



PlantPal

SMART ASSISTANCE FOR THE FLORICULTURE INDUSTRY

2023-133

MEET OUR TEAM



IT19994406

Basnayake N.S.N.



IT19169736

Gamage M.G.U.D.



IT20017088

Prabhashi P.A.N.



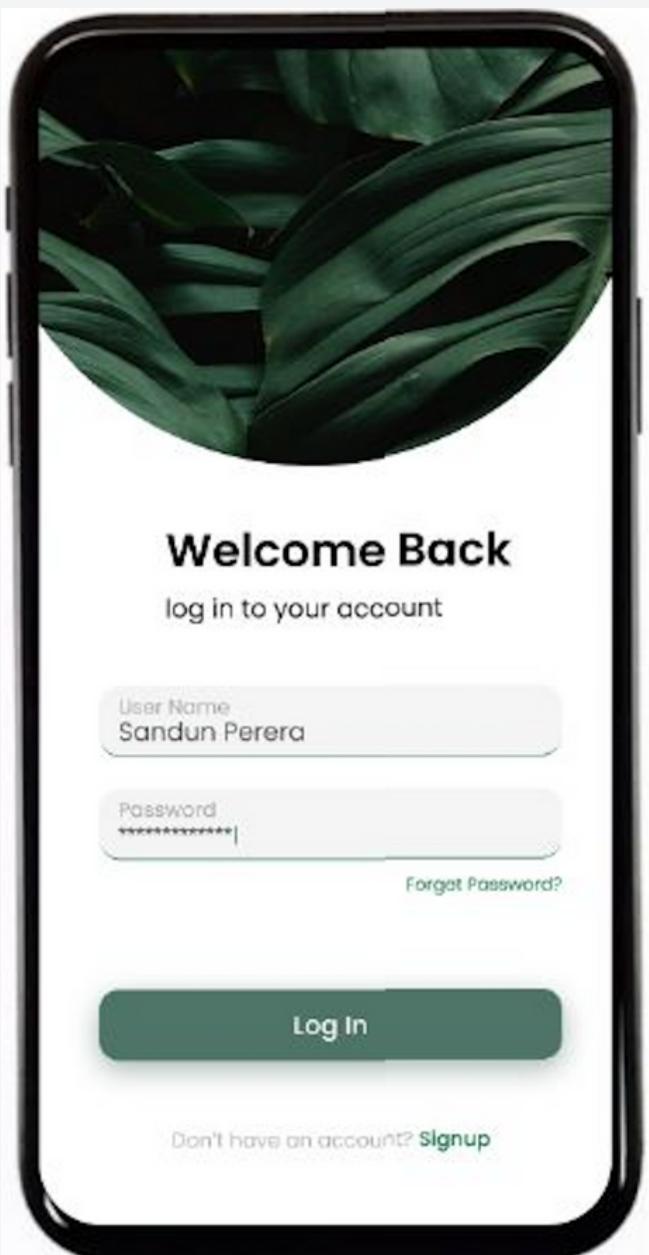
IT20005726

Liyanage S.R.

INTRODUCTION

- Sri Lanka's floriculture industry faces technology inefficiencies.
- Manual plant monitoring makes it difficult to estimate growth.
- Detecting deficiencies or mite attacks through visual inspection is also challenging.
- Beginners lack knowledge sources for plant selection and care.
- Demand forecasting relies on past experience, risking incorrect decisions and resource waste.

MAIN OBJECTIVE

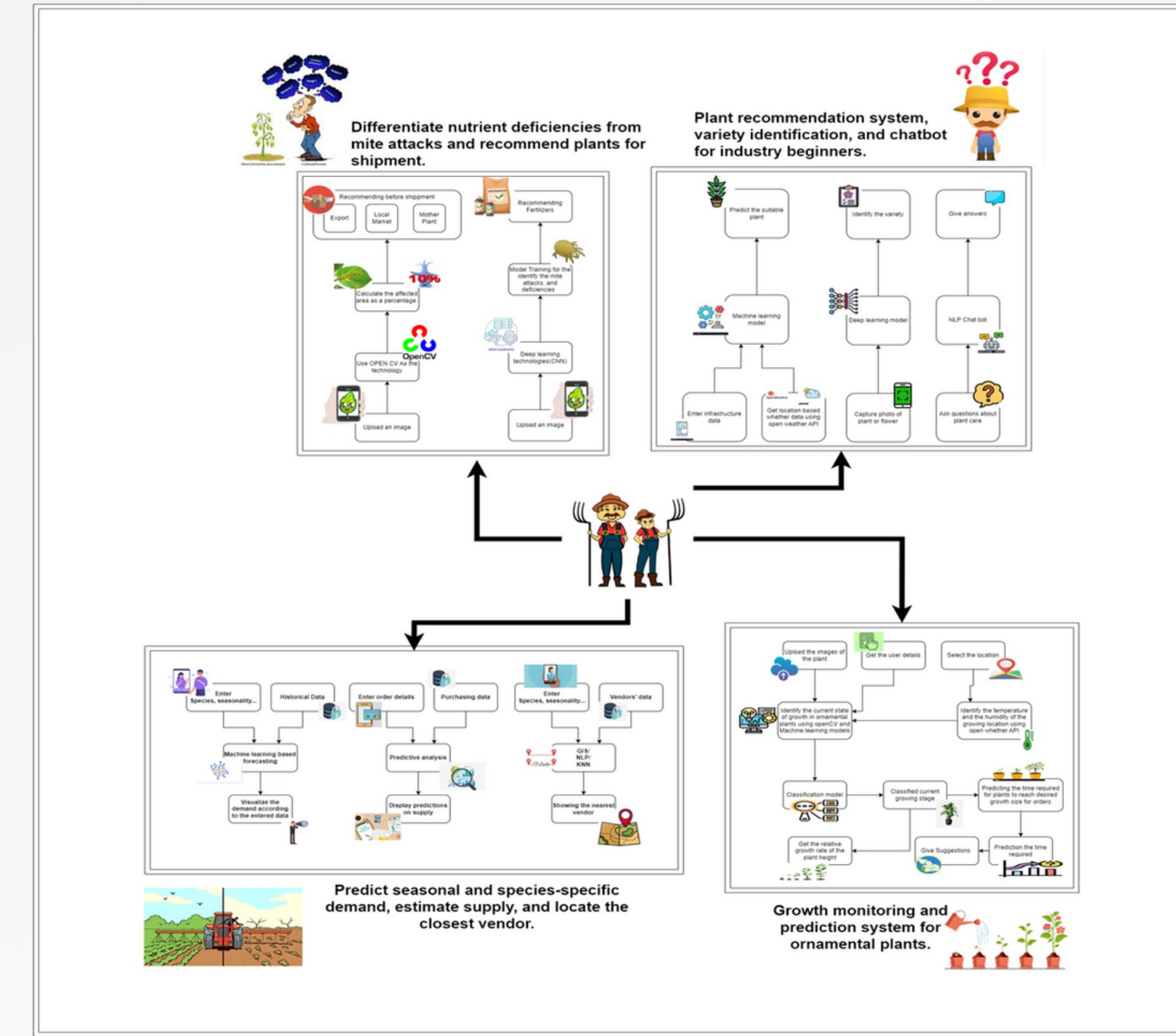


Automated mobile application to provide smart assistance for the floriculture industry

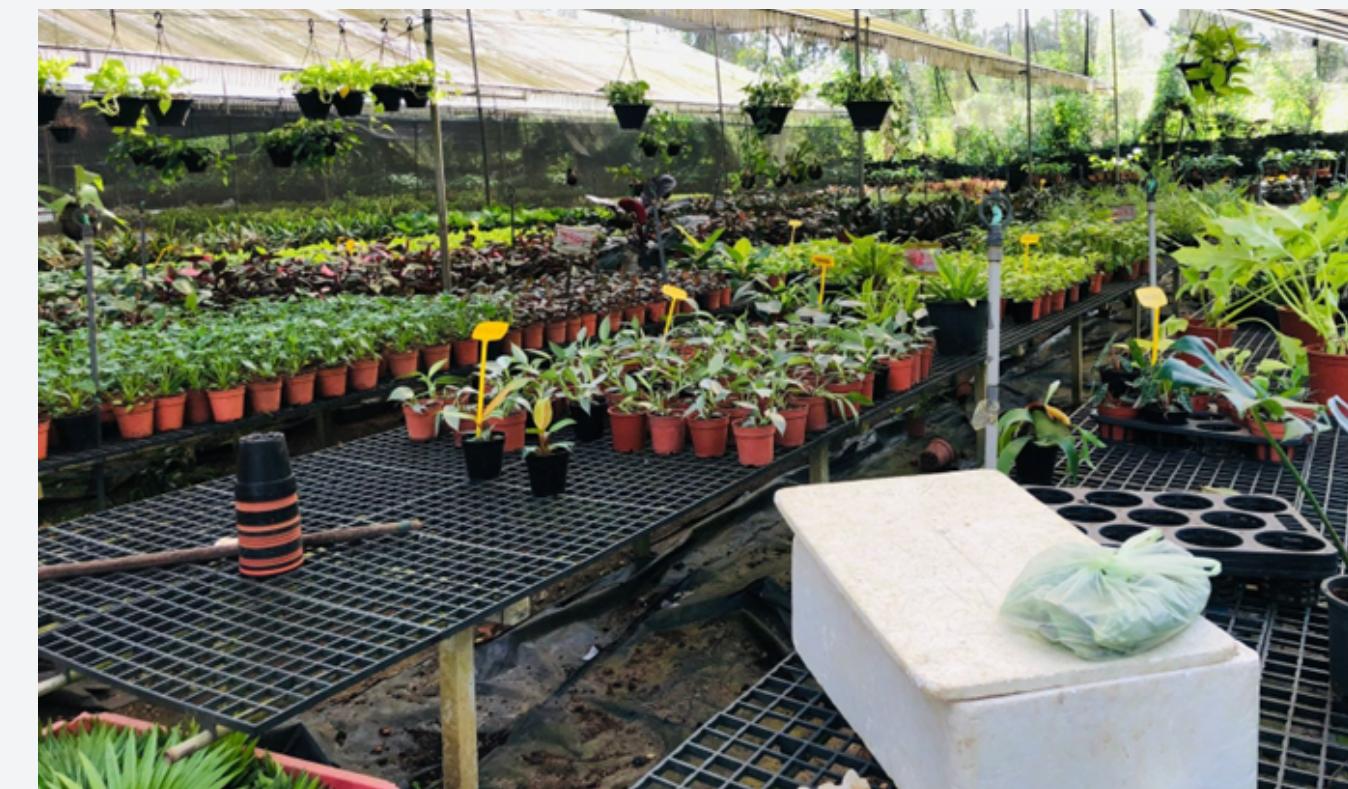
SUB OBJECTIVES

- Growth monitoring and prediction system for ornamental plants.
- Differentiate nutrient deficiencies and recommend plants for shipment.
- Plant recommendation system and variety identification for industry beginners.
- Predict seasonal and species-specific demand, estimate supply, and locate the closest vendor.

OVERALL SYSTEM ARCHITECTURE



FIELD VISIT TO OMEGA GREEN (PVT) LTD, BADALGAMA



DIFFERENTIATE NUTRIENT DEFICIENCIES AND RECOMMEND PLANTS FOR SHIPMENT

IT19994406
Basnayake N.S.N.



Specialization : Data Science

RESEARCH PROBLEM

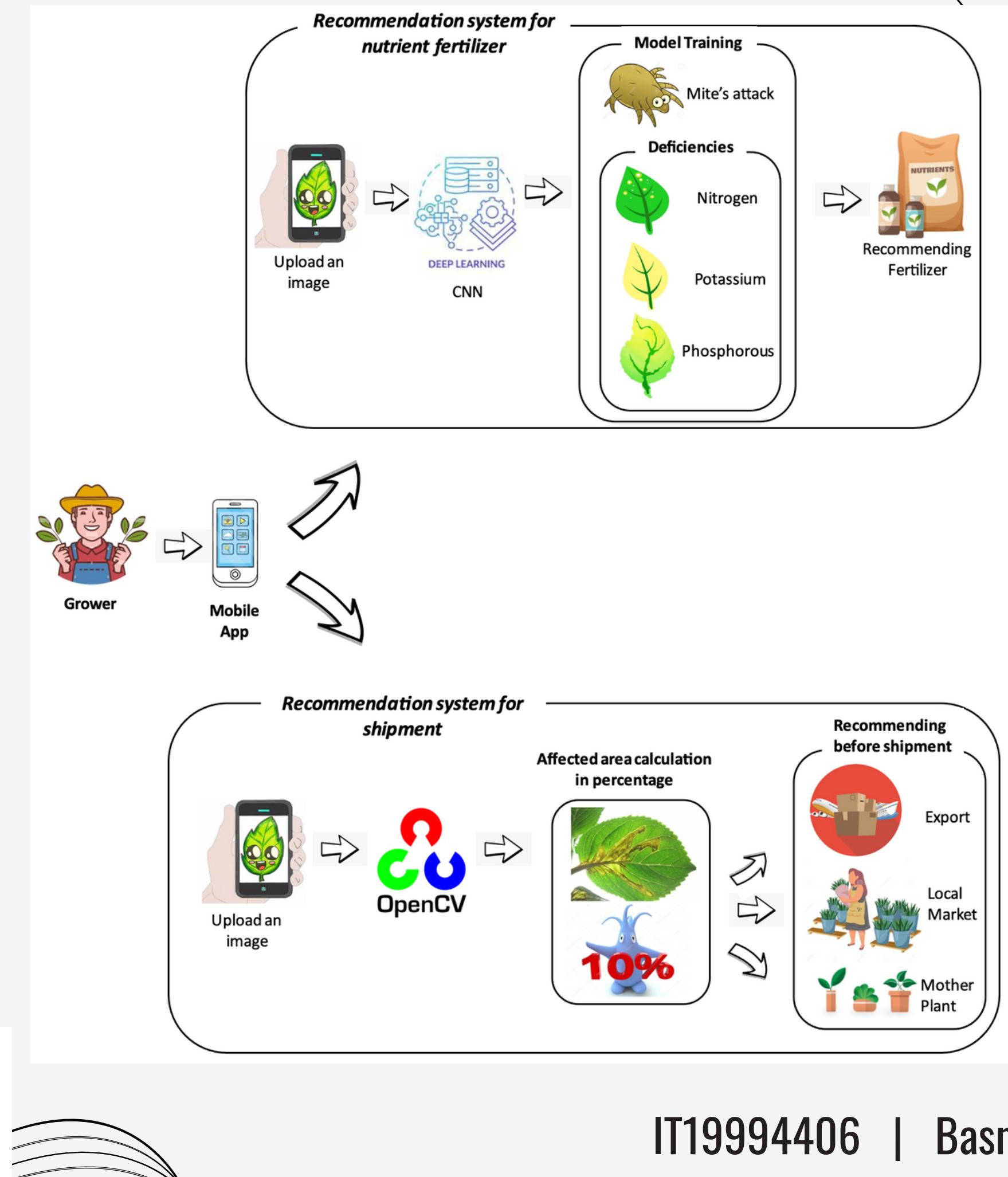
How to determine the correct nutrient fertilizer for the affected plant?

- Nitrogen
- Phosphorous
- Potassium

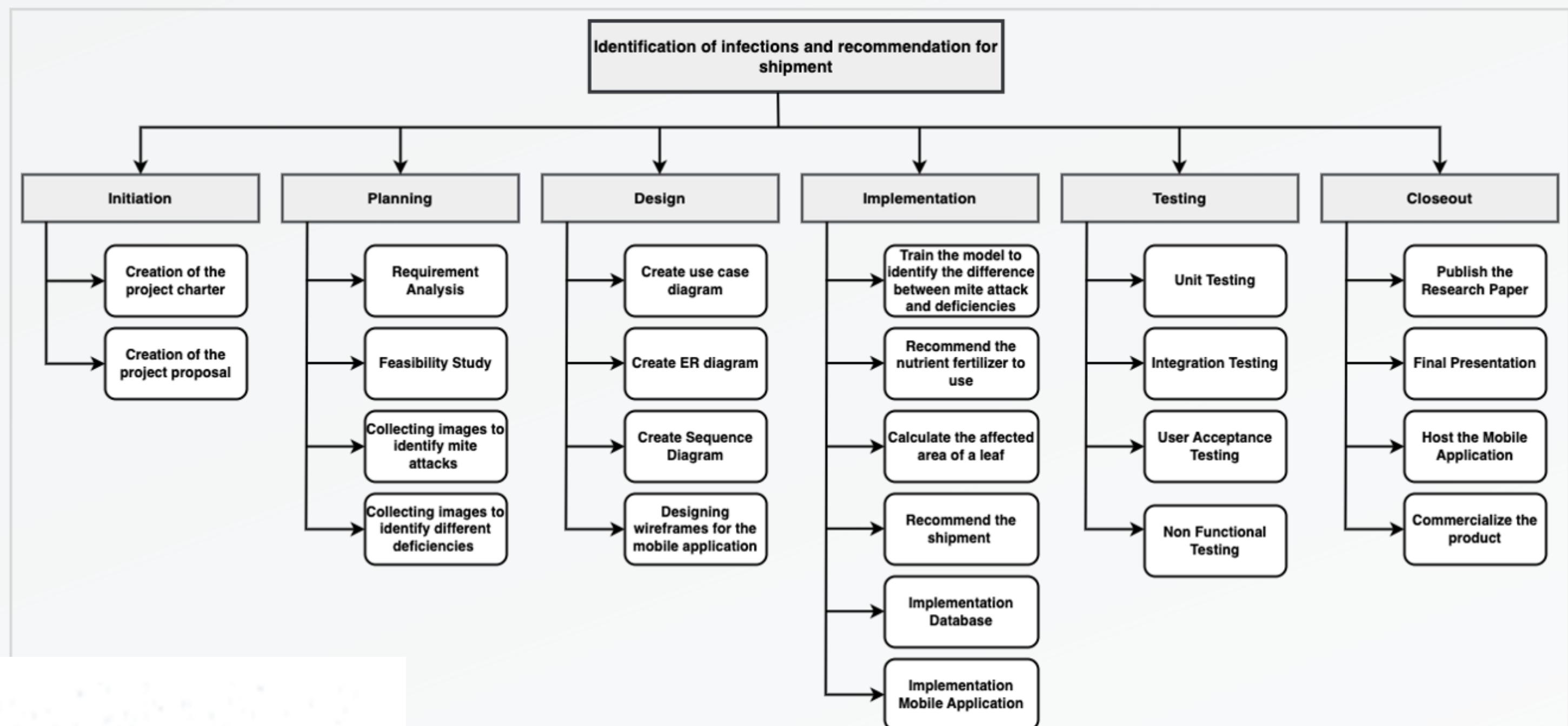
How to determine whether to ,

- Export leaves
- Sell them locally
- Keep them for the mother plant

INDIVIDUAL SYSTEM DIAGRAM



WORK BREAKDOWN



SAMPLE DATA



IMPLEMENTATION

IDENTIFYING THE DEFICIENCIES

Data Preprocessing

```
import splitfolders  
  
input_folder = "/content/drive/MyDrive/images/deficiencies"  
output_folder = "/content/drive/MyDrive/output"  
  
splitfolders.ratio(input_folder, output=output_folder, seed=1337, ratio=(.75, .2, .05), group_prefix=None, move=False)  
  
Copying files: 1170 files [00:50, 23.00 files/s]
```

IMPLEMENTATION

Exploratory Data Analysis

```
# Folder location
data_dir = pathlib.Path('/content/drive/MyDrive/output/train')
train_dir = '/content/drive/MyDrive/output/train'
val_dir = '/content/drive/MyDrive/output/val'
test_dir = '/content/drive/MyDrive/output/test'

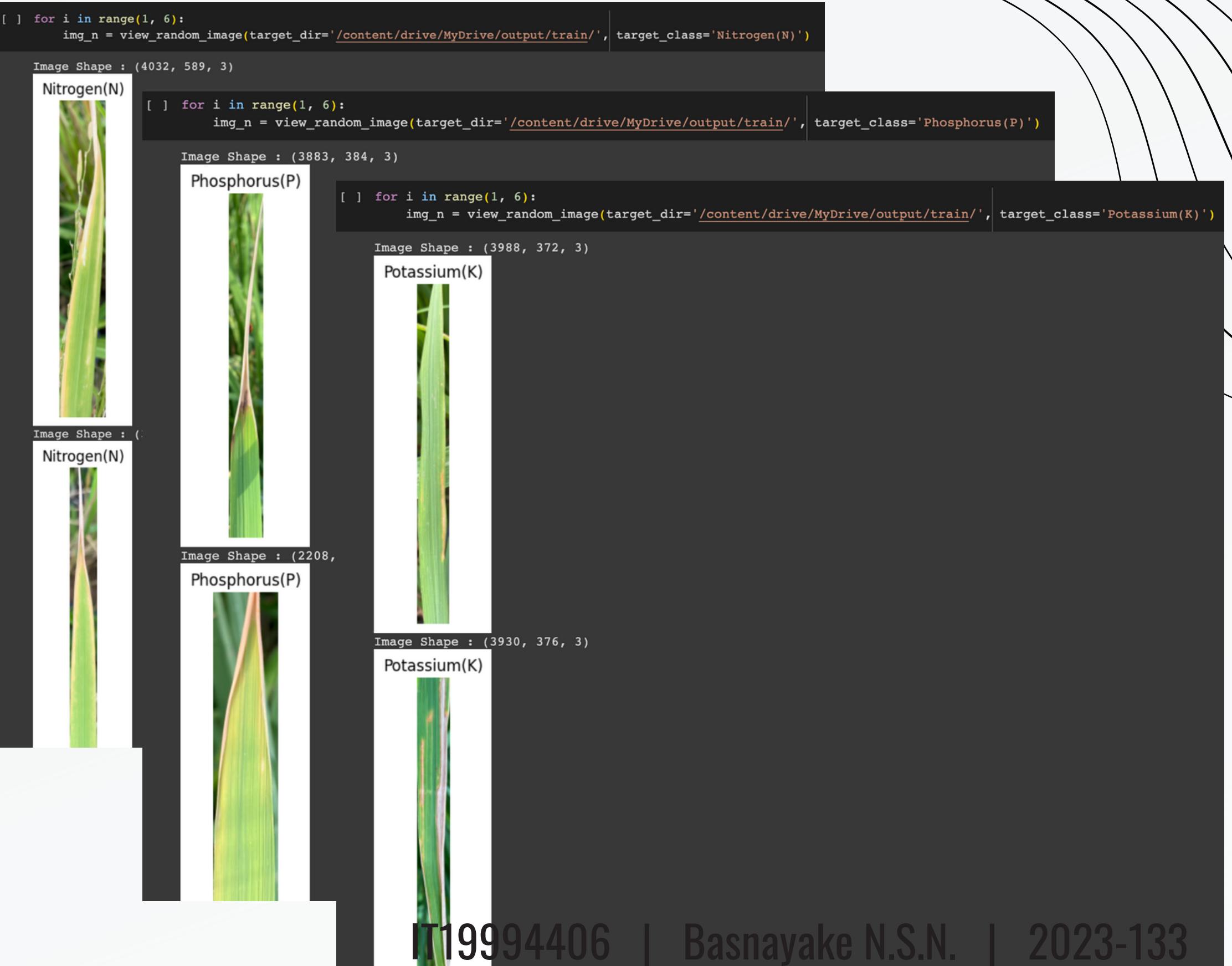
[ ] # Show classes
class_name = np.array(sorted([item.name for item in data_dir.glob("*")]))
print(class_name)

['Nitrogen(N)' 'Phosphorus(P)' 'Potassium(K)']

[ ] # Show total data in folder

for dirpath, dirnames, filenames in os.walk('/content/drive/MyDrive/output'):
    print(f"{len(dirnames)} folders and {len(filenames)} images in {dirpath}")

3 folders and 0 images in /content/drive/MyDrive/output
3 folders and 0 images in /content/drive/MyDrive/output/train
0 folders and 287 images in /content/drive/MyDrive/output/train/Potassium(K)
0 folders and 340 images in /content/drive/MyDrive/output/train/Nitrogen(N)
0 folders and 249 images in /content/drive/MyDrive/output/train/Phosphorus(P)
3 folders and 0 images in /content/drive/MyDrive/output/val
0 folders and 76 images in /content/drive/MyDrive/output/val/Potassium(K)
0 folders and 90 images in /content/drive/MyDrive/output/val/Nitrogen(N)
0 folders and 66 images in /content/drive/MyDrive/output/val/Phosphorus(P)
3 folders and 0 images in /content/drive/MyDrive/output/test
0 folders and 20 images in /content/drive/MyDrive/output/test/Potassium(K)
0 folders and 24 images in /content/drive/MyDrive/output/test/Nitrogen(N)
0 folders and 18 images in /content/drive/MyDrive/output/test/Phosphorus(P)
```



