Digital Naturalist - AI Enabled tool for Biodiversity Researchers

A PROJECT REPORT

Submitted by

SYED ABRAR S A

VIMOSH AASI A

SIVA D

SIVAKUMAR C

for the course

HX8001 – Professional Readiness for Innovation, Employability and Entrepreneurship

in

COMPUTER SCIENCE AND ENGINEERING

LOYOLA INSTITUTE OF TECHNOLOGY

ANNA UNIVERSITY : CHENNAI 600 025

NOV & 2022

ANNA UNIVERSITY : CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"Digital Naturalist - AI Enabled tool for Biodiversity Researchers"** is the bonafide work of **"SYED ABRAR S A, VIMOSH AASI A, SIVA D, SIVAKUMAR C"** who carried out the project work under my supervision.

Dr.G.BHUVANESWARI

HEAD OF THE DEPARTMENT

professor

Computer Science and Engineering

Loyola institute of technology,
Palanchur,
Chennai – 600 123

Mr S.Noor Mohammed

MENTOR

Assistant professor

Computer Science and Engineering

Loyola institute of technology,
Palanchur,
Chennai – 600 123

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Project Overview

A naturalist is someone who studies the patterns of nature, identifies a different kind of flora and fauna in nature. Being able to identify the flora and fauna around us often leads to an interest in protecting wild spaces, and collecting and sharing information about the species we see on our travels is very useful for conservation groups like NCC.

When venturing into the woods, field naturalists usually rely on common approaches like always carrying a guidebook around everywhere or seeking help from experienced ornithologists. There should be a handy tool for them to capture, identify and share the beauty to the outside world.

In this project, we are creating a web application which uses a deep learning model, trained on different species of birds, flowers and mammals (2 subclasses in each for a quick understanding) and get the prediction of the bird when an image is been given.

1.2 Purpose

- Augmenting a dataset to virtually increase the size of small datasets in order to make our machine learning models work better.
- Preprocess the images to a machine-readable format.
- Applying CNN algorithm on the dataset.
- How deep neural networks are predicting the class and subclass of a given image.
- You will be able to know how to find the accuracy of the model.
- You will be able to build web applications using the Flask framework.

2. LITERATURE SURVEY

2.1 Existing problem

Biases in our data arise in part from differences between the aims of the original data collectors (i.e., the photographers) and our aims as biodiversity researchers and ecologists. For example, the spatial distribution of our images was biased toward areas where extensive managed gardens or other displays exhibited large collections of flowering plants. These biases could be addressed by choosing alternative sources, changing the search terms used, or pre-filtering images. Images may also be biased taxonomically or in terms of certain traits, for example, toward species that are typically considered more photogenic due to large colorful flowers or leaves. Search terms could be modified to either focus on a specific sub-group, e.g., searching using scientific names, or to exclude non-target images, e.g., excluding images that include the words “show” or “garden” in their metadata. Finally, high-level image classifiers could be trained to remove images that are clearly not plants, e.g: removing images of animals, paintings. High-level classifiers developed to separate images that contain plants from those that do not, without looking to identify species, could be used to find images worthy of further examination in large datasets that do not have metadata (such as titles and descriptions), removing the need for keyword searches, such as that used in this study.

2.2 References

[1] Aldhebiani AY (2018) Species concept and speciation. Saudi J Biol Sci 25:437–440.

[2] AI naturalists might hold the key to unlocking biodiversity data in social media imagery,

TA August, OL Pescott, A Joly, P Bonnet - Patterns, 2020 - Elsevier.

[3] Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet, ME Mondejar, R Avtar, HLB Diaz, RK Dubey… - Science of the Total …, 2021 - Elsevier

[4] The real-world use of big data, M Schroeck, R Shockley, J Smart, D Romero-Morales… - IBM Global Business …, 2012

[5] Computer Age Statistical Inference, Student Edition: Algorithms, Evidence, and Data Science, B Efron, T Hastie - 2021

2.3 Problem Statement Definition

- **Main Problem statement (common):**

i. How might we help both experienced and inexperienced user to identify species of plants and animals and their characteristics with related information?

- **Specific problem statement:**

i. Inexperienced users need to know about poisonous plants and dangerous animals so that they can stay away from it.

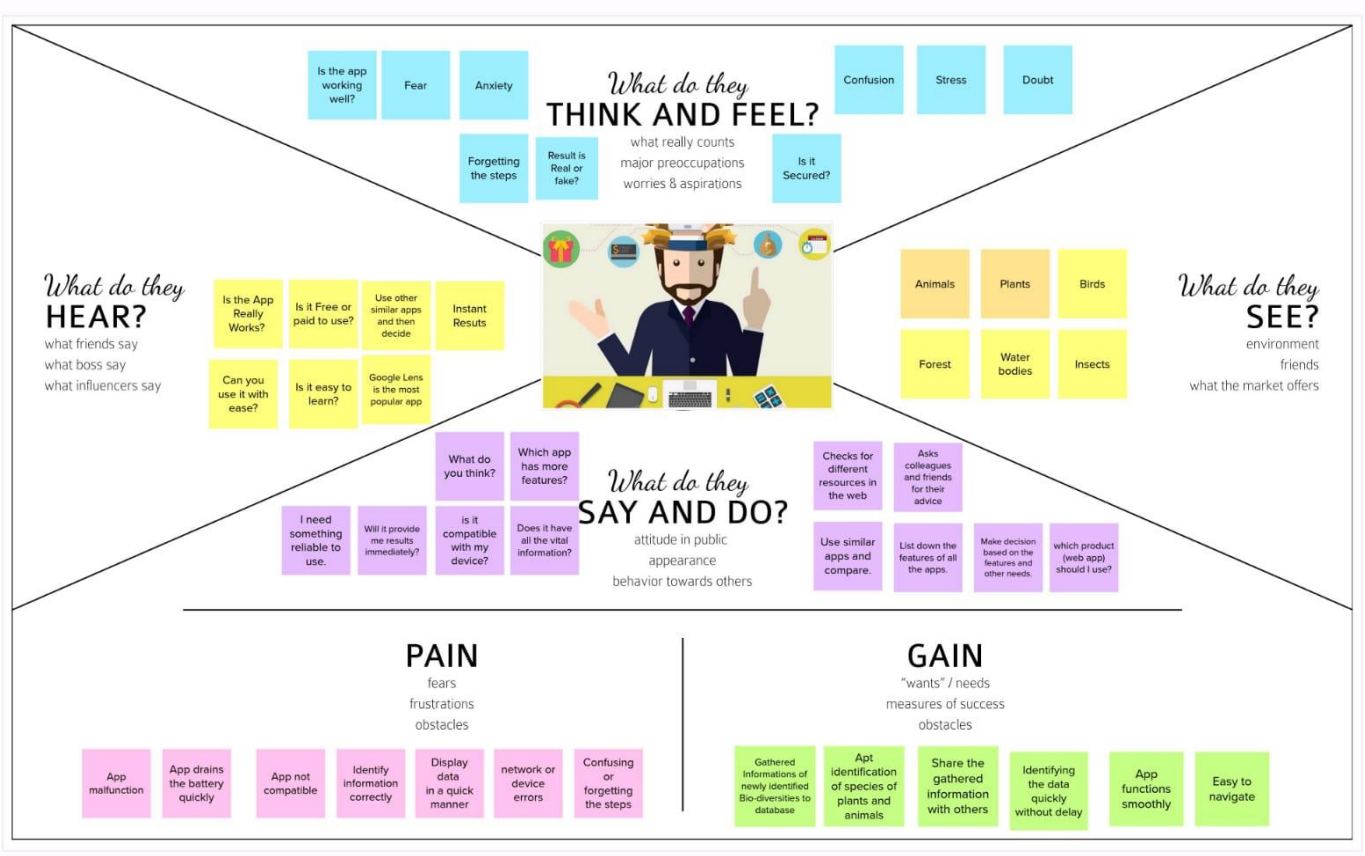
ii. Both experienced and inexperienced users need to know about the medicinal values of a plant because they need to use it in case of emergencies.

iii. All the users need to know the types of species of birds, plants and animals so that they can learn about it in more detailed manner.

