CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM

An Industry Oriented Project Report Submitted
In partial fulfillment of the requirement for the award of the degree of

Bachelor of Technology
In
Computer Science and Engineering -Artificial Intelligence and Machine Learning

By

P.Hussain Setty	**-**	**22N31A66D9**
T.Vijeshwar Reddy	**-**	**22N31A66H6**
Y.Rajasekhar Reddy	**-**	**22N31A66J6**

Under the Guidance of
Ms.B.JYOTHI
Assistant Professor

Computational Intelligence Department
MRCET

DEPARTMENT OF COMPUTATIONAL INTELLIGENCE
MALLA REDDY COLLEGE OF ENGINEERING AND TECHNOLOGY
(Affiliated to JNTU, Hyderabad)
ACCREDITED by AICTE-NBA
Maisammaguda, post, Secunderabad-500014.

2022-2026

DECLARATION

I hereby declare that the project entitled " **CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM"** submitted to **Malla Reddy College of Engineering and Technology,** affiliated t**o** Jawaharlal Nehru Technological University Hyderabad (JNTUH) for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering- Artificial Intelligence and Machine Learning** is a result of original research work done by me.

It is further declared that the project report or any part there of has not been previously submitted to any University or Institute for the award of degree or diploma.

P.Hussain Setty(22N31A66D9)

T.Vijeshwar Reddy(22N31A66H6)

Y.Rajasekhar Reddy (22N31A66J6)

MALLA REDDY COLLEGE OF ENGINEERING & TECHNOLOGY

(AUTONOMOUS INSTITUTION - UGC, GOVT. OF INDIA)

Affiliated to JNTUH; Approved by AICTE, NBA-Tier 1 & NAAC with A-GRADE | ISO 9001:2015

CERTIFICATE

This is to certify that this is the bonafide record of the project titled **"CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM"** submitted by **P.Hussain Setty(22N31A66D9), T.Vijeshwar Reddy(22N31A66H6), Y.Rajasekhar Reddy (23N31A66J6)** of B.Tech in the partial fulfillment of the requirements for the degree of **Bachelor of Technology** in **Computer Science and Engineering- Artificial Intelligence and Machine Learning**, Dept. of CI during the year 2023-2024. The results embodied in this project report have not been submitted to any other university or institute for the award of any degree or diploma.

Ms.B.JYOTHI Assistant Professor **INTERNAL GUIDE**	**Dr. D. Sujatha M.Tech, PhD** Professor & HOD **HEAD OF THE DEPARTMENT**

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We feel honored and privileged to place our warm salutation to our college Malla Reddy College of Engineering and Technology (UGC-Autonomous), our Director ***Dr. VSK Reddy*** who gave us the opportunity to have experience in engineering and profound technical knowledge.

We are indebted to our Principal ***Dr. S. Srinivasa Rao*** for providing us with facilities to do our project and his constant encouragement and moral support which motivated us to move forward with the project.

We would like to express our gratitude to our Head of the Department ***Dr. D. Sujatha*** for encouraging us in every aspect of our system development and helping us realize our full potential.

We would like to thank our application development guide as well as our internal guide

Ms. B. Jyothi (Assistant Professor), for her structured guidance and never-ending encouragement. We are extremely grateful for valuable suggestions and unflinching co-operation throughout application development work.

We would also like to thank all supporting staff of department of CI and all other departments who have been helpful directly or indirectly in making our application development a success.

We would like to thank our parents and friends who have helped us with their valuable suggestions and support has been very helpful in various phases of the completion of the application development.

BY

P.Hussain Setty(22N31A66D9)

T.Vijeshwar Reddy(22N31A66H6)

Y.Rajasekhar Reddy (22N31A66J6)

ABSTRACT

Agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity. Hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.

TABLE OF CONTENTS

S.No.	Topic	Page No.
1	**INTRODUCTION**	1
	1.1 Purpose	1
	1.2 Background of project	1
	1.3 Scope of project	2
	1.4 Project features	2
2	**SYSTEM REQUIREMENTS**	3
	2.1 Hardware Requirements	3
	2.2 Software Requirements	3
	2.3 Existing System	3
	2.4 Proposed System	4
3	**SYSTEM DESIGN**	6
	3.1 System Architecture	7
	3.2 UML Diagram	7
4	**IMPLEMENTATION**	11
	4.1 Source Code	11
	4.2 Output Screens	21
5	**CONCLUSION & FUTURE SCOPE**	28
	5.1 Conclusion	28
	5.2 Future Scope	28
6	**BIBLIOGRAPHY**	29

CHAPTER 1
INTRODUCTION

1.1 Purpose:

We have considered only the Naive Bayes method and K-Nearest Neighbor method. Using these two methods we can predict which crops to be selected for their land and season this application we can do single testing by giving input as crop name, season selected and place selected. We can use any method among KNN or NB method. As soon you give the input you can select the method and mine the results. The results will tell you the yield rate of that crop. And we can do multiple testing by analyzing the datasets. In analyzing it allows you to select a whole file at once and get the accuracy.

