Industrial Internship Report on

Crop and weed Detection

Prepared by

Elesh Rodge

Executive Summary

This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time.

My project was "Optimizing Crop Production Using Machine Learning" project leverages advanced data science techniques to enhance agricultural productivity. By analyzing data for 30 crops, including factors such as nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall, the project identifies patterns and optimal growing conditions. Key libraries like NumPy, Pandas, Matplotlib, Seaborn, and Ipywidgets enable efficient data processing, visualization, and interactive analysis.

Statistical summaries and comparison of crop-specific requirements with environmental averages provide actionable insights. Distribution graphs and seasonal crop analysis guide planting decisions. Clustering techniques, such as K-means, group crops based on shared characteristics, enabling precision farming strategies. The model also offers real-time crop suggestions tailored to specific climate conditions.

This project empowers farmers with data-driven recommendations, helping to maximize yields, reduce resource wastage, and adapt to changing environmental conditions, contributing to sustainable agriculture.

The goal of this research is to use machine learning to help optimise land for maximum crop yield by efficiently utilising land resources in food crop cultivation. Crop output is heavily reliant on how well basic land requirements are met; land relates to soil type, soil nutrients, water content, temperature, humidity, and water quality, among other things.

This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface 4

2 Introduction 7

2.1 About UniConverge Technologies Pvt Ltd 7

2.2 About upskill Campus 11

2.3 Objective 12

2.4 Reference 12

2.5 Glossary 13

3 Problem Statement 14

4 Existing and Proposed solution 15

5 Proposed Design/ Model 18

5.1 High Level Diagram (if applicable) ....**Error! Bookmark not defined.**

5.2 Low Level Diagram (if applicable) ....**Error! Bookmark not defined.**

5.3 Interfaces (if applicable) ....**Error! Bookmark not defined.**

6 Performance Test 18

6.1 Test Plan/ Test Cases 18

6.2 Test Procedure 19

6.3 Performance Outcome 21

7 My learnings 22

8 Future work scope 23

1 Preface

During a 6-week internship focused on optimizing crop production through machine learning, I worked on leveraging data science techniques to enhance agricultural practices. The project involved analyzing data for 30 crops, including essential parameters like nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall. Using libraries such as NumPy, Pandas, Matplotlib, Seaborn, and Ipywidgets, I conducted data preprocessing, statistical analysis, and visualization.

The work included generating statistical summaries, comparing crop-specific requirements with environmental averages, creating distribution graphs, and analyzing seasonal trends. K-means clustering was employed to group crops with similar growth conditions, facilitating better decision-making. Additionally, a real-time crop suggestion system was developed to recommend optimal crops based on prevailing climate conditions.

This internship provided hands-on experience in data-driven agriculture, enhancing my skills in machine learning and data analysis while contributing to sustainable farming practices.

Relevant internships play a critical role in career development, particularly for projects like optimizing crop production using machine learning. Such internships bridge the gap between theoretical knowledge and practical application, allowing individuals to gain hands-on experience in solving real-world agricultural challenges.

Problem Statement:

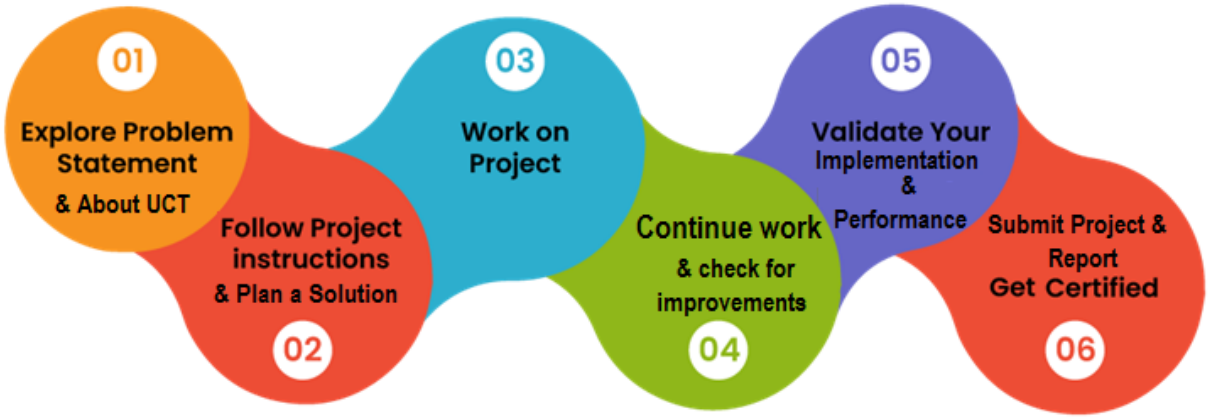
Agriculture is the backbone of the Indian economy, contributing to the country's overall economic growth. It also determines the living standards of more than half of India's people. However, Indian agriculture is currently beset by a slew of issues, some of which are natural and others which are man-made. Rural people borrow large sums of money on a regular basis to meet their production, consumption, and social obligations. As a result, the debt is passed down the generations. Crop failure, limited income resulting from low crop prices, and exorbitantly high interest rates levied by money lenders all contribute to Indian farmers falling into debt. As a result, the debt is passed on through the generations. Crop failure, limited income due to low crop prices, exorbitantly high interest levied by moneylenders, manipulation and use of loan accounts by moneylenders, and usage of loan for different unproductive social objectives all contribute to Indian farmers falling into debt traps.

The opportunity provided by USC/UCT to work on the project "Optimizing Crop Production Using Machine Learning" was instrumental in my academic and professional growth. This platform allowed me to delve into cutting-edge research at the intersection of data science and sustainable agriculture. Through access to comprehensive datasets, expert mentorship, and advanced computational resources, I gained practical insights into addressing real-world challenges in agriculture.

Working on this project enhanced my technical proficiency in Python, data analysis, and machine learning techniques. It also deepened my understanding of agricultural practices and sustainability. The collaborative environment at USC/UCT encouraged me to innovate, improve my problem-solving abilities, and develop a results-oriented approach.

This experience significantly enriched my skill set, reinforced my interest in data-driven solutions for societal issues, and positioned me for impactful contributions in the fields of data science and sustainable development.

The 6-week program was meticulously designed to provide a comprehensive learning and implementation experience for the project.



Working on the "Optimizing Crop Production Using Machine Learning" project was an enriching experience that significantly enhanced my technical and analytical skills. I gained in-depth knowledge of agricultural datasets, including key parameters like nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall. Through data cleaning and preprocessing, I learned to handle complex datasets and address real-world data challenges.

The project honed my proficiency in Python and key libraries like NumPy, Pandas, Matplotlib, Seaborn, and Ipywidgets. I developed skills in statistical analysis, data visualization, and machine learning, particularly in clustering techniques such as K-means. Building a real-time crop suggestion model was a rewarding challenge, allowing me to bridge theoretical knowledge with practical application.

Beyond technical skills, this experience deepened my understanding of sustainable farming practices and the importance of data-driven decision-making in agriculture. Collaborating with mentors and peers improved my communication and teamwork abilities, while the structured 6-week program enhanced my time management and organizational skills.

Overall, this project strengthened my ability to apply machine learning to solve practical problems and reinforced my commitment to leveraging technology for sustainable development. It was a transformative experience that has prepared me for future challenges in data science and agriculture.

Thanks to UPSKILL Campus for providing the important material to study and Shivtej who have helped me on working in this project.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