IMPLEMENTATION

Feature Engineering

```
[ ]  
  
# Sets the global random seed.  
tf.random.set_seed(46)  
  
# preprocess data  
train_datagen = ImageDataGenerator(rescale=1/255.0,  
                                     rotation_range=0.2,  
                                     zoom_range=0.2,  
                                     width_shift_range=0.2,  
                                     height_shift_range=0.2,  
                                     vertical_flip=True,  
                                     horizontal_flip=True)  
valid_datagen = ImageDataGenerator(rescale=1/255.0)  
test_datagen = ImageDataGenerator(rescale=1/255.0)  
  
[ ] # flow from directory using datagen (Pipeline)  
train_data = train_datagen.flow_from_directory(train_dir,  
                                              batch_size=16,  
                                              target_size=(224, 224),  
                                              class_mode='categorical',  
                                              shuffle=True,  
                                              seed=46)  
  
valid_data = valid_datagen.flow_from_directory(val_dir,  
                                              batch_size=16,  
                                              target_size=(224, 224),  
                                              class_mode='categorical',  
                                              shuffle=False,  
                                              seed=46)  
  
test_data = test_datagen.flow_from_directory(test_dir,  
                                              batch_size=16,  
                                              target_size=(224, 224),  
                                              class_mode='categorical',  
                                              shuffle=False,  
                                              seed=46)  
  
Found 876 images belonging to 3 classes.  
Found 232 images belonging to 3 classes.  
Found 62 images belonging to 3 classes.
```

```
[ ] # Check class_indices  
train_data.class_indices  
  
{'Nitrogen(N)': 0, 'Phosphorus(P)': 1, 'Potassium(K)': 2}  
  
[ ] # Set labels  
train_y=train_data.classes  
val_y=valid_data.classes  
test_y=test_data.classes  
  
[ ] # Check shapes of labels  
print("train_y.shape: ", train_y.shape)  
print("val_y.shape: ", val_y.shape)  
print("test_y.shape: ", test_y.shape)  
  
train_y.shape: (876,)  
val_y.shape: (232,)  
test_y.shape: (62,)
```

IMPLEMENTATION

Modeling

```
[ ] # Model Definition
base_model = tf.keras.applications.DenseNet121(input_shape=(224, 224, 3),
                                               include_top=False,
                                               weights='imagenet')

base_model.trainable = False

Downloading data from https://storage.googleapis.com/tensorflow/keras-applications/densenet121_weights_tf_dim_ordering_tf_kernels_notop.h5
29084464/29084464 [=====] - 0s 0us/step

[ ] # Model Preparation
global_average_layer = GlobalAveragePooling2D()
prediction_layer = Dense(3)
softmax = Activation('softmax')

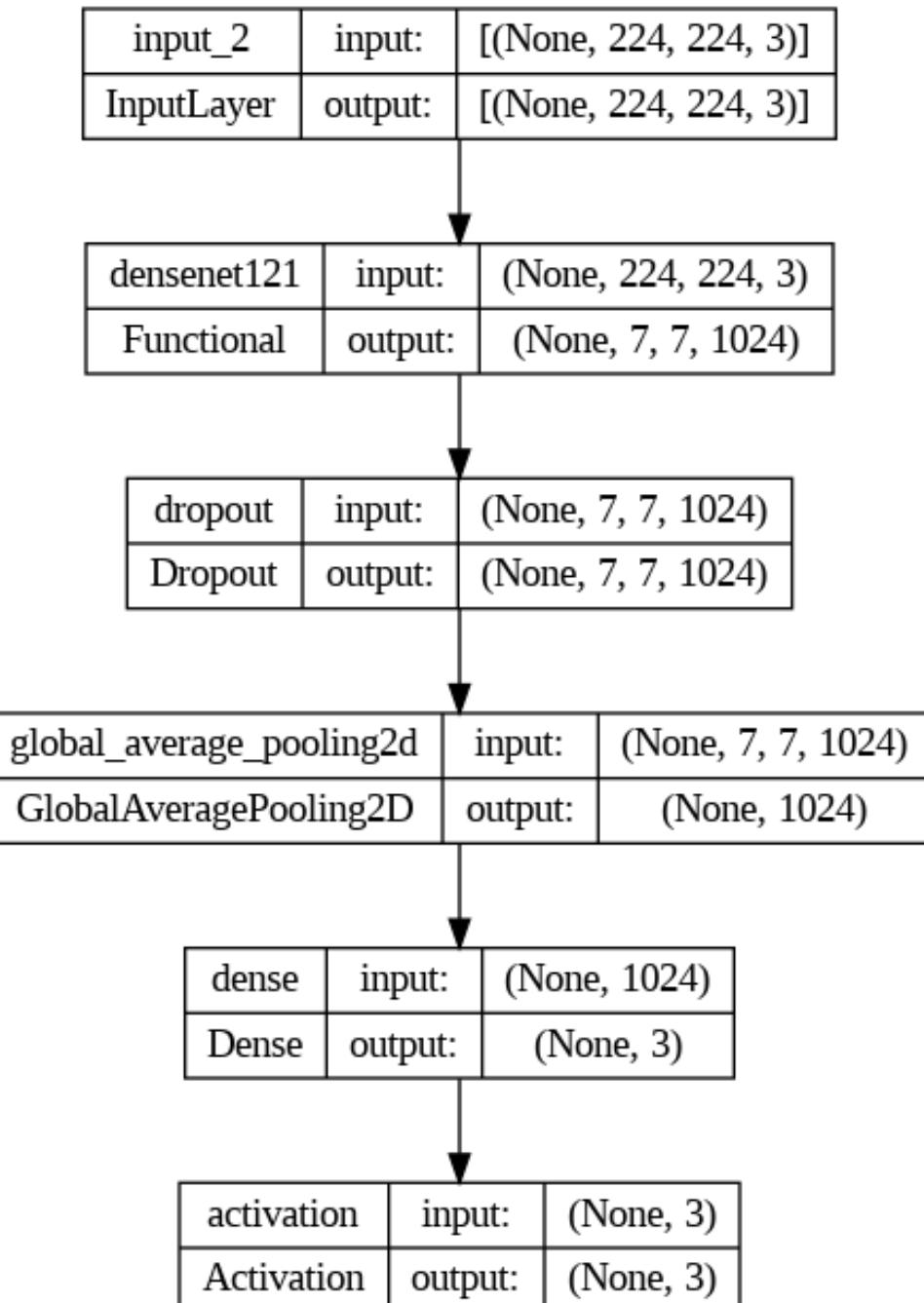
inputs = Input(shape=(224, 224, 3))
x = base_model(inputs, training=False)
x = Dropout(0.25)(x)
x = global_average_layer(x)
outputs = prediction_layer(x)
outputs = softmax(outputs)

model = Model(inputs, outputs)

[ ] # Model Summary
model.summary()

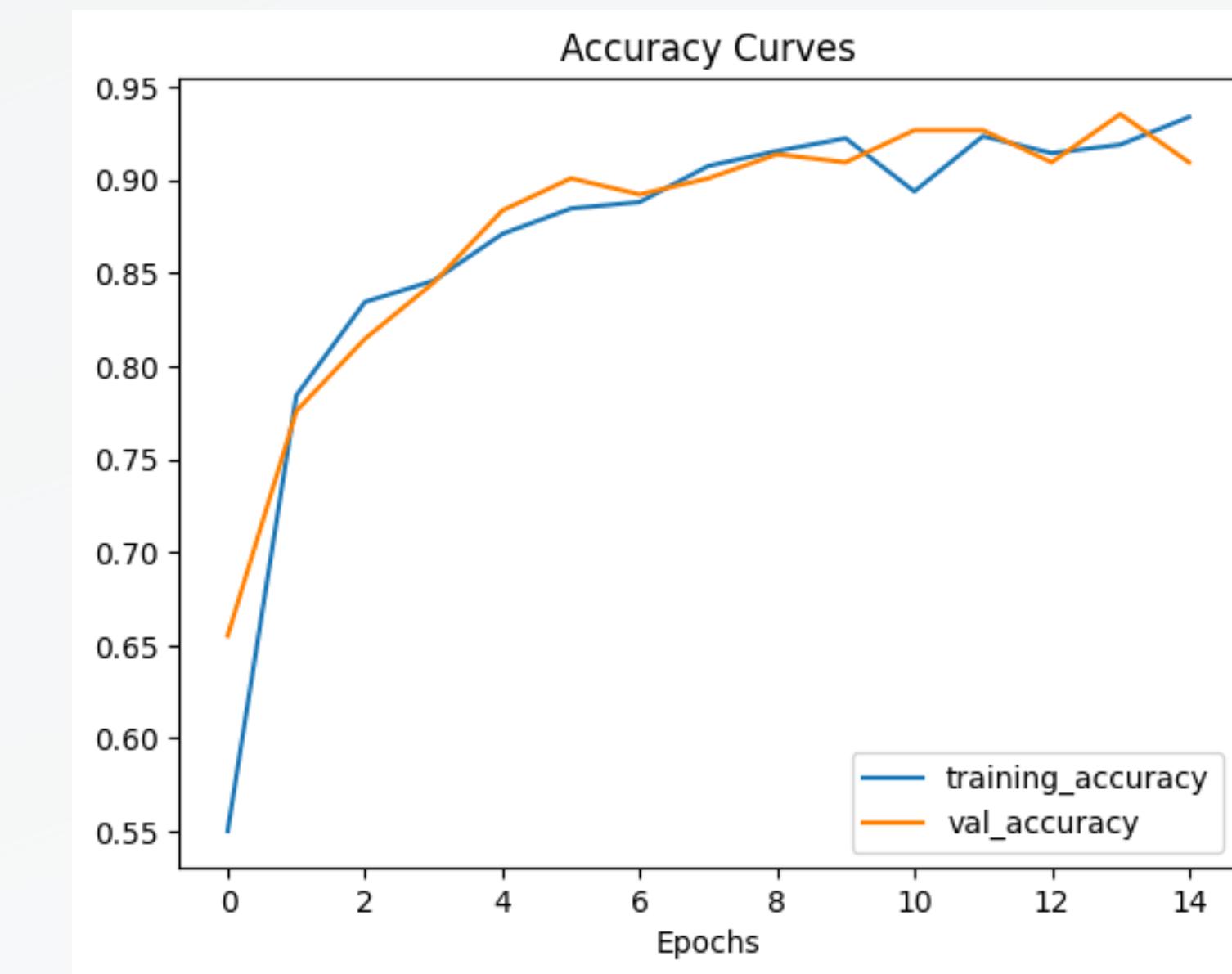
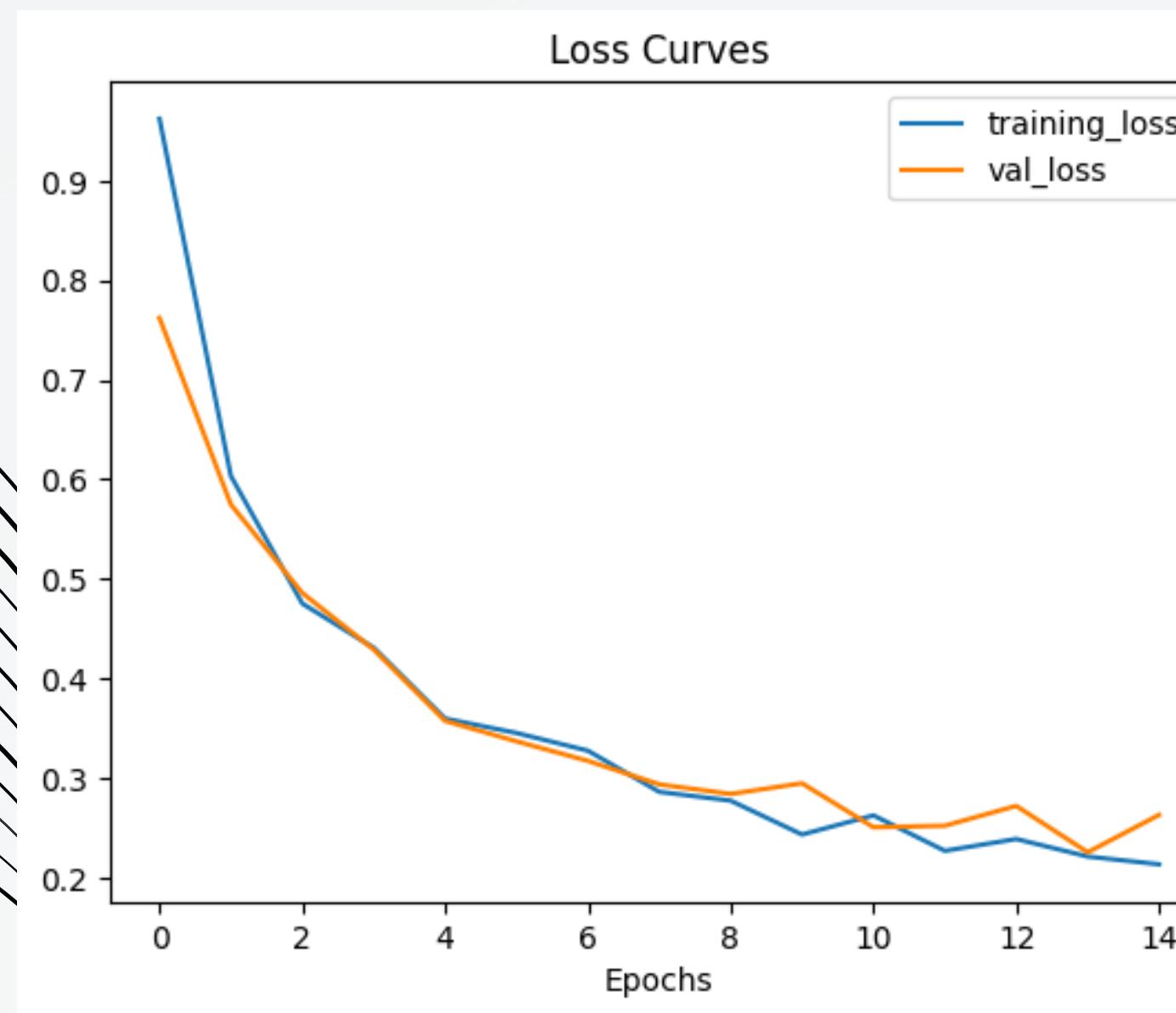
Model: "model"
Layer (type)          Output Shape         Param #
=====
input_2 (InputLayer)  [(None, 224, 224, 3)]  0
densenet121 (Functional) (None, 7, 7, 1024)  7037504
dropout (Dropout)      (None, 7, 7, 1024)  0
global_average_pooling2d (GlobalAveragePooling2D) (None, 1024)  0
dense (Dense)          (None, 3)            3075
activation (Activation) (None, 3)            0
=====
Total params: 7,040,579
Trainable params: 3,075
Non-trainable params: 7,037,504
Screenshot
```

Model Summary



IMPLEMENTATION

Model Evaluation

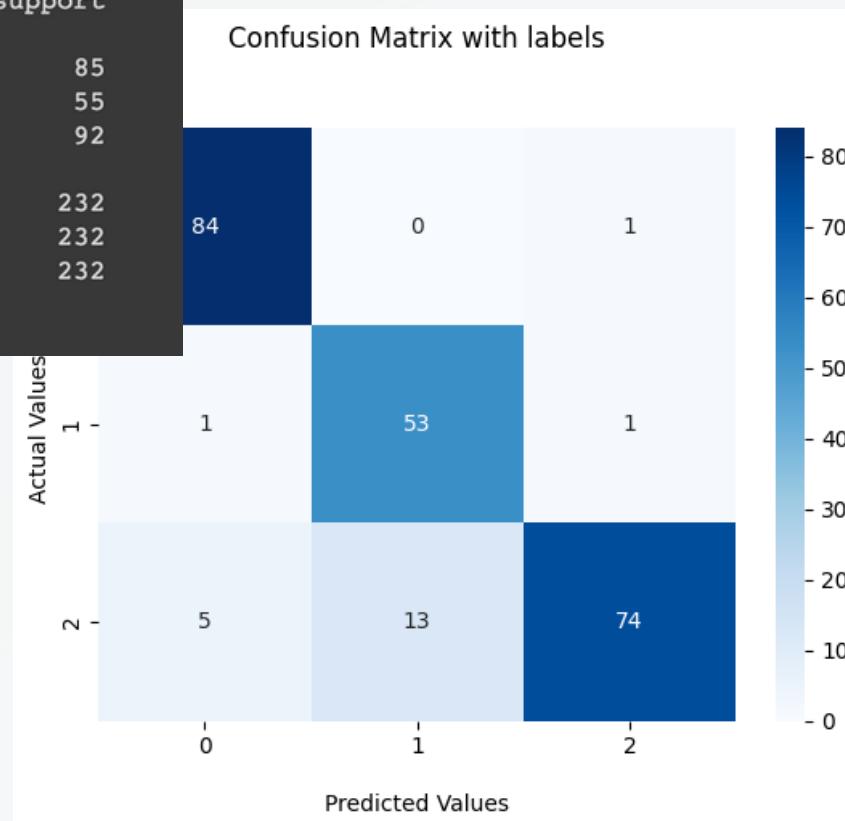


IMPLEMENTATION

Classification Report for Validate Data

```
[ ] # classification report
val_pred = model.predict(valid_data)
val_pred = val_pred.argmax(axis=1)
print(classification_report(val_pred, val_y))
```

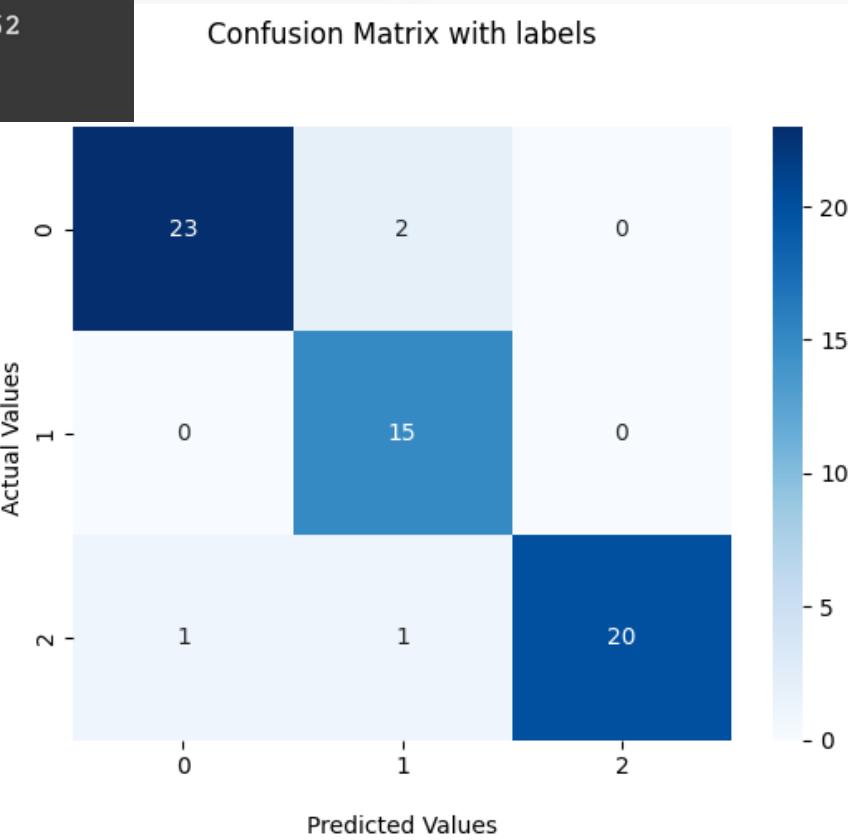
```
15/15 [=====] - 32s 2s/step
      precision    recall   f1-score   support
          0       0.93     0.99     0.96     85
          1       0.80     0.96     0.88     55
          2       0.97     0.80     0.88     92
accuracy           0.91    232
macro avg         0.90     0.92     0.91    232
weighted avg      0.92     0.91     0.91    232
```



Classification Report for Test Data

```
[ ] # classification report
test_pred = model.predict(test_data)
test_pred = test_pred.argmax(axis=1)
print(classification_report(test_pred, test_y))

4/4 [=====] - 9s 2s/step
      precision    recall   f1-score   support
          0       0.96     0.92     0.94     25
          1       0.83     1.00     0.91     15
          2       1.00     0.91     0.95     22
accuracy           0.94    62
macro avg         0.93     0.94     0.93    62
weighted avg      0.94     0.94     0.94    62
```



IMPLEMENTATION

CALCULATING THE EFFECTED AREA

```
[ ] import cv2
import numpy as np
#from config import input_path, leaf_area

input_path = '/content/drive/MyDrive/leaf_images/test.jpeg'
leaf_area = '/content/drive/MyDrive/leaf_images/mask_test.jpeg'

# Open a simple image
img = cv2.imread(input_path)
filename_mask = leaf_area

# converting from gbr to hsv color space
img_HSV = cv2.cvtColor(img, cv2.COLOR_BGR2HSV)

# leaf color range for hsv color space
HSV_mask = cv2.inRange(img_HSV, (0, 15, 0), (17, 170, 255))

HSV_mask = cv2.morphologyEx(HSV_mask, cv2.MORPH_OPEN, np.ones((3, 3), np.uint8))

# converting from gbr to YCbCr color space
img_YCrCb = cv2.cvtColor(img, cv2.COLOR_BGR2YCrCb)

# leaf color range for hsv color space
YCrCb_mask = cv2.inRange(img_YCrCb, (0, 135, 85), (255, 180, 135))

YCrCb_mask = cv2.morphologyEx(YCrCb_mask, cv2.MORPH_OPEN, np.ones((3, 3), np.uint8))
global_mask = cv2.bitwise_and(YCrCb_mask, HSV_mask)
global_mask = cv2.medianBlur(global_mask, 3)
global_mask = cv2.morphologyEx(global_mask, cv2.MORPH_OPEN, np.ones((4, 4), np.uint8))
cv2.imwrite(filename_mask, global_mask)

# counting the number of pixels
number_of_white_pix = np.sum(global_mask == 255)
number_of_black_pix = np.sum(global_mask == 0)
total_pix = number_of_white_pix + number_of_black_pix
white_percentage = (number_of_white_pix / total_pix)

print("Leaf affected percentage: " + "{:.2%}".format(white_percentage));
```

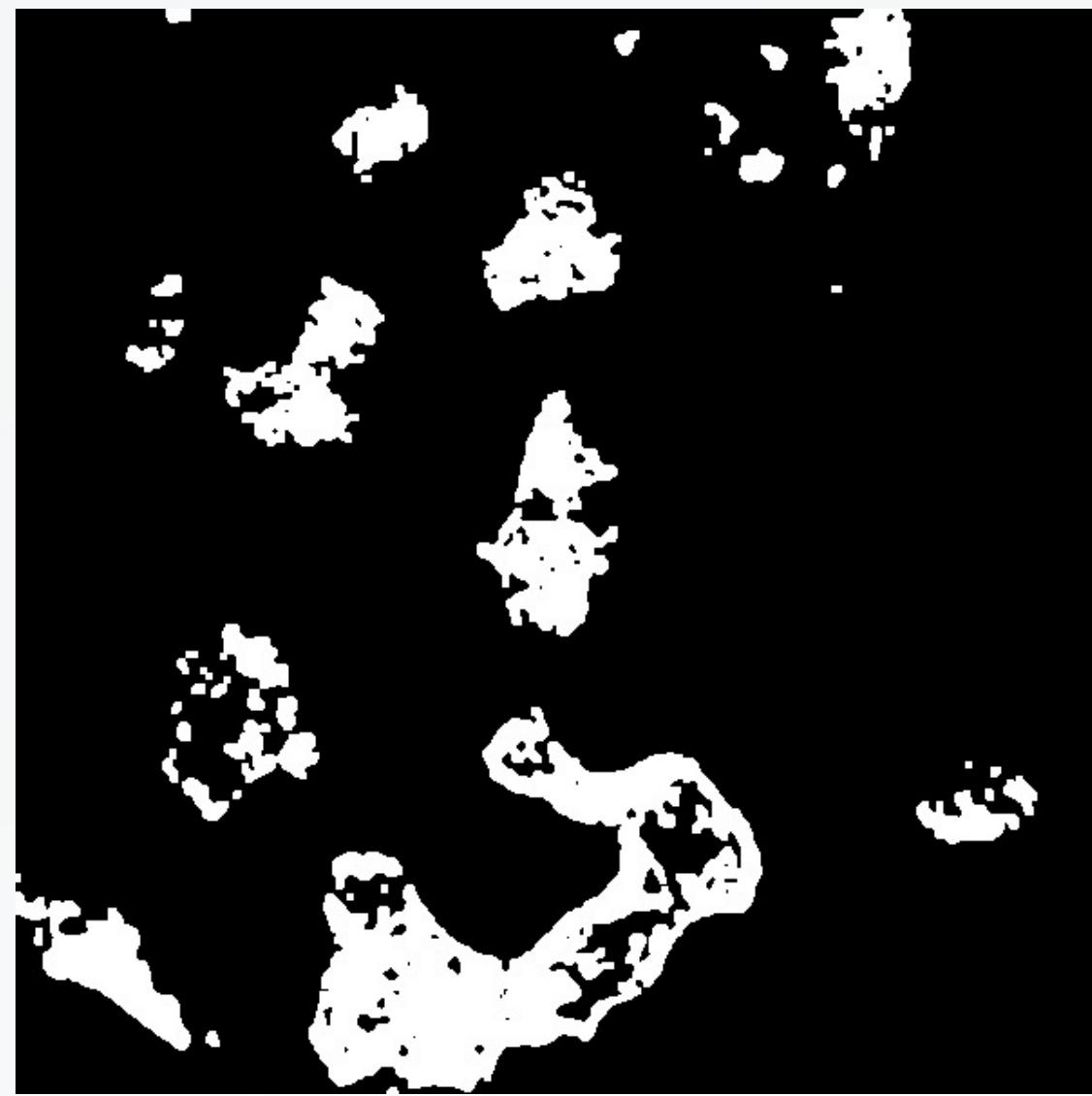
Leaf affected percentage: 12.04%

IMPLEMENTATION

Actual image



Mask image



UI DESIGNS



FUTURE WORKS

IMPROVEMENTS

- Differentiate the nutrient deficiencies from mite attacks
- Recommend the shipment for export, local or as the mother plant