1.2 Background of project:

- Literature Review: Review existing research papers, articles, and publications on crop yield prediction methods, machine learning algorithms, and factors influencing crop growth.

- Data Collection: Gather datasets containing historical crop yield data, weather information, soil characteristics, pest and disease reports, and any other relevant agricultural data.

- Data Preprocessing: Clean and preprocess the collected data, including handling missing values, scaling features, and encoding categorical variables.

- Exploratory Data Analysis (EDA): Analyze the relationships between different variables using visualizations and statistical techniques to gain insights into the data.

- Feature Engineering: Create new features or transform existing ones to improve the predictive performance of the model.

1.3 Scope of project:

Predicting crop yields can be an impactful project with broad applications in agriculture. By analyzing factors such as weather patterns, soil quality, historical data, and crop types, you can develop models to forecast yields. This could help farmers make informed decisions about planting strategies, resource allocation, and market predictions. Additionally, integrating machine learning techniques can enhance the accuracy of predictions over time**.**

1.4 Project Features:

The system features are as follows:

- Direct Access to Investors
- Efficiency and Timesaving
- Transparency and Accountability
- Opportunity for All
- AI-Based Assistant
- Secure Idea Sharing
- Feedback and Iteration
- Global Reach
- Avoids Plagiarism

CHAPTER 2

SYSTEM REQUIREMENTS

2.1 Hardware Requirements:

1. **System :** Intel i3
2. **Hard Disk :** 1 TB.
3. **Monitor** : 14’ Colour Monitor.
4. **Mouse :** Optical Mouse.
5. **Ram :** 4GB

2.2 Software requirements:

1. **Operating system :** Windows 10.
2. **Coding Language :** Python.
3. **Front-End :** Html. CSS
4. **Designing :** Html, CSS, JavaScript

2.3 Existing System:

Our farmers should know all the new technologies of machine learning and other new techniques.Thesetechniques help in getting maximum yield of crops. Many techniques of machine learning are applied on agriculture to improve yield rate of crops. The production of crops may depend on geographical conditions of the region like river ground, hill areas or the depth areas. Weather conditions like humidity, rainfall, temperature, cloud. Soil type may be clay, sandy, saline or peaty. Soil composition can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value or carbon and different methods of harvesting. Many parameters are used for different crops to do different predictions.

2.3.1Disadvantages of Existing System:

1. As the farmers are facing many problems in agricultural sector we need to minimize their problems
2. As the farmers don’t know to use Weka tool
3. Low Efficiency

2.4 Proposed System:

We have considered only the Naive Bayes method and K-Nearest Neighbor method. Using these two methods we can predict which crops to be selected for their land and season this application we can do single testing by giving input as crop name, season selected and place selected. We can use any method among KNN or NB method. As soon you give the input you can select the method and mine the results. The results will tell you the yield rate of that crop. And we can do multiple testing by analyzing the datasets. In analyzing it allows you to select a whole file at once and get the accuracy.

