped it to become a reliable and stable system.

System Testing

In this phase of methodology, testing was carried out on the several application modules. Different kind of testing was done on the modules which are described in the following sections. Generally, tests were done against functional and non-functional requirements of the application following the test cases. Testing the application again and again helped it to become a reliable and stable system.

Unit Testing

Before you can test an entire software program, make sure the individual parts work properly on their own. Unit testing validates the function of a unit, ensuring that the inputs (one to a few) result in the lone desired output. This testing type provides the foundation for more complex integrated software. When done right, unit testing drives higher quality application code and speeds up the development process. Developers often execute unit tests through test automation.

Integration Testing

Integration testing is often done in concert with unit testing. Through integration testing, QA professionals verify that individual modules of code work together properly as a group. Many modern applications run on microservices, self-contained applications that are designed to handle a specific task. These microservices must be able to communicate with each other, or the application won't work as intended. Through integration testing, testers ensure these components operate and communicate together seamlessly.

CODING

CODING

```
import os
from flask import Flask, render_template, Response, redirect, request, session, abort, url_for
import cv2
import shutil
import datetime
import random
from random import seed
from random import randint
import PIL.Image
from PIL import Image
import mysql.connector
from werkzeug.utils import secure_filename
mydb = mysql.connector.connect(
host="localhost",
  user="root",
  password="",
  charset="utf8",
  database="credit_card"
)
app = Flask(__name__)
app.secret_key = 'abcdef'
UPLOAD_FOLDER = 'static/upload'
ALLOWED_EXTENSIONS = { 'png', 'jpg', 'jpeg', 'gif'}
app.config['UPLOAD_FOLDER'] = UPLOAD_FOLDER
@app.route('/')
def index():
@app.route('/login_admin', methods=['GET', 'POST'])
def login_admin():
    msg=""
    if request.method=='POST':
        uname=request.form['uname']
```

```
    pwd=request.form['pass']
    cursor = mydb.cursor()
    cursor.execute('SELECT * FROM admin WHERE username = %s AND password = %s',
(uname, pwd))
    account = cursor.fetchone()
    if account:
      #session['loggedin'] = True
      #session['username'] = account['username']
      # Redirect to home page
      return redirect(url_for('admin'))
    else:
      # Account doesnt exist or username/password incorrect
      msg = 'Incorrect username/password!'
  return render_template('login_admin.html',msg=msg)
@app.route('/login_bank', methods=['GET', 'POST'])
def login_bank():
  msg=""
  if request.method=='POST':
    uname=request.form['uname']
    pwd=request.form['pass']
    cursor = mydb.cursor()
    cursor.execute('SELECT * FROM admin WHERE username = %s AND password = %s',
(uname, pwd))
    account = cursor.fetchone()
    if account:
      #session['loggedin'] = True
      #session['username'] = account['username']
      # Redirect to home page
      return redirect(url_for('bank'))
    else:
      # Account doesnt exist or username/password incorrect
      msg = 'Incorrect username/password!'
  return render_template('login_bank.html',msg=msg)
@app.route('/admin', methods=['GET', 'POST'])
```

```
def admin():
    cursor = mydb.cursor()
    cursor.execute('SELECT * FROM register')
    data = cursor.fetchall()
    return render_template('admin.html',data=data)
@app.route('/bank', methods=['GET', 'POST'])
def bank():
    cursor = mydb.cursor()
    cursor.execute('SELECT * FROM customer')
    data = cursor.fetchall()
    return render_template('bank.html',data=data)
@app.route('/add_cus', methods=['GET', 'POST'])
def add_cus():
    msg=""
    if request.method=='POST':
        name=request.form['name']
        card=request.form['card']
        mobile=request.form['mobile']
        email = request.form['email']
        mycursor = mydb.cursor()
        mycursor.execute("SELECT max(id)+1 FROM customer")
        maxid = mycursor.fetchone()[0]
        if maxid is None:
            maxid=1
            sql = "INSERT INTO customer(id, name, card, mobile, email) VALUES (%s, %s, %s,
%s, %s)"
        val = (maxid, name, card, mobile, email)
        mycursor.execute(sql,val)
        mydb.commit()
        msg="Added success"ursor.execute('select * from product where product like %s || details
like %s',(val, val ))
    else:
        cursor.execute('SELECT * FROM product')
    data = cursor.fetchall()
```

```
    if request.method=='POST':
    txt=request.form['txt']
    cursor1 = mydb.cursor()
    likeString = "%" + txt + "%"
    cursor1.execute('SELECT * FROM product where product like %s',(likeString, ))
    data = cursor1.fetchall()
  return render_template('home2.html',data=data)
@app.route('/analyze', methods=['GET', 'POST'])
def analyze():
  txt=0
  f2=open("mess.txt","r")
  ms=f2.read()
  return render_template('analyze.html',txt=ms)
@app.route('/register', methods=['GET', 'POST'])
def register():
  #import student
  msg=""
  if request.method=='POST':
    name=request.form['name']
    mobile=request.form['mobile']
    email=request.form['email']
    address=request.form['address']
    uname=request.form['uname']
    pass1=request.form['pass']
    cursor = mydb.cursor()
    sql = "INSERT INTO register(name,mobile,email,address,uname,pass) VALUES (%s, %s,
%s, %s, %s, %s)"
    val = (name,mobile,email,address,uname,pass1)
    cursor.execute(sql, val)
    mydb.commit()
    print(cursor.rowcount, "Registered Success")
    result="sucess"
    if cursor.rowcount==1:
      return redirect(url_for('login'))
```

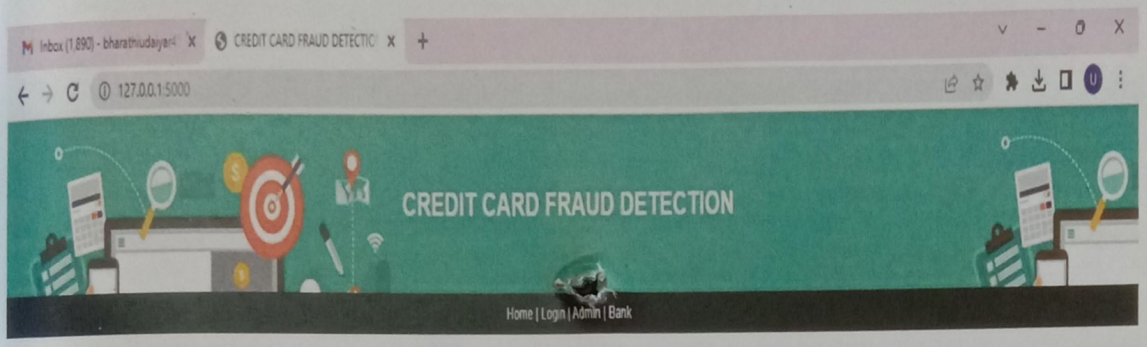
```
    else:
        msg='Already Exist'
    return render_template('register.html',msg=msg)
def view():
    act=""
    pid=""
    if 'username' in session:
        uname = session['username']
        if request.method=='GET':
        act = request.args.get('act')
        pid = request.args.get('pid')
        cursor1 = mydb.cursor()
        cursor1.execute('SELECT count(*) FROM selected WHERE uname=%s && pid = %s',
(uname, pid))
        num = cursor1.fetchone()[0]
        if num==0:
            #print("s")
            mycursor = mydb.cursor()
            mycursor.execute("SELECT max(id)+1 FROM selected")
            maxid = mycursor.fetchone()[0]
            if maxid is None:
                maxid=1 if 'username' in session:
        uname = session['username']
    if request.method=='GET':
        act = request.args.get('act')
        pid = request.args.get('pid')
        if act is None:
            act="1"
        ss=int(act)+1
        if ss==5:
            act="1"
        else:
            act=str(ss)
        cursor1 = mydb.cursor()
```

```
    cursor1.execute('SELECT count(*) FROM selected WHERE uname=%s && pid = %s',
(uname, pid))
    num = cursor1.fetchone()[0]
    if num==0:
      #print("s")
      mycursor = mydb.cursor()
      mycursor.execute("SELECT max(id)+1 FROM selected")
      maxid = mycursor.fetchone()[0]
      if maxid is None:
        maxid=1
      sql = "INSERT INTO selected(id, uname, pid, status) VALUES (%s, %s, %s, %s)"
      val = (maxid, uname, pid, '0')
      mycursor.execute(sql,val)
      mydb.commit()
    else:
      f2=open("mess.txt","r")
      ms=f2.read()
      if ms!="":
        cursor2 = mydb.cursor()
        cursor2.execute('update selected set status=%s WHERE uname=%s and pid = %s', (ms,
uname, pid))
        mydb.commit()
    cursor = mydb.cursor()
    cursor.execute('SELECT * FROM product WHERE id = %s', (pid, ))
    data = cursor.fetchall()
  return render_template('view2.html',act=act, pid=pid, data=data)
########################################
def TND(mean = 0.0, std = 1.0, low = 0.0, high = 10.0):
 a = ( low - mean ) / std
 b = ( high - mean ) / std
 return scst.truncnorm(a, b, loc = mean, scale = std)
```