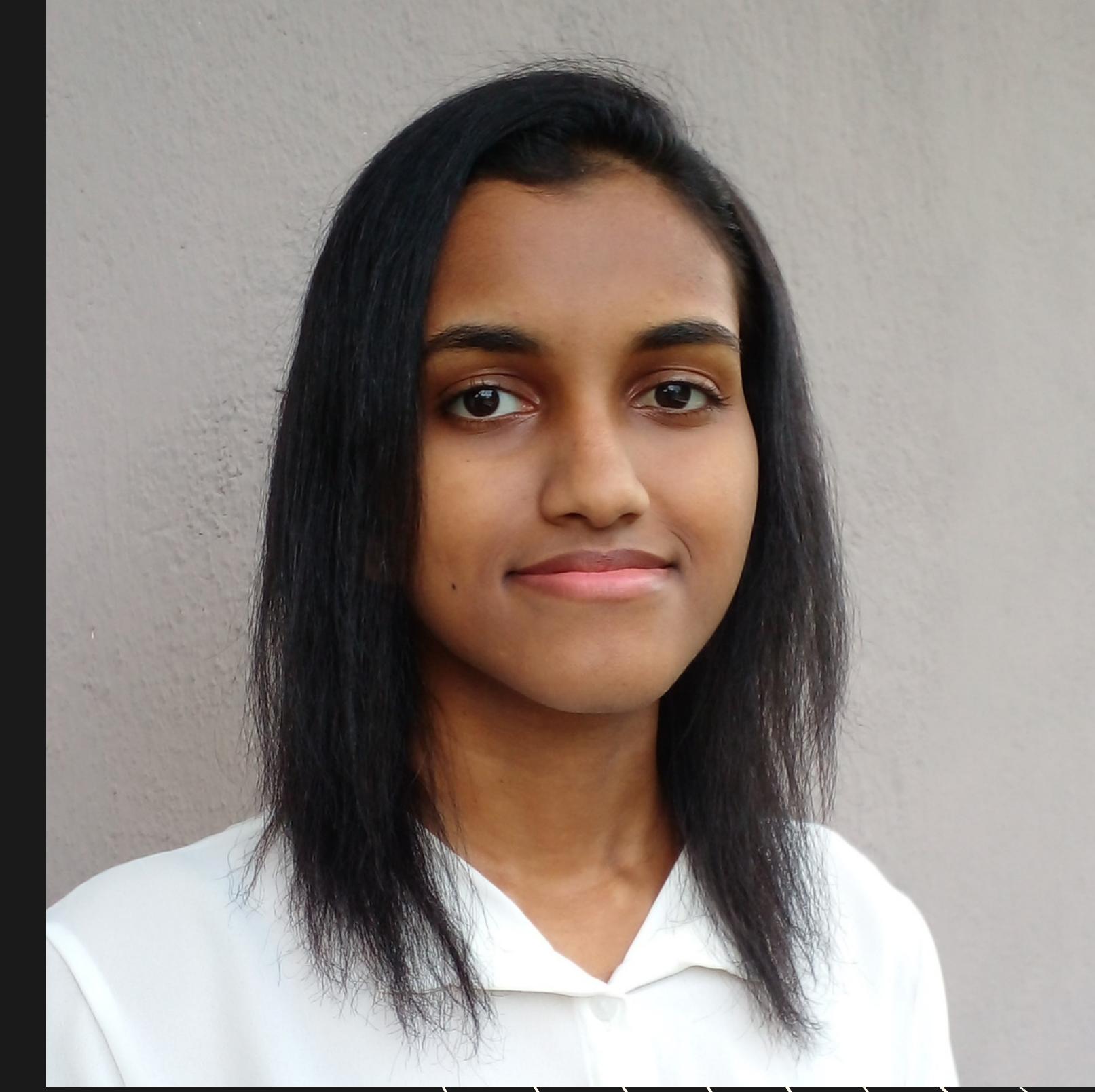
TARGETS

- Implement the frontend
- Build the mobile application for the working function



IT19169736
Gamage M.G.U.D.

SELECTING THE BEST CUT FLOWER PLANT TO GROW BASED ON WEATHER CONDITIONS AND AVAILABLE RESOURCES

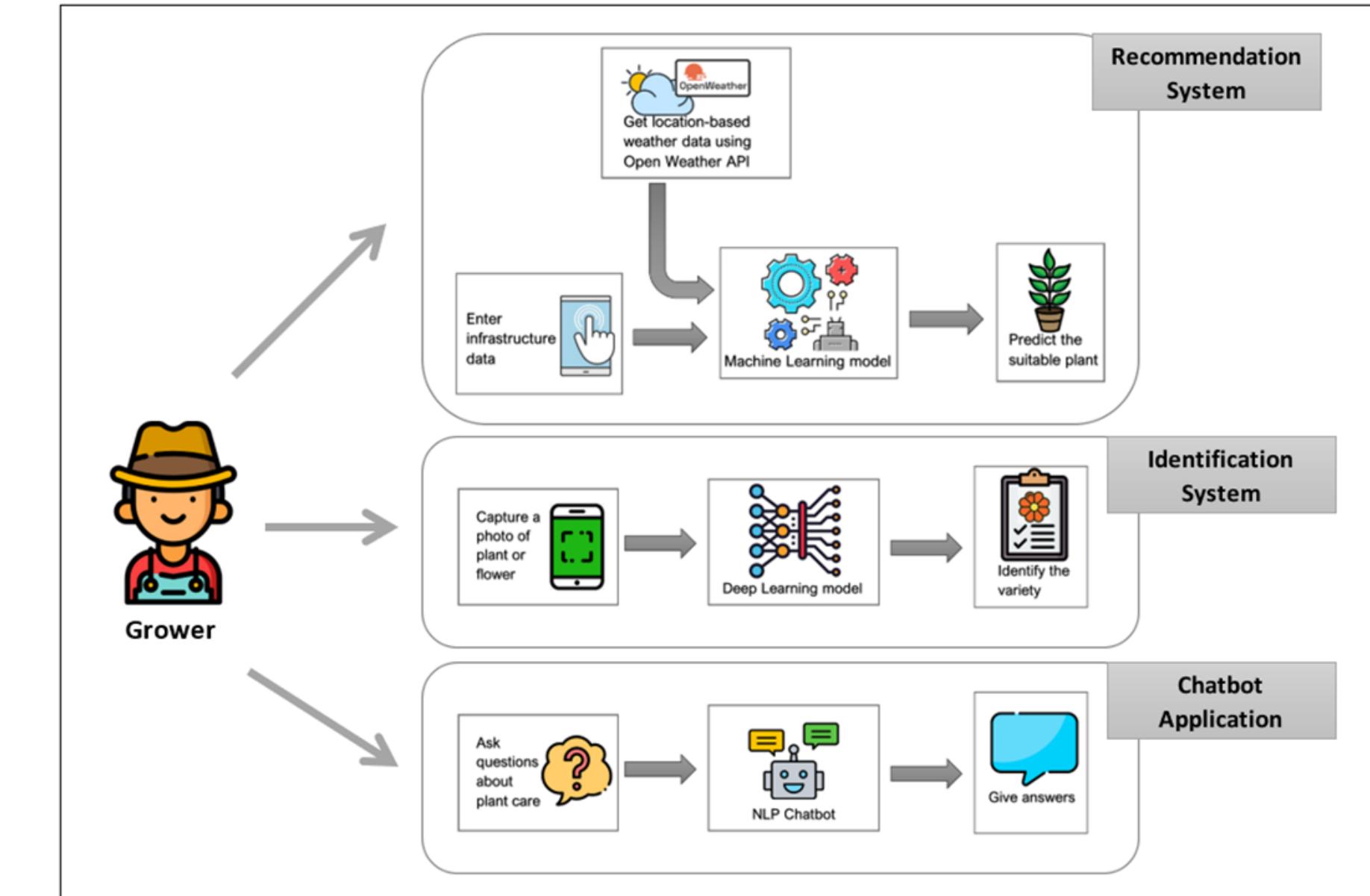


Specialization : Data Science

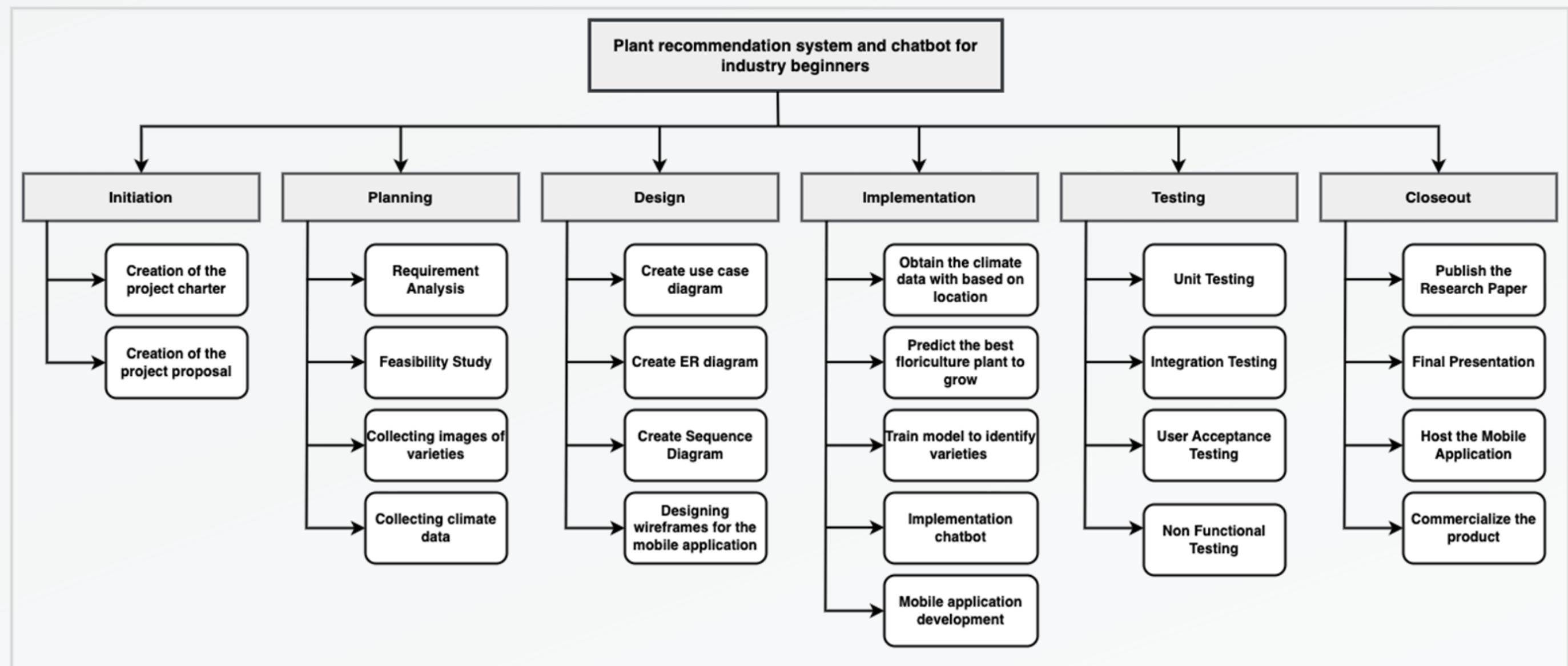
RESEARCH PROBLEM

- How to select the best floriculture plant to grow according to weather and resource factors?
- How to identify varieties of diverse floriculture crops?

INDIVIDUAL SYSTEM DIAGRAM



WORK BREAKDOWN



DATA COLLECTING

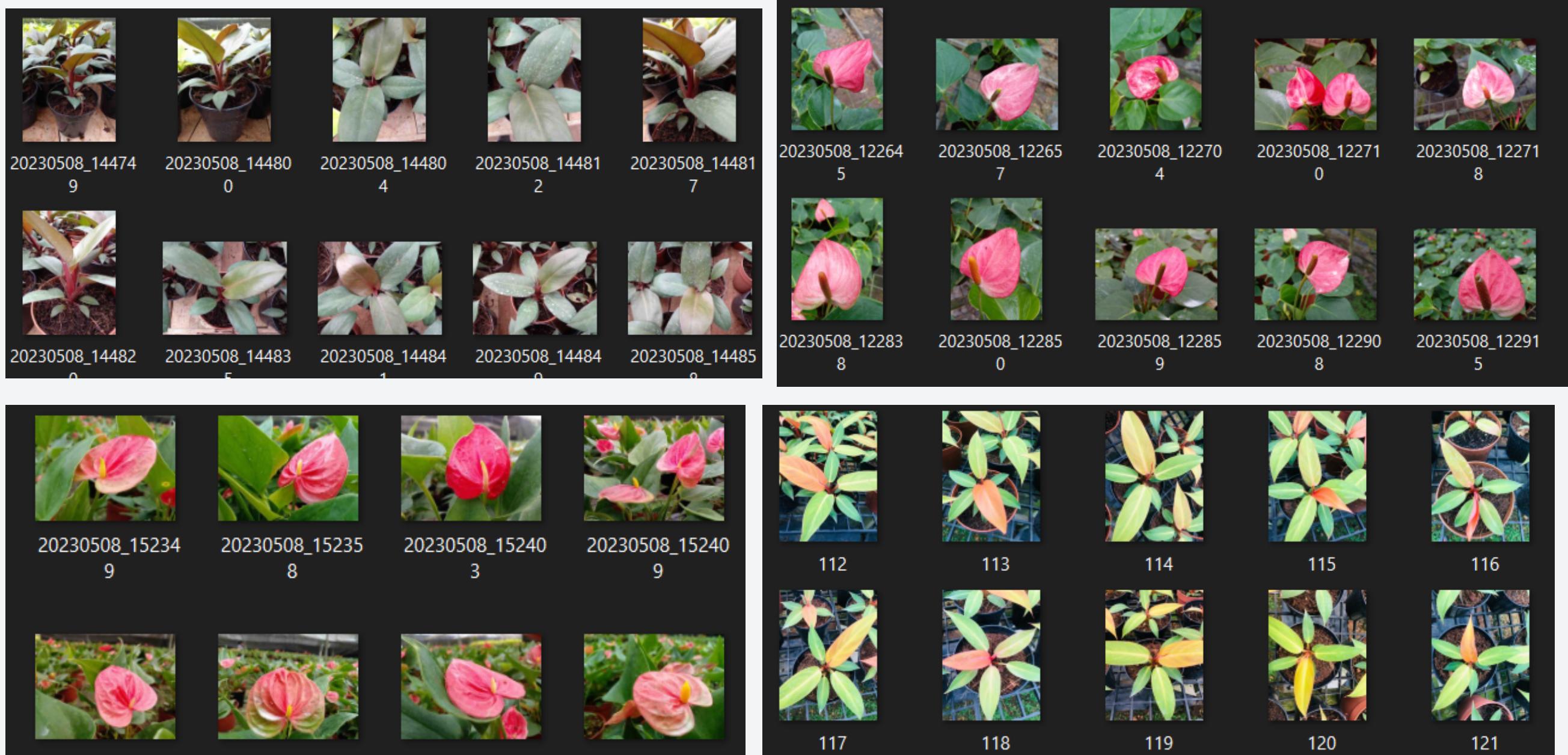
| Date | Areal | Net | Variety | Fertiliser | Ratio | Rate | Weather condition | Quantity | Name of sprayer user | Recommended Officer |
|------------|-----------|-----|-----------------|-------------------|----------|----------|-------------------|----------|----------------------|---------------------|
| 06/01/2021 | H | H | Ferns / Orchids | Yaromila Osmocote | 18-9-10 | kg/plant | Sunny | 5 kg | Renuka | |
| 06/01/2021 | H | H | Ferns | Osmocote | 18-9-10 | kg/plant | Sunny | 5 kg | Nadeeka | |
| 07/01/2021 | H | H | Ferns | Osmocote | 11-11-10 | kg/plant | " | 5 kg | Nadeeka | |
| 12/01/2021 | H,B,F,F | H | Orchids / Ferns | Grow more | 20-20-20 | kg/plant | " | 5 kg | Kelum / Sanju | |
| 12/01/2021 | H | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 500 g | Kelum / Sanju | |
| 15/01/2021 | B,F | H | Orchids | Osmocote | 18-9-10 | kg/plant | " | 25 kg | Kusuma | |
| 18/01/2021 | H,B,F | H | Orchids / Ferns | Grow more | 20-20-20 | kg/plant | " | 5 kg | Kelum / Sanju | |
| 18/01/2021 | H | H | Dendrobium | Basfoliar | 18-40-10 | kg/plant | " | 500 g | Kelum / Sanju | |
| 21/01/2021 | H,F,B | H | Orchids / Ferns | Grow more | 30-10-10 | kg/plant | " | 1 kg | Kelum / Sanju | |
| 21/01/2021 | B,F | H | Orchids | Grow more | 8-30-30 | kg/plant | " | 1.5 kg | Kelum / Sanju | |
| 21/01/2021 | B,F | H | Orchids | Osmocote | 11-11-10 | kg/plant | " | 15 kg | Kusuma | |
| 25/01/2021 | H,F,D | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 250 g | Kelum / Sanju | |
| 25/01/2021 | H,F,B,F,F | H | Orchids / Ferns | Basfoliar | 20-20-20 | kg/plant | " | 2 kg | Kelum / Sanju | |
| 29/01/2021 | H,F,D | H | Orchids | Grow more | 30-10-10 | kg/plant | " | 1 kg | Kelum / Sanju | |
| 29/01/2021 | B,F,D | H | Orchids | Grow more | 8-30-30 | kg/plant | " | 1.5 kg | Kelum / Sanju | |
| 01/02/2021 | H,F,D,B,F | H | Orchids / Ferns | Grow more | 20-20-20 | kg/plant | " | 500 g | Kelum / Sanju | |
| 01/02/2021 | H,F,D | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 500 g | Kelum / Sanju | |
| 01/02/2021 | B,F,D | H | Orchids | Osmocote | 18-5-18 | kg/plant | " | 15 kg | Prema | |
| 01/02/2021 | B,F,D | H | Orchids | Yaromila | 18-5-18 | kg/plant | " | 500 g | Kelum / Sanju | |
| 05/02/2021 | B,F,D | H | Orchids | Grow more | 6-30-30 | kg/plant | " | 500 g | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Orchids | Grow more | 30-10-10 | kg/plant | " | 2 kg | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 250 g | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Orchids / Ferns | Grow more | 20-20-20 | kg/plant | " | 2 kg | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Orchids / Ferns | Osmocote | 18-9-10 | kg/plant | " | 2 kg | Sunma / Sthu | |
| 05/02/2021 | B,F,D | H | Orchids | Grow more | 6-30-30 | kg/plant | " | 1.5 kg | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Orchids | Grow more | 30-10-10 | kg/plant | " | 1 kg | Kelum / Sanju | |
| 05/02/2021 | B,F,D | H | Orchids / Ferns | Grow more | 8-30-30 | kg/plant | " | 2 kg | Kelum / Sanju | |
| 05/02/2021 | B,F,D | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 250 g | Kelum / Sanju | |
| 05/02/2021 | H,F,D | H | Orchids | Basfoliar | 30-10-10 | kg/plant | " | 1 kg | Kelum / Sanja | |
| 05/02/2021 | H,F,D | H | Orchids / Ferns | Grow more | 20-20-20 | kg/plant | " | 500 g | Kelum / Sanja | |
| 05/02/2021 | H,F,D | H | Dendrobium | Basfoliar | 15-40-10 | kg/plant | " | 250 g | Kelum / Sanja | |
| 05/02/2021 | H,F,D | H | Orchids | Grow more | 6-30-30 | kg/plant | " | 1.5 kg | Kelum / Sanja | |

| OMEGA GREEN (PVT) LTD., | | | | | | | | | | |
|--|----------------------|--|--|------|---------|-----------|-----------------|--|--|--|
| Mulleray East, Godagama, Badulla, Sri Lanka. Tel: +94-11-2791840, Fax: 94-11-2788951 E-mail: omega@omega.lk GLOBAL G.A.P. Certified No. 4090373878430 | | | | | | | | | | |
| CONSIGNEE | | SHIPMENT DETAILS | | | | | | | | |
| Tomaszewski Sp. Z.o.o Al Krakowska 205/211, 02-180 Warsaw Poland | | DATE 14.05.2020 FLIGHT NO UL 1205 CMB - FRA 18.05.2020 0 | | | | | | | | |
| Vat nr PL 5220100226 | | A.W.B. NO 003 4624 3996 H.A.W.B.NO S1462001784 | | | | | | | | |
| REX NUMBER LKREX114328235DC0305 | | Notify Party Nagel Airfreight GmbH Flughafen Frankfurt/ Main Tor 26/Geb 454, 60549, Frankfurt/Germany | | | | | | | | |
| DELIVERED UNDER: | | PAYMENT TERMS TELEGRAPHIC TRANSFER (TT) | | | | | | | | |
| PACKING LIST | | | | | | | | | | |
| CARTON NO | DESCRIPTION OF GOODS | | | QTY | PCS/BOX | NO/ BOXES | TOTAL WEIGHT KG | | | |
| 1 | 353 | Livistona rotundifolia 1P 30/35 cm | | 9884 | 28 | 353 | 2 E | | | |
| 354 | 357 | Livistona rotundifolia 1P 30/35 cm | | 110 | 29 | 4 | | | | |
| TOTAL 10000 357 | | | | | | | | | | |
| OMEGA GREEN (PVT) LTD Byceee' (W.M.S.Weerasekara) EXPORT MANAGER | | | | | | | | | | |

| Date | Heir. | Bed. | Op. | Sp. | 10p. | order |
|----------|-------|------|--------|-------|--------|-------|
| 06.07.20 | 58 | 4T | | | 43 | |
| | | 6B | 303 | | 10 | |
| | | 5T | 28 | | 25 | |
| | | 8T | | | 131 | |
| 11.07.20 | 31 | 1T | | | 10 | |
| 28.07.20 | | | | | 18 | |
| | | | | | 10 | |
| 01.09.20 | 26 | 2T | | | 58 | |
| | | 3T | 1T | | 25 | |
| | | 2T | 445 | 8cm | 570 | |
| | | 2T | 68 | " | 270 | |
| | | 6T | 445 | " | 147 | |
| | | 6T | 0.1980 | 8cm | 59 | |
| | | 10T | 0.02 | " | 10 | |
| 30.10.20 | 40 | 1T | | | 58 | |
| | | 40A | 360 | 8cm | 270 | |
| | | 39A | 820 | 8cm | 270 | |
| | | 41A | 266 | 20/25 | 270 | |
| | | 24 | 8cm | " | 22.50 | |
| | | 40A | 326 | 8cm | 250 | |
| | | 41A | 392 | " | 250 | |
| | | " | 96 | 20/20 | 250 | |
| 03.11.20 | 45 | 27B | 10 | | 126 | |
| 01.12.20 | 49 | | | | 74 | |
| | | 0.02 | 300 | " | 250 | |
| 07.12.20 | 50 | T | 80 | 2P | 135 | |
| 21.12.20 | 52 | 17A | | | 70 | |
| | | 10TA | | | 24.20 | |
| | | 6T2 | | | 124.20 | |
| | | 10TA | | | 99.50 | |
| | | 10TA | | | 47 | |
| | | | | | 91 | |
| | | | | | Mango | |

Written records about expenses

DATA COLLECTING



Images taken from nurseries

DATA CLEANING

```
# handling null values  
df.isna().sum()
```

```
Temperature           1  
Relative_Humidity    0  
Water_liters_per_yr   0  
Land_size_in_sqft     1  
Nethouse_land_preperation_cost 1  
No_of_plants          2  
Planting_cost          1  
Maintaining_cost       0  
Labour_cost            1  
Flower_yield_per_yr    1  
Crop_name              1  
dtype: int64
```

```
# drop null values  
df = df.dropna()  
df.isna().sum()
```

```
Temperature           0  
Relative_Humidity    0  
Water_liters_per_yr   0  
Land_size_in_sqft     0  
Nethouse_land_preperation_cost 0  
No_of_plants          0  
Planting_cost          0  
Maintaining_cost       0  
Labour_cost            0  
Flower_yield_per_yr    0  
Crop_name              0  
dtype: int64
```

Drop rows that contain null values

DATA SUMMARIZING

```
df.shape  
(2379, 11)  
  
df.dtypes  


|                                |         |
|--------------------------------|---------|
| Temperature                    | float64 |
| Relative_Humidity              | int64   |
| Water_liters_per_yr            | int64   |
| Land_size_in_sqft              | float64 |
| Nethouse_land_preperation_cost | float64 |
| No_of_plants                   | float64 |
| Planting_cost                  | float64 |
| Maintaining_cost               | int64   |
| Labour_cost                    | float64 |
| Flower_yield_per_yr            | float64 |
| Crop_name                      | object  |
| dtype: object                  |         |

  
df.head(10)  


|   | Temperature | Relative_Humidity | Water_liters |
|---|-------------|-------------------|--------------|
| 0 | 30.9        | 74                |              |
| 1 | 30.4        | 78                |              |


```

```
df.describe()  


|       | Temperature | Relative_Humidity | Wa |
|-------|-------------|-------------------|----|
| count | 2379.000000 | 2379.000000       |    |
| mean  | 22.388146   | 80.076923         |    |
| std   | 5.017900    | 8.065671          |    |
| min   | 13.000000   | 60.000000         |    |
| 25%   | 17.200000   | 74.000000         |    |

  
df.groupby('Crop_name').size()  


| Crop_name     | size |
|---------------|------|
| Alstroemeria  | 299  |
| Anthurium     | 294  |
| Carnation     | 300  |
| Chrysanthemum | 296  |
| Gerbera       | 298  |
| Lily          | 295  |
| Orchid        | 300  |
| Rose          | 297  |

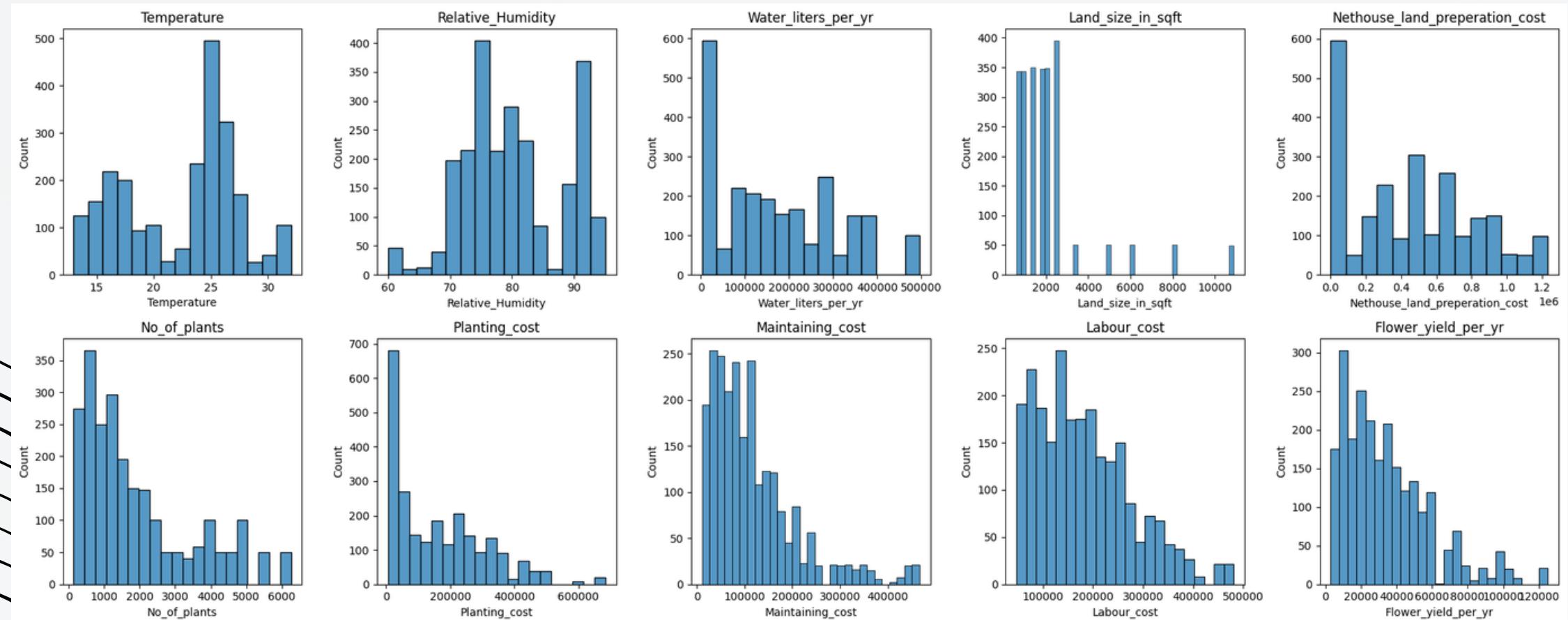


```
dtype: int64
```


```

Gain insights about dimensions, basic statistical properties and each class's size of the dataset.