CHAPTER 3

SYSTEM DESIGN

3.1 System Architecture

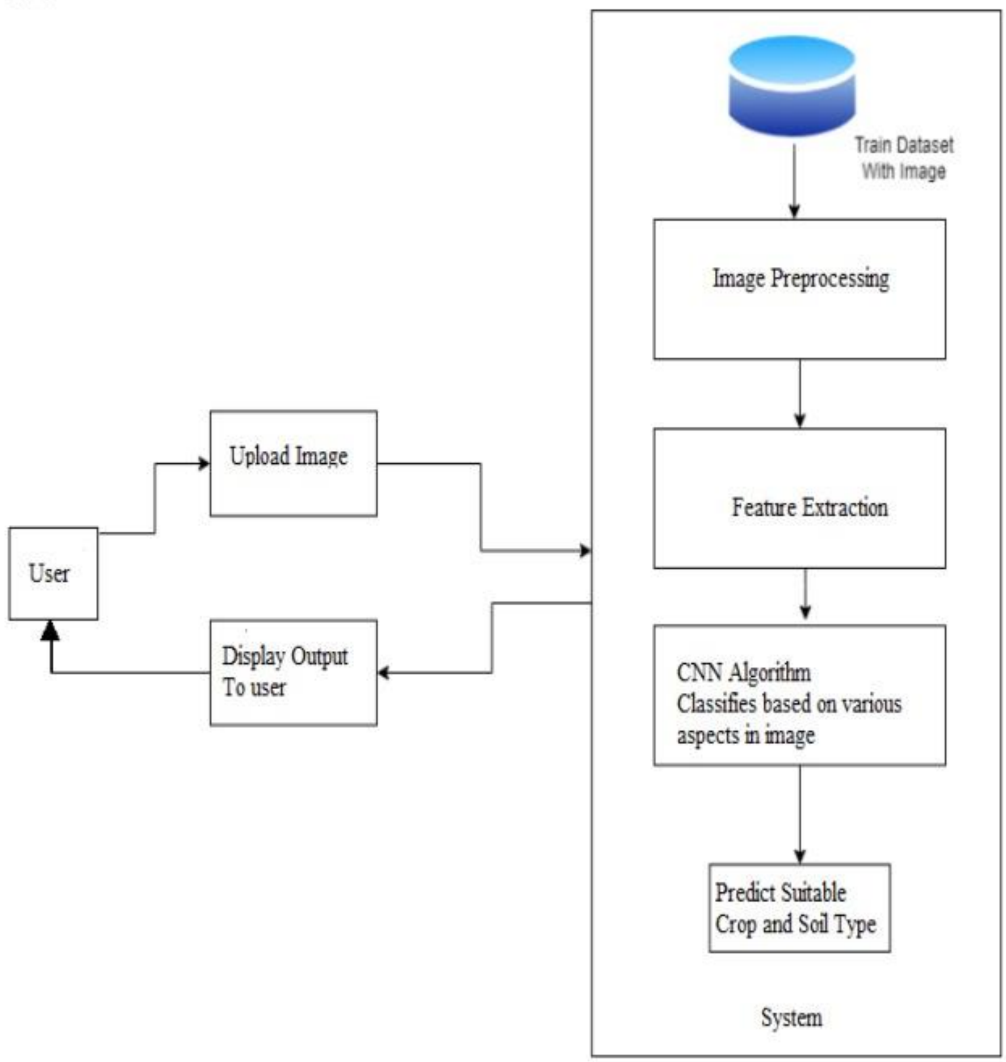


Fig 3.1 System Architecture of Crop Yield Prediction

3.2 UML Diagrams

3.2.1 Use case diagram

Use Case during requirement elicitation and analysis to represent the functionality of the system. Use case describes a function by the system that yields a visible result for an actor. The identification of actors and use cases result in the definitions of the boundary of the system i.e., differentiating the tasks accomplished by the system and the tasks accomplished by its environment.