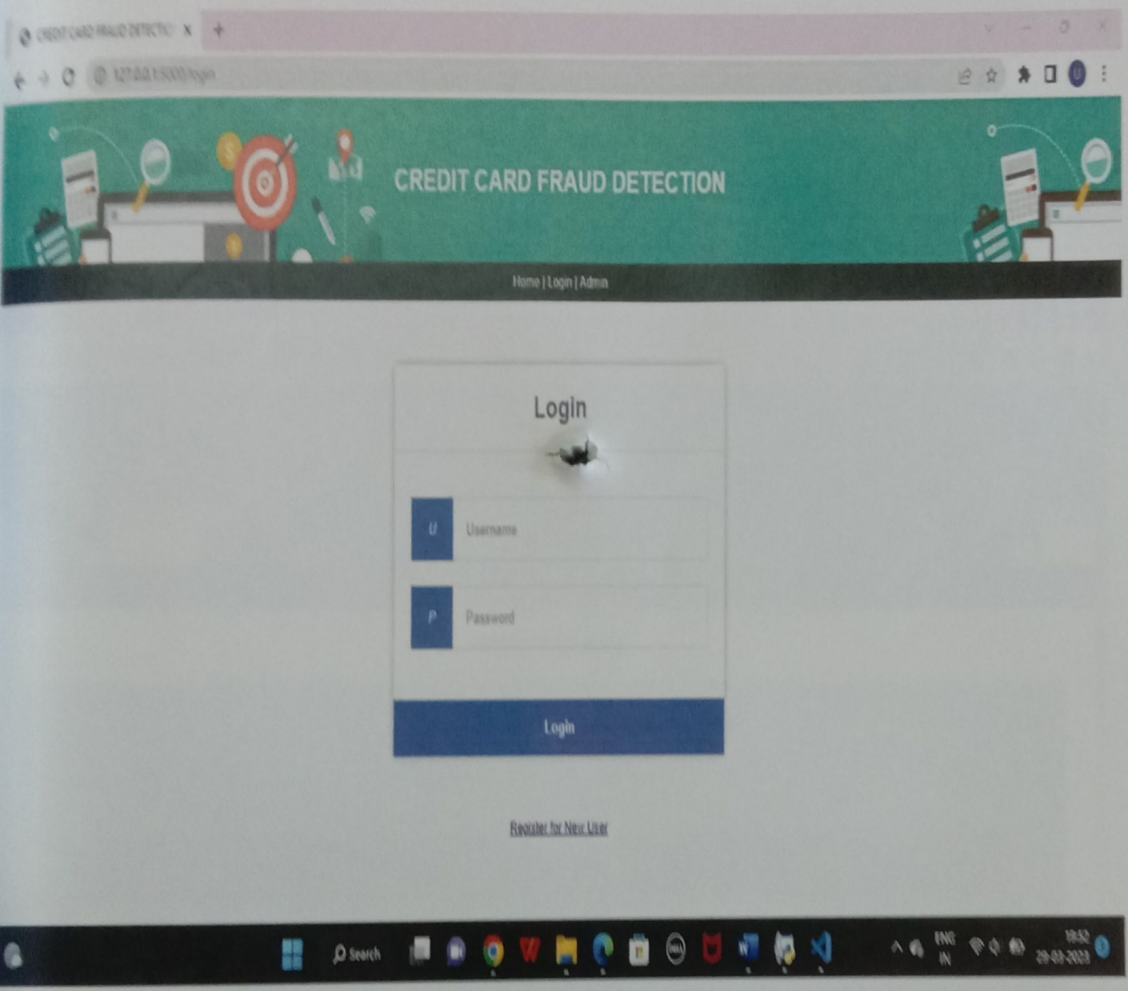
SCREENSHOTS

SCREENSHOTS

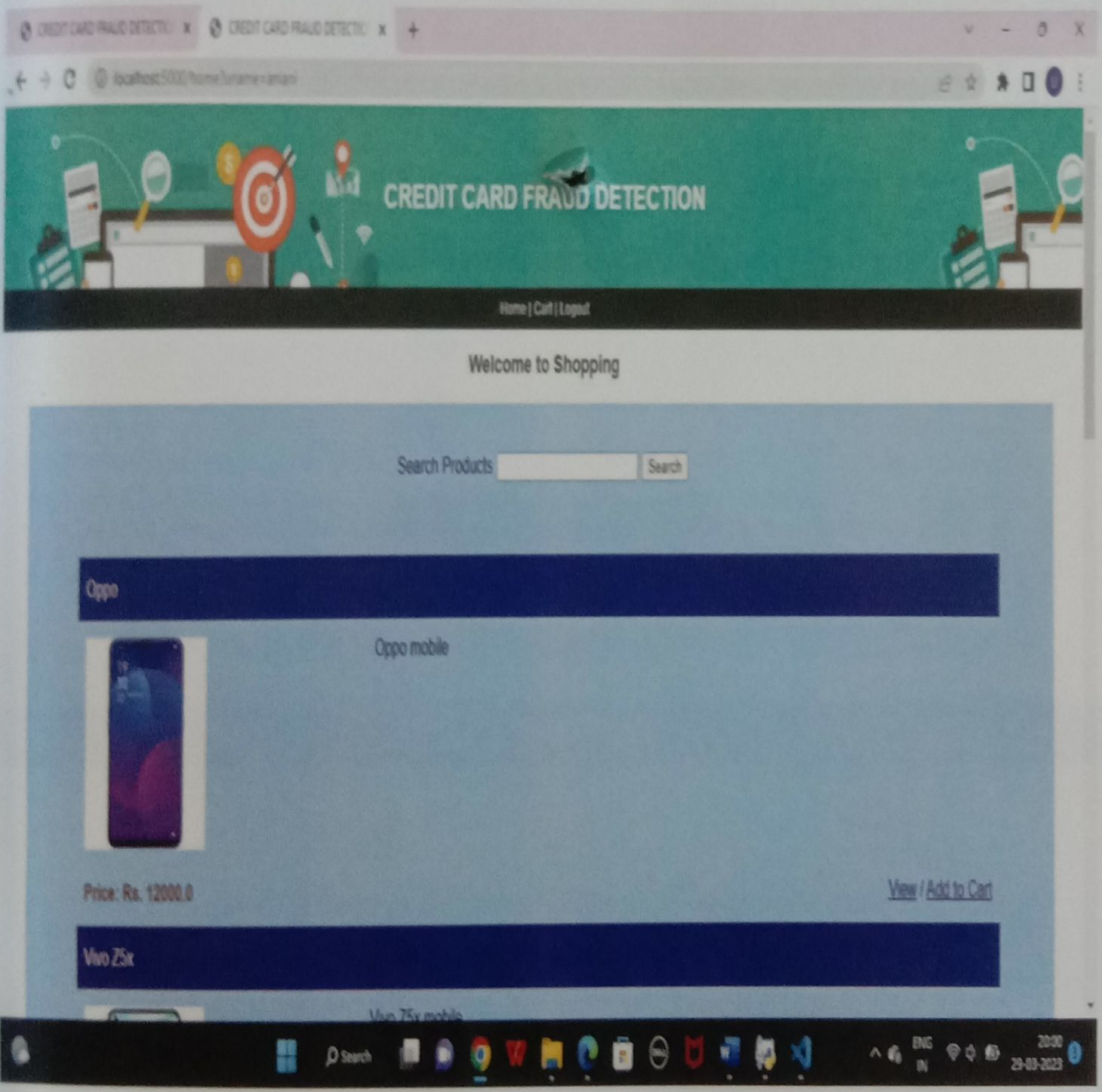
Home page:



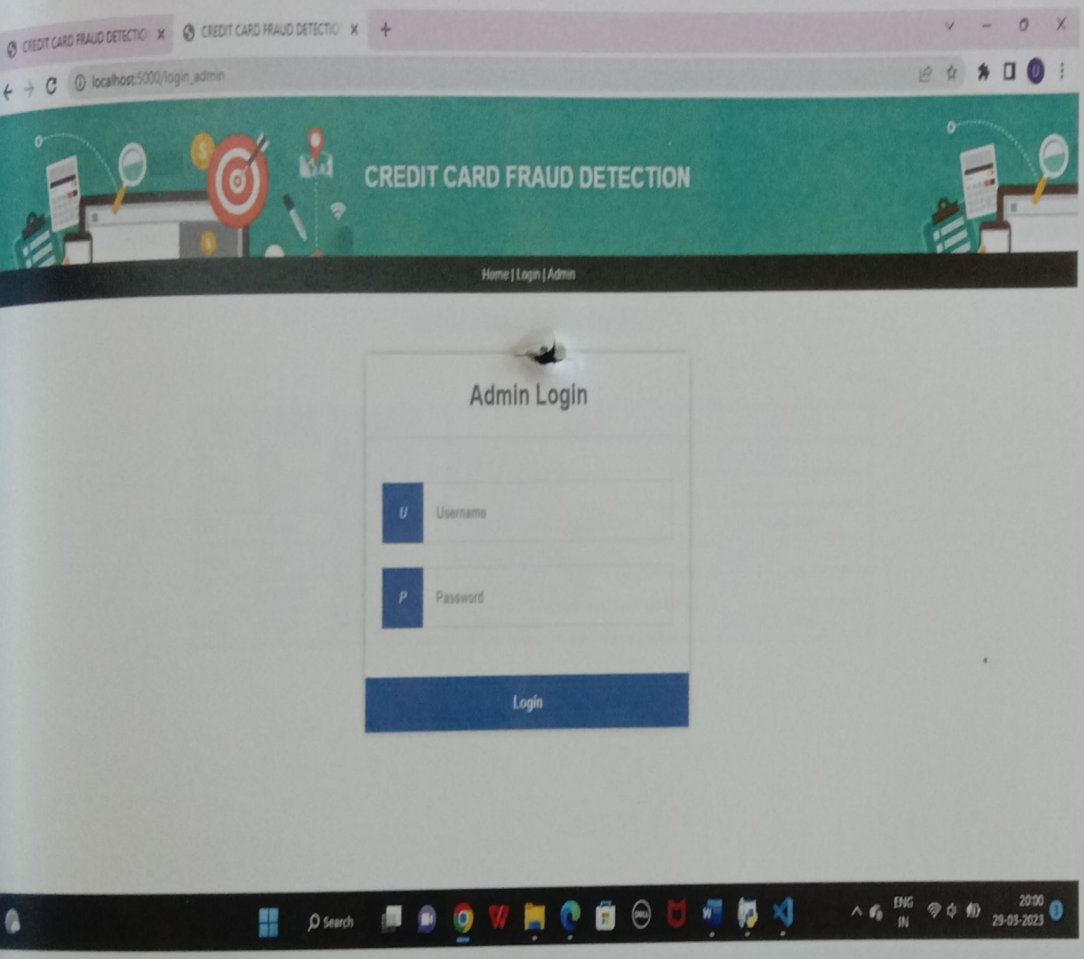
Login page:


CREDIT CARD FRAUD DETECTION
Home | Login | Admin
Login
Username
Password
Login
Register for New User

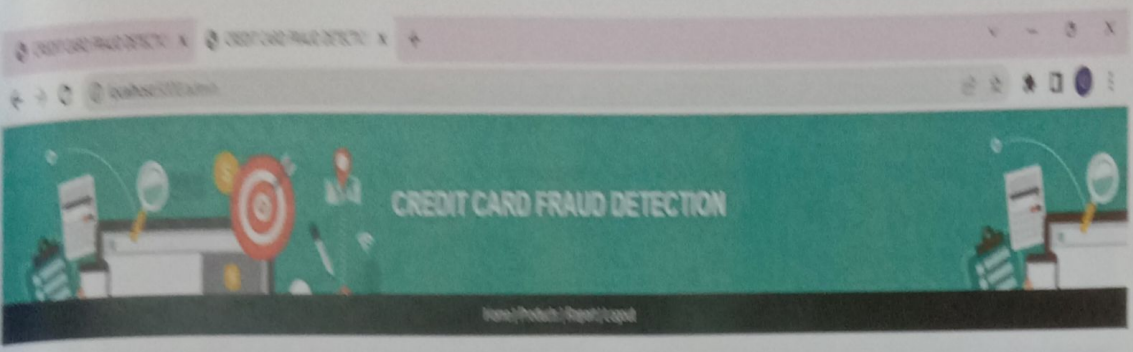
Shopping page:



Admin page:


CREDIT CARD FRAUD DETECTION
localhost:5000/login_admin
CREDIT CARD FRAUD DETECTION
Home | Login | Admin
Admin Login
Username
Password
Login
Search
ENG
IN
20:00
29-03-2023

User info:

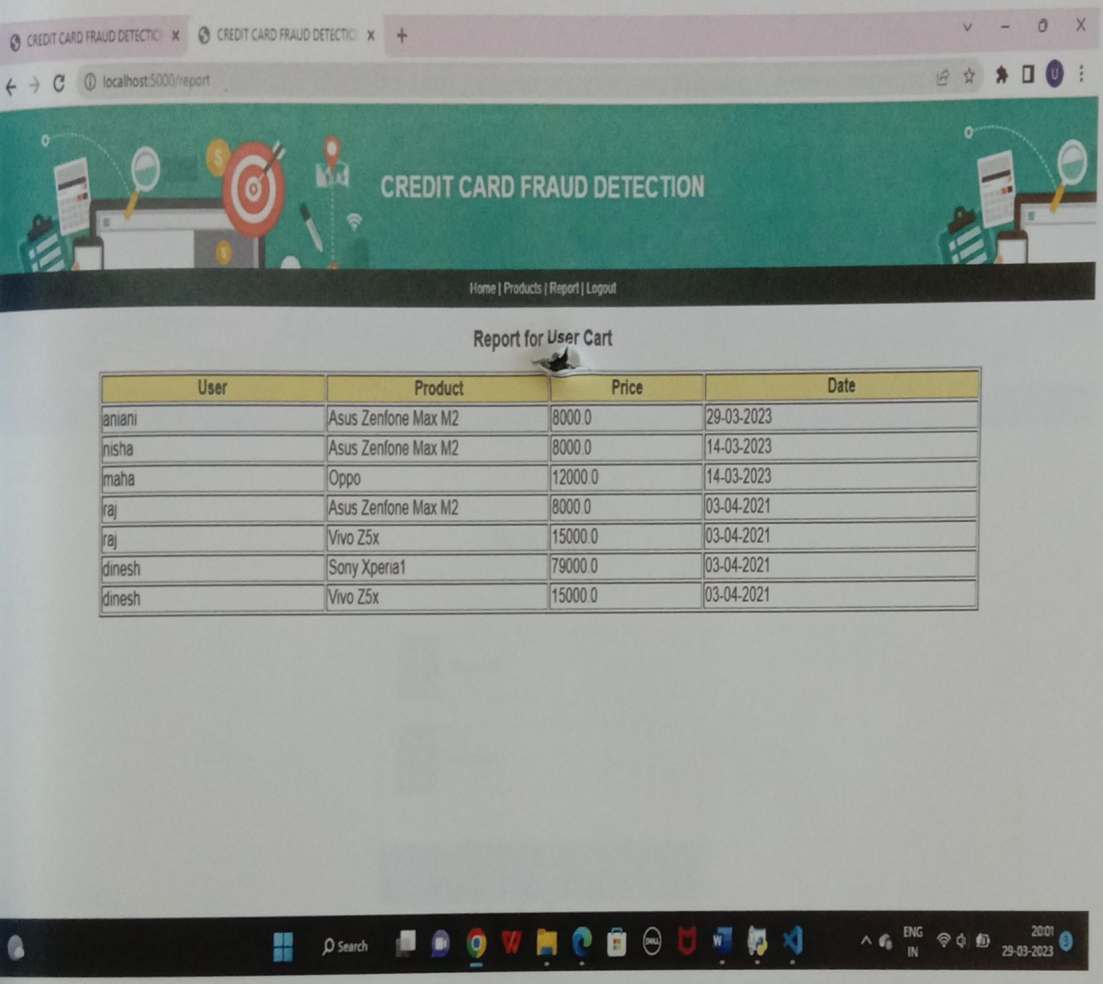


Welcome to Administrator

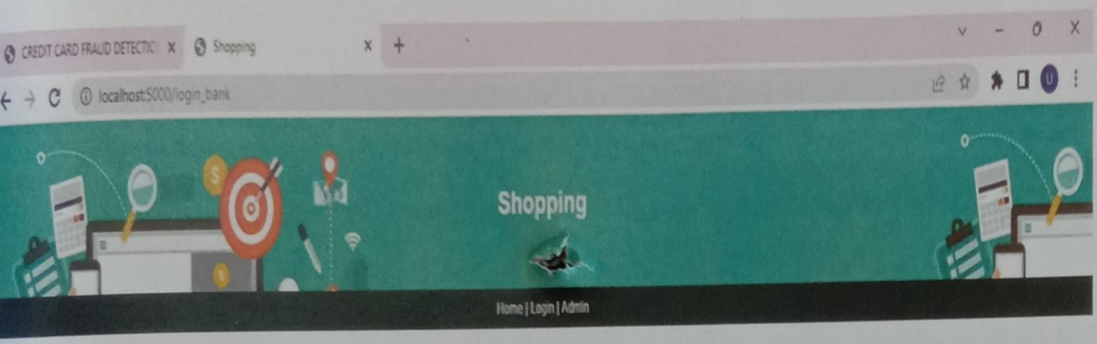
User Information

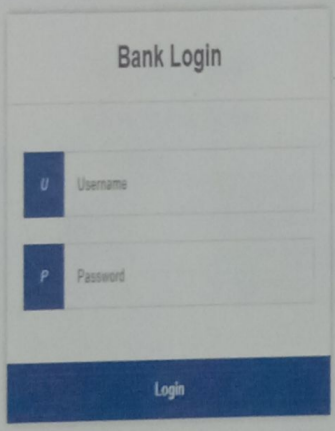
User	Name	Mobile No.	E-mail	Address
raj	Raj	9874573435	raj@gmail.com	22 chetram
dinesh	Dinesh	8369867387	dinesh@gmail.com	Chennai
ishi	Insha	6379631722	bharathudaiyar455@gmail.com	north beach road
maha	bhar	6379631722	bharathudaiyar455@gmail.com	north road
bbb	bharath	9500875475	msharalee3@gmail.com	hgp colony
anian	Anitha	9856732192	ani@gmail.com	annanagar west