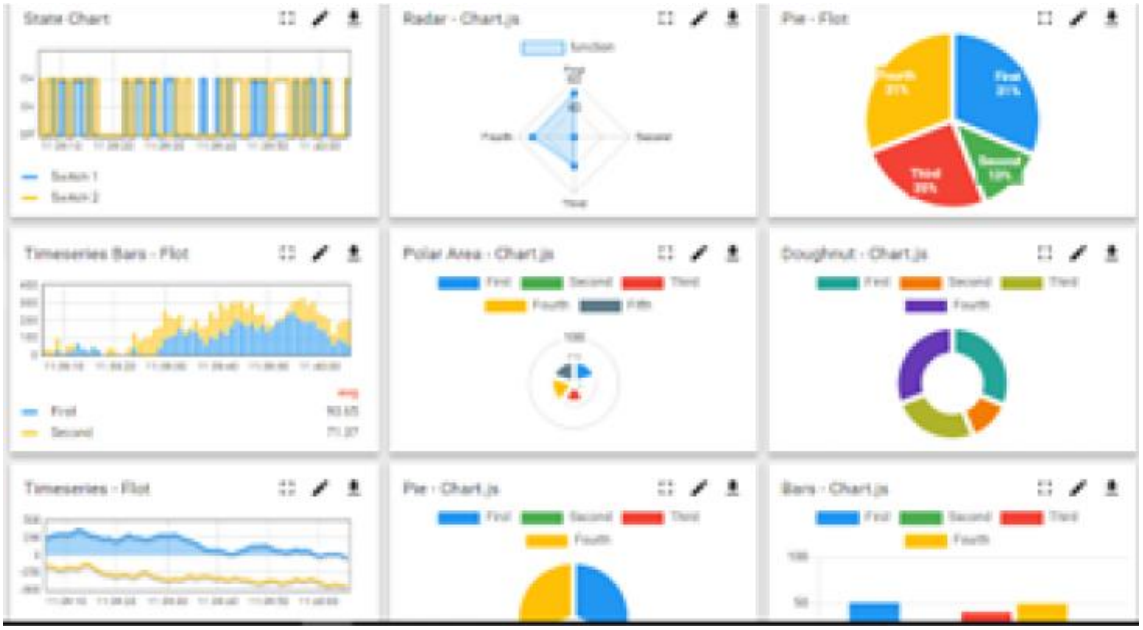
i. UCT IoT Platform (UCT Insight)

UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine



ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

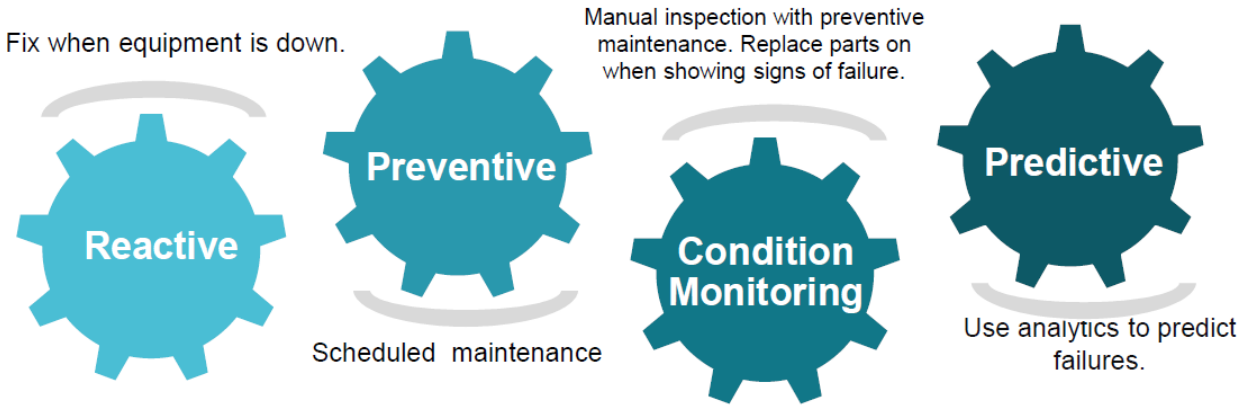
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