DATA VISUALIZING

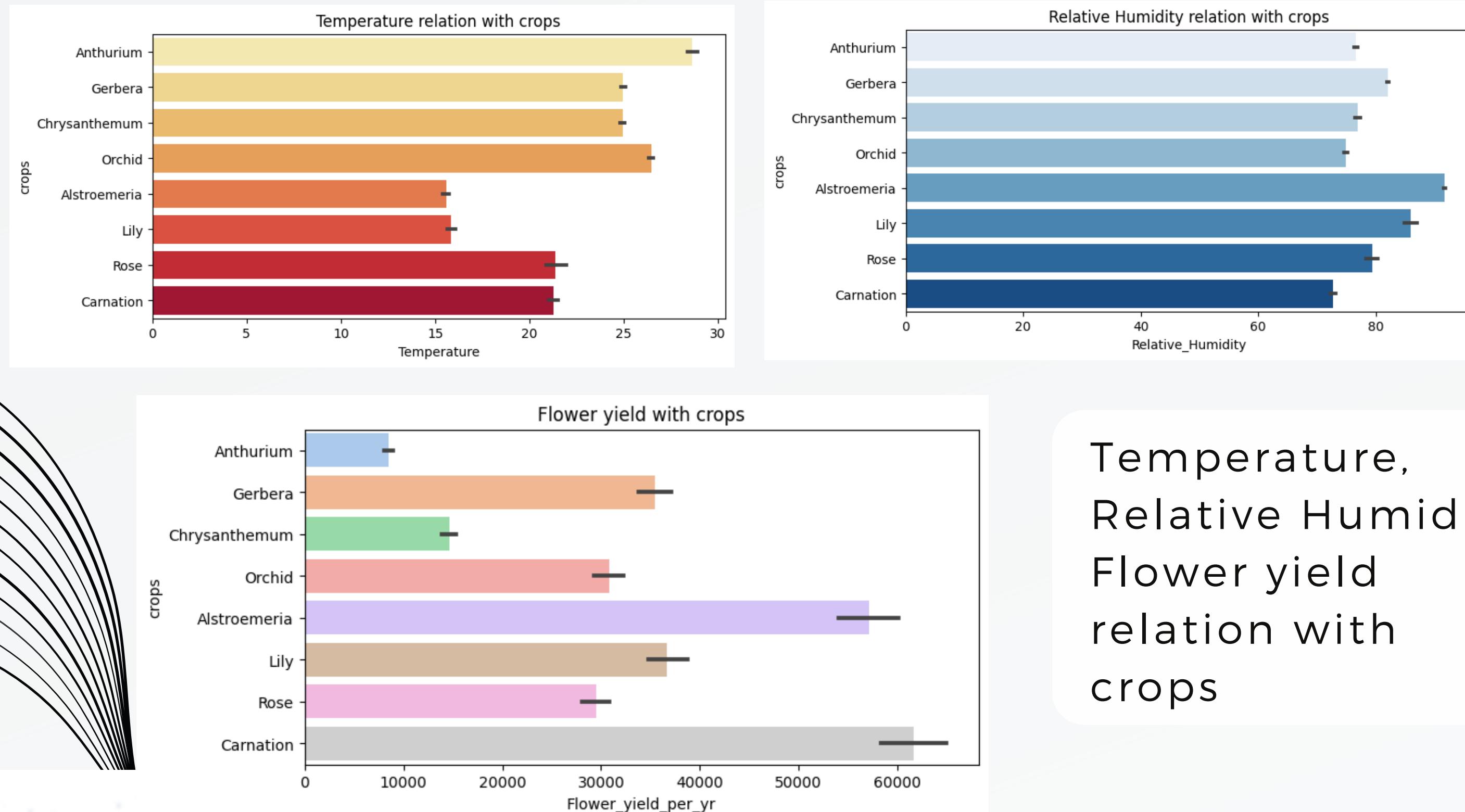


Histograms of features

| | Temperature | Relative_Humidity | Water_liters_per_yr | Land_size_in_sqft | Nethouse_land_preperation_cost | No_of_plants | Planting_cost | Maintaining_cost | Labour_cost | Flower_yield_per_yr |
|--------------------------------|-------------|-------------------|---------------------|-------------------|--------------------------------|--------------|---------------|------------------|-------------|---------------------|
| Temperature | 1 | -0.55 | -0.44 | -0.44 | 0.42 | 0.1 | -0.18 | 0.1 | 0.14 | -0.49 |
| Relative_Humidity | -0.55 | 1 | 0.057 | 0.43 | -0.42 | -0.35 | 0.36 | -0.087 | -0.15 | 0.14 |
| Water_liters_per_yr | -0.44 | 0.057 | 1 | 0.12 | 0.34 | 0.37 | -0.04 | 0.63 | 0.27 | 0.6 |
| Land_size_in_sqft | -0.44 | 0.43 | 0.12 | 1 | -0.2 | 0.063 | 0.64 | -0.052 | 0.14 | 0.64 |
| Nethouse_land_preperation_cost | 0.42 | -0.42 | 0.34 | -0.2 | 1 | 0.62 | -0.31 | 0.63 | 0.46 | 0.16 |
| No_of_plants | 0.1 | -0.35 | 0.37 | 0.063 | 0.62 | 1 | -0.074 | 0.29 | 0.7 | 0.45 |
| Planting_cost | -0.18 | 0.36 | -0.04 | 0.64 | -0.31 | -0.074 | 1 | 0.15 | 0.46 | 0.32 |
| Maintaining_cost | 0.1 | -0.087 | 0.63 | -0.052 | 0.63 | 0.29 | 0.15 | 1 | 0.58 | 0.39 |
| Labour_cost | 0.14 | -0.15 | 0.27 | 0.14 | 0.46 | 0.7 | 0.46 | 0.58 | 1 | 0.45 |
| Flower_yield_per_yr | -0.49 | 0.14 | 0.6 | 0.64 | 0.16 | 0.45 | 0.32 | 0.39 | 0.45 | 1 |

Correlation Matrix

DATA VISUALIZING



Temperature,
Relative Humidity,
Flower yield
relation with
crops

DATA PREPROCESSING

Select features and target values

```
# Features Selection  
selected_features = ['Temperature', 'Relative_Humidity', 'Water_liters_per_yr',  
    'Land_size_in_sqft', 'Nethouse_land_preperation_cost', 'No_of_plants',  
    'Planting_cost', 'Maintaining_cost', 'Labour_cost',  
    'Flower_yield_per_yr']
```

```
# feature values and target values  
y = df['Crop_name']  
X = df[selected_features]
```

Split the dataset into training & testing sets
with 33% of the data allocated for testing

```
# data splitting  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=2)
```

MODEL EVALUATION & COMPARISON

3 classification models are defined:

1. Decision Tree Classifier
2. Gaussian Naive Bayes
3. Gradient Boosting Classifier

```
# Spot Check Algorithms
models = []
models.append(('CART', DecisionTreeClassifier()))
models.append(('NB', GaussianNB()))
models.append(('GB', GradientBoostingClassifier()))
```

MODEL EVALUATION & COMPARISON

For each model
Stratified K-Fold
cross-validation with
3 folds is set up

It evaluates the
model's performance
using accuracy as the
scoring metric

```
# evaluate each model in turn
results = []
names = []
for name, model in models:
    kfold = StratifiedKFold(n_splits=3, random_state=2, shuffle=True)
    cv_results = cross_val_score(model, X_train, y_train, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    print('%s: %f (%f)' % (name, cv_results.mean(), cv_results.std()))
```

MODEL EVALUATION & COMPARISON

Mean accuracy across the cross-validation folds = 0.997489

standard deviation as a measure of the performance consistency = 0.002349

CART: 0.997489 (0.002349)
NB: 0.916510 (0.017007)
GB: 0.997489 (0.002349)

CART & GB both shows high level of accuracy and consistency

MODEL TRAINING

Gradient Boosting Classifier is selected as the best model

```
model = GradientBoostingClassifier(n_estimators=100, learning_rate=0.001)
model.fit(X_train, y_train)
```

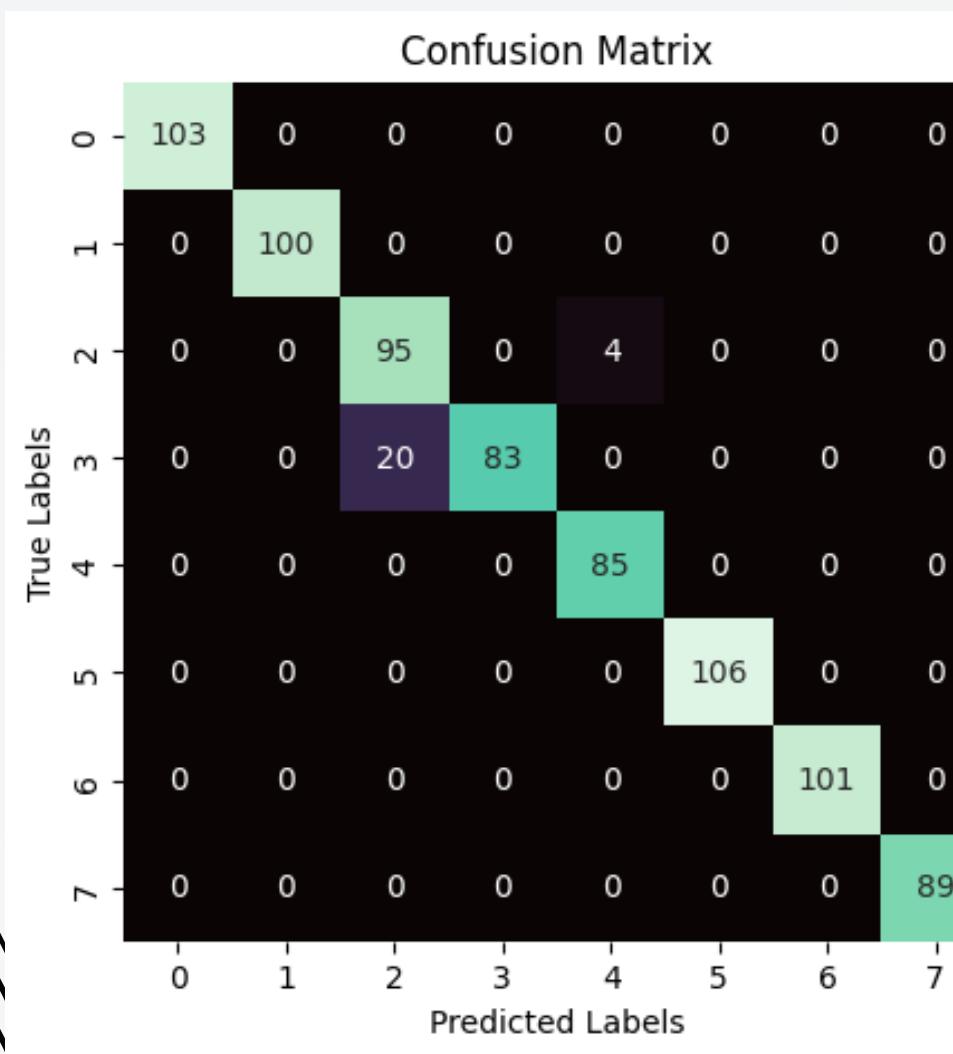
```
# make predictions on test dataset
predictions = model.predict(X_test)
```

Accuracy of 96.95% on the test dataset

```
# evaluate predictions
print("Accuracy : ", accuracy_score(y_test, predictions))
```

```
Accuracy : 0.9694656488549618
```

EVALUATION METRICS



Confusion Matrix

```
print(classification_report(y_test, predictions))
```

| | precision | recall | f1-score | support |
|---------------|-----------|--------|----------|---------|
| Alstroemeria | 1.00 | 1.00 | 1.00 | 103 |
| Anthurium | 1.00 | 1.00 | 1.00 | 100 |
| Carnation | 0.83 | 0.96 | 0.89 | 99 |
| Chrysanthemum | 1.00 | 0.81 | 0.89 | 103 |
| Gerbera | 0.96 | 1.00 | 0.98 | 85 |
| Lily | 1.00 | 1.00 | 1.00 | 106 |
| Orchid | 1.00 | 1.00 | 1.00 | 101 |
| Rose | 1.00 | 1.00 | 1.00 | 89 |
| accuracy | | | 0.97 | 786 |
| macro avg | 0.97 | 0.97 | 0.97 | 786 |
| weighted avg | 0.97 | 0.97 | 0.97 | 786 |

Classification Report

OBTAI N WEATHER DATA

OpenWeather API
to retrieve
Temperature &
Relative Humidity
for a specific city

```
# Parameters
api_key = "71d1945ed6fb6c764f9334a8b0c1dc4a"
city_name = "Malabe"

# Construct the API call
url = f'http://api.openweathermap.org/data/2.5/weather?q={city_name}&appid={api_key}'
print(url)

# Send the request
response = requests.get(url)

# Handle the response
if response.status_code == 200:
    data = json.loads(response.text)
    temperature_kelvin = data['main']['temp']
    temperature_celsius = temperature_kelvin - 273.15
    humidity = data['main']['humidity']
    print(f"Temperature: {temperature_celsius:.2f} °C")
    print(f"Humidity: {humidity}%")
else:
    print("Error:", response.status_code)

http://api.openweathermap.org/data/2.5/weather?q=Malabe&appid=71d1945ed6fb6c764f9334a8b0c1dc4a
Temperature: 28.94 °C
Humidity: 75%
```

PREDICT THE BEST CROP

- Temperature
- Relative Humidity
- No of plants
- Flower yield
- Land and water
- Planting, maintain, nethouse, labor cost

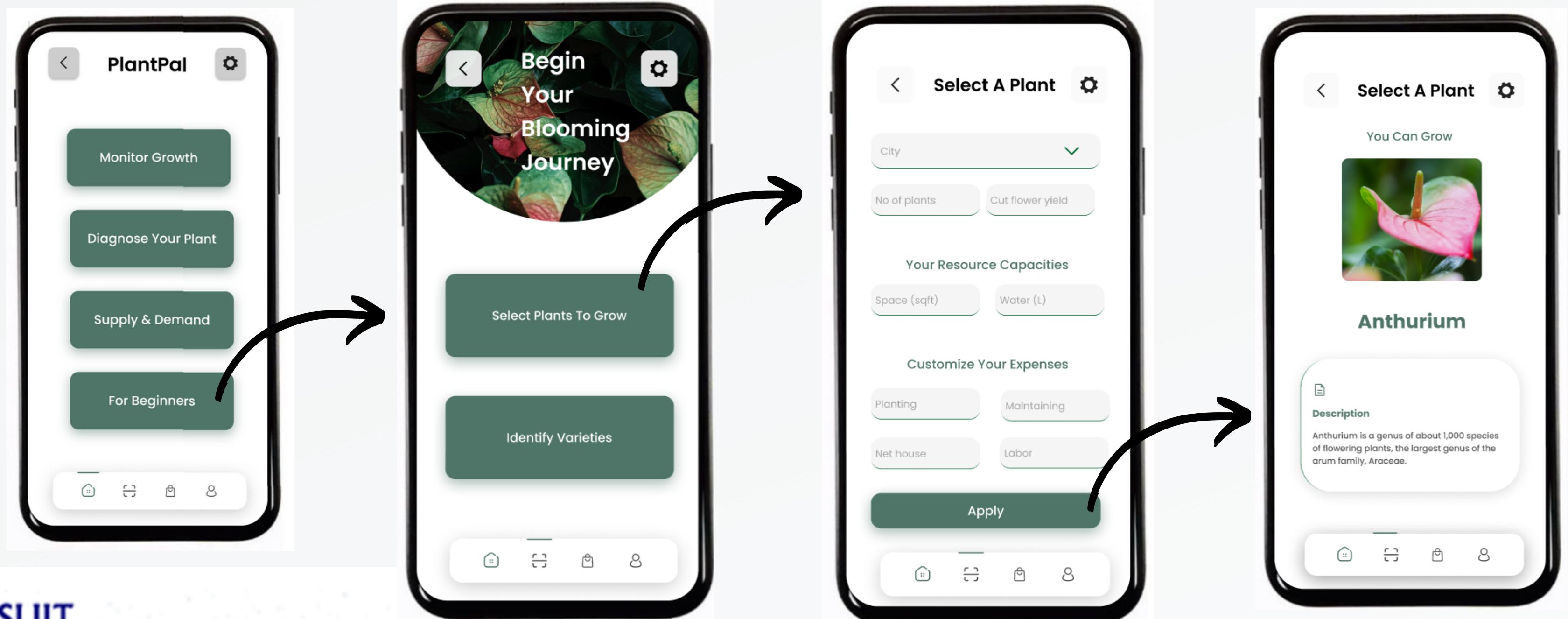
are fed into the model to get predictions

```
# predictions on the input array
prediction = pickled_model.predict(X_new)

# Print the predictions
print("The best crop to grow :", prediction[0])
```

The best crop to grow : Lily

UI DESIGNS



FUTURE WORKS

Improve the plant
selection model

Implement the Frontend

Develop the variety
identification

Complete the mobile
application development



IT20017088
Prabhshi P.A.N.

FORECAST THE FLUCTUATING DEMAND PATTERNS BASED ON SEASONS AND SPECIFIC PRODUCT PREFERENCES



Specialization : Data Science

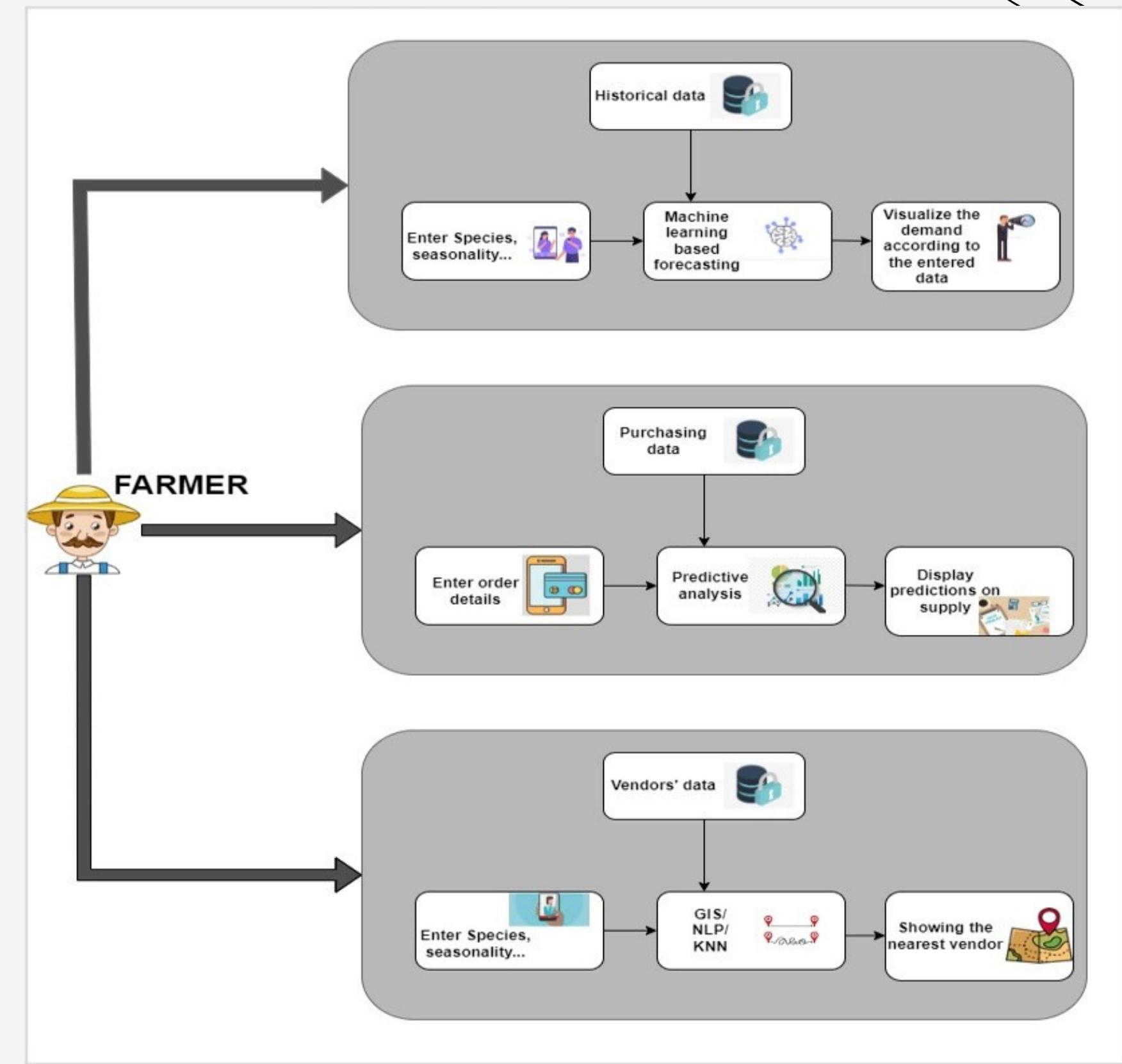
RESEARCH PROBLEM

How to predict the future demand accurately according to various factors?

