A

MINI PROJECT REPORT

ON

"CROP RECOMMENDATION SYSTEM"

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERITY, PUNE
FOR

THE PARTIAL FULFILMENT FOR THE AWARD OF THE DEGREE

OF BACHELOR OFENGINEERING IN

BY

JAYASH GAIKWAD -33318

ABHIJIT KHYADE -33337

DEPARTMENT OF INFORMATION TECHNOLOGY

PUNE INSTITUTE OF COMPUTER TECHNOLOGY, PUNE

Sr. No. 27, Trimurti Chowk, Dhankwadi, Pune-43

AY- 2023-24

CERTIFICATE

This is to certify that Mini project report entitled

"CROP RECOMMENDATION SYSTEM"

Submitted by

Jayash Gaikwad
Abhijit Khyade

is a Mini Project carried out by them in the third year of Engineering (A. Y. 2023-24) under the supervision of **Mr. Vineet Tribhuvan**, and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Information Technology).

Mr. Vineet Tribhuvan	Dr. A. S. Ghotkar	Dr. S. T. Gandhe
Internal Guide	Head of the Department	Principal

ABSTRACT

Our Crop Prediction System leverages the capabilities of machine learning to forecast crop yields, aiding farmers, and agricultural stakeholders in making informed decisions. Analogous to a sports team manager carefully selecting players based on diverse factors, our system classifies crops into different yield categories. These categories enable users to assess whether a crop's yield is likely to be high, low, or remain stable within a specified time frame.

To accomplish this, we have explored various classification algorithms, including Naïve Bayes, Random Forest, Multiclass SVM, and Decision Trees. Our rigorous analysis has demonstrated that the Random Forest classifier consistently delivers the most precise predictions, establishing it as a vital tool for crop yield forecasting.

In the dynamic realm of agriculture, where environmental conditions and factors affecting crop yields are continually changing, the capacity to make data-driven decisions is of paramount importance. Our system equips users with the insights required to optimize their farming strategies and adapt to the ever-evolving agricultural landscape. Whether you're an experienced farmer or new to agriculture, our Crop Prediction System is your pathway to success.

KEYWORDS: Crop Prediction, Machine Learning, Random Forest, Classification Algorithms, Agriculture

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to the individuals who guided and supported us throughout our project. Special thanks to our internal guide, **Mr. Vineet Tribhuvan**, and the dedicated staff of the Information Technology Department for their invaluable assistance.

We acknowledge the support of our Head of Department (HOD), our classmates, and our parents, whose unwavering encouragement was instrumental in the successful completion of our project. We're also thankful to our friends and everyone who, directly or indirectly, contributed to our project's journey.

Your collective guidance and support have illuminated our path, and we are genuinely grateful for your contributions.

This version retains the essence of gratitude while being more concise.

Jayash Gaikwad (33318)

Abhijit Khyade (33337)

CONTENT

LIST OF FIGURES

Chapter 1
Introduction

In a world driven by data and innovation, agriculture represents a dynamic and ever-evolving landscape that sustains communities and economies. Crop production serves as a fundamental pillar of global food security, where the potential for abundance or scarcity is influenced by a multitude of factors.

The Crop Prediction Project enters this realm, offering a holistic solution for forecasting crop yields. It harnesses the capabilities of machine learning, drawing inspiration from various predictive analytics techniques to provide invaluable insights for farmers and agricultural stakeholders.

Much like a skilled cricket team relies on data-driven strategies for victory, this project equips the agricultural community with the tools needed to make informed decisions about crop production. It delves into the intricacies of environmental conditions, soil quality, historical yield data, and crop-specific factors, with the goal of assisting users in understanding and predicting crop yields.

In a world where agricultural decisions can be as time sensitive as a stock market trade, accuracy and efficiency are paramount. As the project addresses these demands, it acknowledges the changing landscape of agriculture, where technology and data analytics continue to play a pivotal role.

The application is designed to cater to farmers, agricultural stakeholders, and enthusiasts, ensuring that they have access to reliable and data-backed predictions for crop yields. The project's primary goal is to explore, analyze, and implement various machine learning algorithms to identify the most accurate model, enhancing the overall precision of crop yield predictions.

Purpose:

The purpose of developing this Crop Prediction Project is to address the increasing demand for accurate and data-driven tools in the world of agriculture. This application aims to provide farmers, agricultural stakeholders, and enthusiasts with a reliable and user-friendly platform for predicting crop yields. By leveraging the power of machine learning and predictive analytics, the project intends to offer valuable insights that empower users to make informed decisions in the complex and dynamic field of agriculture.