Report page:


CREDIT CARD FRAUD DETECTION
localhost:5000/report
CREDIT CARD FRAUD DETECTION
Home | Products | Report | Logout
Report for User Cart
User
Product
Price
Date
aniani
Asus Zenfone Max M2
8000.0
29-03-2023
nisha
Asus Zenfone Max M2
8000.0
14-03-2023
maha
Oppo
12000.0
14-03-2023
raj
Asus Zenfone Max M2
8000.0
03-04-2021
raj
Vivo Z5x
15000.0
03-04-2021
dinesh
Sony Xperia1
79000.0
03-04-2021
dinesh
Vivo Z5x
15000.0
03-04-2021
Search
ENG
IN
20:01
29-03-2023

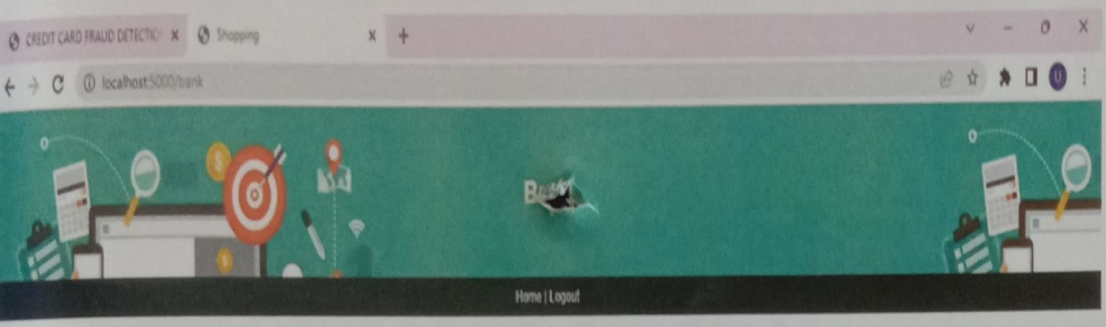
Bank:


CREDIT CARD FRAUD DETECTIC
Shopping
localhost:5000/login_bank
Shopping
Home | Login | Admin


Bank Login
U
Username
P
Password
Login


Search
ENG
IN
20:01
29-03-2023

Credit Card Details:



Welcome to Bank

Customer Information (Add New)

Name	Card No.	Mobile No.	E-mail
Raj	111222333	9974573435	raj@gmail.com
Dinesh	111222101	8248656967	diensh@gmail.com
nisha	123456	9500875475	nisharafee3@gmail.com
Anitha	5678	9856732190	ani@gmail.com

CONCLUSION

CONCLUSION

Credit card fraud is a rapidly growing problem due which financial institutions are losing huge amount of money. Researchers are implementing various new techniques to enhance the credit card fraud detection systems so that credit card frauds can be decreased. Many machine learning methods have been implemented to prevent credit card frauds. Deep learning is a branch of machine learning that is used in many fields like image recognition, speech recognition and many more. Deep learning provides a way to explore complex features within the data so that the model can learn better to predict frauds more efficiently with less false alarms.

FUTURE ENHANCEMENT

FUTURE ENHANCEMENT

- Preventing known and unknown fraud in real time is not easy but it is feasible.
- The proposed architecture was originally designed to detect credit card fraud in online payments.
- To emphasize on providing a fraud prevention mechanism to verify the transaction is fraudulent or legitimate.
- To keep your data safe, delete spam emails asking for personal information and keep your antivirus and anti-spyware software up to date.
- Shop online only with secure web pages.

BIBLIOGRAPHY

BIBLIOGRAPHY

REFERENCES

[1]. Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5916–5923, 2013.

[2]. A. O. Adewumi and A. A. Akinyelu, "A survey of machine-learning and nature-inspired based credit card fraud detection techniques," *International Journal of System Assurance Engineering and Management*, vol. 8, pp. 937–953, 2017.

[3]. A. Srivastava, A. Kundu, S. Sural, A. Majumdar, "Credit card fraud detection using hidden Markov model," *IEEE Transactions onDependable and Secure Computing*, vol. 5, no. 1, pp. 37–48, 2008.

[4]. J. T. Quah, and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1721–1732, 2008.

[5]. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C., "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.

Website link:

- https://github.com/ecloudvalley/Credit-card-fraud-detection-with-SageMaker-using-TensorFlow-estimators
- https://data-flair.training/blogs/credit-card-fraud-detection-python-machine-learning/
- https://www.ibm.com/in-en/topics/deep-learning
- https://memgraph.com/blog/how-to-develop-a-credit-card-fraud-detection-application-using-memgraph-flask-and-d3js
- https://github.com/devrajkataria/Credit-Card-Fraud-Detection-ML-WebApp/blob/master/app.py

REAL TIME SENTIMENT ANALYSIS

A project submitted to

ST. MARY'S COLLEGE (Autonomous), THOOTHUKUDI

Affiliated to

MANONMANIAM SUNDARANAR UNIVERSITY,

TIRUNELVELI

In partial fulfilment of the award of the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

Submitted by

R. RESHMA

Reg.No.:21SPCS06

Under the Supervision and Guidance of

Dr. A. Vithya Vijayalakshmi MCA., M.Phil., Ph.D.,

PG DEPARTMENT OF COMPUTER SCIENCE(SSC)

St. Mary's College (Autonomous), Thoothukudi-628001

APRIL 2023

CERTIFICATE

This is to certify that this project work entitled **"REAL TIME SENTIMENT ANALYSIS"** is submitted to St. Mary's College (Autonomous), Thoothukudi affiliated to Manonmaniam Sundaranar University, Tirunelveli, in partial fulfilment for the award of the degree of Master of Science in Computer Science for the work done during the year 2022-2023 by **R.RESHMA** (Reg. No.: 21SPCS06).

Signature of the Guide

Signature of the Co-ordinator

Signature of the Director

Director
Self Supporting Courses
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Principal

Principal
St. Mary's College (Autonomous)
Thoothukudi - 628 001.

Signature of the Examiner

DECLARATION

I do hereby declare that the project entitled **"REAL TIME SENTIMENT ANALYSIS"** submitted for the degree of Master of Science in Computer Science is my original work carried out under the guidance of **Dr. A. Vithya Vijayalakshmi MCA., MPhil., Ph.D.**, Assistant Professor, PG Department of Computer Science (SSC), St. Mary's College (Autonomous), Thoothukudi.

R. Reshma

Place: Thoothukudi **Signature of the Student**

Date: 10.04.2023

ACKNOWLEDGEMENT

I express my first and foremost thanks to God Almighty for his gracious help and shower of blessings for having rendered us the strength and support to finish our project successfully.

My sincere thanks to Dr. Sr. A. S. J. Lucia Rose M.Sc., PGDCA., M.Phil., Ph.D., Principal, Dr. Sr. C. Shibana M.Sc., B.Ed., M.Phil., Ph.D., Secretary, Rev. Sr. S. Josephine Jeyarani, Director of SSC, St. Mary's College (Autonomous), Thoothukudi, for giving permission to work on this project.

I convey my heartfelt thanks to Ms. C. Nayanthra Mascarenhas M.Sc., M.Phil., SET., Assistant Professor and Coordinator, PG Computer Science (SSC), for her support and counsel.

I express my hearty thanks to my guide Dr. A. Vithya Vijayalakshmi MCA., Mphil., Ph.D., Assistant professor, PG Department of Computer Science (SSC) for her valuable suggestions, gentle guidance, enthusiastic ideas, to carry out andcomplete my work entirely effortlessly.

I also express my boundless thanks to Ms. A. Jenitta Jebamalar M.Sc (IT)., M.Sc (CS)., M.Phil., B.Ed., Assistant Professor, PG Department of Computer Science (SSC), for her encouragement and support.

I am much indebted to Dr. P. Johnson Durai Raj, Director, Postulate InfoTech for his untiring effort, Immense Knowledge and priceless contribution without which I couldn't have finished my work effectively on time.

I thank my family members especially my parents for their encouragement and support both morally and financially which helped me to finish the project successfully.

POSTULATE

01.04.2023

To

The Head of the Department

PG Department of Computer Science (SSC)

St. Mary's College(Autonomous)

Thoothukudi

Dear Sir/Madam,

Sub: Project Work Completion-Reg

On behalf of **POSTULATE**, we are pleasure to inform you that **Ms. R. Reshma, Reg No: 21SPCS06** studying **Master of Computer Science** Final year has been done the project work at our concern Postulate Info Tech Private Limited in the title "**Real Time Sentiment Analysis**" during the period from December 2022 to March 2023.

During this period, the candidate was found diligent and sincere. We wish all the best for future endeavors.