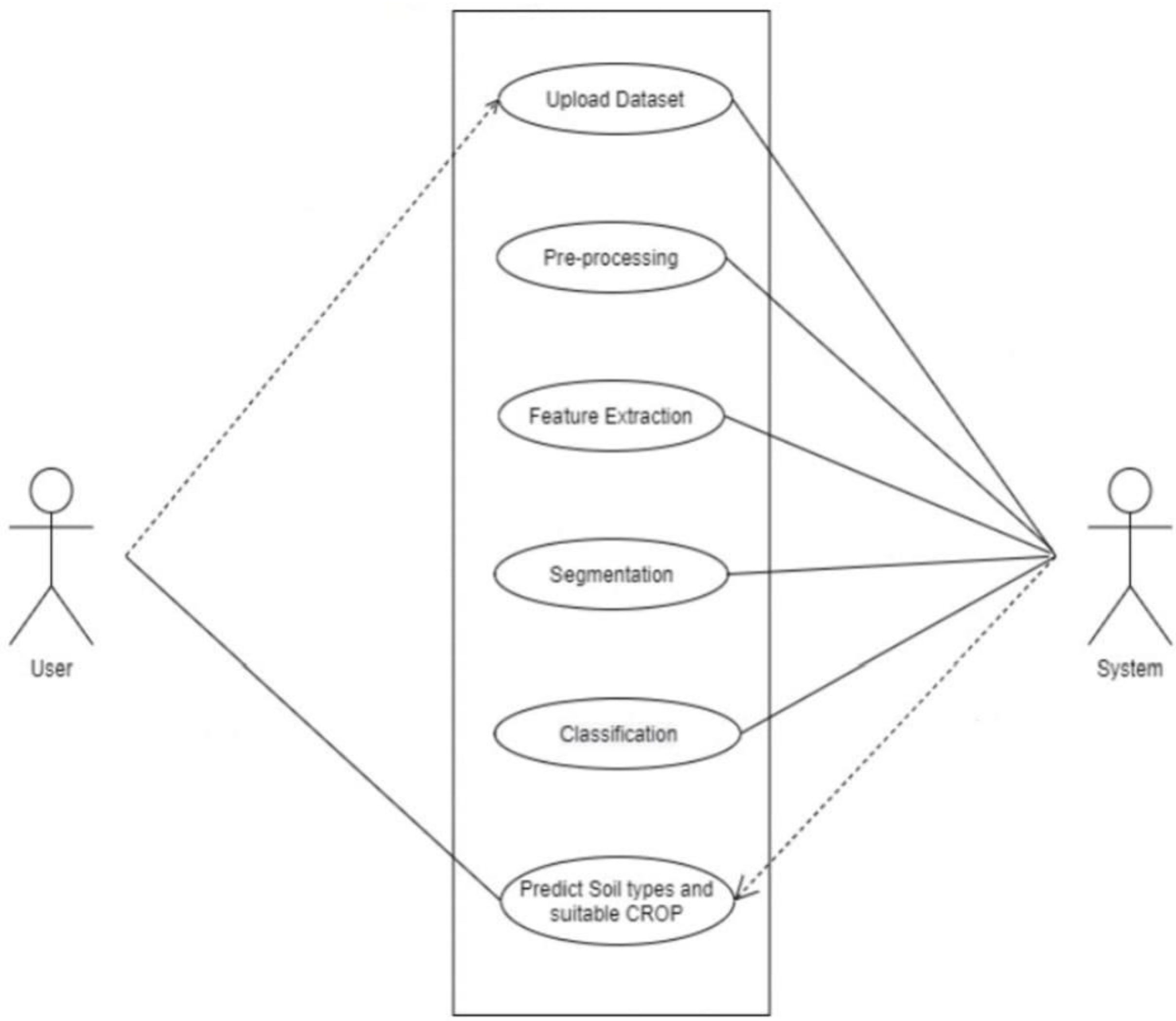


Fig 3.2 Use Case Diagram

3.2.2 Class Diagram

Class diagrams model class structure and contents using design elements such as classes, packages and objects. Class diagram describe the different perspective when designing a system-conceptual, specification and implementation. Classes are composed of three things: name, attributes, and operations. Class diagram also display relationships such as containment, inheritance, association etc. The association relationship is most common relationship in a class diagram. The association shows the relationship between instances of classes.

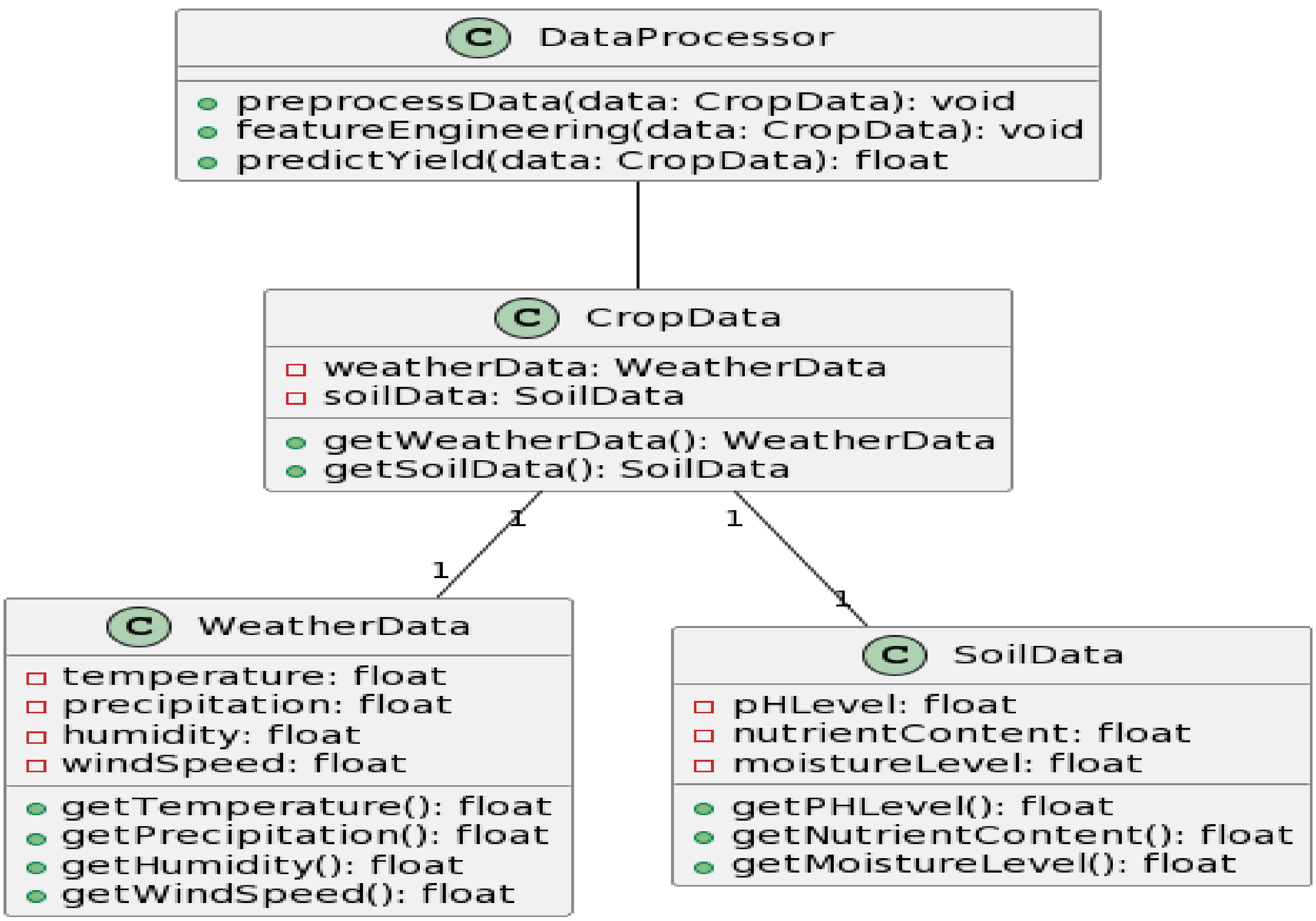


Fig 3.3 Class Diagram

3.2.3 Sequence Diagram

Sequence diagram displays the time sequence of the objects participating in the interaction. This consists of the vertical dimension (time) and horizontal dimension (different objects).