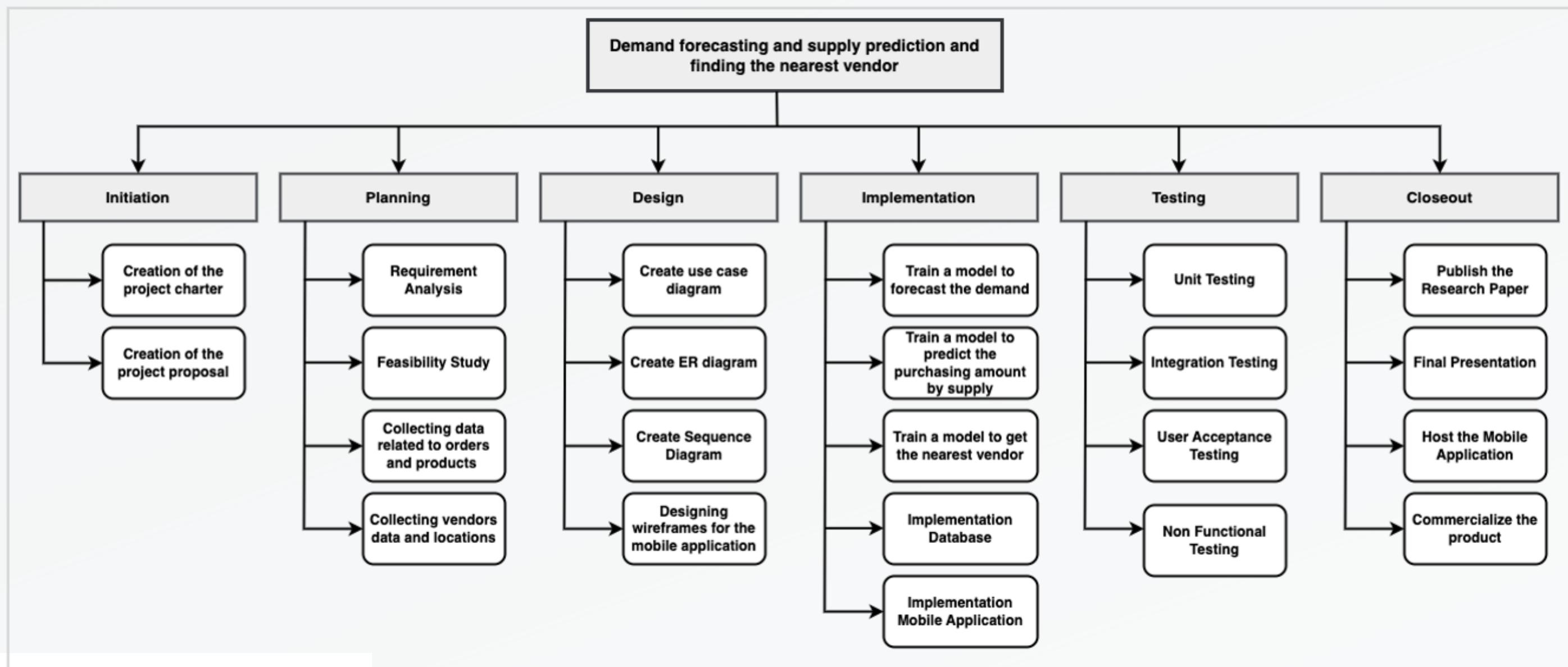
Various factors such as,

- Country / Region
- Season
- Variety

INDIVIDUAL SYSTEM DIAGRAM



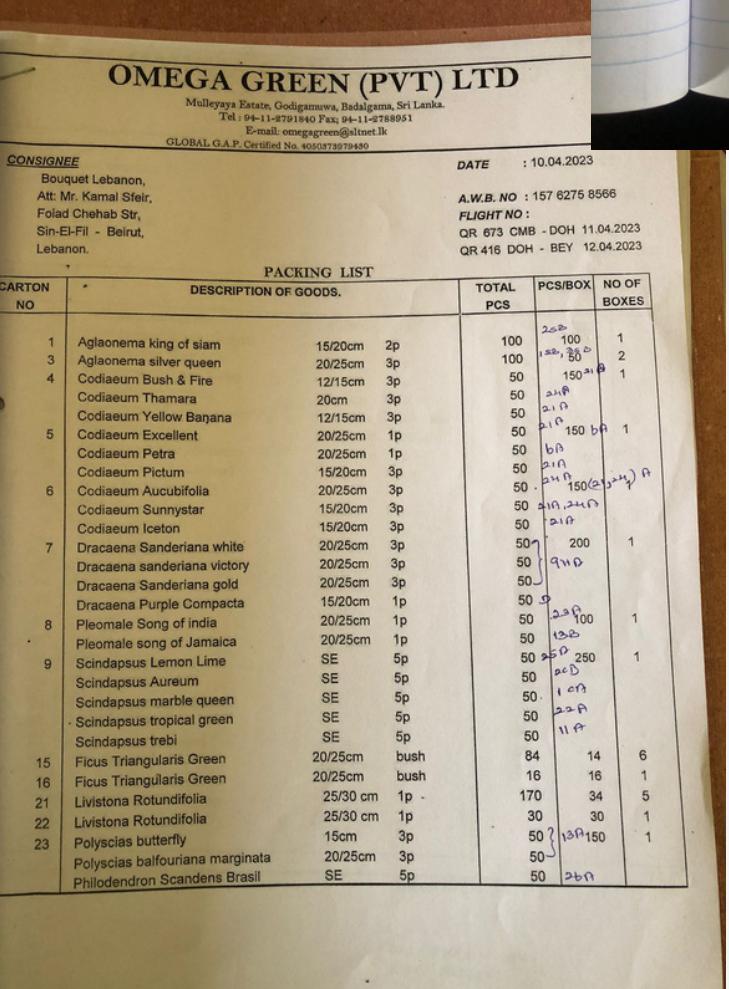
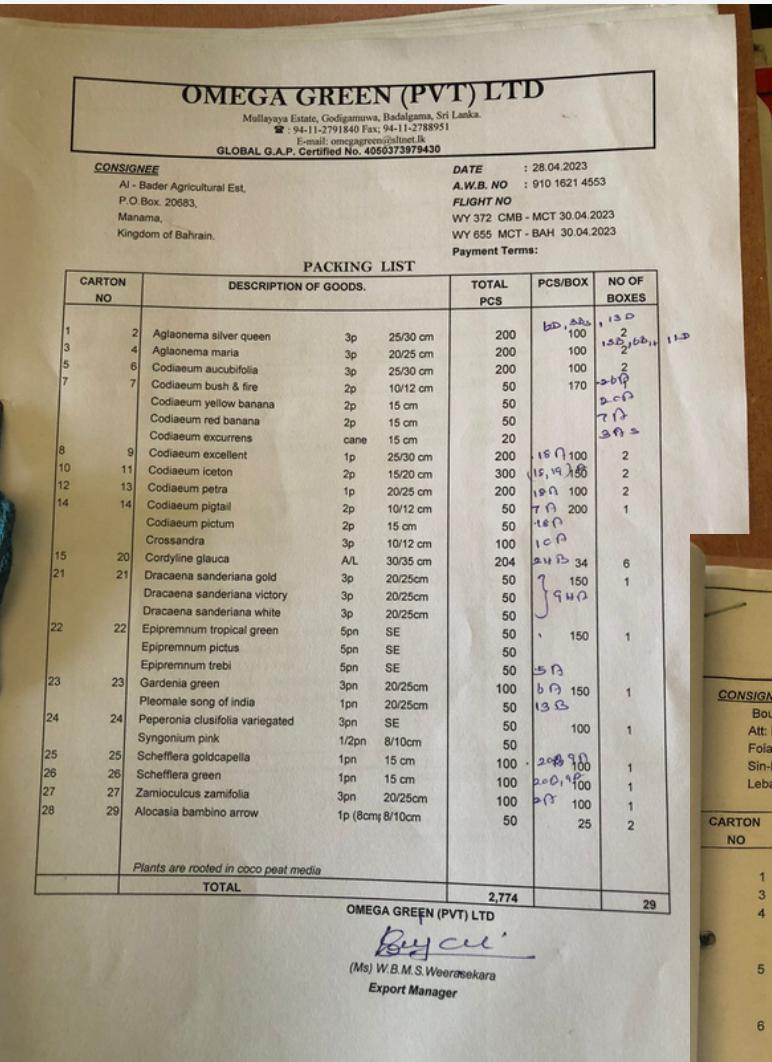
WORK BREAKDOWN



FIELD DATA

These data is collected
from the omega green
Pvt Ltd

And all data are collected
in manual way



| Date | Week No | Box No | Quantity | Size | order/shipments |
|----------|---------|--------|----------|------|-----------------------------|
| 14.02.18 | 07 | 0.B | 2128 | 1p | 10cm 0.10 |
| | | | 2184 | 1p | 15cm 0.10 |
| | | | 346 | 1p | " |
| 19.02.18 | 08 | 12B | 455 | 1p | 10cm 0.10 |
| | | 9.A | 401 | 1p | 15cm 0.10 |
| | | | = | | |
| 08.03.18 | 09 | 14A | 255 | 2p | 20cm |
| | | 13A | 293 | 3p | |
| | | | 14+ 71 | 3p | |
| 02.03.18 | 10 | 10B | 383 | 3p | 15/20cm |
| | | | | | AL 44 th (cont.) |
| 25.09.18 | 29 | 26A | 880 | 3p | |
| | | 26A | 1385 | | |
| | | 9.A | 522 | | |
| | | 16A | 71 | | |
| | | 29A | 280 | | |
| | | | 2788 | | |
| 10.12.18 | 50 | 29A | 51 | 3p | Bam pat |

| Customer Name | Address | Contact No. |
|-----------------------|-------------------------|-------------------------|
| Piami | Negombo | 077-6018494 |
| Anju flora | Pothuhera | 011-6490135 |
| Wasana Nursery | Mathara | 071-7384569 |
| Wenura | Kuliyapitiya | 077-2741224 |
| Suranjika Paradise | Galle | 077-3119080 |
| Dias - stopped | Makola | 076-7641010 |
| Paragan Plant Nursery | Anuradhapura | 071-8210452 |
| Nishani | Mathara | 077-8677183 |
| Dileepa | Higurakgoda | 077-282827 |
| Buwankaka | Negombo | 077-9814438 |
| Rathnayake (M/s) | Negombo | 077-3083724 |
| Dilsha (SAHA 2000) | Galle | 0777-49755 |
| Damroka | Galle | 0777-458421 |
| Latan | Kandy | 071-8303957 |
| Wasantha | Kandy | 076-7072130 |
| Thilina | Kandy | 071-5305957 |
| Darshana Gardens | Dehatottilai | 070-2638430/077-9886911 |
| Hirudi plant Nursery | Uva | 077-7915260/097-9894929 |
| La Dulce | 091-9256978/078-3795854 | |
| Ashoka (Meerigama) | 091-9256978/078-3795854 | |
| Wasana Nursery | Piliyandala | 0173-746772/0777-796772 |
| Withanage | Beralanda | 077-0063610 |
| Mahesh | Mahangama | 076-1093986 |
| Ruwan | Wonnappuwa | 072-6622616/077-614772 |
| Tharangi | Negombo | 076-6393027 |
| Anil - Amburum | O | 077-8083378 |
| Wasana Nursery | Wasana - Mathara | 071-4656150 |

IMPLEMENTATION

Data Preprocessing

DATA PREPROCESSING

Identify the null values

```
[11] 1 # Handle missing values
     2 # Check for missing values
     3 Demand_df.isnull().sum()
```

| | |
|-----------------|-----|
| County / Region | 0 |
| Year | 0 |
| Season | 0 |
| Variety | 0 |
| Quantity | 613 |
| 1 pcs | 0 |
| Price | 0 |
| dtype: int64 | |

Selecting the mean / median to use for fill

```
[12] 1 column_quantity = Demand_df['Quantity'].mean()
```

```
[13] 1 column_quantity
```

1674070.7251551538

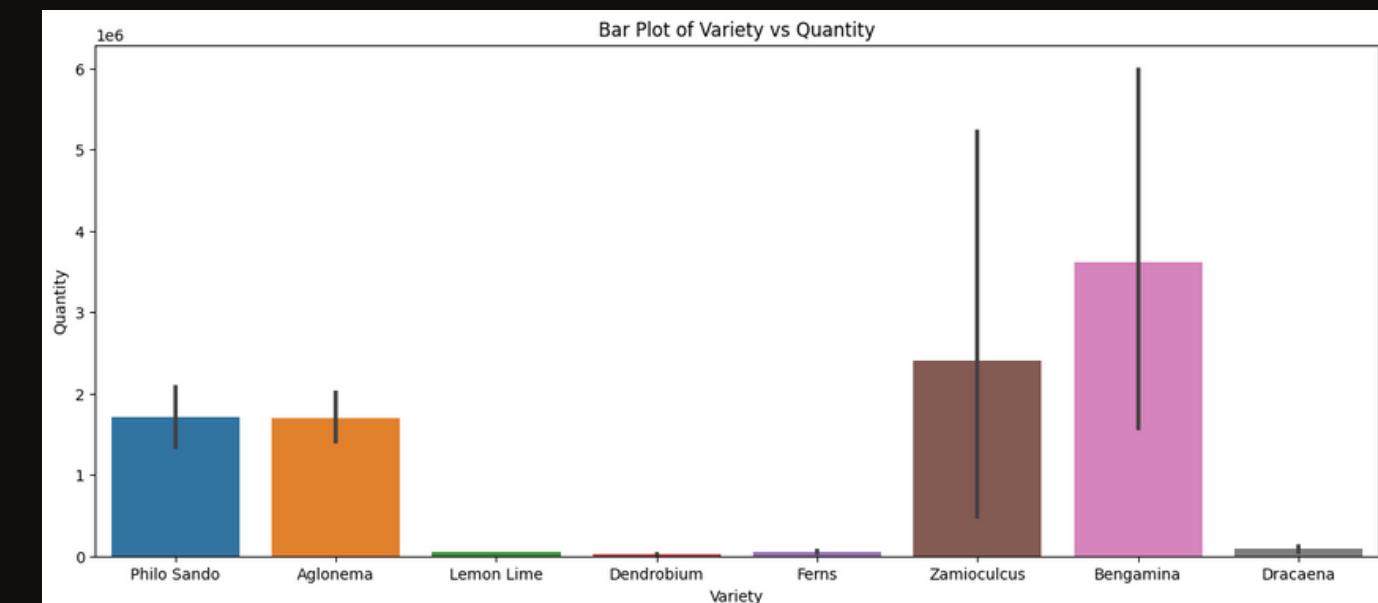
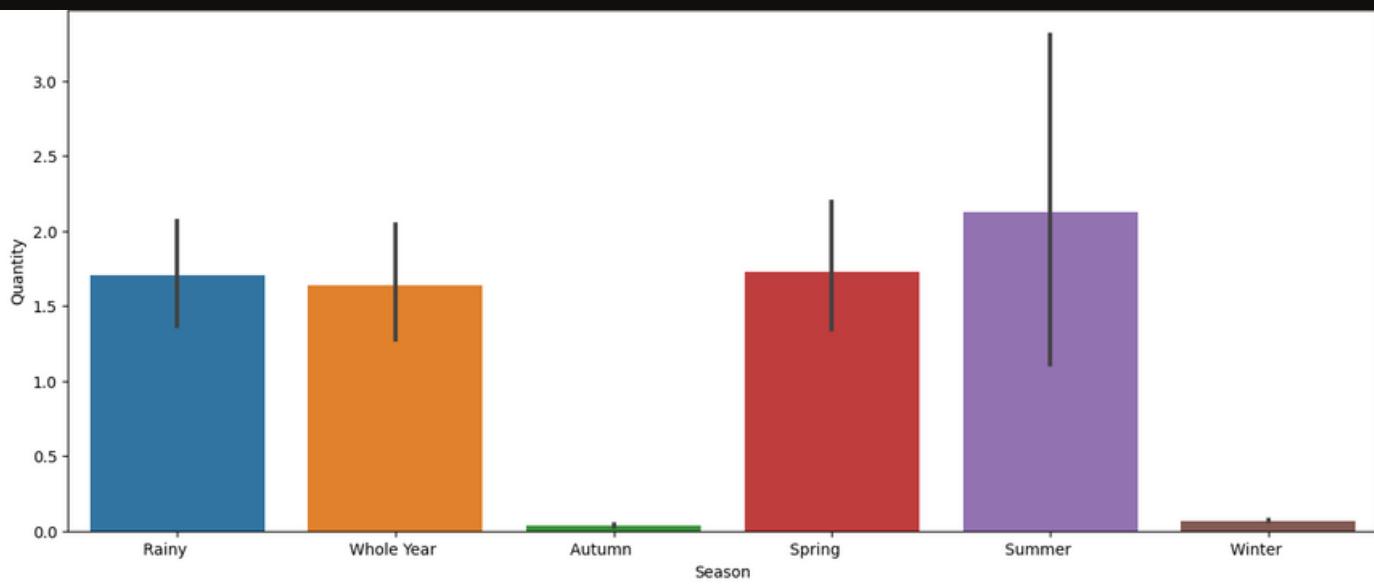
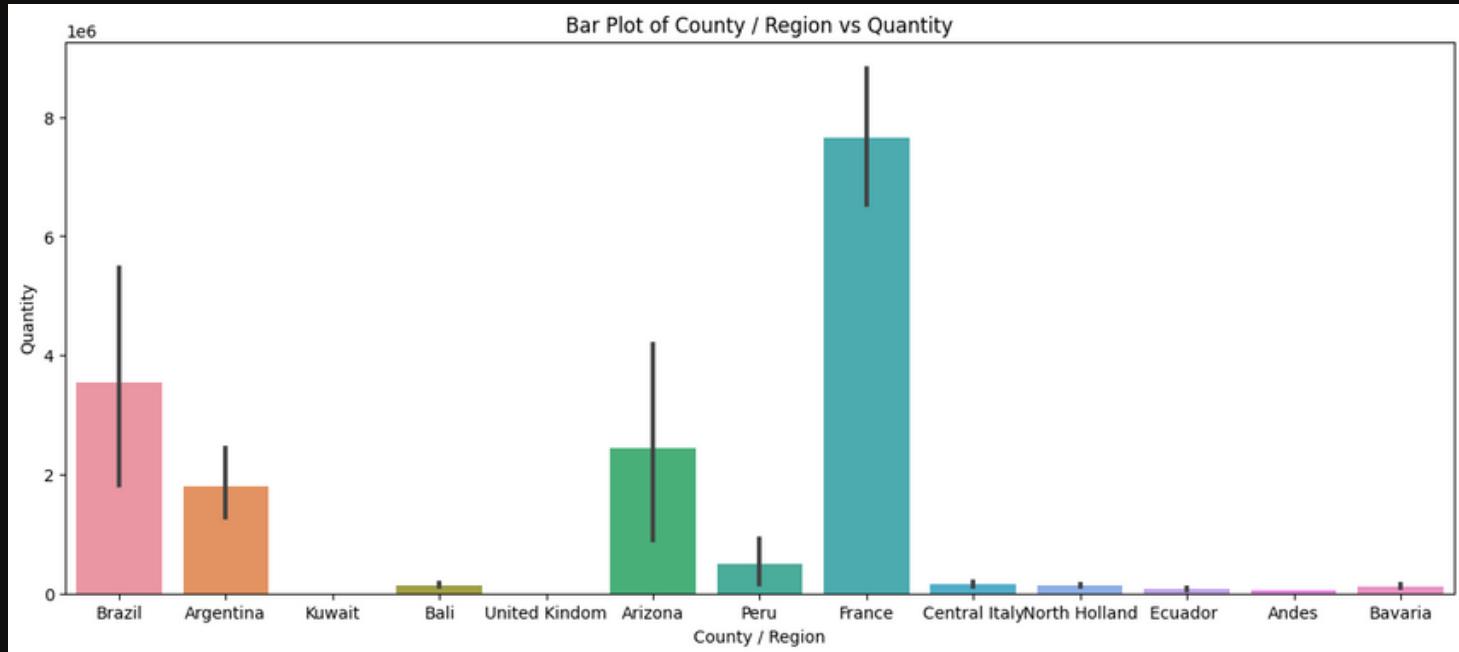
```
[14] 1 column_quantity_median = Demand_df['Quantity'].median()
```

```
[15] 1 column_quantity_median
```

1043.0

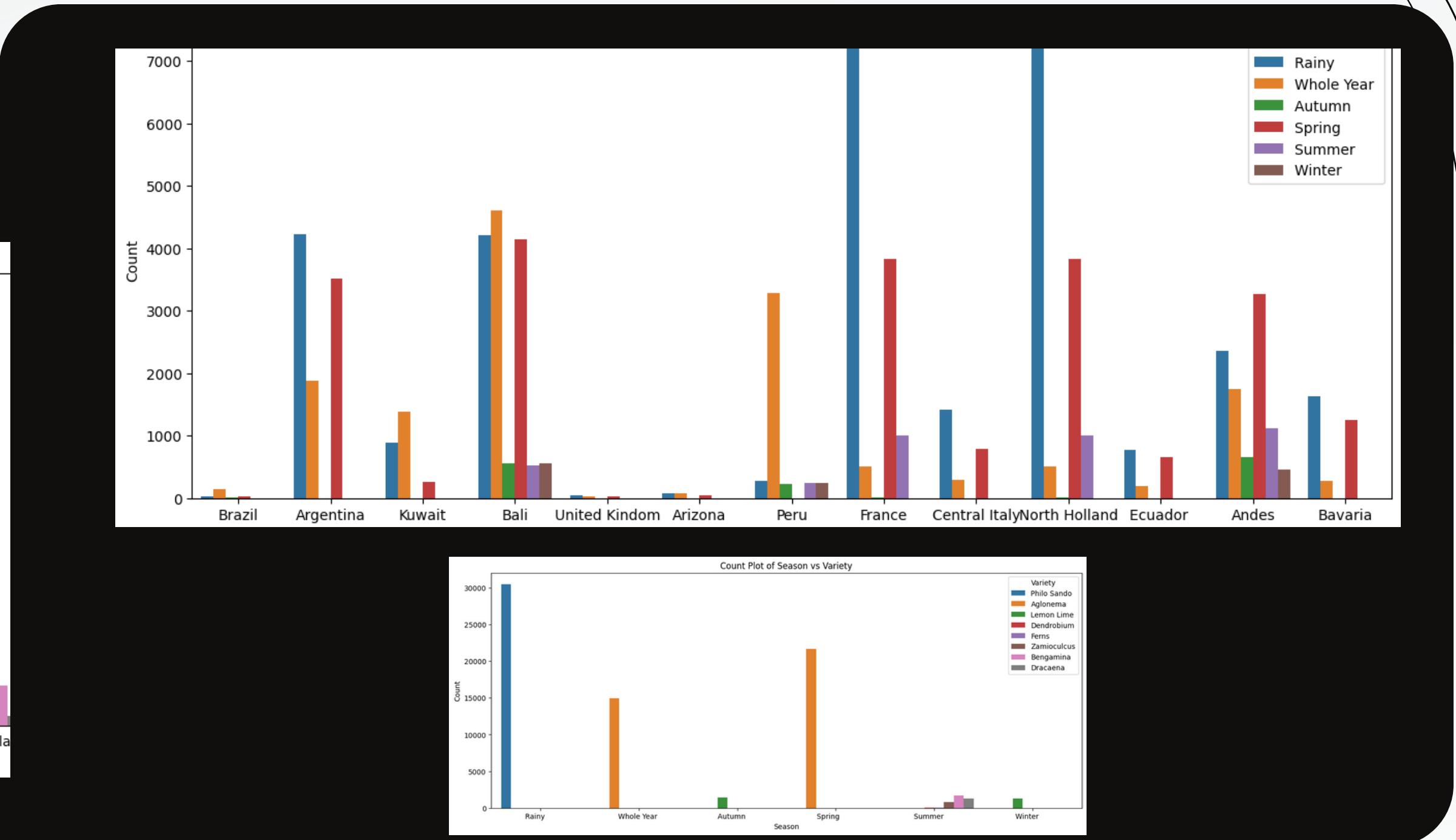
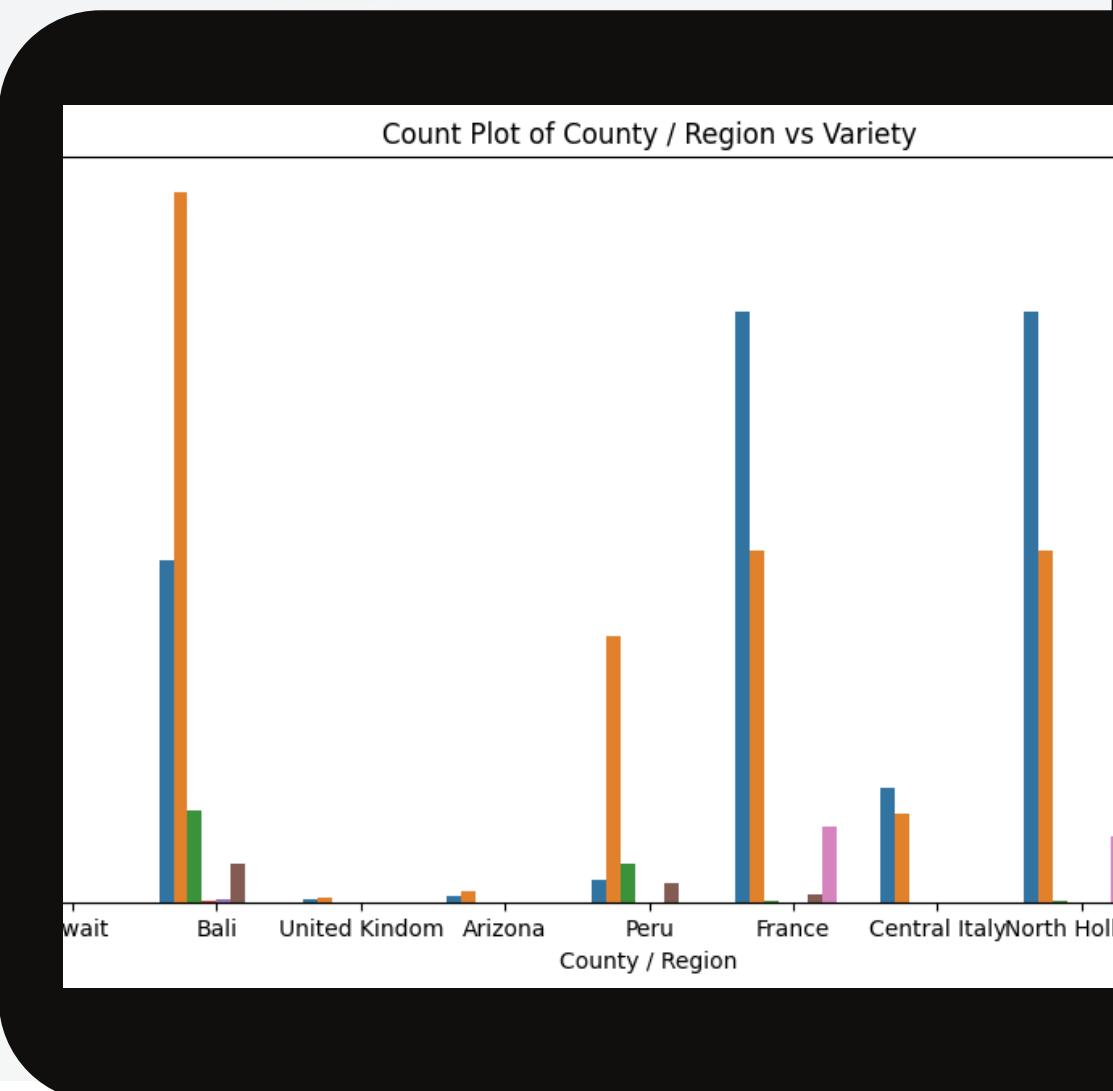
IMPLEMENTATION