Scope:

The scope of the project encompasses the creation of a comprehensive web application designed to forecast crop yields. This application will cater to a wide range of users, from small-scale farmers to large agricultural enterprises, providing them with a platform for making data-driven decisions in agriculture. It will consider various factors that influence crop yields, such as environmental conditions, soil quality, historical yield data, and crop-specific variables. The application will focus on predictive accuracy and the selection of the most suitable machine learning algorithms to enhance crop yield predictions.

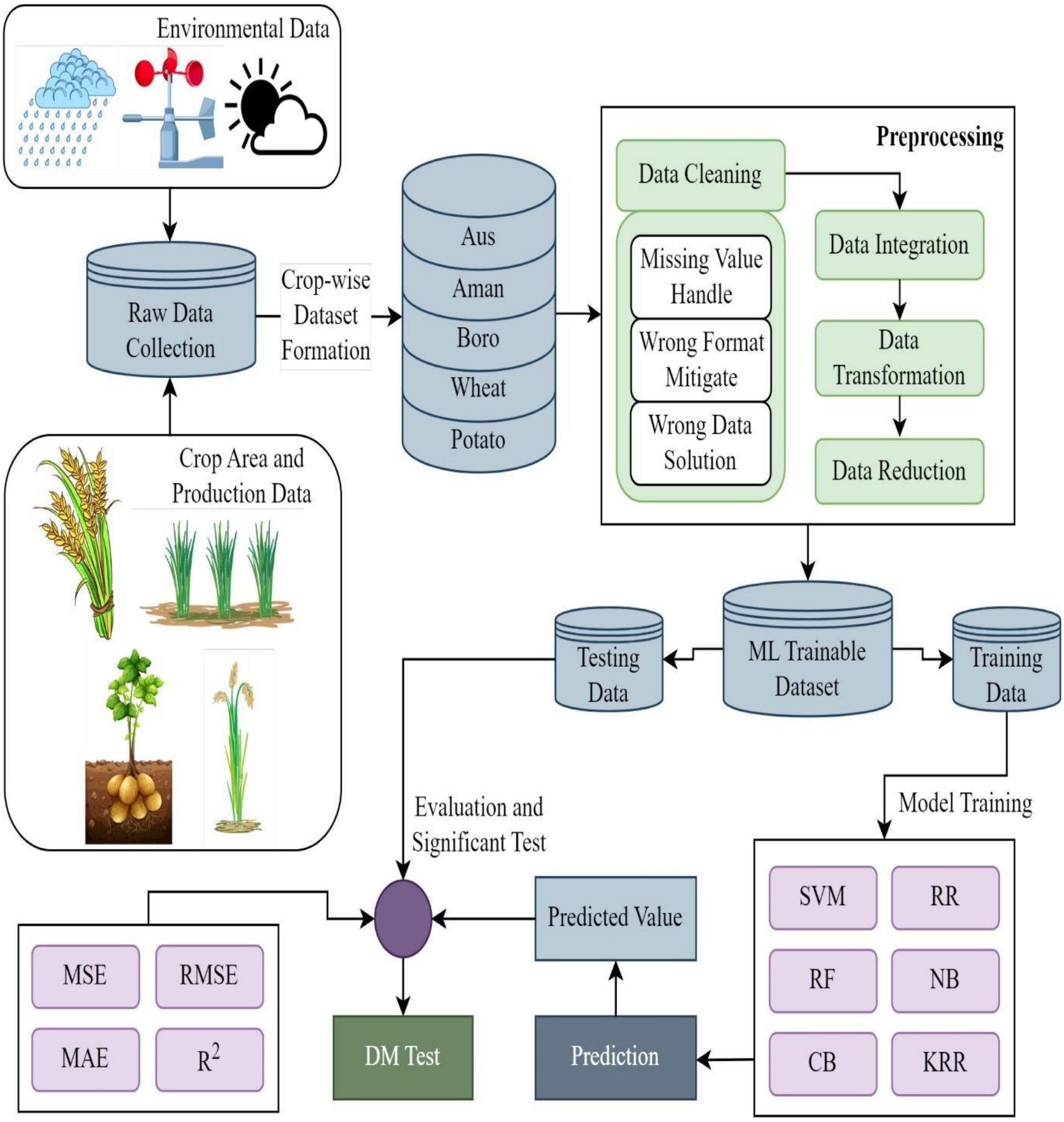
Background and Motivation:

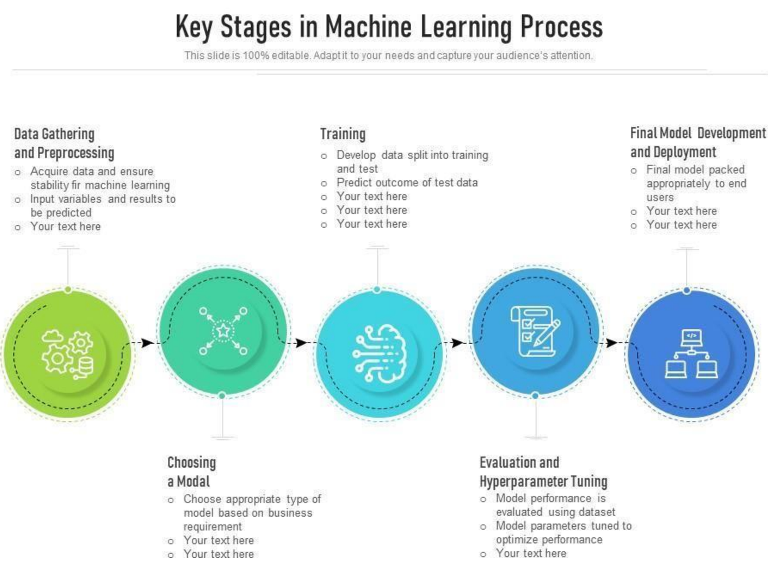
The background of this project lies in the transformative influence of technology and data analytics on the field of agriculture. As agriculture continues to evolve, there is a growing need for sophisticated tools that can offer insights into crop yields. The motivation for this project arises from the desire to address this demand and provide a solution that allows users to navigate the complexities of crop production with confidence. In an era where technology plays an increasingly pivotal role in agriculture, this project aims to harness the potential of machine learning to make crop yield predictions accessible and accurate.

Chapter 2

Process Flow, Methodology, and Application

2.1 Process Flow





1. **Data Collection:**
    - The project begins with the collection of historical agricultural data. This data includes information on factors such as weather conditions, soil quality, crop types, and past yield records. Data sources may include government databases, weather stations, and agricultural research institutions.
2. **Data Preprocessing:**
    - The collected data is preprocessed to ensure it is clean and structured for analysis. This involves tasks like handling missing values, data cleaning, and normalization. Data quality is crucial for accurate predictions.
3. **Feature Selection:**
    - Feature selection techniques are applied to identify the most relevant attributes that influence crop yields. These features could include weather patterns, soil moisture levels, crop varieties, and historical yield data.
4. **Model Selection:**
    - Various machine learning models suitable for regression tasks are considered. Models such as linear regression, decision trees, random forests, and support vector machines are evaluated to determine the best-performing model for crop yield prediction.
5. **Training and Testing:**

- The chosen machine learning model is trained on a portion of the historical data. The model's performance is then assessed on a separate validation dataset to ensure it can generalize well to unseen data. Hyperparameter tuning may be performed to optimize model performance.

6. **Deployment:**
   - The most accurate and effective model is deployed within the Crop Prediction Web Application. Users can access this model via the application's user-friendly interface.

7. **User Input:**
   - Users provide input to the application, specifying the crop they want to predict, the location, historical data range, and any other relevant parameters.

8. **Prediction:**
   - The application leverages the trained model to predict the future yield of the specified crop. Predictions are presented to the user in a user-friendly and easily understandable format, such as yield estimates for specific time periods.

9. **Feedback Loop**:
   - Users are encouraged to provide feedback on the accuracy of the predictions. This feedback can be used to fine-tune the model and improve the application's overall performance. Continuous user input and data updates are essential for maintaining prediction accuracy.

2.2 Methodology

The methodology employed in this project is centered on the application of supervised machine learning techniques. Given labeled historical agricultural data, the primary objective is to predict crop yields, effectively addressing a regression problem.

Data Preparation: The initial phase includes data collection, preprocessing, and feature selection. Historical agricultural data is collected, cleaned, and made ready for machine learning.

Feature Engineering: Key features are identified, emphasizing factors significantly affecting crop yields. This step ensures that the input data fed into the models is pertinent and contributes to accurate predictions.

Model Selection: Multiple machine learning models are evaluated, including Random Forest, Naïve Bayes, Support Vector Machines, and Decision Trees. Model selection is based on predictive performance, with a focus on accuracy and generalization.

Training and Testing: The selected model is trained using historical agricultural data, and its performance is assessed on separate validation and testing datasets.

Deployment: The most accurate model is seamlessly integrated into the web application, empowering users to make real-time predictions for crop yields.