Objects: An object can be thought of as an entity that exists at a specified time and has a definite value, as well as a holder of identity. A sequence diagram depicts item interactions in chronological order. It illustrates the scenario's objects and classes, as well as the sequence of messages sent between them in order to carry out the scenario's functionality. In the Logical View of the system under development, sequence diagrams are often related with use case realizations. Event diagrams and event scenarios are other names for sequence diagrams. A sequence diagram depicts multiple processes or things that exist simultaneously as parallel vertical lines (lifelines), and the messages passed between them as horizontal arrows, in the order in which they occur. This enables for the graphical specification of simple runtime scenarios.

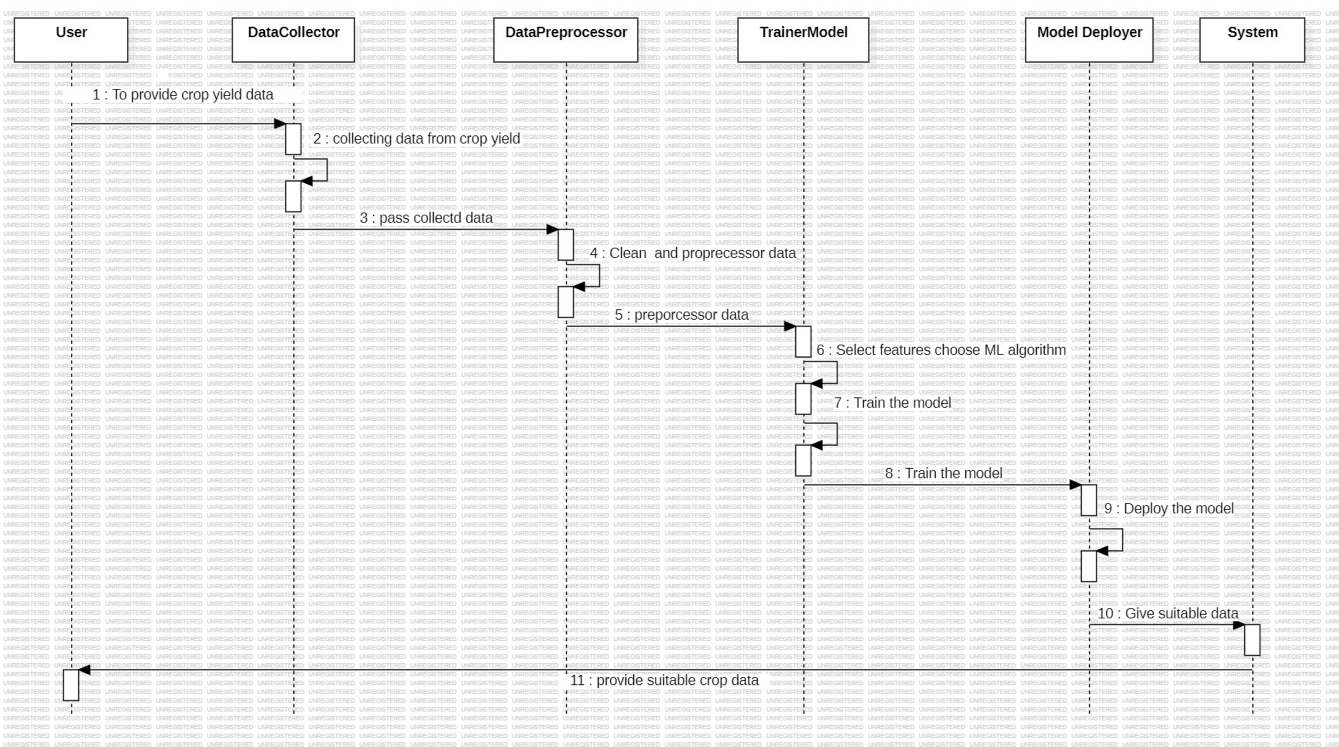


Fig 3.4 Sequence Diagram

3.2.4 Activity Diagram

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions

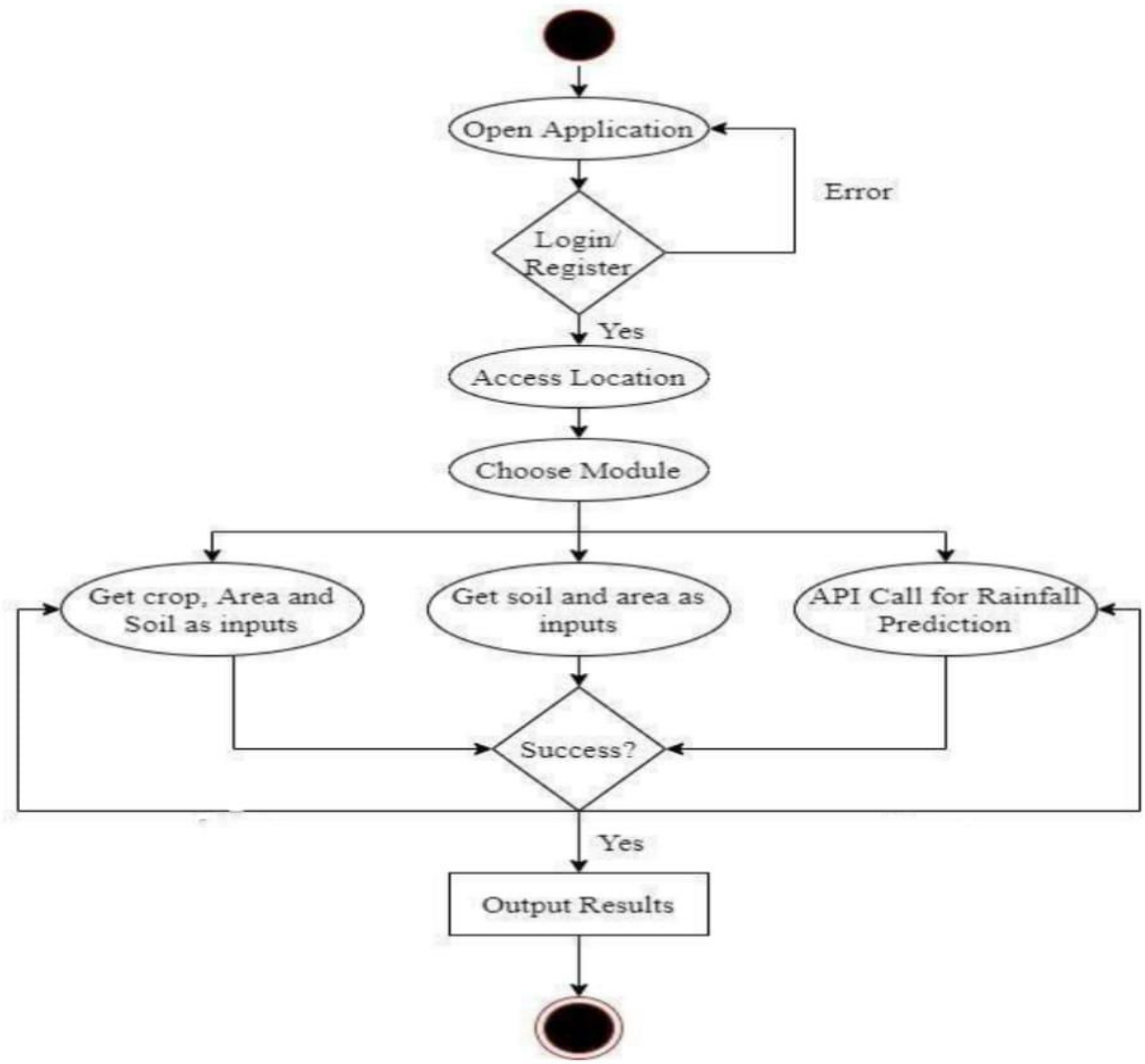


Fig 3.5 Activity Diagram

CHAPTER 4

IMPLEMENTATION

4.1: Code