Data Visualizing
with target



IMPLEMENTATION

Data Visualizing
with each features



IMPLEMENTATION

Data Preprocessing

Numerical data scaled using the standard scaler and the categorical variables handled by one hot encoding

Handling numerical data using Standard Scaler

```
[38] 1 # Extract the "Quantity" column
2 quantity_data = Demand_Filled_df_Qty["Quantity"].values.reshape(-1, 1)

[39] 1 # Initiating dummy values for categorical variables
2 scaler =

[40] 1 # Create dummy variables for the categorical columns
2 dummy_columns = pd.get_dummies(Demand_Filled_df_cat[categorical_columns], prefix=categorical_columns, drop_first=True)

[41] 1 # Create scaled data
2 scaled_data = np.concatenate((scaled_quantity_data, dummy_columns), axis=1)
```

| | C_County / Region_Argentina | C_County / Region_Arizona | C_County / Region_Bali | C_County / Region_Bavaria | C_County / Region_Brazil | C_County / Region_Central Italy | C_County / Region_Ecuador | C_County / Region_Peru |
|---|-----------------------------|---------------------------|------------------------|---------------------------|--------------------------|---------------------------------|---------------------------|------------------------|
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

IMPLEMENTATION

Splitting data
in to x & y

The data set has
split in to two as
independent
and dependent
variables by
20% of the data
taken as the test
size

Splitting training and testing data set

```
[63] 1 X = shuffled_df.drop("Scaled_Quantity", axis=1)
2 y = shuffled_df["Scaled_Quantity"]
3 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_st
```

| | C_County / Region_Argentina | C_County / Region_Arizona | C_County / Region_Bali | C_County / Region_Bavaria | C_County / Region_Brazil | Region_Croatia |
|-------|-----------------------------|---------------------------|------------------------|---------------------------|--------------------------|----------------|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 73729 | 0 | 0 | 0 | 0 | 0 | 0 |
| 73730 | 1 | 0 | 0 | 0 | 0 | 0 |
| 73731 | 0 | 0 | 0 | 0 | 0 | 0 |
| 73732 | 1 | 0 | 0 | 0 | 0 | 0 |
| 73733 | 0 | 0 | 1 | 0 | 0 | 0 |

| | y |
|--|-----------|
| 0 | -0.056071 |
| 1 | -0.056018 |
| 2 | -0.055635 |
| 3 | -0.056129 |
| 4 | -0.056089 |
| ... | ... |
| 73729 | -0.055686 |
| 73730 | -0.056129 |
| 73731 | -0.056123 |
| 73732 | -0.056128 |
| 73733 | -0.056114 |
| Name: Scaled_Quantity, Length: 73734, dtype: float64 | |

IMPLEMENTATION

After checking outliers models are defined I have selected 3 types of regression models to train and they are

1.Linear regression

2.Extreme gradient boosting

3.Random forest

MODEL TRAINING

01.Random forest regression

```
[67] 1 # Random Forest Regression
2 rf = RandomForestRegressor()
3 rf.fit(X_train, y_train)
4 rf_pred_train = rf.predict(X_train)
5 rf_pred_test = rf.predict(X_test)
6 rf_mse_train = mean_squared_error(y_train, rf_pred_train)
7 rf_mse_test = mean_squared_error(y_test, rf_pred_test)
8 print("Random Forest - Train MSE:", rf_mse_train)
9 print("Random Forest - Test MSE:", rf_mse_test)
```

```
Random Forest - Train MSE: 0.9526527655989198
Random Forest - Test MSE: 1.13248521470431
```

03.Lineare Regression

```
[73] 1 # Linear Regression
2 lr = LinearRegression()
3 lr.fit(X_train, y_train)
4 lr_pred_train = lr.predict(X_train)
5 lr_pred_test = lr.predict(X_test)
6 lr_mse_train = mean_squared_error(y_train, lr_pred_train)
7 lr_mse_test = mean_squared_error(y_test, lr_pred_test)
8 print("Linear Regression - Train MSE:", lr_mse_train)
9 print("Linear Regression - Test MSE:", lr_mse_test)
```

```
Linear Regression - Train MSE: 0.9544264394774482
Linear Regression - Test MSE: 1.1360368381360995
```

02.XGBoost Regression

```
[70] 1 # XGBoost Regression
2 xgb = XGBRegressor()
3 xgb.fit(X_train, y_train)
4 xgb_pred_train = xgb.predict(X_train)
5 xgb_pred_test = xgb.predict(X_test)
6 xgb_mse_train = mean_squared_error(y_train, xgb_
7 xgb_mse_test = mean_squared_error(y_test, xgb_
8 print("XGBoost - Train MSE:", xgb_mse_train)
9 print("XGBoost - Test MSE:", xgb_mse_test)
```

```
XGBoost - Train MSE: 0.9526493973594272
XGBoost - Test MSE: 1.1325992861367322
```

IMPLEMENTATION

After model evaluation part according to MAE, MSE and R squared values Extreme Gradient Boost is the best model for this

| Model | Train MSE | Test MSE | Train MAE | Test MAE | Train R-squared | Test R-squared |
|-------------------|-----------|----------|-----------|----------|-----------------|----------------|
| Linear Regression | 0.954426 | 1.13604 | 0.112089 | 0.120399 | 0.00936312 | 0.0088274 |
| XGBoost | 0.952649 | 1.1326 | 0.104581 | 0.112726 | 0.0112076 | 0.0118267 |
| Random Forest | 0.952653 | 1.13249 | 0.104738 | 0.112885 | 0.0112041 | 0.0119262 |

And accuracy is also calculated for all three models

Additionaly done the cross validation parts for avoid overfitting
as a part of hyper parameter tuning

IMPLEMENTATION

Accuracy for each model

```
[87] 1 threshold = 0.1 # Define the threshold for considering predictions as accurate  
2  
3 # Calculate the percentage of predictions within the threshold  
4 accurate_predictions = np.abs(y_test - lr_pred_test) <= threshold  
5 accuracy_percentage = round((np.sum(accurate_predictions) / len(y_test)) * 100, 2)  
6  
7 print("Linear Regression - Accuracy (% within threshold):", accuracy_percentage, "%")
```

Linear Regression - Accuracy (% within threshold): 81.95 %

```
[88] 1 # Calculate the percentage of predictions within the threshold  
2 accurate_predictions = np.abs(y_test - xgb_pred_test) <= threshold  
3 accuracy_percentage = round((np.sum(accurate_predictions) / len(y_test)) * 100, 2)  
4  
5 print("XGBoost - Accuracy (% within threshold):", accuracy_percentage, "%")
```

XGBoost - Accuracy (% within threshold): 79.63 %

```
[89] 1 # Calculate the percentage of predictions within the threshold  
2 accurate_predictions = np.abs(y_test - rf_pred_test) <= threshold  
3 accuracy_percentage = round((np.sum(accurate_predictions) / len(y_test)) * 100, 2)  
4  
5 print("Random Forest - Accuracy (% within threshold):", accuracy_percentage, "%")
```

Random Forest - Accuracy (% within threshold): 79.63 %

```
[79] 1 # XGBoost Regression
```

```
2 xgb = XGBRegressor()  
3 xgb_pipeline = make_pipeline(SimpleImputer(), xgb)  
4 xgb_cv_scores = cross_val_score(xgb_pipeline, X_train, y_train, cv=5)  
5 xgb_cv_mse = -np.mean(xgb_cv_scores)  
6 print("XGBoost - Cross-validated MSE:", xgb_cv_mse)
```

XGBoost - Cross-validated MSE: 0.9538598053789

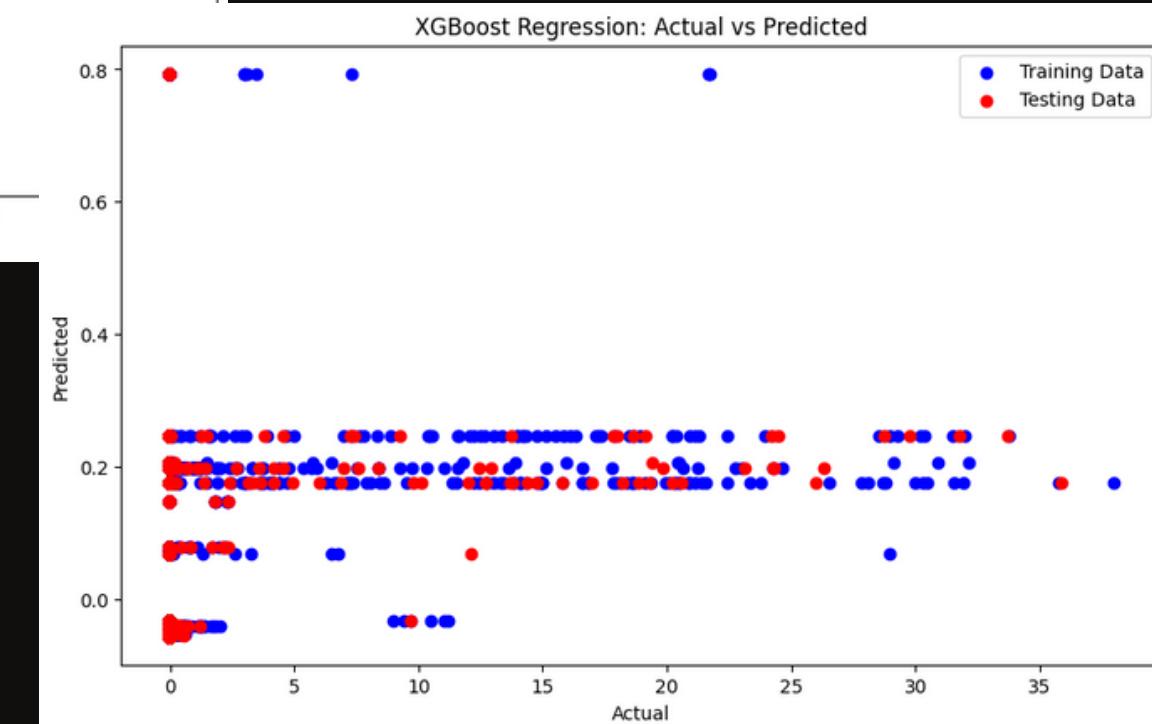
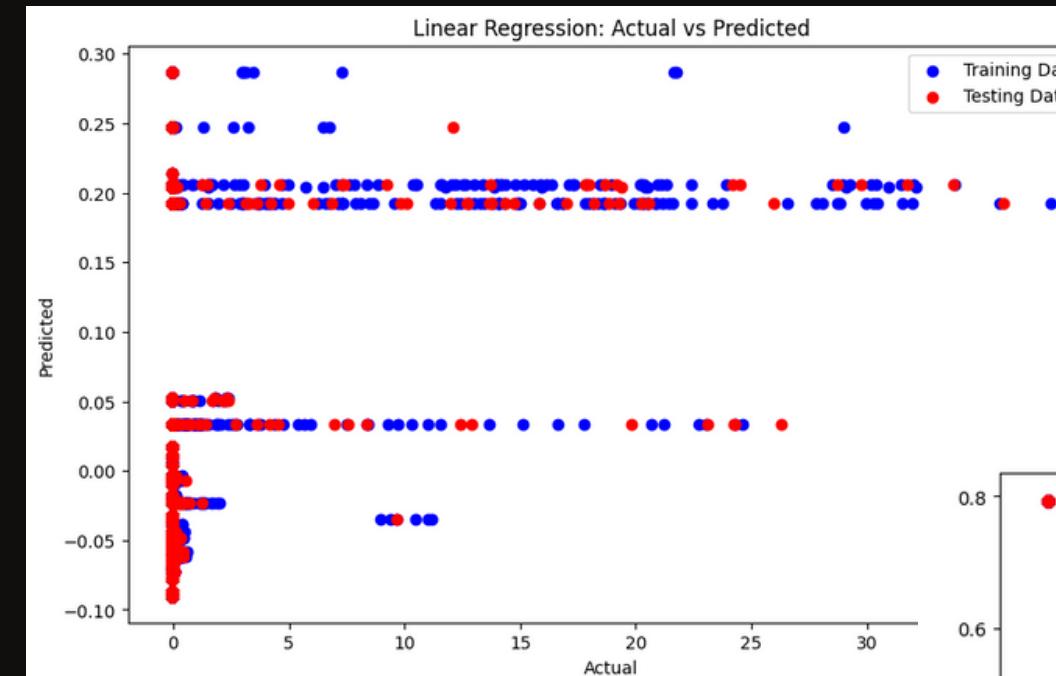
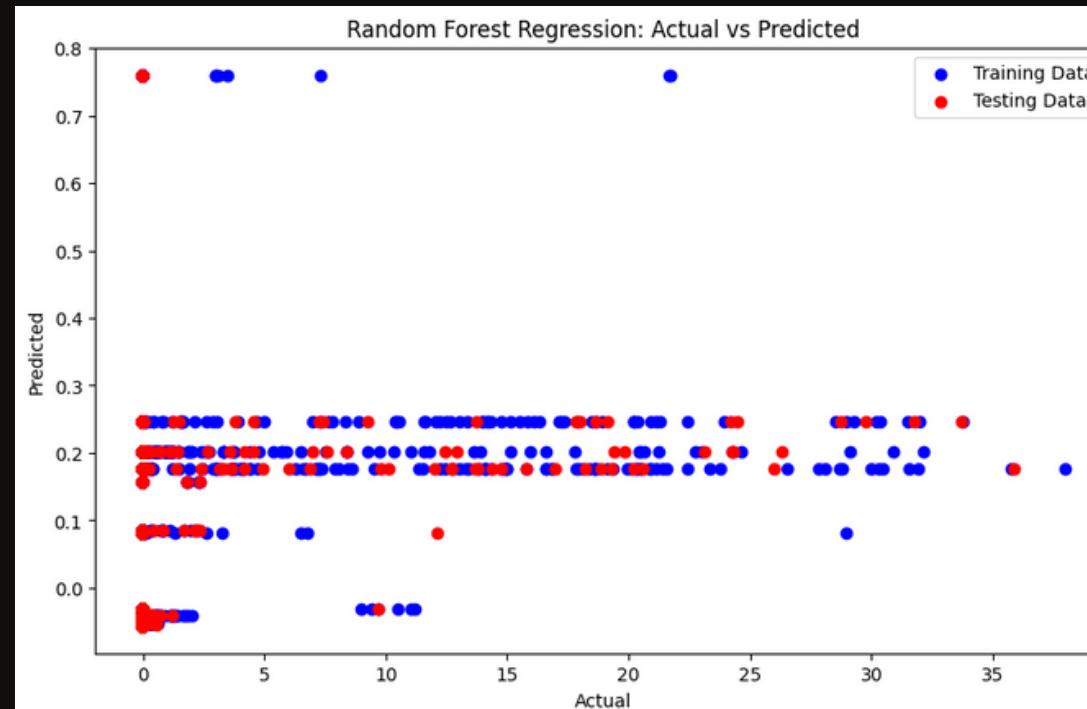
```
[80] 1 # Cross-validation for Linear Regression
```

```
2 lr_cv_scores = cross_val_score(lr, X_train, y_train, cv=5)  
3 lr_cv_mse = -np.mean(lr_cv_scores)  
4 print("Linear Regression - Cross-validated MSE:", lr_cv_mse)
```

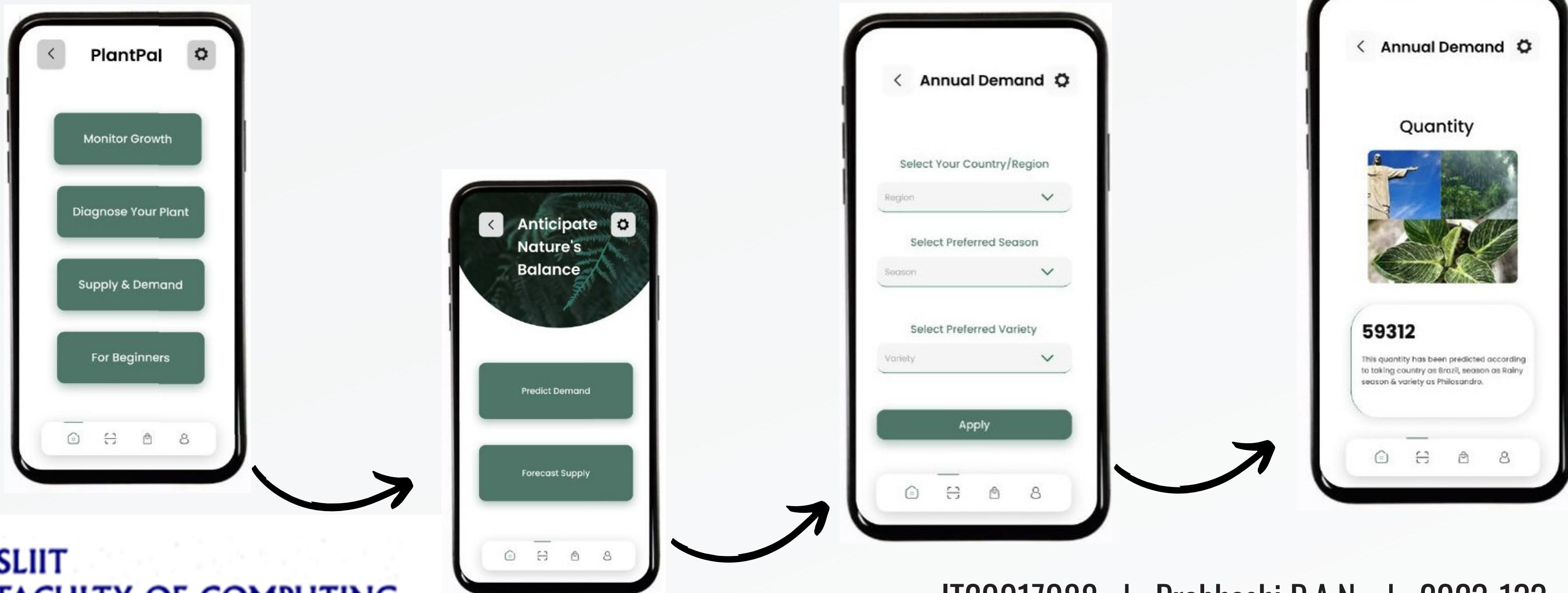
Linear Regression - Cross-validated MSE: 0.9540000000000001

IMPLEMENTATION

Scatter plots for
actual and predicted



UI



FUTURE

IMPROVE THE CURRENT
MODEL AND APPLY THE
SUPPLY PREDICTION

DEVELOPING THE MODEL
TO FIND THE NEAREST
VENDOR

DEVELOP THE FRONT
END AND INTEGRATE THE
APPLICATION

IT20005726
Liyanage S.R

GROWTH MONITORING AND PREDICTION SYSTEM FOR ORNAMENTAL PLANTS

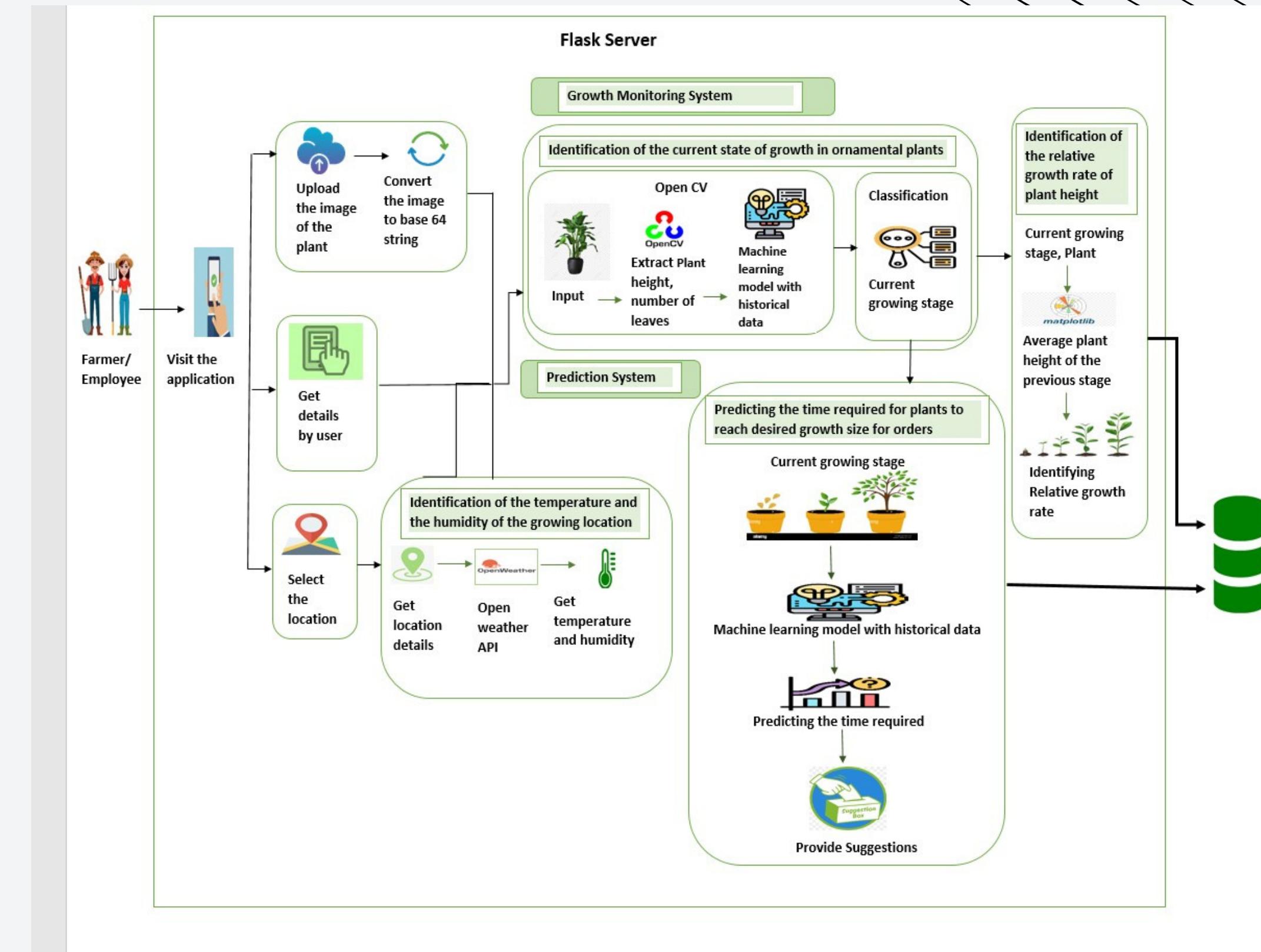


Specialization : Data Science

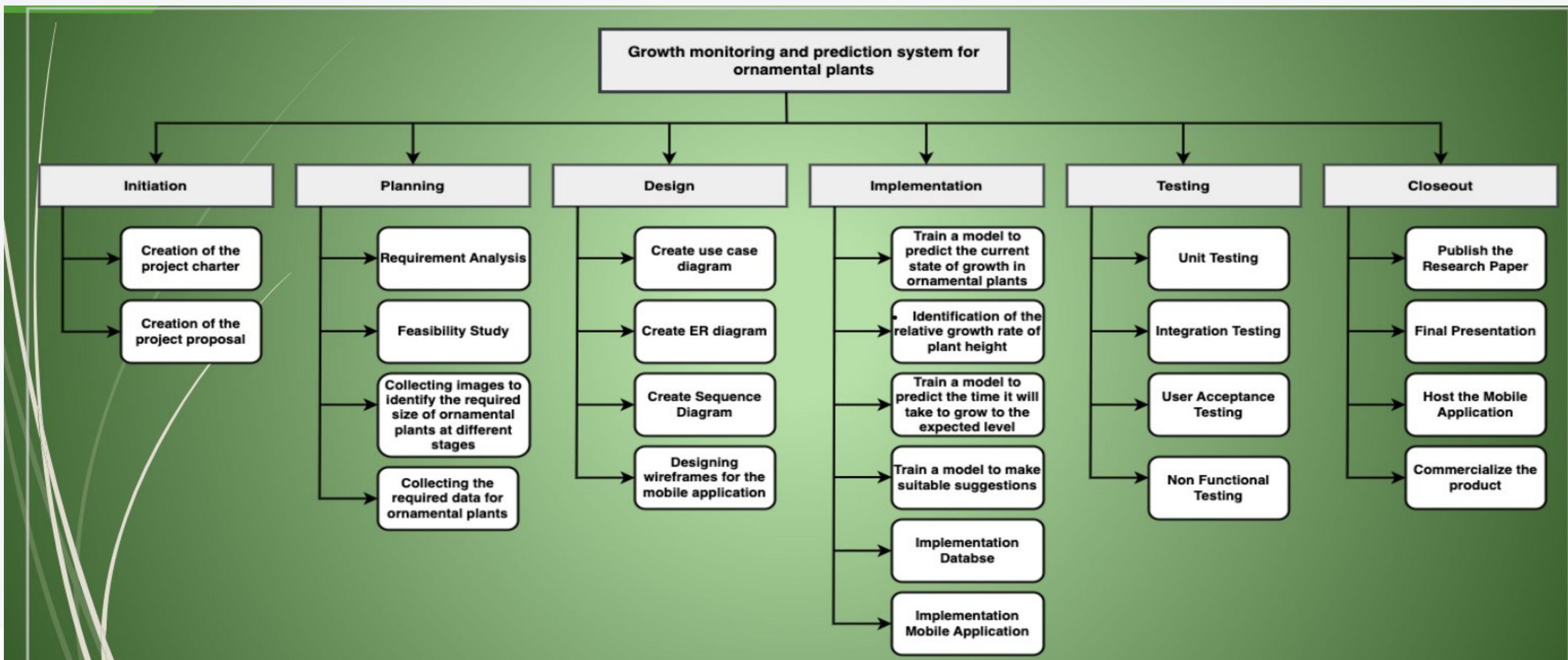
RESEARCH PROBLEM

- How can the growth of ornamental plants be monitored, and their current state of growth predicted?
- How can the time it takes for ornamental plants to reach the required growth size for an order be predicted?