2.3 Application

The Crop Prediction Web Application stands as a valuable tool for farmers and agricultural stakeholders seeking data-backed insights into crop production. It offers an intuitive interface where users can specify the crop they wish to predict, set relevant parameters, and receive accurate yield predictions. This application presents a data-driven approach to agricultural decision-making, allowing users to enhance their farming strategies and adapt to changing environmental conditions effectively.

User feedback is strongly encouraged, as it plays a pivotal role in continually improving the prediction accuracy and overall performance of the application. The incorporation of machine learning into the realm of agriculture underscores the growing importance of data-driven insights in optimizing crop production, ensuring food security, and sustainable agriculture.

Chapter 3

Implementation

In [1]:
```
import numpy as np
import pandas as pd
```

Read the Data

In [2]:
```
#Importing the data
crop = pd.read_csv('Crop_recommendation.csv')
```

In [3]:
```
crop
```

Out[3]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 8 columns

In [4]:
```
crop.head()
```

Out[4]:

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

In [5]:
```
#Shape of Data
crop.shape
```

Out[5]: (2200, 8)

In [10]:
```
correlation = crop.corr()
correlation
```

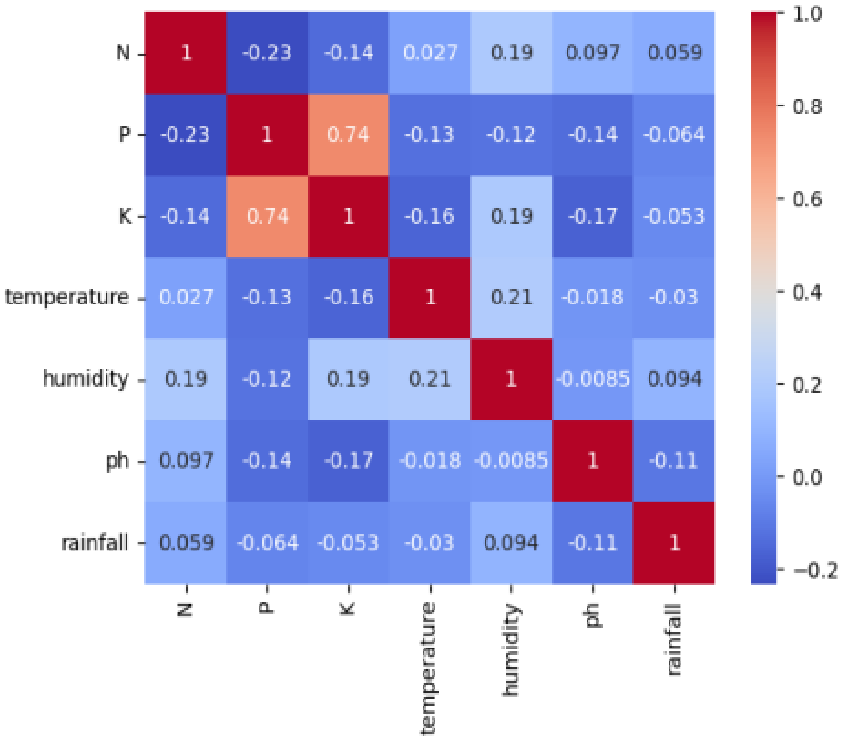
```
C:\Users\Dell\AppData\Local\Temp\ipykernel_11968\1129341823.py:1: FutureW
arning: The default value of numeric_only in DataFrame.corr is deprecate
d. In a future version, it will default to False. Select only valid colum
ns or specify the value of numeric_only to silence this warning.
  correlation = crop.corr()
```

Out[10]:

	N	P	K	temperature	humidity	ph	rainfall
N	1.000000	-0.231460	-0.140512	0.026504	0.190688	0.096683	0.059020
P	-0.231460	1.000000	0.736232	-0.127541	-0.118734	-0.138019	-0.063839
K	-0.140512	0.736232	1.000000	-0.160387	0.190859	-0.169503	-0.053461
temperature	0.026504	-0.127541	-0.160387	1.000000	0.205320	-0.017795	-0.030084
humidity	0.190688	-0.118734	0.190859	0.205320	1.000000	-0.008483	0.094423
ph	0.096683	-0.138019	-0.169503	-0.017795	-0.008483	1.000000	-0.109069
rainfall	0.059020	-0.063839	-0.053461	-0.030084	0.094423	-0.109069	1.000000

In [11]:
```
import seaborn as sns
sns.heatmap(correlation, annot=True, cbar=True, cmap='coolwarm')
```

Out[11]: <Axes: >



In [16]:
```
crop.sample(2)
```

Out[16]:

	N	P	K	temperature	humidity	ph	rainfall	label	crop_num
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice	1
714	51	56	18	28.127878	64.209777	6.706506	70.863408	blackgram	16