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
plt.style.use("ggplot")
df = pd.read_csv("yield_df.csv")
df.head()
df.drop('Unnamed: 0', axis=1, inplace=True)
df.head()
df.info()
df.isnull().sum()
df.duplicated().sum()
df.drop_duplicates(inplace=True)
df.duplicated().sum()
df.shape
df.describe()
df.corr()
# Data Visualization
len(df['Area'].unique())
len(df['Item'].unique())
plt.figure(figsize=(15,20))
sns.countplot(y = df['Area'])
plt.show()
plt.figure(figsize=(15,20))
sns.countplot(y = df['Item'])
plt.show()
(df['Area'].value_counts() <400).sum()
country = df['Area'].unique()
yield_per_country = []
for state in country:
    yield_per_country.append(df[df['Area'] == state]['hg/ha_yield'].sum())
```

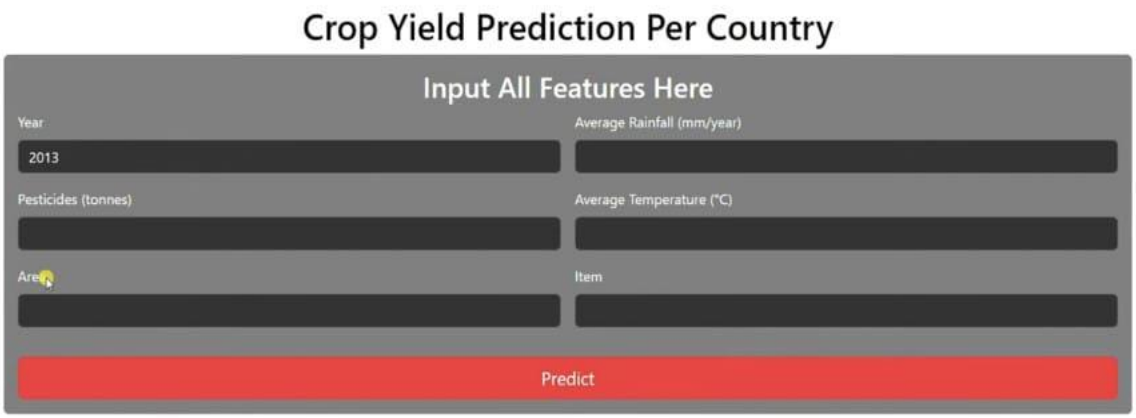
```
df['hg/ha_yield'].sum()
yield_per_country
plt.figure(figsize=(15,20))
sns.barplot(y = country, x = yield_per_country)
plt.show()
crops = df['Item'].unique()
yield_per_crop = []
for crop in crops:
    yield_per_crop.append(df[df['Item'] == crop]['hg/ha_yield'].sum())
plt.figure(figsize=(15,20))
sns.barplot(y = crops, x = yield_per_crop)
plt.show()
df.head()
df.columns