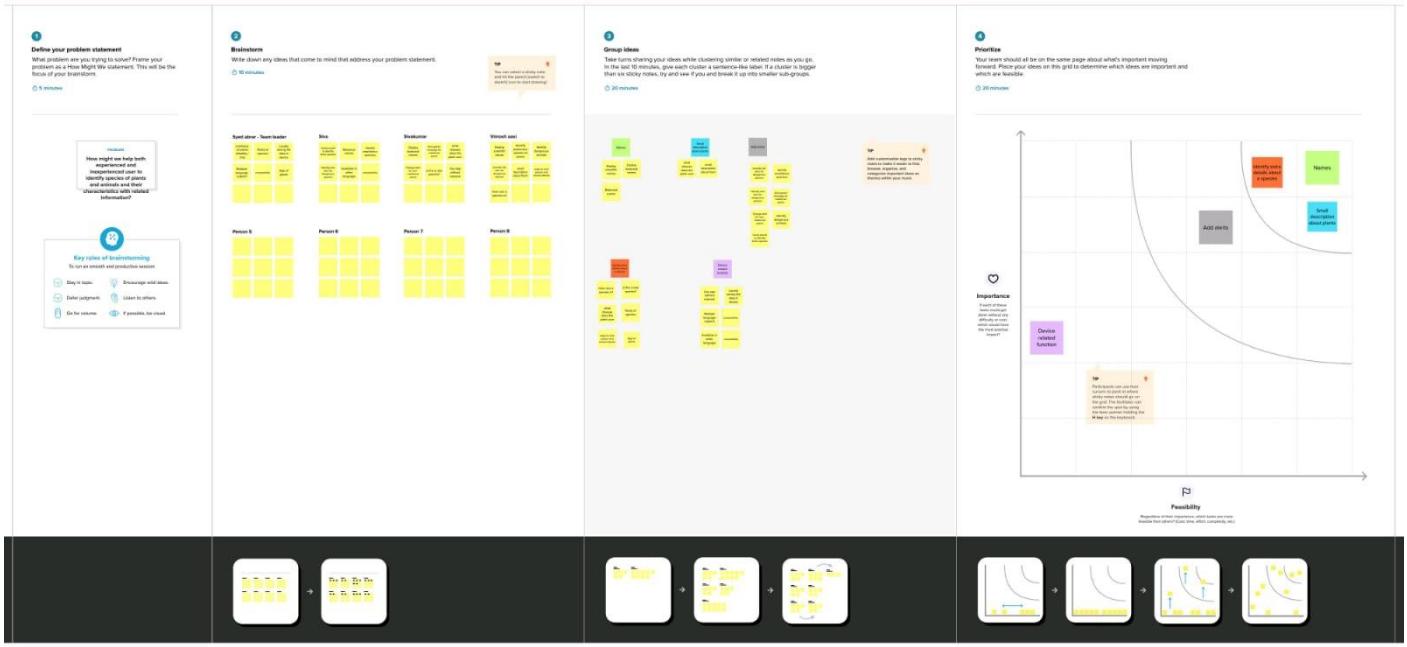
iv. All the users need to know about the rarity of the species of birds, animals or plants so that they can preserve and save it.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming



Note:

For full view and understanding of our idea and brainstorming click the pdf icon

Adobe Acrobat Document

3.3 Proposed Solution:

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	**i)** How might we help both experienced and inexperienced user to identify species of plants and animals and their characteristics with related information? **ii)** Inexperienced users need to know about poisonous plants and dangerous animals so that they can stay away from it. **iii)** Both experienced and inexperienced users need to know about the medicinal values of a plant because they need to use it in case of emergencies. **iv)** All the users need to know about the rarity of the species of birds, animals or plants so that they can preserve and save it.
2.	Idea / Solution description	**i)** Display Botanical names **ii)** Display alert messages for plants/animals using different colours **iii)** small description about them **iv)** Rarities of the species **v)** What disease does the plant cure
3.	Novelty / Uniqueness	**i)** Providing alerts based on if a species is harmful or not **ii)** Alerting the user on the rarity of the species **iii)** Gives the complete description about the species being viewed **iv)** If the plant being viewed has a medicinal value , it gives a description about it. **v)** Display the scientific name of the species.
4.	Social Impact / Customer Satisfaction	Being able to identify the **flora** and **fauna** around us often leads to an interest in protecting wild spaces.
5.	Business Model (Revenue Model)	**i)** Can make money through subscription based. **ii)** Partnership with many laboratories and scientists around the world
6.	Scalability of the Solution	**i)** As the usage and user base of this application grows more feature can be added to the premium or subscription model. **ii)** We can introduce subscription models like free plan, business plan, educational plan and many more based on its usage **iii)** As the usage increase we can scale the application by releasing more languages based on the geographical usage.

3.4 Problem Solution fit

Project title: Digital Naturalist - AI Enabled tool for Biodiversity Researchers

Team ID: PNT2022TMID25592

Define CS, fit into

1. CUSTOMER SEGMENT(S) CS

1)People who go for hikes or trips to the forest areas and mountains.

2)Amateurs or Students or people who like to learn more about the biodiversity

3)Experienced professionals and Inexperienced people who are willing to learn about bio diversity.

6. CUSTOMER CC

- Cannot remember everything
- Inexperience
- Fear
- No knowledge about bio diversity
- Confusion
- Not able to identify the plants and animals

5. AVAILABLE SOLUTIONS AS

1. Carry a guidebook to identify the species.
2. Get help from experienced people
3. Internet
4. Other apps

There are various disadvantages in these methods such as books can be damaged or weigh a lot, cannot find the right person

Explore AS, Focus on J&P, tap into BE, understand

Focus on J&P, tap into BE, understand

2. JOBS-TO-BE-DONE / PROBLEMS J&P

1) Cannot remember everything as it is too much
2) No knowledge or experience about bio diversity as the user is just starting to learn which can lead to confusion or fear
3) Fear of misidentification.
4) Need to know about at least the basics.