Train Test Split

In [17]:
```
X = crop.drop(['crop_num','label'],axis=1)
y = crop['crop_num']
```

In [18]:
```
X
```

Out[18]:

	N	P	K	temperature	humidity	ph	rainfall
0	90	42	43	20.879744	82.002744	6.502985	202.935536
1	85	58	41	21.770462	80.319644	7.038096	226.655537
2	60	55	44	23.004459	82.320763	7.840207	263.964248
3	74	35	40	26.491096	80.158363	6.980401	242.864034
4	78	42	42	20.130175	81.604873	7.628473	262.717340
...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507
2196	99	15	27	27.417112	56.636362	6.086922	127.924610
2197	118	33	30	24.131797	67.225123	6.362608	173.322839
2198	117	32	34	26.272418	52.127394	6.758793	127.175293
2199	104	18	30	23.603016	60.396475	6.779833	140.937041

2200 rows × 7 columns

In [19]:
```
X.shape
```

Out[19]: (2200, 7)

In [20]:
```
y.shape
```

Out[20]: (2200,)

In [21]:
```
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, ranc
```

In [22]:
```
X_train.shape
```

Out[22]: (1760, 7)

Scale the features using MinMaxScaler

In [25]:
```
from sklearn.preprocessing import MinMaxScaler
ms = MinMaxScaler()

X_train = ms.fit_transform(X_train)
X_test = ms.fit_transform(X_test)
```

In [26]:
```
X_train
```

Out[26]:
```
array([[0.12142857, 0.07857143, 0.045     , ..., 0.9089898 , 0.48532225,
        0.29685161],
       [0.26428571, 0.52857143, 0.07      , ..., 0.64257946, 0.56594073,
        0.17630752],
       [0.05      , 0.48571429, 0.1       , ..., 0.57005802, 0.58835229,
        0.08931844],
       ...,
       [0.07857143, 0.22142857, 0.13      , ..., 0.43760347, 0.46198144,
        0.28719815],
       [0.07857143, 0.85      , 0.995     , ..., 0.76763665, 0.44420505,
        0.18346657],
       [0.22857143, 0.52142857, 0.085     , ..., 0.56099735, 0.54465022,
        0.11879596]])
```


In [29]:
```
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import ExtraTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.metrics import accuracy_score

# create instance of all models
models = {
    'Logistic Regression': LogisticRegression(),
    'Naive Bayes': GaussianNB(),
    'Support Vector Machine': SVC(),
    'K-Nearest Neighbors': KNeighborsClassifier(),
    'Decision Tree': DecisionTreeClassifier(),
    'Random Forest': RandomForestClassifier(),
    'Bagging': BaggingClassifier(),
    'AdaBoost': AdaBoostClassifier(),
    'Gradient Boosting': GradientBoostingClassifier(),
    'Extra Trees': ExtraTreeClassifier(),
}

for name, md in models.items():
    md.fit(X_train, y_train)
    ypred = md.predict(X_test)

    print(f"{name} with accuracy : {accuracy_score(y_test,ypred)}")
```

```
Logistic Regression with accuracy : 0.9568181818181818
Naive Bayes with accuracy : 0.9931818181818182
Support Vector Machine with accuracy : 0.9704545454545455
K-Nearest Neighbors with accuracy : 0.9568181818181818
Decision Tree with accuracy : 0.9818181818181818
Random Forest with accuracy : 0.990909090909091
Bagging with accuracy : 0.9818181818181818
AdaBoost with accuracy : 0.1409090909090909
Gradient Boosting with accuracy : 0.9636363636363636
Extra Trees with accuracy : 0.925
```

Chapter 4
Conclusion

The Crop Recommendation Project represents a significant stride in utilizing data-driven insights to empower agriculture. By leveraging machine learning and predictive analytics, we've provided farmers and agricultural stakeholders with a practical tool to make informed decisions in the ever-changing world of crop production.

As we conclude this project, our commitment to continual improvement remains strong. We recognize the transformative potential of machine learning in agriculture and its role in shaping a data-driven and sustainable future. User feedback remains invaluable in enhancing the application's performance and ensuring the agricultural community has access to the latest tools for accurate crop yield predictions. Together, we move forward toward more resilient and data-informed agriculture.

Chapter 5

REFERENCE

https://www.geeksforgeeks.org/visualization-and-prediction-of-crop-production-data-using-python/

https://www.analyticsvidhya.com/blog/2023/06/crop-yield-prediction-using-machine-learning-and-flask-deployment/

https://www.mdpi.com/2076-3417/13/16/9288