POSTULATE INFO TECH PRIVATE LIMITED

DIRECTOR

ABSTRACT

ABSTRACT

A sentiment refers to the reflection of emotions of people. Today's world stands on the strings of emotions. People express happiness, sadness, love, hatred etc. through some actions. Division of emotions i.e., positive, neutral and negative, is called sentiment analysis. Nowadays there is a sentiment rich data in the form of tweets, status updates, blog posts, reviews, comments, forums for discussion etc. If we efficiently work upon this bucket full of sentiment rich data, it gives way in apprehending the opinions, views or perspective of masses in a specific functional area. Moreover, the result of this analysis will aid people in taking suitable actions or corrective measures for their growth. This effort of ours is like a drop in the ocean to try to analyze the reviews posted by people. Further, the reviews are processed and analyzed using machine learning procedures, algorithms and other related aspects. Finally, the conclusion is derived by finding the polarity of a particular review whether it is poor, average or excellent. This entire task was performed using Python.

TABLE OF CONTENTS

INTRODUCTION

1. INTRODUCTION

Sentiment Analysis is a use case of Natural Language Processing (NLP) and comes under the category of text classification. To put it simply, Sentiment Analysis involves classifying a text into various sentiments, such as positive or negative, Happy, Sad or Neutral, etc. Thus, the ultimate goal of sentiment analysis is to decipher the underlying mood, emotion, or sentiment of a text. This is also known as Opinion Mining.

The main purpose of sentiment analysis is to analyze the opinions of users of a particular product or service, which helps customers understand the quality of the product. For example, every time Apple releases a new iPhone, we see a lot of people giving their opinion on it, some like it and some criticize it, in the end, all people's opinions help us decide whether we should buy the new iPhone or not.

Opinions are nothing more than people's feelings about a particular product, which helps a business improve its product and helps customers decide whether or not to buy the product. Now, what if you want to analyze people's feelings in real-time, i.e. ask a user about your product and understand your product in real-time.

Python is one of the most powerful tools when it comes to performing data science tasks it offers a multitude of ways to perform sentiment analysis. The most popular ones are enlisted here:

1. Using Text Blob
2. Using Vader
3. Using Bag of Words Vectorization-based Models
4. Using LSTM-based Models
5. Using Transformer-based Models

Using LSTM-Based Models

Though we were able to obtain a decent accuracy score with the Bag of Words Vectorization method, it might fail to yield the same results when dealing with larger datasets. This gives rise to the need to employ deep learning-based models for the training of the sentiment analysis model.

For NLP tasks are generally use RNN-based models since they are designed to deal with sequential data. Here, I'll train an LSTM (Long Short Term Memory) model using TensorFlow with Keras. The steps to perform sentiment analysis using LSTM-based models are as follows:

1. Pre-Process the text of training data (Text pre-processing involves Normalization, Tokenization, Stopwords Removal, and Stemming/Lemmatization.)
2. Import Tokenizer from Keras. preprocessing.text and create its object. Fit the tokenizer on the entire training text (so that the Tokenizer gets trained on the training data vocabulary). Generated text embeddings using the texts_to_sequence() method of the Tokenizer and store them after padding them to an equal length. (Embeddings are numerical/vectorized representations of text. Since I cannot feed my model with the text data directly, first need to convert them to embeddings)
3. After having generated the embeddings this is ready to build the model. Build the model using TensorFlow—add Input, LSTM, and dense layers to it. Add dropouts and tune the hyperparameters to get a decent accuracy score. Generally, I tend to use ReLU or LeakyReLU activation functions in the inner layers of LSTM models as it avoids the vanishing gradient problem. At the output layer, I use Softmax or Sigmoid activation function.

Python

Python is a high level interpreted language used for general purpose programming. It is widely used for scientific computing and can be used for a wide variety of general tasks from data mining to software development. Python is the main language used for this project. This allows for quick prototyping and debugging, and makes it easier to use for non-programmers. It has a wide range of libraries and frameworks for developing applications, from web development to data science. Python is an object-oriented language, meaning that it is composed of objects and classes. This makes it easier to develop complex applications, since objects can be reused and modified. The language features an easy-to-learn syntax and a large standard library, which is the collection of functions and modules that come with the language. Python is also used for scripting and automation. This makes it possible to automate repetitive tasks, such as collecting data from websites and running simulations. The language also provides powerful tools for data analysis and machine learning. Python is often used to develop artificial intelligence and deep learning

applications. Python is an open source language, meaning that anyone can contribute to it and all of its source code is freely available. This makes it easy to learn and use, and encourages collaboration and development. Python is also platform independent, meaning that applications written in Python can be used on any operating system.

Python is an excellent choice for developing applications and automating tasks. It is easy to learn and use, and provides a wide range of libraries and frameworks for developing applications. It is also an open source language, meaning that anyone can contribute to its development. Finally, it is platform independent, making it possible to use applications written in Python on any operating system. Python is a high level interpreted language used for general purpose programming. It is widely used for scientific computing and can be used for a wide variety of general tasks from data mining to software development. Python is the main language used for this project.

Pandas

Pandas is also a library in Python, like numpy is also used for data pre-processing and preparation. One of the main features about pandas is the DataFrame and Series data structure. These data structures are optimized and contain fancy indexing that allow a variety of features such as reshaping, slicing, merging, joining and etc to be available. Pandas and Numpy are extremely powerful when used together for manipulating data. Pandas is a Python library that provides high-level data structures and a wide variety of tools for data analysis. It is designed to make data manipulation and analysis easy and intuitive. It is a powerful tool for manipulating, analyzing, and transforming data. Pandas is built on top of the popular Python programming language, so it is easy to use and understand. It is a data analysis library that provides a variety of functions and methods to manipulate data. It provides a wide range of data structures and tools for working with structured data. It also provides a high-level interface for creating and manipulating data frames, which are collections of data. Pandas is used for data manipulation and analysis in many different fields, including finance, economics, biology, and social science. It is a powerful tool for working with large datasets and has become the industry standard for data analysis in Python. Pandas is highly efficient at performing data manipulation tasks. It can quickly and easily read in and process structured data, such as CSV and Excel files. It also provides an intuitive interface for data cleaning and wrangling. It can also be used to create data visualizations and perform statistical analysis. Pandas is a powerful library for data analysis and is used by many companies and data scientists. It is a great tool for working with large datasets and can be used to

quickly and easily process and analyze data. Pandas is the perfect tool for data science projects, allowing users to quickly and effectively manipulate data and perform statistical analysis.

Matplotlib

Matplotlib is a Python plotting library that allows programmers to create a wide variety of graphs and visualizations with ease of use. The great feature about Matplotlib is that it integrates very well with Jupyter Notebook and creating visualizations is simplified. Matplotlib also works very well with pandas and numpy. Matplotlib's high-level plotting API provides a convenient way to create and customize plots. For example, the pyplot module allows users to quickly create line graphs, histograms, and scatter plots. The library is highly customizable, with numerous settings and parameters that can be adjusted to create a customized plot. For example, users can set the figure size, add labels and annotations, set the font size and style, and adjust the colors used in the plot. Matplotlib also supports a variety of backends, allowing the user to select which backend to use in order to generate a plot. Matplotlib is used in a wide variety of applications, ranging from data analysis to scientific visualization. It is often used in conjunction with other libraries such as NumPy, SciPy, and pandas to create powerful visualizations. It is also commonly used in machine learning applications to visualize the accuracy of a model's predictions. Overall, Matplotlib is an incredibly powerful and versatile plotting library for Python. It is easy to use and highly customizable, making it a great tool for creating professional plots and charts.Matplotlib is a Python plotting library that allows programmers to create a wide variety of graphs and visualizations with ease of use. The great feature about Matplotlib is that it integrates very well with Jupyter Notebook and creating visualizations is simplified. Matplotlib also works very well with pandas and numpy.

Anaconda

Anaconda is a distribution of the 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. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012.[8] As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and

Anaconda Enterprise Edition, both of which are not free.

Tensor Flow

TensorFlow is an open-source deep learning library by Google. It was originally developed by Google's engineers who were working on Google Brain and has been used for research on machine learning and deep learning. TensorFlow at it is core is about computations of multidimensional arrays called tensors but what makes TensorFlow great is its ability to be flexible to deploy computations on different devices such as CPU's and GPU's.

TensorFlow Architecture

TensorFlow architecture works in three parts:

- Preprocessing the data
- Build the model
- Train and estimate the model

It is called TensorFlow because it takes input as a multi - dimensional array, also known as tensors. You can construct a sort of flowchart of operations (called a Graph) that you want to perform on that input. The input goes in at one end, and then it flows through this system of multiple operations and comes out the other end as output.

This is why it is called TensorFlow because the tensor goes in it flows through a list of operations, and then it comes out the other side.

Keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Keras contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on Git Hub, and community support forums include the Git Hub issues page, and a Slack channel. Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of

distributed training of deep learning models on clusters of Graphics Processing Units (GPU).

TKinter

Tkinter is a python based graphical user interface (GUI) library. It is the most commonly used library for the development of GUI applications. Tkinter is a standard package in python and is very easy to learn. It provides a powerful object-oriented interface to the Tk GUI toolkit. Tkinter is used to create a variety of GUI applications such as games, dialog boxes, and more.

Tkinter provides a wide range of widgets. These widgets can be used to create graphical user interfaces (GUIs) for applications. Widgets are the basic building blocks of a GUI application. They provide the user with a way to interact with the application. The most commonly used widgets are buttons, labels, entry fields, radio buttons, check boxes, scrollbars, and spin boxes.