9. PROBLEM ROOT CAUSE RC

1. Too much data cannot be stored by any human or they may forget or other due to any other problems like age.
2. Users may not be a naturalist or just a student who just started to learn so they may not know any information.
3. Inexperience or no education
4. Need to learn as they can gain knowledge and experience

7. BEHAVIOUR BE

Direct

1)Carry guide books or other notes to identify species.
2)Get help from experienced professionals
3)Try to remember the species based on its feature based on the characteristics or common knowledge
4)Notes, both digital or notebooks
5)Try to use other apps for identification

Indirect

Either start learning about bio-diversity or research on the internet

Identify strong TR & EM

3. TRIGGERS TR

They hear about new app with best features by their friend or colleagues, or read about app in news, or search the internet on their own.

4. EMOTIONS: BEFORE / AFTER EM

1. Inexperienced → start to gain knowledge and experience
2. Fear→confident
3. Confusion → Clear

10. YOUR SOLUTION SL

- Display Botanical names
- Display alert messages for plants/animals using different colors
- Small description about them
- Rarities of the species
- What disease does the plant cure?

With these solutions they gain knowledge and identify the species correctly without anyones help.

8. CHANNELS of BEHAVIOUR CH

ONLINE

1. Search using the internet about the species
2. Apps or websites

OFFLINE

1. Get help from friends or professionals
2. Guidebook or they even take their own notes

Extract online & offline CH of

4. REQUIREMENT ANALYSIS

(Following are the functional & non-functional requirements of the proposed solution)

4.1 Functional requirement.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Navigation Service	GPS
FR-4	Database	My SQL, IBM Cloud
FR-5	Premium features	Location sharing, Adding information of new data by User
FR-6	Updating and bug fixing	Updating the application based on user feedback
FR-7	Final Output	Final description of the image (species) captured.
FR-8	Alerts	System should alert about dangerous plants and animals

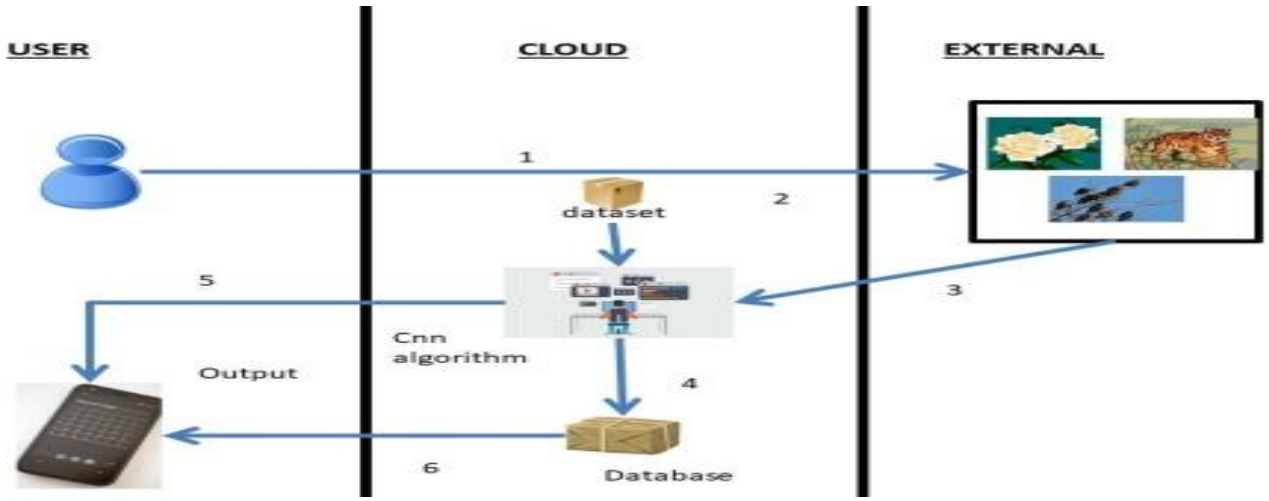
4.2 Non-functional Requirements.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The users must be able to use the application without any problems and difficulties. The app is easy to view and does not strain the eyes. All information are in simple terms. The error rate of the final output must not be more than 20%
NFR-2	**Security**	SHA-256, Encryptions, AES etc.
NFR-3	**Reliability**	The system must perform without failure in 80 percent of the time.
NFR-4	**Performance**	Under normal load, the system must show the results within 15 seconds, and under maximum it can take up to load 30 seconds
NFR-5	**Availability**	The application will be available 99 % of the time in a month.
NFR-6	**Scalability**	The system must be able to support 10,000 users while using it. As the usage and user base of this application grows, more features can be added like languages based on the geographical usage, premium or subscription model, etc.

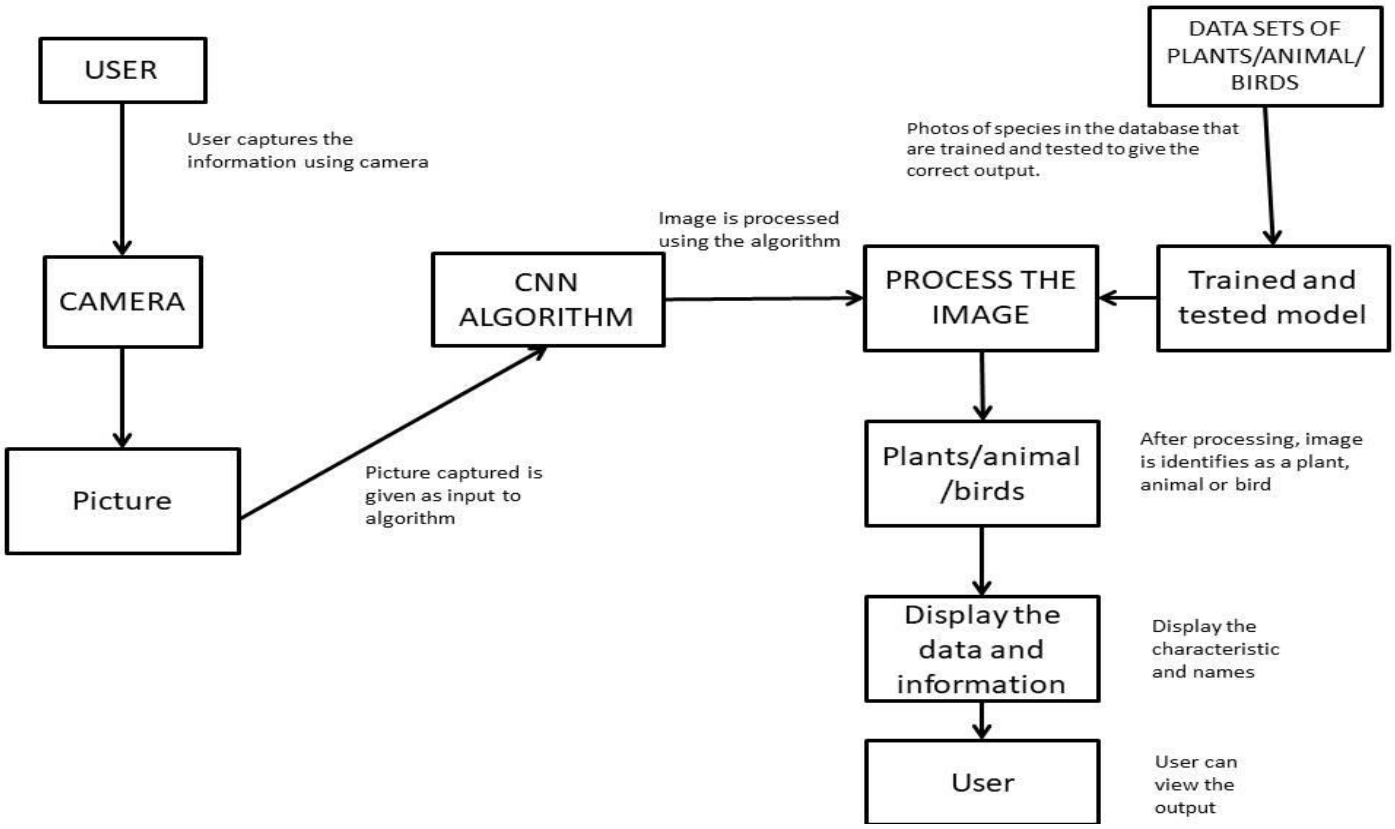
5. PROJECT DESIGN

5.1 Data Flow Diagrams

Example: (Simplified)