INDIVIDUAL SYSTEM DIAGRAM



WORK BREAKDOWN



DATA GATHERING

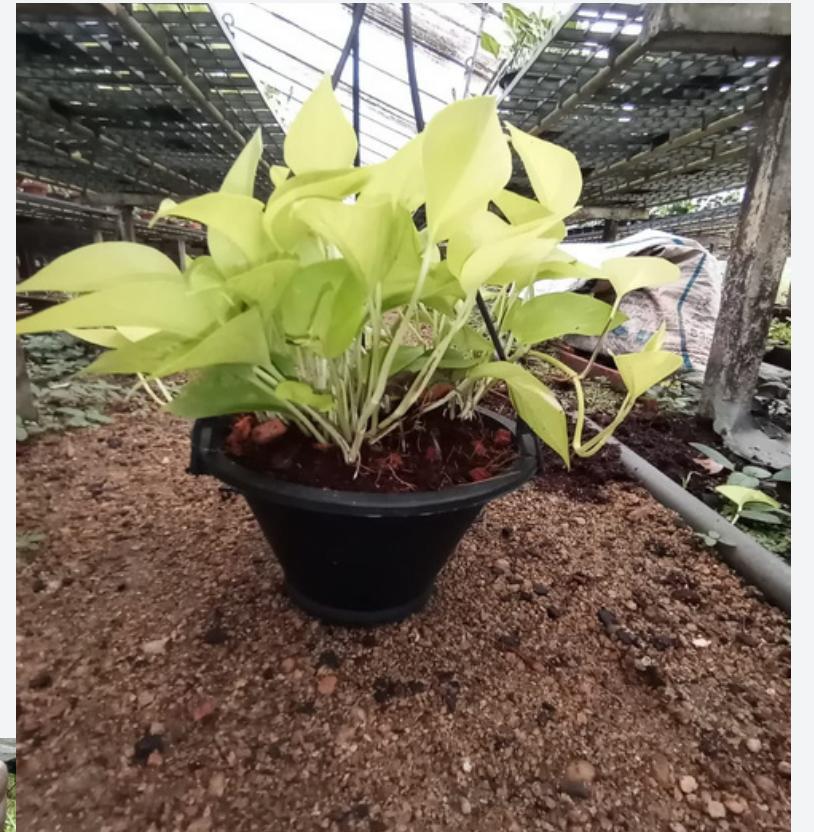
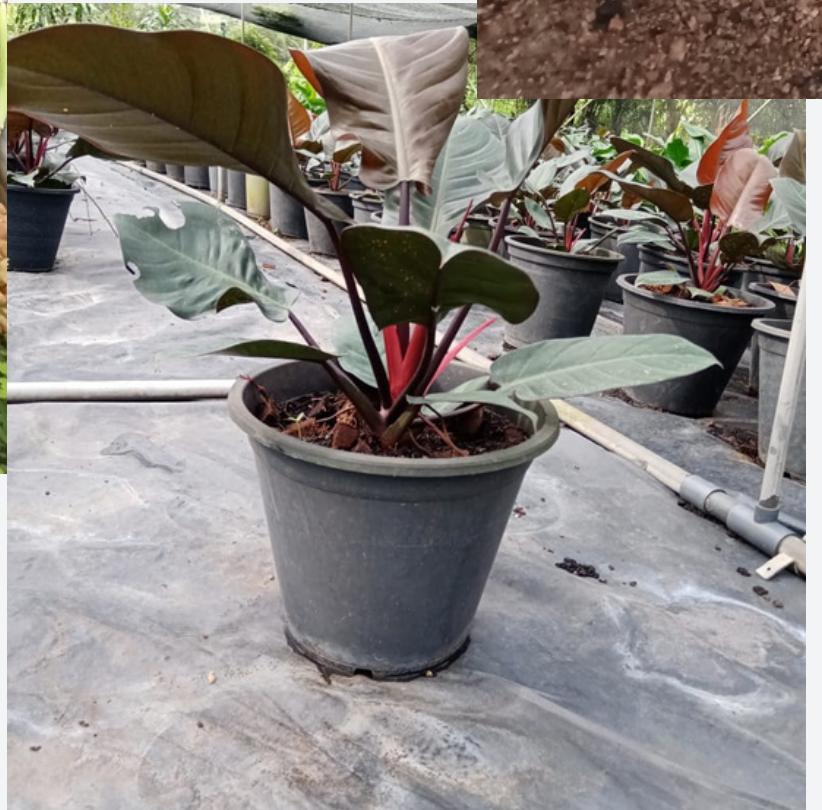
These data is collected
from the omega green
Pvt Ltd

And all data related to
growth monitoring
collected in manual way

| Date | Plant Stage | Net house | Reason | chemical | Name |
|-------|-------------|-----------|------------------|----------------|--|
| 06/01 | Large | E | Prevent fungus | Zira | Active Ing. Propiconazole 10ml/16L |
| 13/01 | " | " | " | " | " |
| 20/01 | " | " | " | " | " |
| 23/01 | All | E | Grass hoppers | Gem propinofos | Propinofos 60ml/16L |
| 11/02 | Large | " | Prevent fungus | Zira | Propiconazole 10ml/16L |
| 19/02 | " | " | " | " | " |
| 25/02 | All | " | Mightes | Mis abamaectin | Abamaectin 125ml/200L |
| 02/3 | All | E | Root rot | Homaï | 100g/200L |
| 03/3 | Large | " | Prevent fungu | Zira | Propiconazole 10ml/16L |
| 04/3 | All | E | Root rot | Homaï | 100g/200L |
| 11/3 | Large | " | Prevent fungu | Zira | Propiconazole 10ml/16L |
| 13/3 | " | E | Root rot | Homaï | 100g/200L |
| 19/3 | Small | B | Root growth | Seradix | SeBA mg |
| 01/4 | " | " | " | " | " |
| 3/4 | " | " | " | " | " |
| 07/05 | Large | E | Grass hoppers | Profenofos | Propenofos 600ml/200L |
| 07/06 | " | " | Prevent fungi | Daconil | chlorothalonil 400/300ml |
| 08/06 | " | H | " | " | " |
| 11/06 | " | E | " | Ronil | " |
| 14/06 | Small | B | Root growth | Seradix | IBA |
| 15/07 | Large | E | " | Profenofos | Profenofos 600ml/200L |
| 17/07 | Small | B | Root stimulation | Seradix | IBA |
| 14/08 | Large | E | Prevent fungi | Daconil | Chlorothalonil 600ml/200L |
| 6/08 | " | " | Insects | Cobra | Imidacloprid 10ML/16L |
| 0/08 | " | " | " | " | " |
| " | " | " | Fungi | Zira | Propiconazole |
| 1/08 | " | " | " | " | " |
| 09 | " | " | " | " | " |
| 7 | " | " | " | " | " |
| 9 | " | B | Root stimulation | Seradix | IBA |

DATA GATHERING

Pictures of main plant types considered



MODEL CREATION

Data Preprocessing

```
#cheking data types of the coulmn  
Growth_Data.info()  
  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2184 entries, 0 to 2183  
Data columns (total 16 columns):  
 # Column Non-Null Count Dtype  
--- ---  
 0 Plant_Bed_number 2184 non-null object  
 1 Plant_number 2184 non-null object  
 2 Plant_type 2184 non-null object  
 3 Variety 2184 non-null object  
 4 Media_Grew 2157 non-null object  
 5 Checked_Date 2184 non-null object  
 6 Temperature 2145 non-null float64  
 7 Humidity 2147 non-null float64  
 8 Fertilizer 2180 non-null object  
 9 Fertilizer_Amount 2184 non-null int64  
 10 Growth_Category 2184 non-null object  
 11 Current_pot_type 2184 non-null object  
 12 Current_life_time 2178 non-null object  
 13 Average_min_height 2184 non-null int64  
 14 Average_max_height 2184 non-null int64  
 15 Planted_Date 2184 non-null object  
dtypes: float64(2), int64(3), object(11)  
memory usage: 273.1+ KB
```

```
[7] #cheking the available columns in the data set  
Growth_Data.columns  
  
Index(['Plant_Bed_number', 'Plant_number', 'Plant_type ', 'variety',  
       'Media_Grew', 'Checked_Date', 'Temperature', 'Humidity', 'Fertilizer',  
       'Fertilizer_Amount', 'Growth_Category', 'Current_pot_type',  
       'Current_life_time', 'Average_min_height', 'Average_max_height'])
```

Identify null values

```
[10] #check null values of the trainin gdata set  
Growth_Data.isnull()  
  
[11] Growth_Data.isnull().sum().sort_values(ascending=False)
```

| | Plant_Bed_number | Plant_number | Plant_type | Variety | Media_Grew | Checked_Date | Temperature | Humidity | Fertilizer | Fertilizer_Amount | Growth_Category | Current_pot_type | Current_life_time | Ave |
|------|------------------|--------------|------------|---------|------------|--------------|-------------|----------|------------|-------------------|-----------------|------------------|-------------------|-------|
| 0 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 1 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 3 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 4 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2179 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2180 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2181 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2182 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2183 | False | False | False | False | False | False | False | False | False | False | False | False | False | False |
| 2184 | rows | x | 16 | columns | | | | | | | | | | |

```
Temperature      39  
Humidity        37  
Media_Grew      27  
Current_life_time    6  
Fertilizer        4
```

MODEL CREATION

Data Preprocessing

```
[20] Growth_Data.loc[Growth_Data['Growth_Category'] == 'coir,cow_dung,coco chips', 'Media_Grew'] = 'coir,cowdung,coco chips'
```

| | Plant_Bed_number | Plant_number | Plant_type | Variety | Media_Grew | Checked_Date | Temperature | Humidity | Fert: |
|------|------------------|--------------|----------------------|--------------|--------------------------|--------------|-------------|----------|---------------|
| 0 | 22-TGN-07 | TGN249 | Tropical Green enjoy | pothos | coir,cow_dung,coco chips | 2022.10.09 | 22.0 | 94.0 | Yaramila, Ap |
| 1 | 20-MQ-01 | MQ-73 | Marble Queen | phothos | coir,cowdung,coco chips | 2021.04.26 | 33.0 | 46.0 | Yaramila, Gro |
| 2 | 20-MQ-01 | MQ-449 | Marble Queen | phothos | coir,cowdung,coco chips | 2020.02.29 | 36.0 | 46.0 | Yaramila, Gro |
| 3 | 20-MQ-08 | MQ-344 | Marble Queen | phothos | coir,cowdung | 2020.10.07 | 28.0 | 85.0 | Yaramila, Gro |
| 4 | 22-LL-27 | LL-292 | Lemon Lime | phothos | coir,cowdung,coco chips | 2022.05.26 | 27.0 | 86.0 | Yaramila, Ap |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2179 | 21-BC-76 | BC-17 | Black Cardinal | philodrendon | coir,cowdung | 2021.12.21 | 30.0 | 64.0 | Yaramila, Ap |

```
#cheking data types of the coulmn  
Growth_Data.info()
```

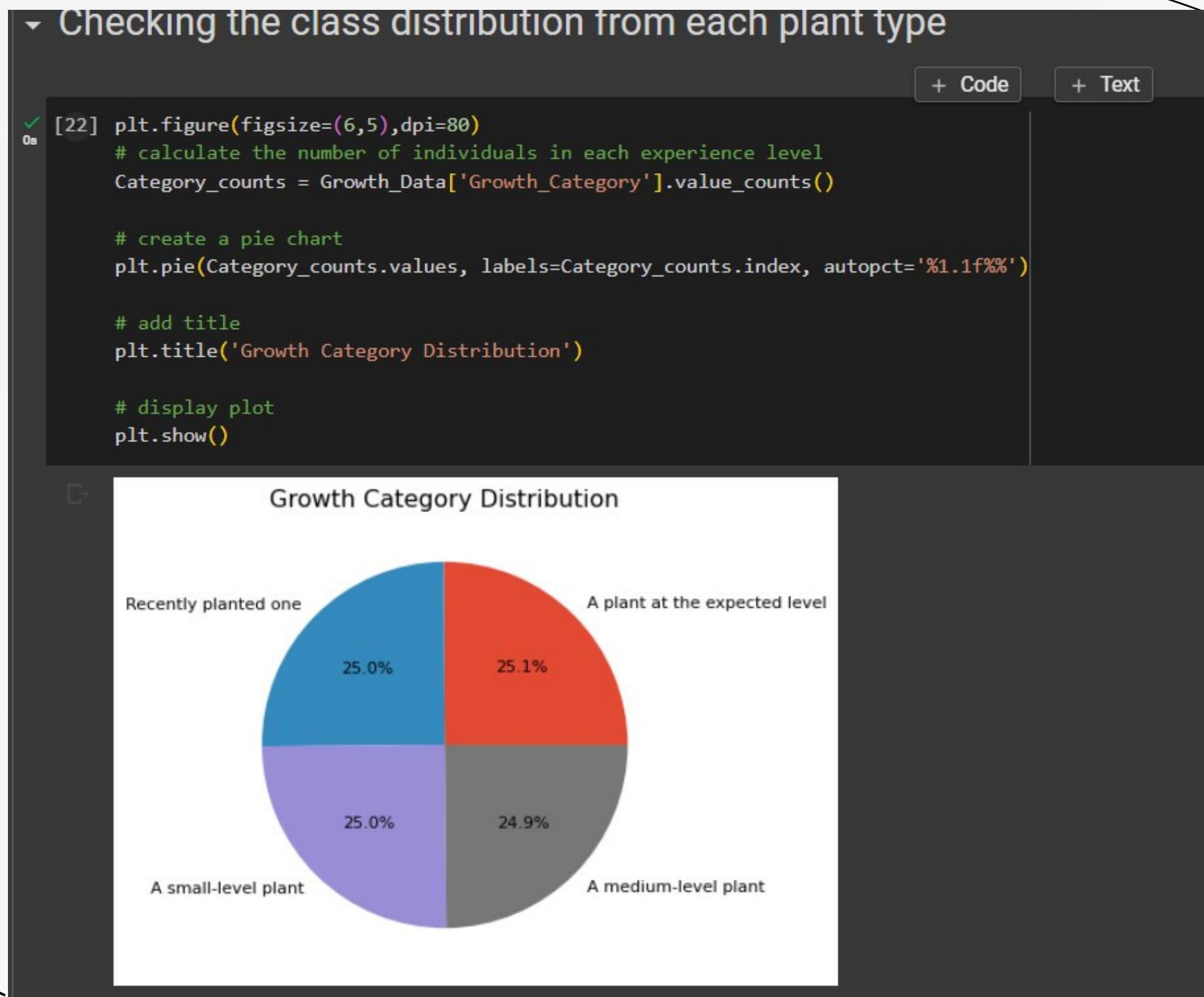
```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 2184 entries, 0 to 2183  
Data columns (total 16 columns):  
 #   Column           Non-Null Count Dtype  
 ---  --  
 0   Plant_Bed_number 2184 non-null  object  
 1   Plant_number     2184 non-null  object  
 2   Plant_type       2184 non-null  object  
 3   Variety          2184 non-null  object  
 4   Media_Grew       2157 non-null  object  
 5   Checked_Date     2184 non-null  object  
 6   Temperature      2145 non-null  float64  
 7   Humidity         2147 non-null  float64  
 8   Fertilizer       2180 non-null  object  
 9   Fertilizer_Amount 2184 non-null  int64  
 10  Growth_Category  2184 non-null  object  
 11  Current_pot_type 2184 non-null  object  
 12  Current_life_time 2178 non-null  object  
 13  Average_min_height 2184 non-null  int64  
 14  Average_max_height 2184 non-null  int64  
 15  Planted_Date    2184 non-null  object  
dtypes: float64(2), int64(3), object(11)  
memory usage: 273.1+ KB
```

```
[7] #cheking the available columns in the data set  
Growth_Data.columns
```

```
Index(['Plant_Bed_number', 'Plant_number', 'Plant_type', 'Variety',  
       'Media_Grew', 'Checked_Date', 'Temperature', 'Humidity', 'Fertilizer',  
       'Fertilizer_Amount', 'Growth_Category', 'Current_pot_type',  
       'Current_life_time', 'Average_min_height', 'Average_max_height',  
       'Planted_Date'])
```