Tkinter also provides a selection of geometry managers which are used to arrange widgets in a window. These geometry managers can be used to create different layouts for your application. The most commonly used geometry managers are pack, grid, and place.

Tkinter also provides a variety of events which can be used to create interactive applications. Events are used to detect user interaction with widgets. For example, a button click event is used to detect when a user clicks on a button. Tkinter also provides a set of tools for creating graphical images. These tools can be used to create simple images or complex animations.

SYSTEM SPECIFICATION

2. SYSTEM SPECIFICATION

HARDWARE REQUIREMENTS:

- Processor : Pentium-IV
- Speed : 1.1 GHZ
- RAM : 512 GB
- Hard Disk : 40GB

SOFTWARE REQUIREMENTS:

- Operating System : Windows 8/7/10
- Tools : Spyder (python)
- Front End : TKinter
- Back End : TensorFlow

PROJECT DESCRIPTION

3. PROJECT DESCRIPTION

Preprocessing

Preprocessing simply refers to perform series of operations to transform or change data. It is transformation applied to our data before feeding it to algorithm. Data processing refers to perform operations on data to retrieve, transform, or change data, especially by computer.

Text preprocessing

Text preprocessing involves transforming text into a clean and consistent format that can then befed into a model for further analysis and learning. Text preprocessing techniques may be general so that they are applicable to many types of applications, or they can be specialized for a specific task.

The various text preprocessing steps are:

1. Tokenization
2. Lower casing
3. Stop words removal
4. Stemming
5. Lemmatization

Tokenization:

Splitting the sentence into words.

Lower casing:

Converting a word to lower case (NLP -> nlp). Words like Book and book meanthe same but when not converted to the lower case those two are represented as two different words in the vector space model (resulting in more dimensions).

Stop words removal:

Stop words are very commonly used words (a, an, the, etc.) in thedocuments. These words do not really signify any importance as they do not help in distinguishing two documents.

Stemming:

It is a process of transforming a word to its root form.

Lemmatization:

Unlike stemming, lemmatization reduces the words to a word existing in thelanguage.

Word Embedding:

Here, I use word embedding process. In natural language processing (NLP), a word embedding is a representation of a word. The embedding is used in text analysis. Typically, the representation is a real-valued vector that encodes the meaning of the word in such a way that words that are closer in the vector space are expected to be similar in meaning.

Build the Text Classifier:

Embedding Layer:

The embedding layer maps the words to their embedding vectors from theembedding matrix. I can keep this layer as trainable, which will learn the word embedding itself.

Spatial Dropout:

Spatial Dropout is a type of dropout for convolutional networks. For a givenconvolution feature tensor of size × height × width, I perform only dropout trials and extend thedropout value across the entire feature map.

LSTM:

LSTM is one of the Recurrent Neural Networks used to efficiently learn long-term dependencies. With LSTM, you can easily process sequential data such as video, text, speech, etc. LSTM modules consist of gate layers that act as key drivers to control information in neural networks.

Uses:

Long Short-Term Memory (LSTM) network is a type of Recurrent Neural Network (RNN) that iswidely used for learning sequential data prediction problems. As every other neural network LSTM also has some layers which help it to learn and recognize the pattern for better performance.

Dropout:

Dropout is one of the most effective and most commonly used regularization techniques for NNs and consists of randomly turning off hidden units during training, that way thenetwork does not rely 100% on all its neurons and instead, forces itself to find more meaningful patterns in the data in order to increase the metric you're trying to optimize.

Dense:

Dense Layer is used to classify image based on output from convolutional layers. EachLayer in the Neural Network contains neurons, which compute the weighted average of its inputand this weighted average is passed through a non-linear function, called as an "activation function".

Train the Model:

Here I train the sentiment analysis model with 5 epochs on the whole dataset. Our data is getting trained within the given epoch limit until it reaches the better accuracy rate. The best accuracy rate should be between the limit 0.9 to 1.

SYSTEM STUDY

4. SYSTEM STUDY

EXISTING SYSTEM

It is challenging to understand the latest trends and summarize the state or general opinions about products due to the big diversity and size of social media data, and this creates the need of automated and real time opinion extraction and mining. Mining online opinion is a form of sentiment analysis that is treated as a difficult text classification task. the role of text pre-processing in sentiment analysis, and report on experimental results that demonstrate that with appropriate feature selection and representation, sentiment analysis accuracies using support vector machines (SVM) in this area may be significantly improved. The level of accuracy achieved is shown to be comparable to the ones achieved in topic categorization although sentiment analysis is considered to be a much harder problem in the literature.

PROPOSED SYSTEM

In this project classify sentiments with the help of machine learning and natural language processing (NLP) algorithms, I use the datasets from Kaggle which was crawled from the internet and labelled positive/negative. The data provided comes with emoticons (emoji), usernames and hashtags which are required to be processed (so as to be readable) and converted into a standard form. I use various machine learning algorithms based on NLP (Natural Language Processing) to conduct sentiment analysis using the extracted features. Finally, I report my experimental results and findings at the end.

SYSTEM ANALYSIS

5. SYSTEM ANALYSIS

Feasibility Study

A feasibility study is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into account various constraints within which the system should be implemented and operated. In this stage, the resource needed for the implementation such as computing equipment, manpower and costs are estimated. The estimated are compared with available resources and a cost benefit analysis of the system is made. The feasibility analysis activity involves the analysis of the problem and collection of all relevant information relating to the project. The main objectives of the feasibility study are to determine whether the project would be feasible in terms of economic feasibility, technical feasibility and operational feasibility and schedule feasibility or not. It is to make sure that the input data which are required for the project are available. Thus I evaluated the feasibility of the system in terms of the following categories:

- Technical feasibility
- Operational feasibility
- Economic feasibility
- Schedule feasibility

Technical Feasibility

Evaluating the technical feasibility is the trickiest part of a feasibility study. This is because, at the point in time there is no any detailed designed of the system, making it difficult to access issues like performance, costs(on account of the kind of technology to be deployed) etc. A number of issues have to be considered while doing a technical analysis; understand the different technologies involved in the proposed system. Before commencing the project, I have to be very clear about what are the technologies that are to be required for the development of the new system. Is the required technology available? Our system is technically feasible since all the required tools are easily available.

Operational Feasibility

Proposed project is beneficial only if it can be turned into information systems that will meet the operating requirements. Simply stated, this test of feasibility asks if the system will work when it is developed and installed. Are there manor barriers to implementation? The proposed was to make a simplified web application. It is simpler to operate and can be used in any webpages. It is free and not costly to operate.

Economic Feasibility

Economic feasibility attempts to weigh the costs of developing and implementing a new system, against the benefits that would accrue from having the new system in place. This feasibility study gives the top management the economic justification for the new system. A simple economic analysis which gives the actual comparison of costs and benefits are much more meaningful in this case. In addition, this proves to be useful point of reference to compare actual costs as the project progresses. There could be various types of intangible benefits on account of automation. These could increase improvement in product quality, better decision making, and timeliness of information, expediting activities, improved accuracy of operations, better documentation and record keeping, faster retrieval of information. This is a web-based application. Creation of application is not costly.

Schedule feasibility

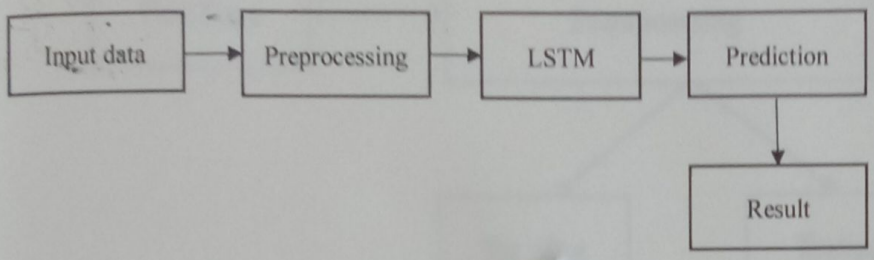
A project will fail if it takes too long to be completed before it is useful. Typically, this means estimating how long the system will take to develop, and if it can be completed in a given period of time using some methods like payback period. Schedule feasibility is a measure how reasonable the project timetable is, given our technical expertise, are the project deadlines reasonable? Some project is initiated with specific deadlines. It is necessary to determine whether the deadlines are mandatory or desirable.

A minor deviation can be encountered in the original schedule decided at the beginning of the project. The application development is feasible in terms of schedule.