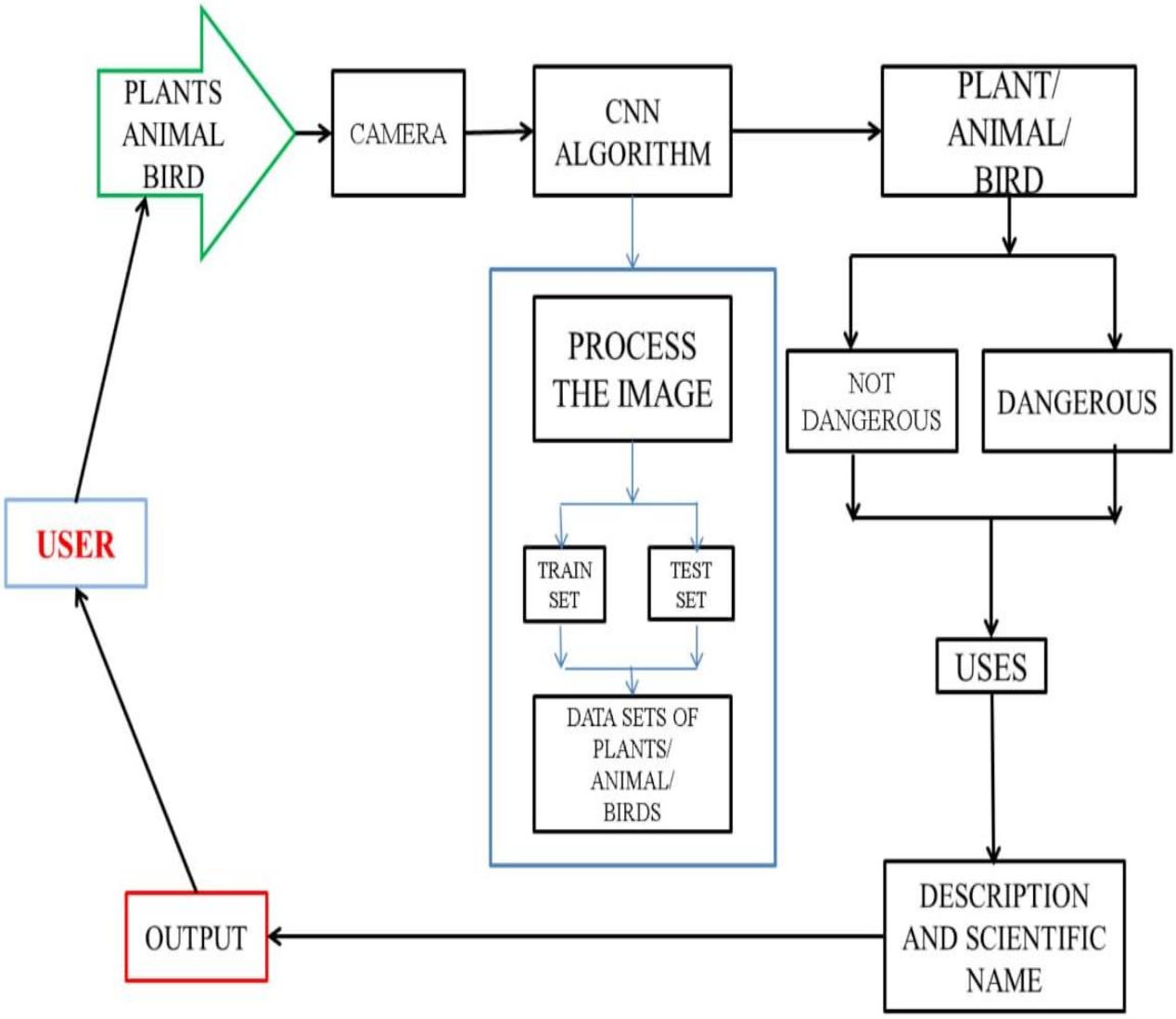


DFD LEVEL 0 (Industry standard)



1. User captures photo using camera
2. Dataset contains photos that are used for training and testing
3. Photo is sent to CNN for classification and processing based on dataset
4. Processed image is checked with database for additional info
5. Photo from CNN algorithm is sent to the user
6. Matching data from database is sent to the user

5.2 Solution & Technical Architecture



5.3 User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority
Customer (Mobile user)	Image capture	USN-1	As a user, I can take photos of the plant life, animals and birds	I can take photos when required	High
		USN-2	As a user, I will receive processed information about the type of species	I can see the type of plant or animal or plant	High
		USN-3	As a user, I can share it with others	I can share using share option	Low
	Data process	USN-4	Data must be trained and tested and CNN algorithm must work properly.	I must see the correct processed information	High
	Output	USN-5	As a user, I can see the scientific name of the species	I must see the correct data	High
		USN-6	As a user, I can see the characteristics and alert messages	I must see the correct data	High
Administrator	Manage	USN-1	As a admin I must add various data and edit information	I must edit the data present	High

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Use the below template to create product backlog and sprint schedule