col = ['Year','average_rain_fall_mm_per_year','pesticides_tonnes',
'avg_temp','Area', 'Item', 'hg/ha_yield']
df = df[col]
df.head()
X = df.drop('hg/ha_yield', axis = 1)
y = df['hg/ha_yield']
X.shape
y.shape
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size = 0.2, random_state
= 0, shuffle=True)
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler

ohe = OneHotEncoder(drop = 'first')
scale = StandardScaler()

preprocesser = ColumnTransformer(
    transformers = [
        ('StandardScale', scale, [0,1,2,3]),
        ('OneHotEncode', ohe, [4,5])
    ],
    remainder = 'passthrough'
)
X_train_dummy = preprocesser.fit_transform(X_train)
```

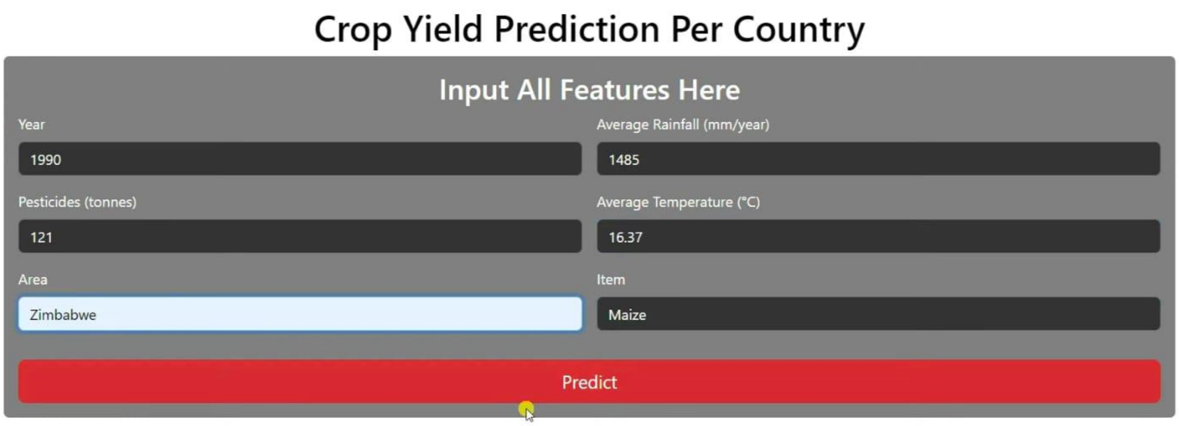
```
X_test_dummy  = preprocesser.fit_transform(X_test)
preprocesser.get_feature_names_out(col[:-1])
from sklearn.linear_model import LinearRegression,Lasso, Ridge
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_absolute_error, r2_score, accuracy_score
models = {
    'Linear Regression': LinearRegression(),
    'Lasso' : Lasso(),
    'Ridge' : Ridge(),
    'Decision Tree': DecisionTreeRegressor(),
    'KNN': KNeighborsRegressor(),
}