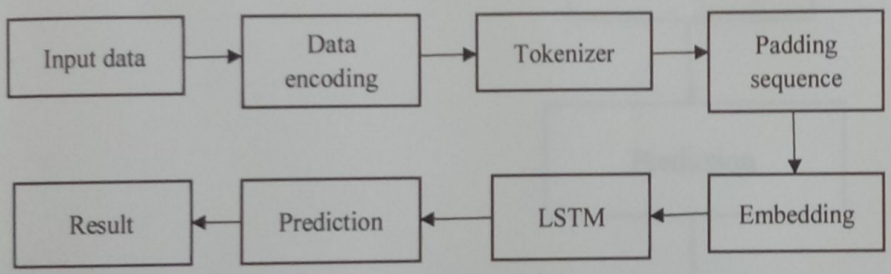
SYSTEM DESIGN

6. SYSTEM DESIGN

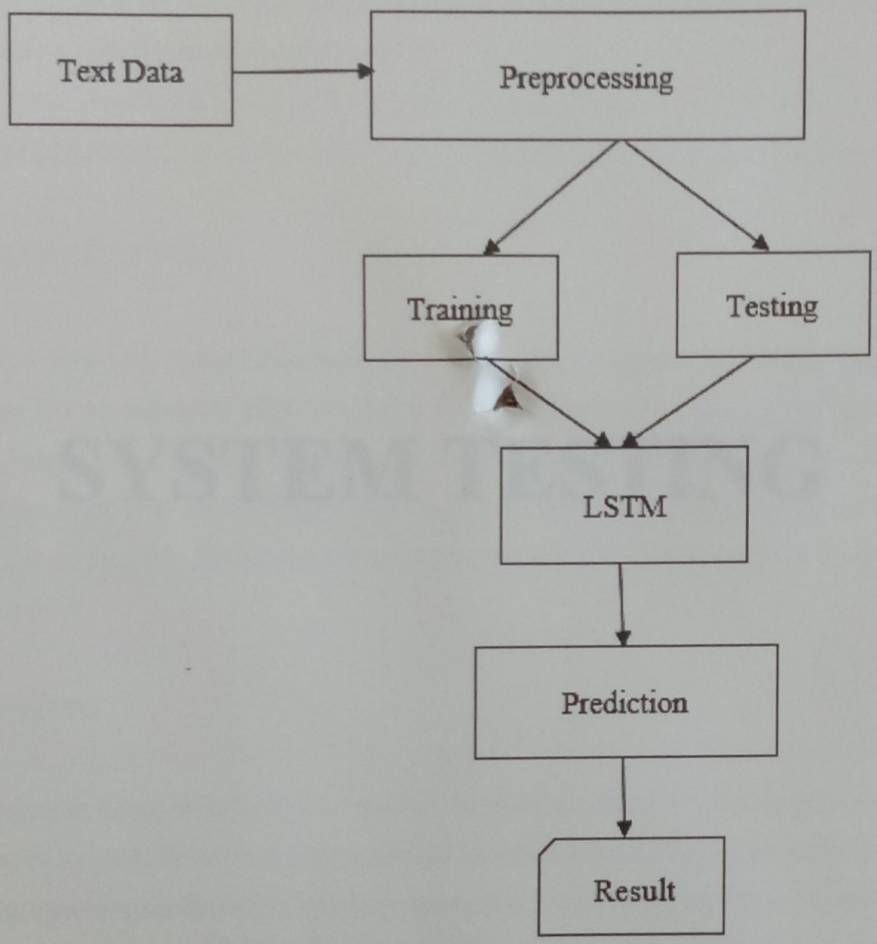
INPUT DESIGN



OUTPUT DESIGN



FLOW CHART


Text Data
Preprocessing
Training
Testing
LSTM
Prediction
Result

SYSTEM TESTING

7. SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, subassemblies, assemblies and/or a finished product it is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTING

REGRESSION TESTING

Regression Testing is defined as a type of software testing to confirm that a recent program or code change has not adversely affected existing features. Regression Testing is nothing but a full or partial selection of already executed test cases that are re-executed to ensure existing functionalities work fine. This testing is done to ensure that new code changes do not have side effects on the existing functionalities. It ensures that the old code still works once the latest code changes are done.

LOAD TESTING

Load testing is a type of performance testing that simulates a real-world load on any software, application, or website. Without it, your application could fail miserably in real-world conditions. That's why I build tools like retrace to help you monitor application performance and fix bugs before your code ever gets to production. Load testing examines how the system behaves during normal and high loads and determines if a system, piece of software, or computing device can handle high loads given a high demand of end-users. This tool is typically applied when a software development project nears completion.

FUNCTIONAL TESTING

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

- Valid Input : identified classes of valid input must be accepted.
- Invalid Input : identified classes of invalid input must be rejected.
- Functions : identified functions must be exercised.
- Output : identified classes of application outputs must be exercised.
- Systems Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing.

RECOVERY TESTING

Recovery Testing is software testing technique which verifies software's ability to recover from failures like software/hardware crashes, network failures etc. The purpose of Recovery Testing is to determine whether software operations can be continued after disaster or integrity loss. Recovery testing involves reverting back software to the point where integrity was known and reprocessing transactions to the failure point.

MIGRATION TESTING

Document related to migration is also reviewed by the testing team who is performing migration testing. Testing team verifies if the document is clear and there is no complexity to perform the activities related to migration. Any kind of documentation error whether it is related to the application, database, server and operating system, can be reported by the testing team. Time taken to migrate the system completely known as downtime is recorded properly and it is very useful when the application is launched in production. Once the migration activities are completed as per migration documentation, a basic check of application functionality is done and basic database check-ups are done to ensure that database is properly connected to the application.

SYSTEM IMPLEMENTATION

Implementation is the most crucial stage in achieving a successful system and giving the user's

confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system. Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly, Initially as a first step the executable form of the application is to be created and loaded in the common server machine which is accessible to the entire user and the server is to be connected to a network. The final stage is to document the entire system which provides components and the operating procedures of the system.

CODING

8. CODING

Realtimesentiment.py

```
import pandas as pd
import matplotlib.pyplot as plt

from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM,Dense, Dropout, SpatialDropout1D
from tensorflow.keras.layers import Embedding

df = pd.read_csv("Tweets.csv")
df.head()
df.columns
tweet_df = df[['text','airline_sentiment']]
print(tweet_df.shape)
tweet_df.head(5)
tweet_df = tweet_df[tweet_df['airline_sentiment'] != 'neutral']
print(tweet_df.shape)
tweet_df.head(5)
tweet_df["airline_sentiment"].value_counts()
sentiment_label = tweet_df.airline_sentiment.factorize()
print(sentiment_label)
tweet = tweet_df.text.values
tokenizer = Tokenizer(num_words=5000)
tokenizer.fit_on_texts(tweet)
vocab_size = len(tokenizer.word_index) + 1
encoded_docs = tokenizer.texts_to_sequences(tweet)
padded_sequence = pad_sequences(encoded_docs, maxlen=200)
print(tokenizer.word_index)
print(tweet[0])
print(encoded_docs[0])
```

```
print(padded_sequence[0])
embedding_vector_length = 32
model = Sequential()

model.add(Embedding(vocab_size, embedding_vector_length, input_length=200) )
model.add(SpatialDropout1D(0.25))
model.add(LSTM(50, dropout=0.5, recurrent_dropout=0.5))
model.add(Dropout(0.2))
model.add(Dense(1, activation='sigmoid'))
model.compile(loss='binary_crossentropy',optimizer='adam', metrics=['accuracy'])
print(model.summary())
history = model.fit(padded_sequence,sentiment_label[0],validation_split=0.2, epochs=5,
    batch_size=32)
plt.plot(history.history['accuracy'], label='acc')
plt.plot(history.history['val_accuracy'], label='val_acc')
plt.legend()
plt.show()
plt.savefig("Accuracy plot.jpg")
plt.plot(history.history['loss'], label='loss')
plt.plot(history.history['val_loss'], label='val_loss')
plt.legend()
plt.show()
plt.savefig("Loss plot.jpg
```

tkinderwindow.py

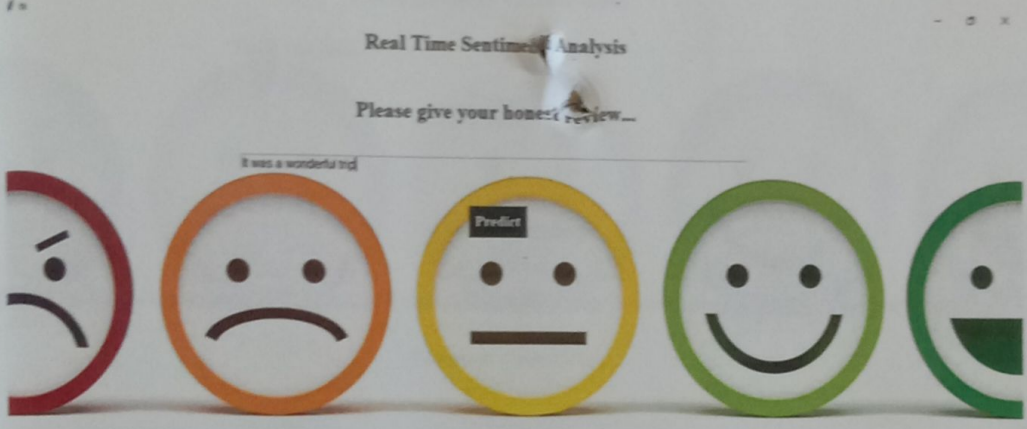
```
import tkinter as tk
import Realtimesentiment as rts
from PIL import Image, ImageTk

root= tk.Tk()

canvas1 = tk.Canvas(root, width=1800, height=1200, relief='raised')
canvas1.pack()

img= tk.PhotoImage(file='image.png')
img_label= tk.Label(root,image=img)
img_label.place(x=0, y=0,relwidth=1,relheight=1)

label1 = tk.Label(root, text='Real Time Sentiment Analysis',bg='white')
label1.config(font=('Times New Roman', 20))
canvas1.create_window(750, 25, window=label1)

label2 = tk.Label(root, text='Please give your hones review...',bg='white')
label2.config(font=('Times New Roman', 20))
canvas1.create_window(750, 125, window=label2)

entry1 = tk.Entry(root,width=90, font=15)
canvas1.create_window(750, 200, window=entry1)
canvas1.pack(padx=22,pady=20)

# def predict_sentiment():
#     tw = rts.tokenizer.texts_to_sequences([text])
#     tw = rts.pad_sequences(tw,maxlen=200)
#     prediction = int(rts.model.predict(tw).round().item())
#     print("Predicted label: ", rts.sentiment_label[1][prediction])
def predict_sentiment():
    x1 = entry1.get()
```

```
    tw = rts.tokenizer.texts_to_sequences([x1])
    tw = rts.pad_sequences(tw,maxlen=200)
    prediction = int(rts.model.predict(tw).round().item())
    print("Predicted label: ", rts.sentiment_label[1][prediction])

    label3 = tk.Label(root, text='Review is : ' + rts.sentiment_label[1][prediction], font=('Times
New Roman', 20),bg='white')
    canvas1.create_window(750, 520, window=label3)

    #label4 = tk.Label(root, text=float(x1)**0.5, font=('helvetica', 10, 'bold'))
    #canvas1.create_window(200, 230, window=label4)

button1 = tk.Button(text='Predict', command=predict_sentiment, bg='black', fg='white',
font=('Times New Roman', 15, 'bold'))
canvas1.create_window(750, 300, window=button1)

root.mainloop()
```