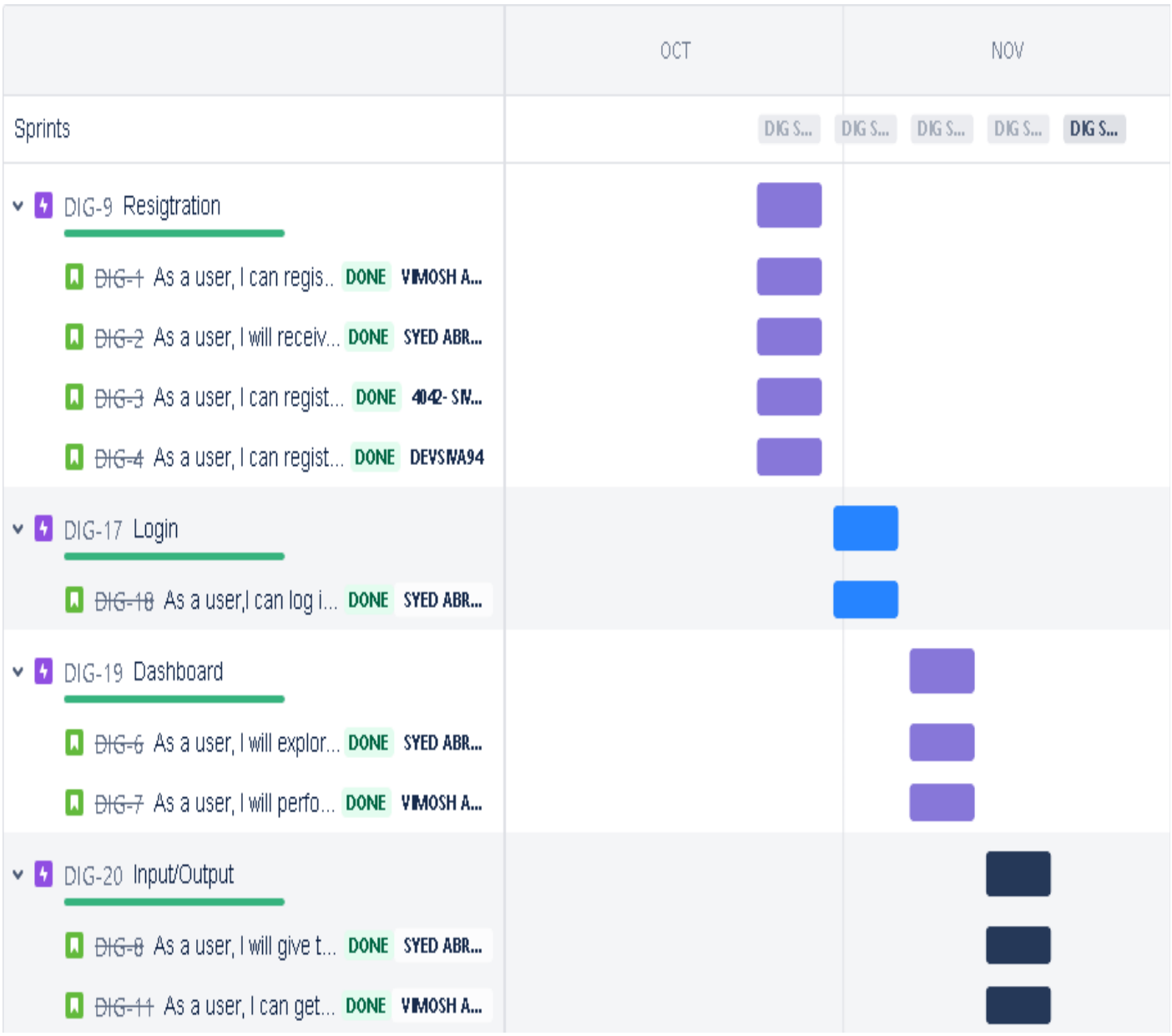
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	A.Vimosh Aasi
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	S.A.Syed Abrar
Sprint-1		USN-3	As a user, I can register for the application through Facebook	2	Low	C.Sivakumar
Sprint-1		USN-4	As a user, I can register for the application through Gmail	2	Medium	D.Siva
Sprint-2	Login	USN-5	As a user, I can log into the application by entering email & password	1	High	S.A.Syed Abrar
Sprint-3	Dashboard	USN-6	As a user, I will explore the functionalities of the software	1	Medium	A.Vimosh Aasi
Sprint-3		USN-7	As a user, I will perform the required task through the software	1	Medium	C.Sivakumar
Sprint-4	Input / Output	USN-8	As a user, I will give the input through the software to get the output information	2	High	S.A.Syed Abrar
Sprint-4		USN-9	As a user, I can get the Final description of the image (species) captured.	1		A.Vimosh Aasi

6.2 Sprint Delivery Schedule

Project Tracker, Velocity & Burndown Chart:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	30 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	06 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

6.3 Reports from JIRA



7. CODING & SOLUTIONING

7.1 Feature 1:

- Display Botanical names
- Display alert messages for plants/animals using different colors
- small description about them
- Rarities of the species

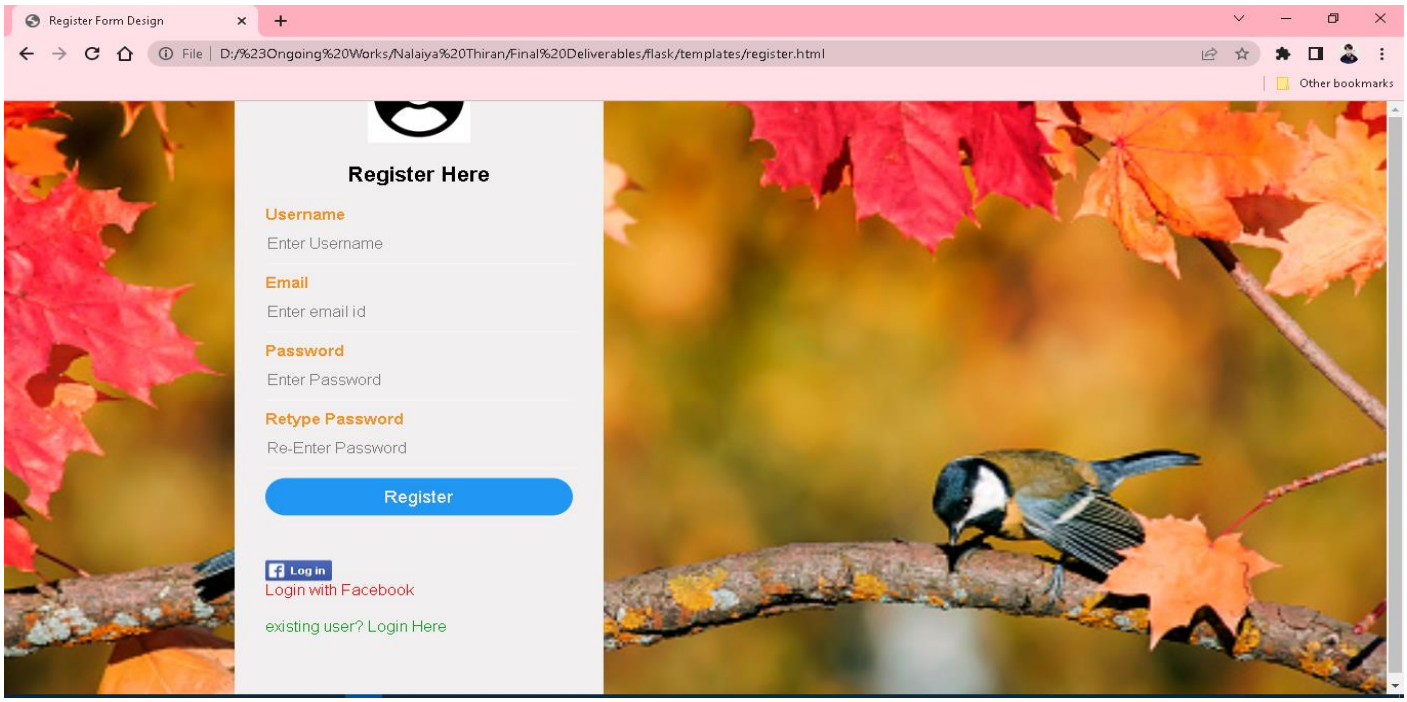
7.2 Feature 2:

- What disease does the plant cure
- Providing alerts based on if a species is harmful or not

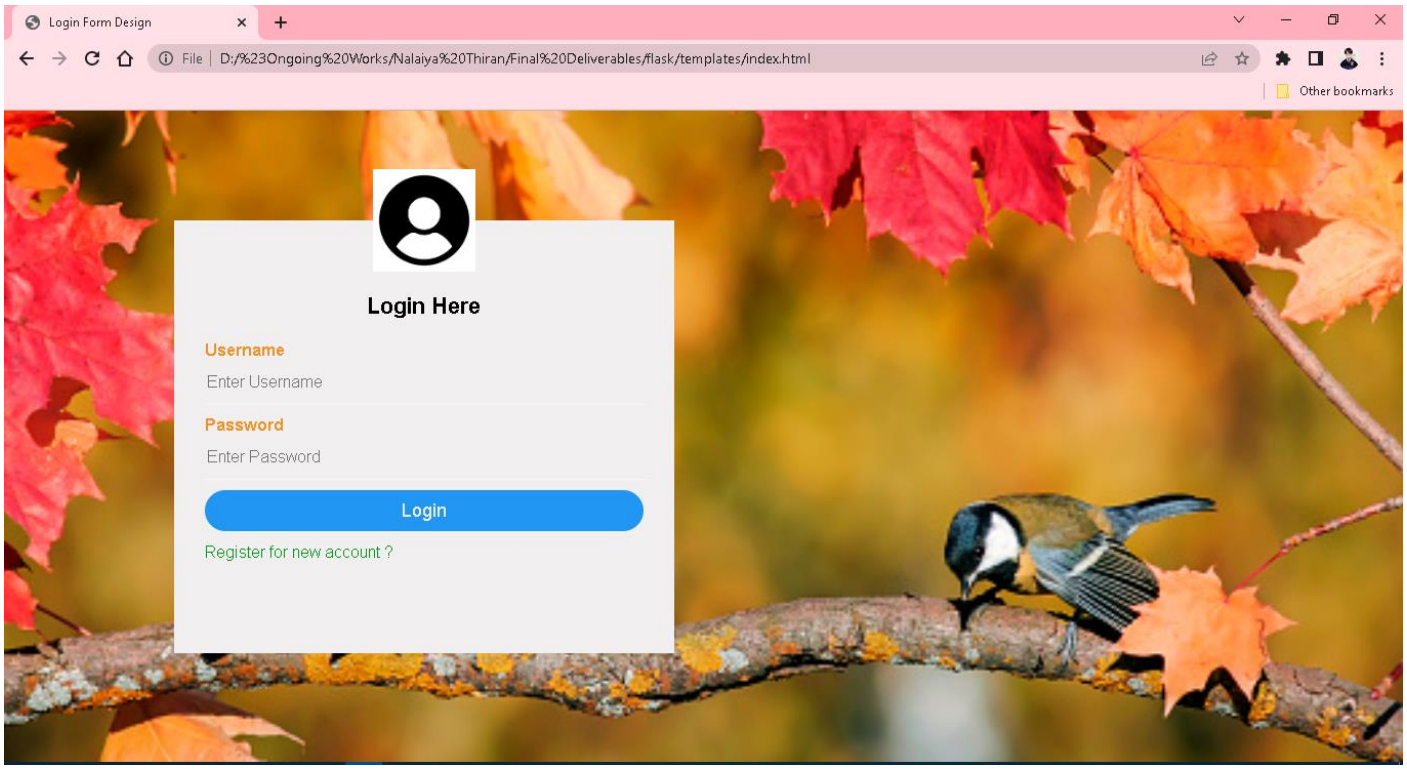
8. RESULTS

8.1 Performance Metrics

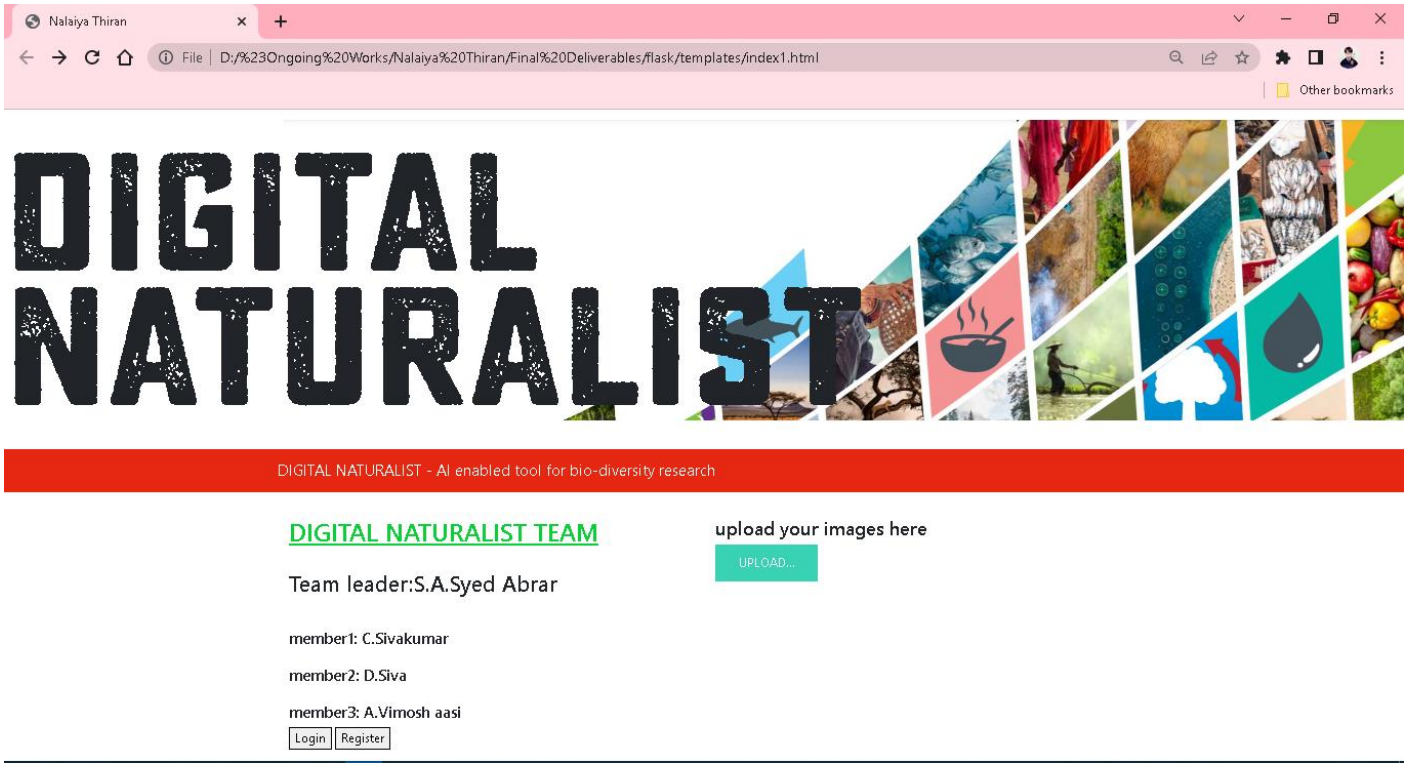
Registration Page:



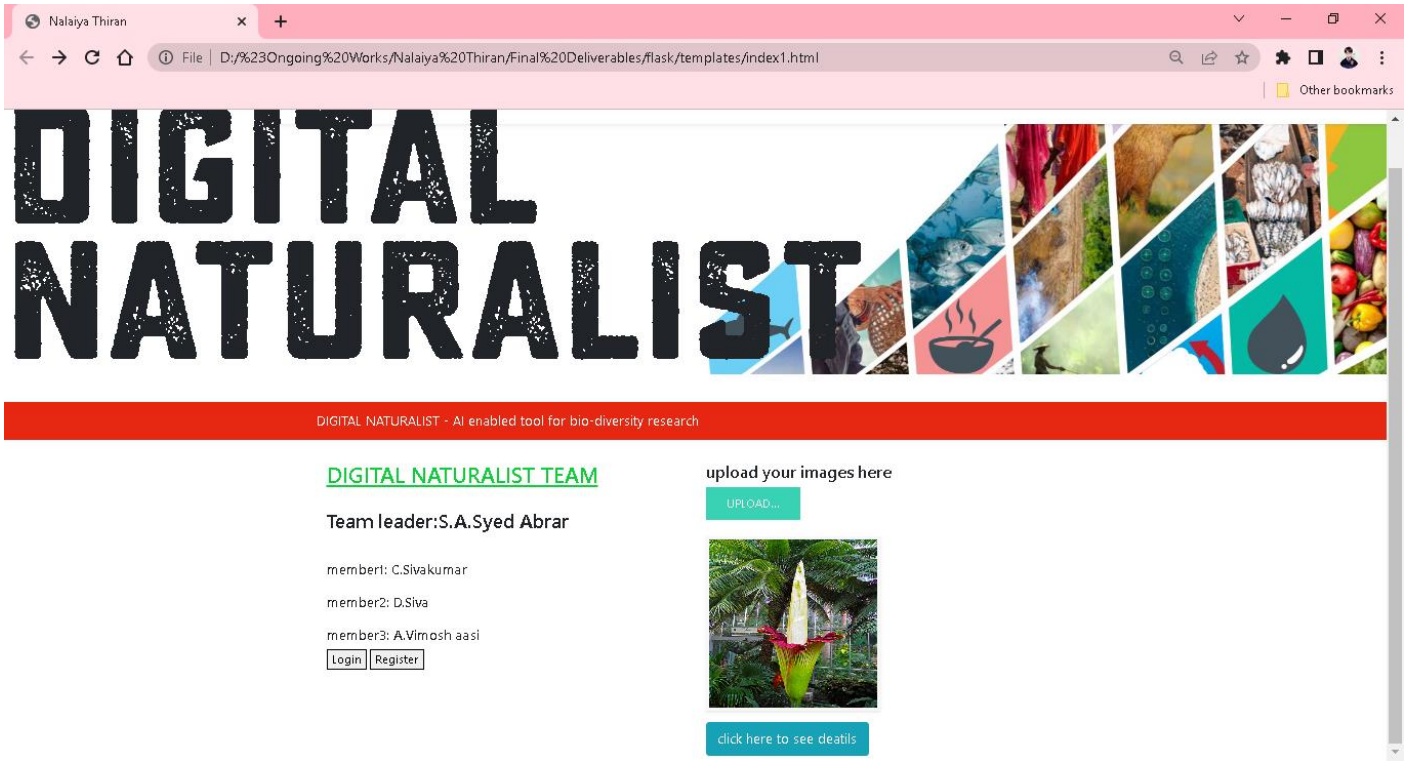
Login Page :



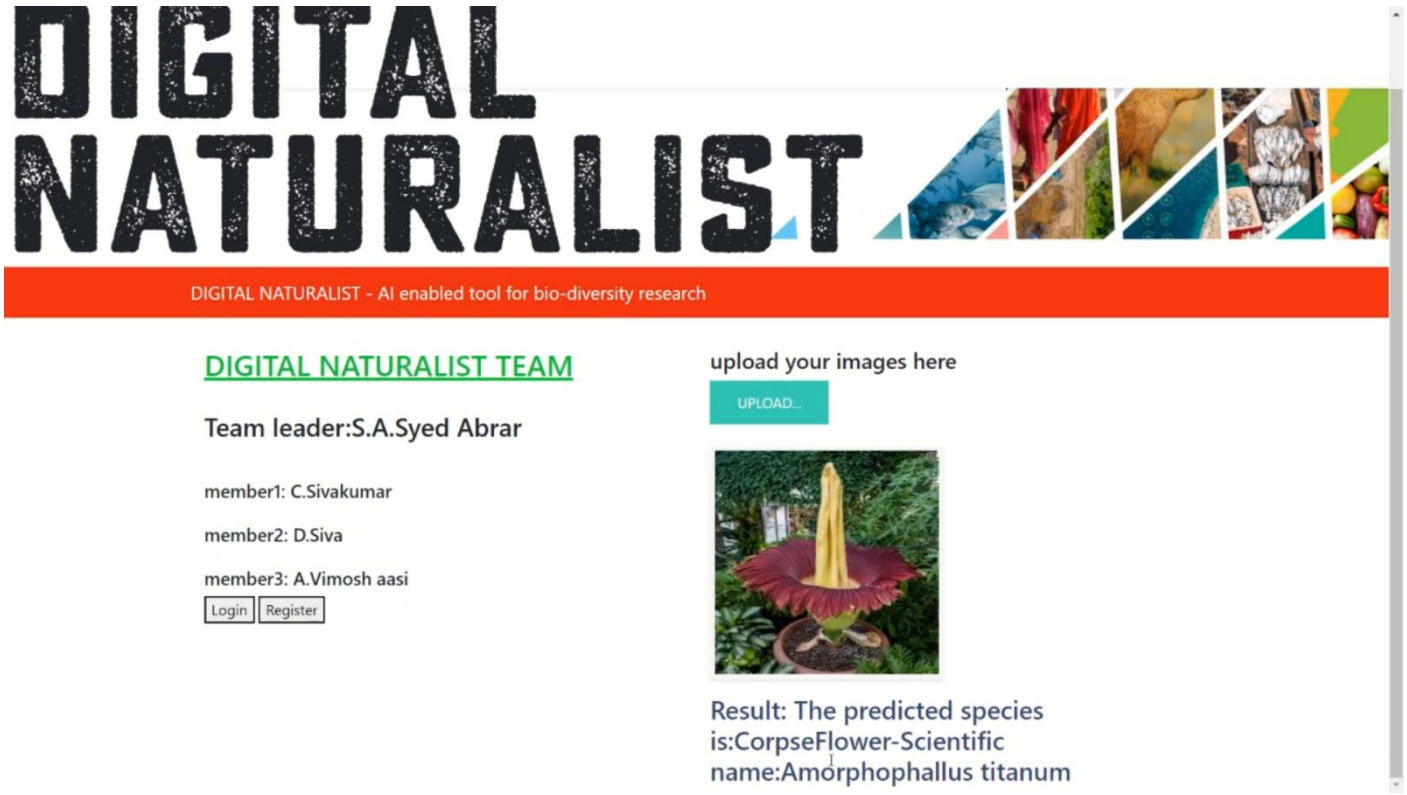
Home Page(Dashboard):



Input:



Output:


DIGITAL NATURALIST
DIGITAL NATURALIST - AI enabled tool for bio-diversity research
DIGITAL NATURALIST TEAM
Team leader:S.A.Syed Abrar
member1: C.Sivakumar
member2: D.Siva
member3: A.Vimosh aasi
Login
Register
upload your images here
UPLOAD...
Result: The predicted species is:CorpseFlower-Scientific name:Amorphophallus titanum

9. ADVANTAGES & DISADVANTAGES

Advantages:

- Allows researchers to study behaviors or situations that cannot be manipulated in a lab due to ethical concerns. For example, it would be unethical to study the effects of imprisonment by actually confining subjects. But researchers can gather information by using naturalistic observation in actual prison settings.