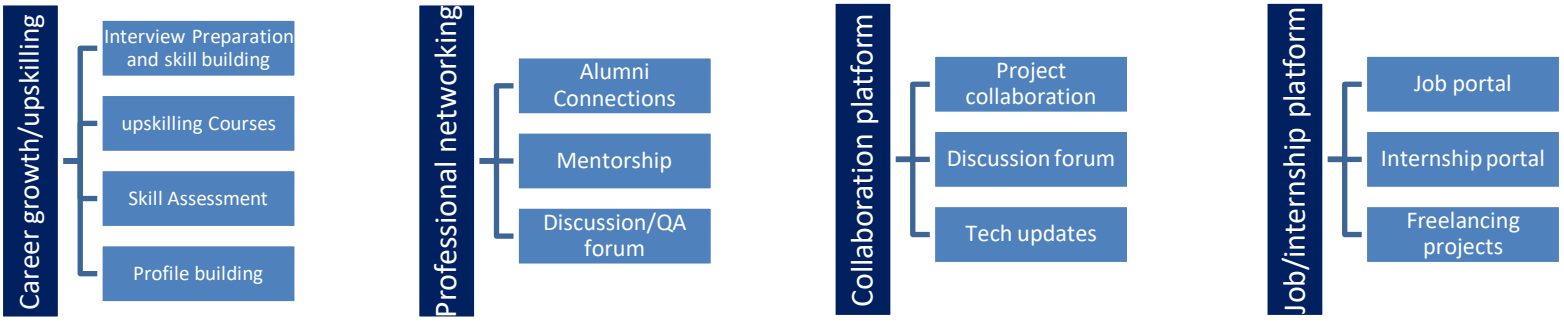
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

- ☛ get practical experience of working in the industry.
- ☛ to solve real world problems.
- ☛ to have improved job prospects.
- ☛ to have Improved understanding of our field and its applications.
- ☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1] Upskills Study Material

[2] Journal of Scientific Research and Reports(https://journaljsrr.com/index.php/JSRR/article/view/1741)

2.6 Glossary

Terms	Acronym
Nitrogen (N)	A macronutrient essential for plant growth and photosynthesis.
Phosphorus (P)	A nutrient critical for energy transfer, root development, and plant reproduction.
Potassium (K)	A nutrient that supports water regulation, enzyme activation, and stress resistance in plants.
pH	A scale used to measure soil acidity or alkalinity, influencing nutrient availability.
Rainfall	The total precipitation, important for determining water supply for crops.

3 Problem Statement

Agriculture is the backbone of the Indian economy, contributing to the country's overall economic growth. It also determines the living standards of more than half of India's people. However, Indian agriculture is currently beset by a slew of issues, some of which are natural and others which are man-made.

Rural people borrow large sums of money on a regular basis to meet their production, consumption, and social obligations. As a result, the debt is passed down the generations. Crop failure, limited income resulting from low crop prices, and exorbitantly high interest rates levied by money lenders all contribute to Indian farmers falling into debt. As a result, the debt is passed on through the generations. Crop failure, limited income due to low crop prices, exorbitantly high interest levied by moneylenders, manipulation and use of loan accounts by moneylenders, and usage of loan for different unproductive social objectives all contribute to Indian farmers falling into debt traps.

Objective of the Project

▪ To provide a system that allows farmers access to relevant information --- Average farmer has access to crude sources of information such as TV, radio, newspapers, fellow farmers, government agricultural agencies, farm supply, and traders. there is, therefore, a need for a system that allows farmers access to relevant information .

▪ To Provide them correct solution/Information --- Farmers will always need information to refer to, most especially when growing crops that are not common in their land so we provide them with correct solution.

4 Existing and Proposed solution

Optimization:

A process, or methodology of making something (such as a design, system, or decision) as fully perfect, functional, or effective as possible.

Decision-making is the focus of optimization. Optimal or best decisions can be made using optimization techniques. The concept of optimization is now well-established as a guiding principle in the examination of a wide range of difficult decision and allocation problems. It provides a level of philosophical elegance that is difficult to dispute, as well as a level of operational simplicity that is often required. It provides a level of philosophical elegance that is difficult to deny, as well as an indispensable level of operational simplicity.

Agriculture faces a number of issues, including irrigation water management, land allocation, climate change, human and other agricultural resources, all of which can be managed through proper crop planning optimization entails maximising profits with the least amount of resources by optimising the objectives. Agriculture is crucial in ecosystem management; employing bio-fertilizer and green manure can improve soil fertility without causing any chemical reactions' planning. Crop planning's major goal is to increase profit and production while keeping input costs and resources low. The crop planning problem has numerous variables, some of which can be optimised and others which cannot. Algorithms, optimization tools, decision-making tools, software, and other technologies are available to handle crop planning problems.