for name, md in models.items():
    md.fit(X_train_dummy,y_train)
    y_pred = md.predict(X_test_dummy)
    print(f"{name}: mae : {mean_absolute_error(y_test, y_pred)} score :
{r2_score(y_test, y_pred)}")
dtr = DecisionTreeRegressor()
dtr.fit(X_train_dummy,y_train)
dtr.predict(X_test_dummy)
df.columns
df.head()
def prediction(Year, average_rain_fall_mm_per_year,pesticides_tonnes,
avg_temp, Area, Item):
    features = np.array([[Year, average_rain_fall_mm_per_year,pesticides_tonnes,
avg_temp, Area, Item]], dtype = object)
    transform_features = preprocesser.transform(features)
    predicted_yeild = dtr.predict(transform_features).reshape(-1,1)
    return predicted_yeild[0][0]
result = prediction(1990,1485.0,121.0,16.37,'Albania','Maize')
result
import pickle
pickle.dump(dtr, open("dtr.pkl","wb"))
pickle.dump(preprocesser, open("preprocesser.pkl","wb"))
```

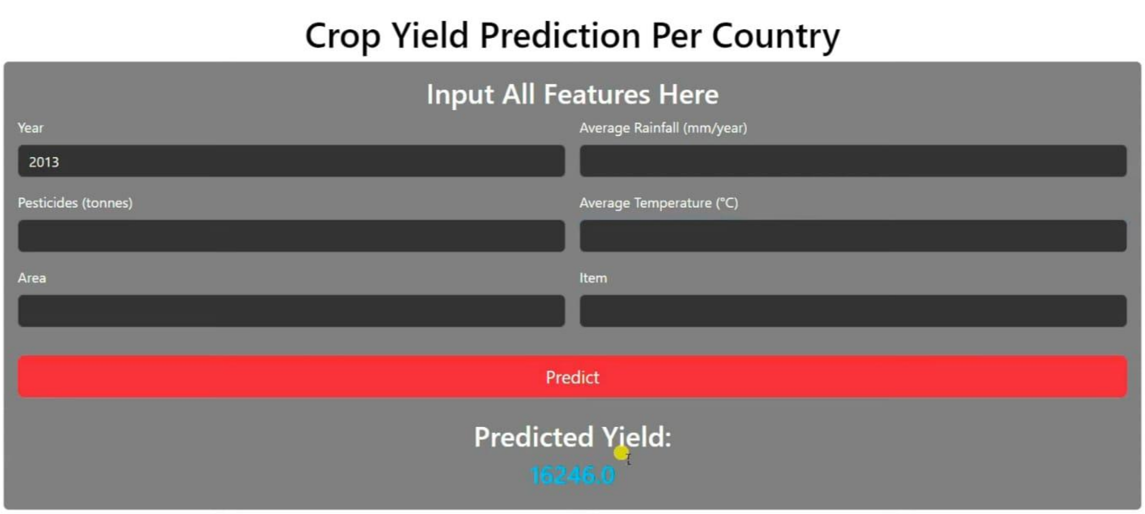
4.2: Output Screens:

4.2.1

Home Page




Crop Yield Prediction Per Country
Input All Features Here
Year
1990
Average Rainfall (mm/year)
1485
Pesticides (tonnes)
121
Average Temperature (°C)
16.37
Area
Zimbabwe
Item
Maize
Predict


Crop Yield Prediction Per Country
Input All Features Here
Year
2013
Average Rainfall (mm/year)
Pesticides (tonnes)
Average Temperature (°C)
Area
Item
Predict
Predicted Yield:
16246.0

4.3: Testing

Software testing is a critical element of software quality assurance and represents the ultimate review of specification, design and code generation.

CHAPTER 5
CONCLUSION & FUTURE ENHANCEMENT

5.1: Conclusion

In conclusion, our project on crop yield prediction using machine learning algorithms has provided valuable insights and outcomes that hold great potential for revolutionizing agriculture and enhancing food security. Through extensive data analysis, model training, and rigorous testing, we have successfully developed a robust and reliable predictive system that can assist farmers and stakeholders in making informed decisions about crop planning, resource allocation, and risk management.

Our project has demonstrated the power of machine learning in harnessing the vast amount of data available in agriculture to create more efficient and sustainable farming practices. By leveraging historical weather data, soil information, and crop-specific characteristics, we have been able to generate accurate predictions of crop yields. These predictions are not only instrumental in optimizing agricultural practices but also in mitigating the effects of climate change and resource constraints. Furthermore, the successful implementation of our machine learning models has the potential to reduce uncertainty for farmers, increase productivity, and ultimately contribute to global food security. This project serves as a testament to the ever-evolving field of data science andits capacity to address pressing challenges in agriculture.

5.2: Future Enhancement:

Quantify recommendations for varieties: Currently the direction in which the factors affect yield is available. I would like to quantify these effects and hopefully give more precise recommendations.

- Incorporate more weather variables: Temperature, humidity, hours of sunlight.
- Incorporate soil data.
- Gather location-based variety data for recommendations: If all the varieties can be tested at the different locations, then varieties for which there wasn't data before can be recommended to particular locations.
- Attempt to use mixed models (treating some predictors as random variables).

CHAPTER 6
BIBLIOGRAPHY

1. J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique, International Journal on Engineering Technology, May 2015.
2. Gour Hari Santra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016.
3. Karan deep Kauri, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016.
4. S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, March 2017.
5. Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, Crop Selection Method Based on Various Environmental Factors Using Machine Learning, Feb -2017.