- Can support the external validity of research. Researchers might believe that the findings of a lab study can be generalized to a larger population, but that does not mean they would actually observe those findings in a natural setting. They may conduct naturalistic observation to make that confirmation.

Disadvantages:

- Inability to draw cause-and-effect conclusions: The biggest disadvantage of naturalistic observation is that determining the exact cause of a subject's behavior can be difficult

- Observer bias: The biases of the people observing the natural behaviors can influence the interpretations that experimenters make.

10. CONCLUSION

Assessment of regional biodiversity based on global scientific consensus is a scientific basis for the whole society and a tool for local to international discussion and decision making. The solution to reach an agreement on policy-related environmental issues is to stay away from the trivial matters. It can also be said that the synthesis of the scientific literature available with the highest objective analysis is aimed at sufficient quality research..

11. FUTURE SCOPE

- **AI image classifiers can create biodiversity datasets from social media imagery**
- **Flickr hosts many images of plants; some can be accurately classified to species by AI**
- **Images are spatially aggregated around tourist sites and under-represent native species**
- **Images focused on a single, non-horticultural, plant are most reliably**

12. APPENDIX

Source Code

<---------- # index1.html ---------->

```
<html lang="en">
<head>
<meta charset="UTF-8">
<meta name="viewport" content="width=device-width, initial-scale=1.0">
<meta http-equiv="X-UA-Compatible" content="ie=edge">
<title>Nalaiya Thiran</title>
<link href="https://cdn.bootcss.com/bootstrap/4.0.0/css/bootstrap.min.css" rel="stylesheet">
<script src="https://cdn.bootcss.com/popper.js/1.12.9/umd/popper.min.js"></script>
<script src="https://cdn.bootcss.com/jquery/3.3.1/jquery.min.js"></script>
<script src="https://cdn.bootcss.com/bootstrap/4.0.0/js/bootstrap.min.js"></script>
<link rel="stylesheet" type="text/css" href="../static/css/main.css">
<style>
.bg-dark {
background-color: #e62813!important;
}
#result {
color: #0a1c4ed1;
}
@font-face {
font-family: 'Sucrose Bold Two';
src: url('https://s3-us-west-2.amazonaws.com/s.cdpn.io/4273/sucrose.woff2') format('woff2');
}
@font-face {
```

```
font-family: 'IM Fell French Canon Pro';

src: url('https://s3-us-west-2.amazonaws.com/s.cdpn.io/4273/im-fell-french-canon-pro.woff2')
format('woff2');

}header {

background: url(../static/images/image1.jpg) no-repeat right;

background-attachment: fixed;

padding-top: 3%;

background-size: 80%;

font-family: 'Sucrose Bold Two';

}

header h1 {

position: static;

top: 2rem;

right: 2rem;

font-size: 12vw;

line-height: .8;

margin-top: 0;

margin-bottom: 2%;

text-align: left;

}

main {

background: rgb(232, 119, 219);

position: relative;

border: 1px solid #fff;

font-family: 'IM Fell French Canon Pro';

font-size: 1.4rem;

padding: 2rem 25%;
```

```
line-height: 1.6;
}
@media all and (max-width: 400px) {
main { padding: 2rem; }
}
</style>
</head>
<body>
<header>
<h1>digital naturalist</h1>
</header>
<nav class="navbar navbar-dark bg-dark">
<div class="container">
<a class="navbar-brand" href="#">DIGITAL NATURALIST - AI enabled tool for bio-diversity research</a>
</div>
</nav>
<div class="container">
<div id="content" style="margin-top:2em">
<div class="container">
<div class="row">
<div class="col-sm-6 bd" >
<h2 id="blink"><u>DIGITAL NATURALIST TEAM </u></h2>
<br>
<h3>Team leader:S.A.Syed Abrar<br><br></h3>
<h5>member1: C.Sivakumar<br><br>
member2: D.Siva<br><br>
```

```
member3: A.Vimosh aasi</h5>

<button onclick="document.location='index.html'">Login</button>

<button onclick="document.location='register.html'">Register</button>

</div>

<div class="col-sm-6">

<div>

<h4>upload your images here</h4>

<form action = "http://localhost:5000/predict" id="upload-file" method="post" enctype="multipart/form-data">

<label for="imageUpload" class="upload-label">

UPLOAD...

</label>

<input type="file" name="image" id="imageUpload" accept=".png, .jpg, .jpeg">

</form>

<div class="image-section" style="display:none;">

<div class="img-preview">

<div id="imagePreview">

</div>

</div>

<div>

<button type="button" class="btn btn-info btn-lg " id="btn-predict">click here  to see deatils</button>

</div>

</div>

<div class="loader" style="display:none;"></div>

<h3>

<span id="result"> </span>

</h3>
```

```
</div>
</div>
</div>
</div>
</div>
</div>
</body>
<footer>
<script src="../static/js/main.js" type="text/javascript"></script>
</footer>
</html>
```

<---------- # app.py---------->

```
import numpy as np
import os
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing import image
from flask import Flask, render_template, request
app=Flask(__name__)
model=load_model("final_model.h5")
@app.route('/')
def index():
return render_template("index1.html")
@app.route('/predict', methods=['GET', 'POST'])
def upload():
```

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

f=request.files['image']

basepath=os.path.dirname(__file__)

filepath=os.path.join(basepath, 'uploads', f.filename)

f.save(filepath)

img=image.load_img(filepath, target_size=(64,64))

x=image.img_to_array(img)

x=np.expand_dims(x,axis=0)

pred=np.argmax(model.predict(x), axis=1)

index=['CorpseFlower-Scientific name: Amorphophallus titanum','Great Indian Bustard Bird','Lady Slipper Orchid Flower-Scientific name:Orchidaceae','Pangolin Mammal-Scientific name: Pholidota','Senenca White Deer Mammal','Spoon Billed Sandpiper Bird-Scientific name: Calidris pygmaea']

text = "The predicted species is:" +str(index[pred[0]])

return text

if __name__=='__main__':

app.run(debug=False)
```

GitHub & Project Demo Link

- **GitHub Link :** https://github.com/IBM-EPBL/IBM-Project-53485-1661411454
- **Project Demo Link :** https://youtu.be/cZvE75BferU