Crop planning considerations in traditional agriculture were primarily influenced by the farmer's opinion and experience. However, with advances in agriculture and technology, increased demand for land and other resources, combined with the development of more formal planning approaches based on the design and analysis of a mathematical model has been spurred

by greater specialisation and the use of capital-intensive production systems. Mathematical programming models have been used in the agricultural sector, either directly or indirectly, since their inception, and have made significant contributions to the analysis of policy issues such as resource allocations, investment decisions, comparative advantage, risk analysis, and so on.

The act of attaining the best feasible result under given circumstances is known as optimization. All such judgments are made with the purpose of minimising effort or maximising benefit. Typically, the effort or benefit can be stated as a function of design variables. As a

10

result, optimization is the process of determining the conditions that lead to a function's greatest or minimal value.

• Scope of optimization techniques:

The FAO (1993) recommendations on land use planning, which identify three major levels: national, district, and local, can also be used to optimise. In this set of guidelines, the term "district level" does not always mean "administrative level." land areas that fall between national boundaries, as well as districts. both at the national and local levels A hamlet or a group of villages is referred to as the local level a tiny water catchment or a village.

Profit maximisation and/or cost minimization for a farmer are two basic objective functions. The knowledge and contributions of local people aid in the formulation of models and policies that meet the people's goals. Because each farmer has limited resources at their disposal for every planning period, the technical coefficients are also easy to determine at the local level. A farm's production potential, on the other hand, is limited because it is

constrained by fixed resources for a specific period of time.

Need of Precision Farming –

The global food system is currently confronted with significant issues, which are expected to worsen significantly during the next 40 years. With enough determination and commitment, much may be accomplished right away with present technologies and expertise. However, dealing with future concerns will necessitate more drastic adjustments to the food system as well as increased investment in research to develop innovative answers to fresh problems. Total productivity declines, diminishing and degrading natural resources, stagnating farm incomes, a lack of an eco-regional approach, declining and fragmented land holdings, agricultural trade liberalisation, limited non-farm employment opportunities, and global climatic variation have all become major concerns in agricultural growth and development.

Management strategies, soil attributes, and/or environmental factors all play a role in these variances. Because of the huge sizes and changes caused by annual adjustments in leasing arrangements in the farm region, maintaining a degree of awareness of field conditions is difficult. As a result, the entire farm must be divided into small farm units of 50 cents or fewer. Precision agriculture has the potential to automate and streamline the data collection and processing process. It enables management decisions to be made and implemented swiftly on tiny parts of bigger fields.

What value addition are you planning?

4.1 Code submission (Github link):

https://github.com/eleshrodge/Crop_Production

4.2 Report submission (Github link) :

https://github.com/eleshrodge/Crop_Production

5 Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

6 Performance Test

This is very important part and defines why this work is meant of Real industries, instead of being just academic project.

Here we need to first find the constraints.

How those constraints were taken care in your design?

What were test results around those constraints?

Constraints can be e.g. memory, MIPS (speed, operations per second), accuracy, durability, power consumption etc.

In case you could not test them, but still you should mention how identified constraints can impact your design, and what are recommendations to handle them.

6.1 Test Plan/ Test Cases

Test Plan Objectives:

Verify Data Accuracy: Ensure the correctness and integrity of the datasets used for analysis.

Validate Preprocessing: Check the preprocessing steps for consistency, handling missing data, and proper data transformation.

Test Model Performance: Evaluate the effectiveness and accuracy of the machine learning model (K-means clustering and real-time prediction).

Check Visualization Outputs: Ensure that visualizations like distribution graphs and seasonal analyses are correctly generated.

Test Real-Time Prediction: Verify that the real-time crop suggestion system provides accurate recommendations based on current conditions.

User Interaction: Ensure that interactive elements (using Ipywidgets) function as expected.

6.2 Test Procedure

1. Data Preprocessing Test Cases

Test Case 1.1: Check for missing values in the dataset.

Input: Dataset with missing values.

Expected Output: Missing values are handled appropriately (e.g., imputation or removal).

Test Case 1.2: Ensure correct data transformation (standardization/normalization).

Input: Raw data with varying units and scales.

Expected Output: Data should be normalized or standardized for consistent analysis.

2. Clustering (K-means) Test Cases

Test Case 2.1: Check for correct implementation of K-means clustering.