SCREENSHOTS

9. SCREEN SHOTS

TEXT POSITIVE COMMENT

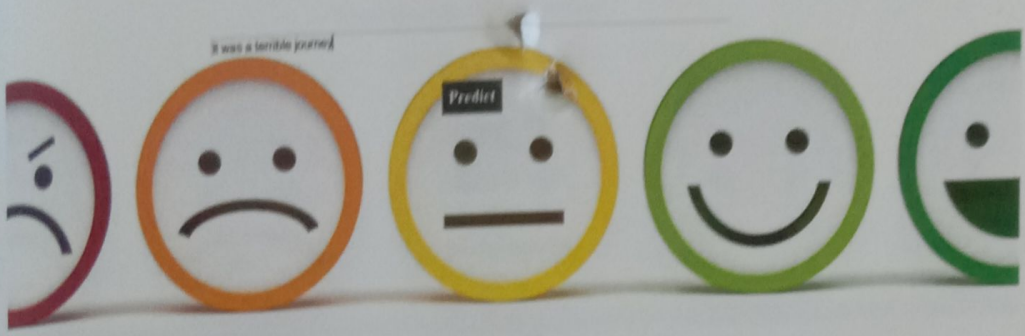


Review is : positive

TEXT NEGATIVE COMMENT

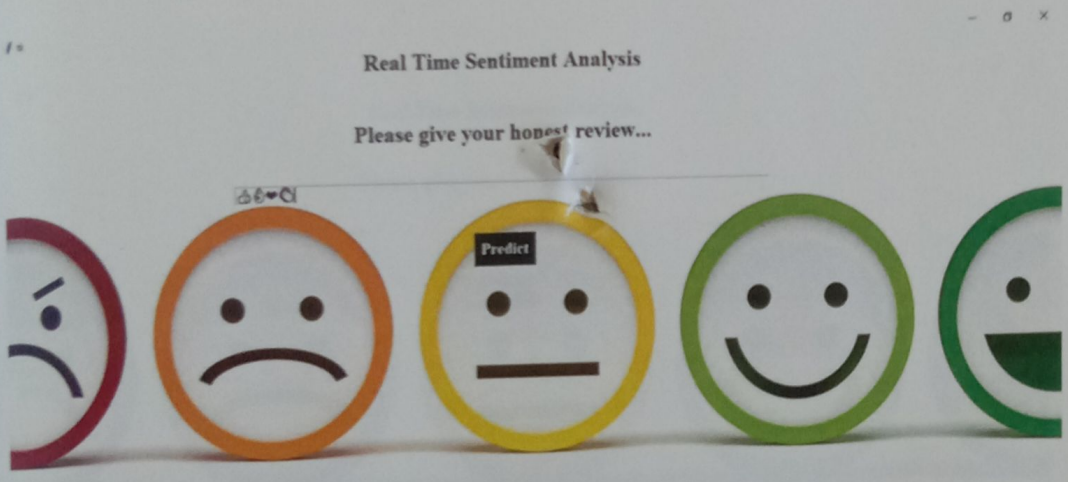
Real Time Sentiment Analysis

Please give your honest review...



Review is : negative

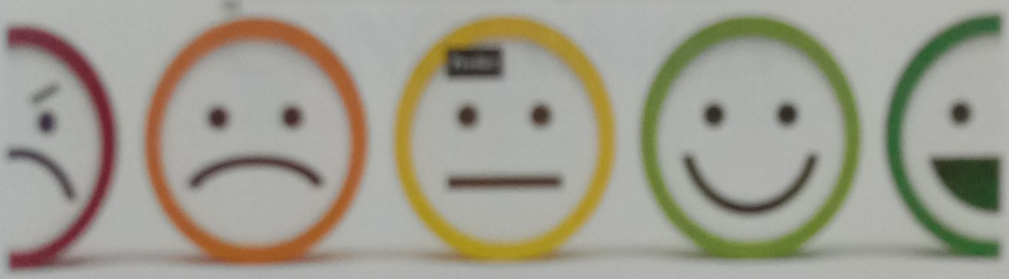
EMOJI POSITIVE COMMENT



2 MOST NEGATIVE COMMENT

Real Time Sentiment Analysis

Please give your honest review...



Review is : negative

CONCLUSION

10. CONCLUSION

Sentiment analysis is a technique used to understand the emotional tone of the text. It can be used to identify positive, negative, and neutral sentiments in a piece of writing. This information can be useful for business owners who want to understand how their customers feel about their company. By understanding the sentiment of your customer's reviews and feedback, you can work to improve those areas that are causing dissatisfaction and increase loyalty among your customer base. Marketers can use sentiment analysis to better understand customer feedback and adjust their strategies accordingly. Additionally, it can be used to determine whether a particular campaign or product resonates with customers in a positive or negative way.

FUTURE ENHANCEMENT

11. FUTURE ENHANCEMENT

Though I ought to have a successfully project, it could be still be improved further, according to some specific needs.

- This system can be improved as a passport forgery system that detects and classifies the forgery by extracting better discriminative features.
- I can propose a system to decrease the false positive and false negative to achieve high performance results and better accuracy.
- Using sentiment analysis datasets from different domains to improve the accuracy of sentiment analysis.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Kabir, M., Kabir, M.M.J., Xu, S. and Badhon, B., 2021. An empirical research on sentiment analysis using machine learning approaches. *International Journal of Computers and Applications*, 43(10), pp.1011-1019.
2. Naresh, A. and Venkata Krishna, P., 2021. An efficient approach for sentiment analysis using machine learning algorithm. *Evolutionary Intelligence*, 14(2), pp.725-731.
3. Jain, P.K., Pamula, R. and Srivastava, G., 2021. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer science review*, 41, p.100413.
4. Rawat, R., Mahor, V., Chirgaiya, S., Shaw, R.N. and Ghosh, A., 2021. Sentiment analysis at online social network for cyber-malicious post reviews using machine learning techniques. *In Computationally intelligent systems and their applications* (pp. 113-130). Springer, Singapore.
5. Kinyua, J.D., Mutigwe, C., Cushing, D.J. and Poggi, M., 2021. An analysis of the impact of President Trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis. *Journal of behavioral and experimental finance*, 29, p.100447.
6. Yadav, N., Kudale, O., Rao, A., Gupta, S. and Shitole, A., 2021. Twitter sentiment analysis using supervised machine learning. *In Intelligent Data Communication Technologies and Internet of Things* (pp. 631-642). Springer, Singapore.
7. Garg, S., 2021, January. Drug recommendation system based on sentiment analysis of drug reviews using machine learning. *In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 175-181). IEEE.
8. Saifullah, S., Fauziah, Y. and Aribowo, A.S., 2021. *Comparison of machine learning for sentiment analysis in detecting anxiety based on social media data*. arXiv preprint arXiv:2101.06353.
9. Malhan, Y., Saxena, S., Mala/, S. and Shankar, A., 2021. Geospatial modelling and trend analysis of coronavirus outbreaks using sentiment analysis and intelligent algorithms. *In Artificial Intelligence in Healthcare (pp. 1-19)*. Singapore: Springer Singapore.
10. Gadri, S., Chabira, S., Ould Mehieddine, S. and Herizi, K., 2022. Sentiment analysis: Developing an efficient model based on machine learning and deep learning approaches. *In Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021* (ICO2021) 3 (pp. 237-247). Springer International Publishing.

11. https://thecleverprogrammer.com/2021/03/09/real-time-sentiment-analysis-using-python/

12. https://towardsdatascience.com/sentiment-analysis-using-lstm-step-by-step-50d074f09948