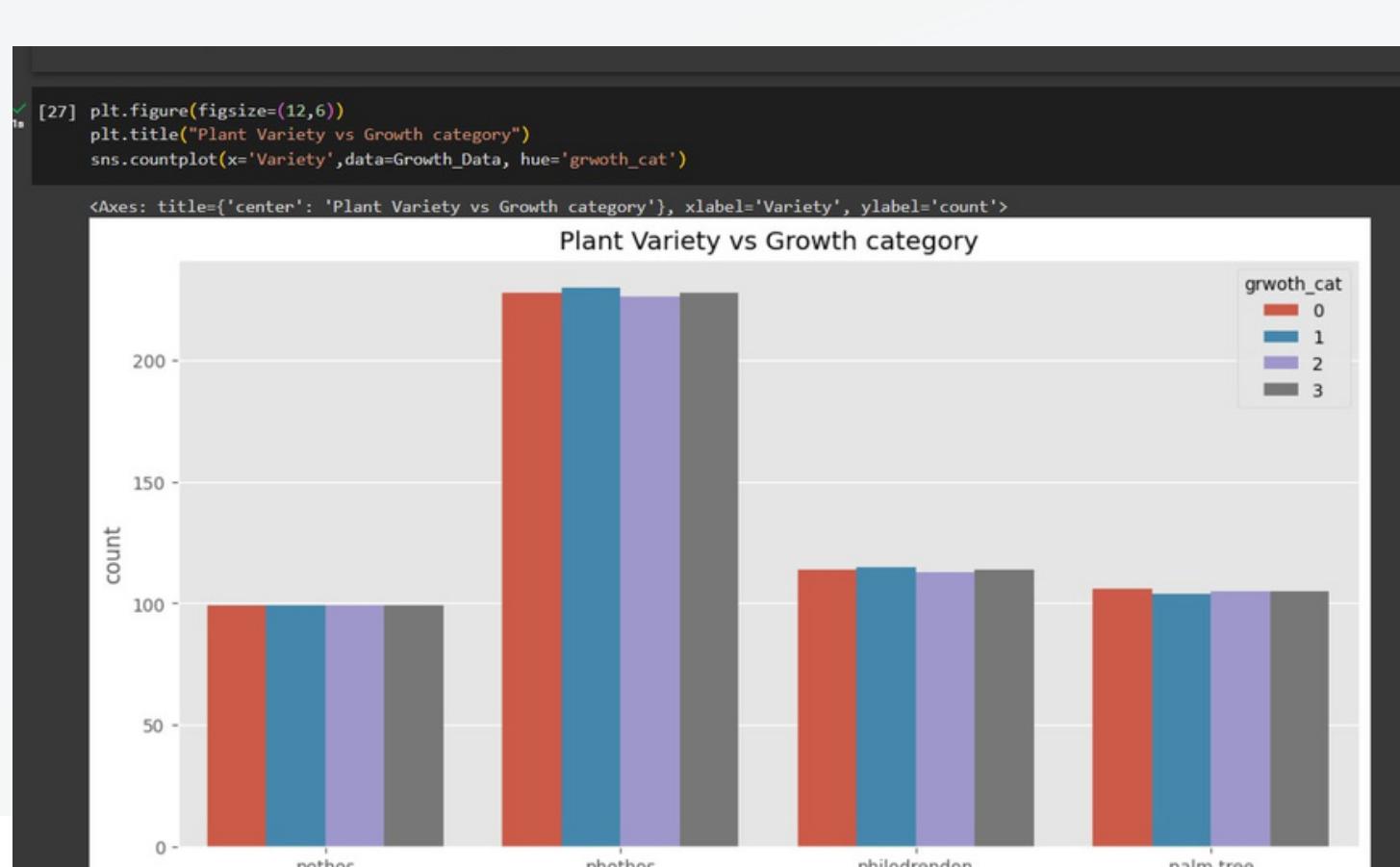
MODEL CREATION

Checking the
balance of class

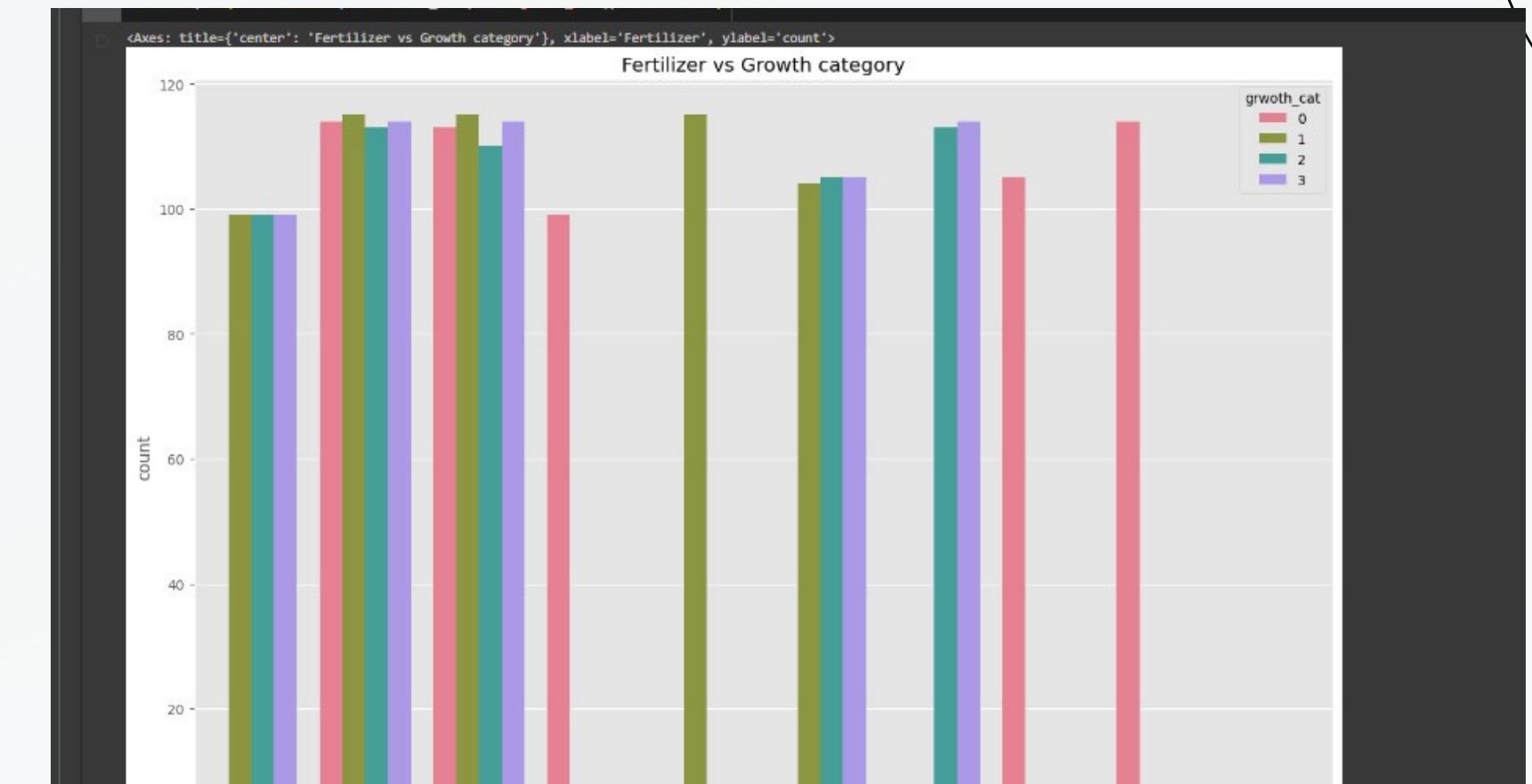


MODEL CREATION

Exploratory data analysis



Categorical data vs Target

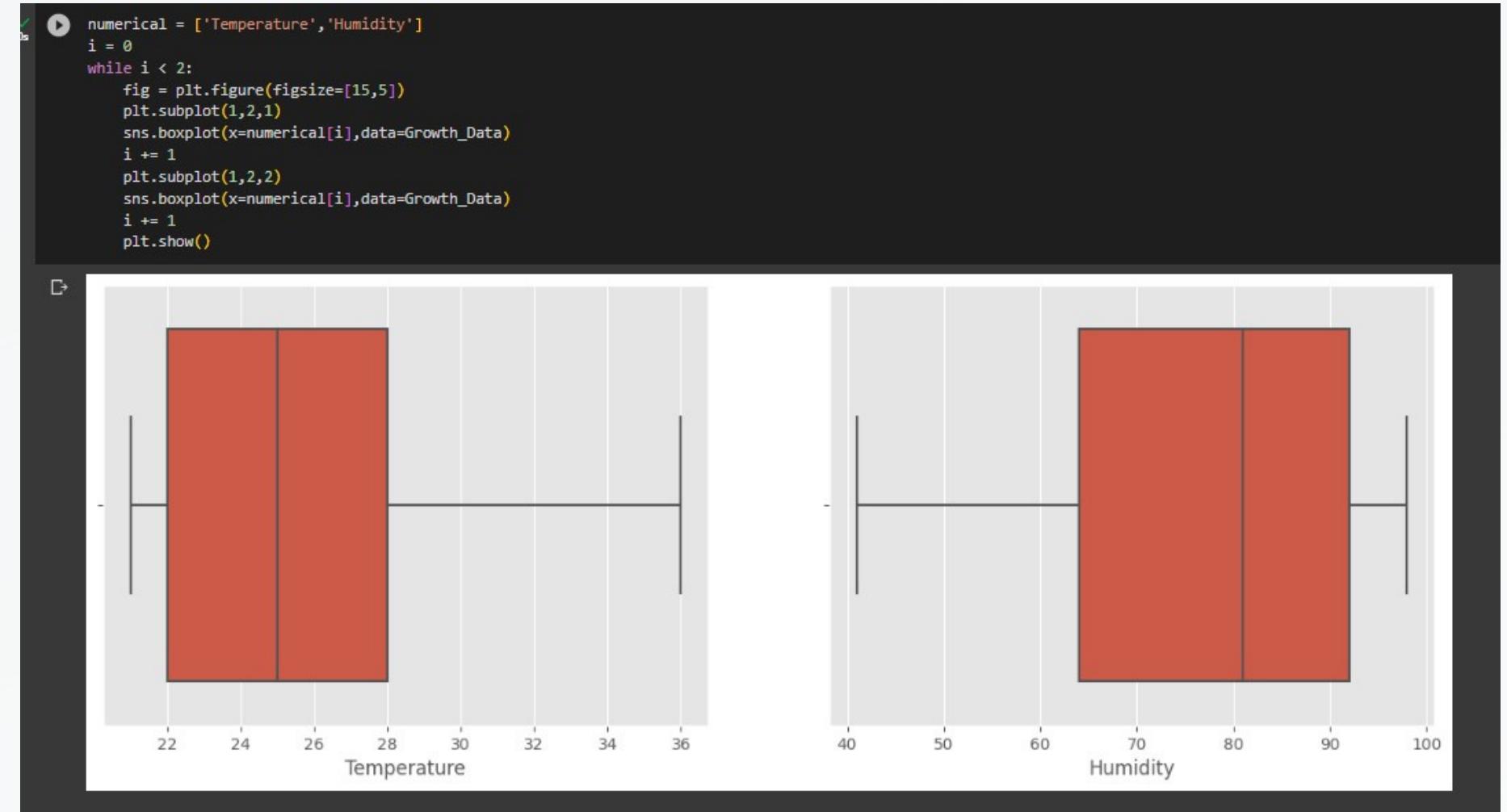


MODEL CREATION

Exploratory data analysis



Numerical data vs Target



MODEL CREATION

One hot encoding

```
Handling Categorical Data with OneHotEncoding

[34] #one hot encoding data
from sklearn.preprocessing import OneHotEncoder

numerical_cols_train = Input_columns.select_dtypes(include=np.number).columns.tolist()
categorical_cols_train = Input_columns.select_dtypes('object').columns.tolist()
categorical_cols_train

['Plant_type', 'Variety', 'Media_Grew', 'Fertilizer']

[35] numerical_cols_train

['Temperature',
 'Humidity',
 'Fertilizer_Amount',
 'Average_min_height',
 'Average_max_height',
 'grwoth_cat']

[36] encoder = OneHotEncoder(sparse=False,handle_unknown='ignore').fit(Input_columns[categorical_cols_train])
encoder

/usr/local/lib/python3.10/dist-packages/sklearn/preprocessing/_encoders.py:868: FutureWarning: `sparse` was renamed to `sparse_output` in version 1.2 and will be
warnings.warn(
    "OneHotEncoder"
OneHotEncoder(handle_unknown='ignore', sparse=False, sparse_output=False)

[37] encoded_cols = list(encoder.get_feature_names_out(categorical_cols_train))
encoded_cols

['Plant_type _Black Cardinal',
 'Plant_type _Lemon Lime',
 'Plant_type _Livistonia',
 'Plant_type _Marble Queen',
 'plant_type _Tropical Green enjoy',
 'Variety_palm tree',
 'Variety_philodrendon',
 'Variety_solidago']
```

MODEL CREATION

Splitting the data set

```
Splitting training and testing data set with ratio of 80:20

[45]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20,random_state=42)

[46]: x_train.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 1747 entries, 812 to 860
Data columns (total 30 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Temperature     1747 non-null    float64
 1   Humidity        1747 non-null    float64
 2   Fertilizer_Amount 1747 non-null    int64  
 3   Average_min_height 1747 non-null    int64  
 4   Average_max_height 1747 non-null    int64  
 5   Plant_type _Black Cardinal 1747 non-null    float64
 6   Plant_type _Lemon Lime    1747 non-null    float64
 7   Plant_type _Livistonia   1747 non-null    float64
 8   Plant_type _Marble Queen 1747 non-null    float64
 9   Plant_type _Tropical Green enjoy 1747 non-null    float64
 10  Variety_palm tree   1747 non-null    float64
 11  Variety_philodrendon 1747 non-null    float64
 12  Variety_photos     1747 non-null    float64
 13  Variety_pothos     1747 non-null    float64
 14  Media_Grew_coir    1747 non-null    float64
 15  Media_Grew_coir,cow_dung 1747 non-null    float64
 16  Media_Grew_coir,cow_dung,coco chips 1747 non-null    float64
 17  Media_Grew_coir,cowdung 1747 non-null    float64
 18  Media_Grew coir,cowdung,coco chips 1747 non-null    float64
 19  Media_Grew_not_given 1747 non-null    float64
 20  Fertilizer_Foliar application k44 1747 non-null    float64
 21  Fertilizer_Foliar Application 1747 non-null    float64
 22  Fertilizer_Gem Profinofus 1747 non-null    float64
 23  Fertilizer_Yaramila   1747 non-null    float64
 24  Fertilizer_Yaramila,Foliar Application 1747 non-null    float64
```

Scaling the data set

```
# Scale the numeric features
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(x_train)
X_test_scaled = scaler.transform(x_test)

[49]: #checking the data set
X_train_scaled

array([[0.          , 0.96491228, 1.          , ... , 0.          , 0.          ,
       0.          ],
       [0.2         , 0.87719298, 0.00401606, ... , 0.          , 0.          ,
       0.          ],
       [0.06666667, 0.85964912, 0.09638554, ... , 0.          , 1.          ,
       0.          ],
       ...,
       [0.06666667, 1.          , 0.0060241 , ... , 0.          , 0.          ,
       0.          ],
       [0.6         , 0.40350877, 0.00200803, ... , 0.          , 0.          ,
       0.          ],
       [0.73333333, 0.29824561, 0.00200803, ... , 0.          , 0.          ,
       0.          ]])
```

MODEL CREATION

Model Training

Random Forest Classifier

```
# Random Forest with regularization
rf_classifier = RandomForestClassifier(n_estimators=100, max_depth=5, min_samples_split=10, min_samples_leaf=5, random_state=42)
rf_classifier.fit(X_train_scaled, y_train)
y_pred_rf_rg = rf_classifier.predict(X_test_scaled)

# Evaluate Random Forest model
acc_rf = accuracy_score(y_test, y_pred_rf_rg)
recall_rf = recall_score(y_test, y_pred_rf_rg, average='micro')
precision_rf = precision_score(y_test, y_pred_rf_rg, average='micro')
f1score_rf = f1_score(y_test, y_pred_rf_rg, average='micro')

print("Random Forest Classifier")
print("Accuracy: {:.2f}".format(acc_rf))
print("Precision: {:.2f}".format(precision_rf))
print("Recall: {:.2f}".format(recall_rf))
print("F1-Score: {:.2f}".format(f1score_rf))

Random Forest Classifier
Accuracy: 0.99
Precision: 0.99
Recall: 0.99
F1-Score: 0.99
```

```
[54] #PRINT CLASSIFICATION REPORT
print(classification_report(y_pred_rf_rg,y_test))

          precision    recall   f1-score   support
0           1.00     1.00     1.00      117
1           1.00     1.00     1.00      109
2           1.00     0.97     0.99      106
3           0.97     1.00     0.99      105

accuracy                           0.99      437
macro avg                           0.99      0.99      437
weighted avg                          0.99      0.99      437
```

MODEL CREATION

Model Training

02.XGBoost Model

```
## ----- XGBoost model v1 -----
## base run of model with default hyperparameters
import xgboost as xgb
xgb_clf = xgb.XGBClassifier(objective='multi:softmax',
                             num_class=4,
                             missing=1,
                             early_stopping_rounds=10,
                             learning_rate=0.001,
                             eval_metric=['merror', 'mlogloss'],
                             seed=42)
xgb_clf.fit(x_train,
             y_train,
             verbose=0, # set to 1 to see xgb training round intermediate results
             eval_set=[(x_train, y_train), (x_test, y_test)])
# XGBoost with regularization
xgb_classifier = xgb.XGBClassifier(reg_alpha=0.1, reg_lambda=0.1, learning_rate=0.001, eval_metric=['merror', 'mlogloss'], random_state=42 )
xgb_classifier.fit(X_train_scaled, y_train, verbose=0, eval_set=[(X_train_scaled, y_train), (X_test_scaled, y_test)])
# # preparing evaluation metric plots
```

XG boost

```
----- Confusion Matrix -----
----- Key Metrics -----
Accuracy: 1.00
Balanced Accuracy: 1.00

Micro Precision: 1.00
Micro Recall: 1.00
Micro F1-score: 1.00

Macro Precision: 1.00
Macro Recall: 1.00
Macro F1-score: 1.00

Weighted Precision: 1.00
Weighted Recall: 1.00
Weighted F1-score: 1.00

----- Classification Report -----
precision    recall   f1-score   support
0           1.00     1.00      1.00      117
1           1.00     1.00      1.00      109
2           1.00     1.00      1.00      103
3           1.00     1.00      1.00      108

accuracy          1.00      1.00      1.00      437
macro avg       1.00     1.00      1.00      437
weighted avg    1.00     1.00      1.00      437

----- XGBoost -----
```

MODEL CREATION

Model Training

03. Decision Tree Classifier

```
[57] # decision tree classifier
  from sklearn.tree import DecisionTreeClassifier
  DT_model = DecisionTreeClassifier()
  #Training the model
  DT_model.fit(X_train_scaled, y_train)

  y_pred = DT_model.predict(X_test_scaled)
  RF_probability = DT_model.predict_proba(X_test_scaled)[:,1]

  dt_classifier_model = DecisionTreeClassifier(max_depth=5, min_samples_split=10, min_samples_leaf=5, random_state=42)

  # Train the decision tree classifier
  dt_classifier_model.fit(X_train_scaled, y_train)

  y_pred_DT_rg = dt_classifier_model.predict(X_test_scaled)
  #getting the accuracy , recall and precision of the model using test set

  acc_DT=accuracy_score(y_test,y_pred_DT_rg)
  recall_DT=recall_score(y_test,y_pred_DT_rg,average='micro')
  precision_DT=precision_score(y_test,y_pred_DT_rg,average='micro')
  f1score_DT=f1_score(y_test,y_pred_DT_rg,average='micro')

  print("Accuracy : ", acc_DT)
  print("Precision:",precision_DT)
  print("Recall:",recall_DT)
  print("F1-Score:",f1score_DT)
```

Accuracy : 1.0
Precision: 1.0
Recall: 1.0
F1-Score: 1.0

Decision Tree Classifier

MODEL CREATION

Model Training

4. Naive Bayes classifier

```
[58] # train a Gaussian Naive Bayes classifier on the training set
      from sklearn.naive_bayes import GaussianNB

      # instantiate the model
      gnb = GaussianNB()

      # fit the model
      gnb.fit(X_train_scaled, y_train)
      y_pred_GNB = gnb.predict(X_test_scaled)

      #GETTING THE scores
      acc_GNB=accuracy_score(y_test,y_pred_GNB)
      recall_GNB=recall_score(y_test,y_pred_GNB,average='micro')
      precision_GNB=precision_score(y_test,y_pred_GNB,average='micro')
      f1score_GNB=f1_score(y_test,y_pred_GNB,average='micro')

      print("Accuracy : ", acc_GNB)
      print("Precision:",recall_GNB)
      print("Recall:",precision_GNB)
      print("F1-Score:",f1score_GNB)
```

```
Accuracy :  0.9725400457665904
Precision: 0.9725400457665904
Recall: 0.9725400457665904
F1-Score: 0.9725400457665903
```

Naive bayes Classifier

```
[59] from sklearn.metrics import confusion_matrix

      cm = confusion_matrix(y_test, y_pred_GNB)

      print('Confusion matrix\n\n', cm)

      print('\nTrue Positives(TP) = ', cm[0,0])
      print('\nTrue Negatives(TN) = ', cm[1,1])
      print('\nFalse Positives(FP) = ', cm[0,1])
      print('\nFalse Negatives(FN) = ', cm[1,0])

      Confusion matrix

      [[117   0   0   0]
       [  0 109   0   0]
       [  0   9  94   0]
       [  0   0   3 105]]

      True Positives(TP) =  117
      True Negatives(TN) =  109
      False Positives(FP) =  0
      False Negatives(FN) =  0
```

MODEL CREATION

Comparison between models

▼ COMPARISON AMONG THE MODELS

```
[60] ind=['XGBoost','Randomforest','Decision Tree','GaussianNB']
0s    data={"Accuracy":[acc_xg,acc_rf,acc_DT,acc_GNB],"Recall":[recall_xg,recall_rf,recall_DT,recall_GNB],"Precision":[precision_xg,precision_rf,precision_DT,precision_GNB],
      'f1_score':[f1score_xg,f1score_rf,f1score_DT,f1score_GNB]}
      result=pd.DataFrame(data=data,index=ind)
      result
```

| | Accuracy | Recall | Precision | f1_score | ✎ |
|---------------|----------|----------|-----------|----------|---|
| XGBoost | 1.000000 | 1.000000 | 1.000000 | 1.000000 | |
| Randomforest | 0.993135 | 0.993135 | 0.993135 | 0.993135 | |
| Decision Tree | 1.000000 | 1.000000 | 1.000000 | 1.000000 | |
| GaussianNB | 0.972540 | 0.972540 | 0.972540 | 0.972540 | |

MODEL CREATION

Checking the model work
flow

```
✓ 0s   import pyowm
      owm = pyowm.OWM(OWM_API_KEY)
      weather_mgr = owm.weather_manager()
      place = 'Gampaha, LK'
      observation = weather_mgr.weather_at_place(place)

✓ 0s   [67] temperature = observation.weather.temperature("celsius")["temp"]
      humidity = observation.weather.humidity
      wind = observation.weather.wind()
      print(f'Temperature(°C): {temperature}')
      print(f'Humidity(%): {humidity}')

Temperature(°C): 26.47
Humidity(%): 86
```

Temperature and
humidity through Open
weather API

MODEL CREATION

Checking the model work flow

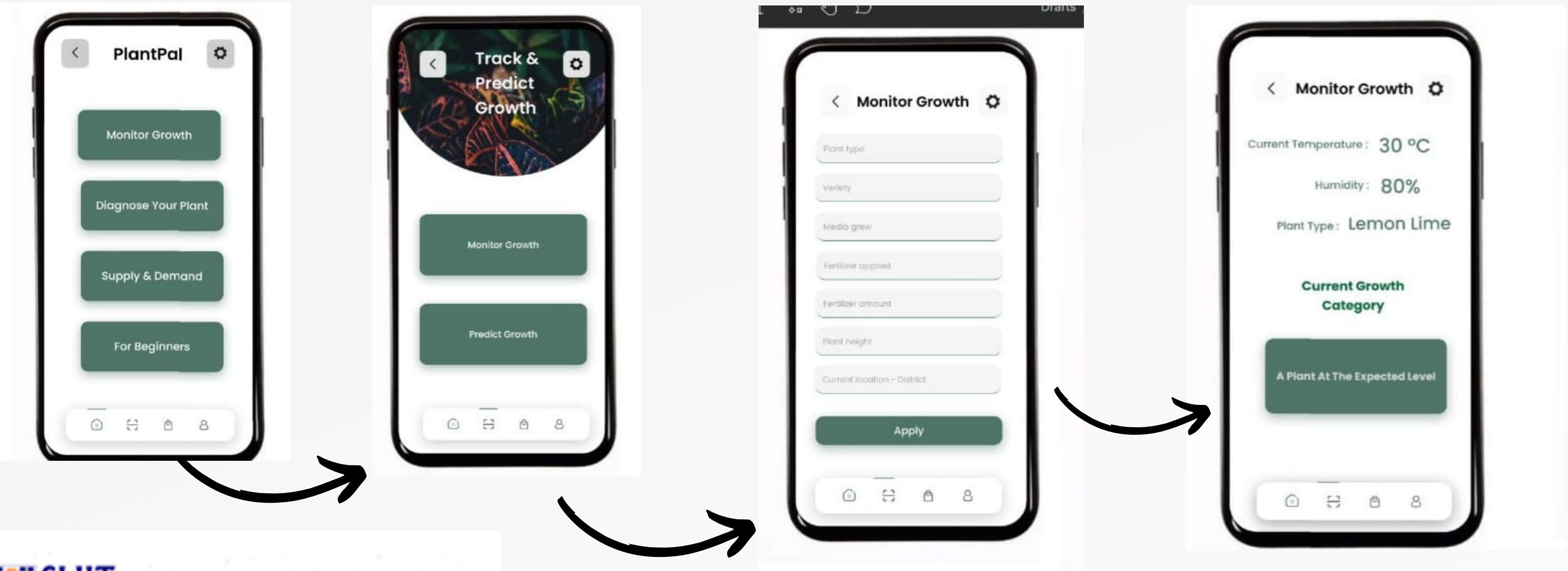
```
[70] # Scale the numeric features
     scaler = MinMaxScaler()
     X_new_scaled = scaler.fit_transform(X_new)

[71] #making the new prediction
     My_predictions = rf_classifier.predict(X_new_scaled)
     if My_predictions ==[0]:
         print('Recently planted one')
     elif My_predictions ==[1]:
         print('A plant at small level')
     elif My_predictions ==[2]:
         print('A plant at medium level')
     elif My_predictions ==[3]:
         print('A plant at expected level')

Recently planted one
```

Model predict the
expected value correctly

UI



FUTURE

Imporove the
growth
monitoring
model

Create a model
for predictions

Developing the
UI's

COMMERCIALIZATION & BUSINESS PLAN

Commodity Version

- Prediction of the current state of growth in ornamental plants
- Identification of the relative growth rate of plant height
- Distinguishing nutrient deficiencies from mite attacks
- Identification of the most suitable floriculture plant to grow based on weather and resource factors
- Forecastination of the demand for local and international markets

Premium Version (100\$annually)

- Prediction of the time it will take for a plant to reach the required growth size for an order
- Providing suggestions to promote growth and advance the ornamental plant to the next level
- Identifying various floriculture crop varieties
- Recommendation of plants for shipments
- Predicting the optimal supply strategy for meeting the required output quantity based on geographical areas and finding the nearest vendor
- Providing packaging recommendations based on the plant's lifespan

COMMERCIALIZATION & BUSINESS PLAN

Target Audience:

- Floriculture businesses (small, medium, & large scale)
- Farmers and gardeners who grow ornamental plants
- Horticulture enthusiasts
- Research and educational institutions

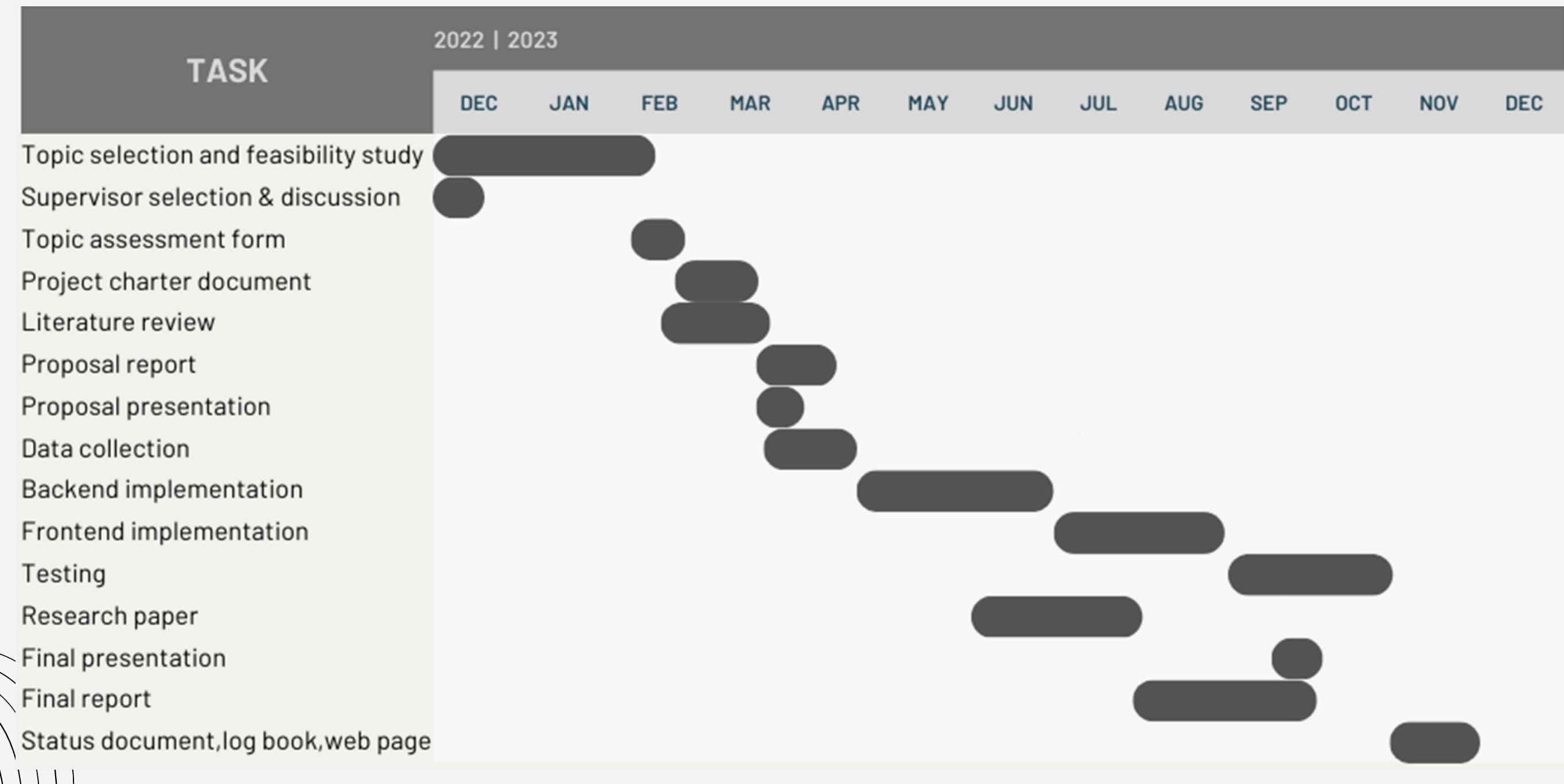
Marketplace:

- Mobile application for Android & iOS devices
- Secure & user-friendly interface for customers & vendors

Marketing Plan:

- Social media marketing
- (Facebook, Instagram, Twitter)
- Email marketing campaigns
- Share leaflets

GANTT CHART





PlantPal

THANK YOU