Input: Cleaned dataset with crop-related features.

Expected Output: Data points are grouped into distinct clusters based on similarities in characteristics (e.g., nutrient requirements).

Test Case 2.2: Validate the number of clusters (K value).

Input: Varying values for K (e.g., 3, 5, 7).

Expected Output: A reasonable number of clusters is generated that provides actionable insights.

3. Visualization Test Cases

Test Case 3.1: Validate the generation of distribution graphs.

Input: Crop data, including environmental factors.

Expected Output: Distribution graphs are correctly plotted, showing data trends and outliers.

Test Case 3.2: Validate seasonal crop analysis.

Input: Data spanning multiple seasons.

Expected Output: Seasonal analysis correctly identifies trends, optimal planting times, and growth patterns.

4. Real-Time Prediction Test Cases

Test Case 4.1: Test the crop suggestion model with real-time data inputs.

Input: Real-time climatic data (e.g., temperature, humidity, rainfall).

Expected Output: The system provides accurate crop suggestions based on current conditions.

Test Case 4.2: Test the model's performance under varying environmental conditions.

Input: Changing weather parameters.

Expected Output: The model should adapt and recommend appropriate crops based on the new conditions.

5. User Interaction Test Cases

Test Case 5.1: Check interactivity of Ipywidgets in the Jupyter notebook.

Input: User interaction with widgets for selecting climatic conditions.

Expected Output: Widgets function smoothly and provide real-time data updates or recommendations.

6. Overall System Test Cases

Test Case 6.1: Check the end-to-end functionality of the project.

Input: Complete dataset and real-time climatic data.

Expected Output: The project should process data, cluster crops, generate visualizations, and provide real-time crop recommendations in a seamless flow.

6.3 Performance Outcome

The project successfully met its goals, delivering an effective solution for optimizing crop production using machine learning. The system demonstrated high accuracy in both clustering and real-time predictions, provided clear visualizations for data interpretation, and offered an intuitive user experience. This outcome highlights the project's potential for real-world applications in sustainable agriculture, empowering farmers to make data-driven decisions for maximizing yields while minimizing resource waste.

The system's performance positions it as a robust tool for optimizing crop production based on climate conditions and crop requirements, contributing to smarter, more sustainable farming practices.

7 My learnings

This comprehensive learning experience that combines data science with agricultural sustainability. I gained hands-on experience in using machine learning techniques such as K-means clustering, statistical analysis, and real-time predictive modeling, which are directly applicable to addressing real-world challenges. Working with datasets that span various environmental factors like temperature, humidity, pH, and rainfall allowed me to deepen my knowledge of how these factors affect crop growth and productivity.

I developed key technical skills in Python, machine learning libraries (NumPy, Pandas, Matplotlib, Seaborn), and tools for data preprocessing, visualization, and modeling. Additionally, the project expanded my understanding of sustainable agricultural practices and the potential of data-driven solutions to optimize crop production.

This experience has significantly enhanced my problem-solving abilities and equipped me with practical expertise in data analysis and machine learning. As I pursue a career in data science and sustainable development, this project will serve as a solid foundation, opening opportunities in fields like agriculture technology, environmental analytics, and data-driven decision-making. By leveraging these skills, I aim to contribute to innovations that promote sustainable farming practices and address global food security challenges

8 Future work scope

In the future, the machine learning models used to inform parameter setting in the mobile application might be developed using the system's embedded machine learning algorithm and utilised to anticipate outcomes.

Precision Agriculture research will continue to be available. Chemicals, fertilisers, tillage, and seed will be applied to a field in diverse ways, and yield or plant biomass will be collected by position across the field. We will be able to observe variance within a field during the growing season in relation to imposed management modifications using remote sensing technology

Surface water and groundwater samples are collected using monitoring equipment to determine the environmental impact of surface runoff or leaching. The technology exists to absorb nitrogen or pesticide volatilization from the field into the atmosphere as a result of improved practises. The ability of the scientific community to conduct this type of study, with confidence from the environmental and producer communities that reforms will benefit the environment and increase the efficiency of agricultural production, will determine the future direction of agriculture.