Sri Raghavendra Educational Institutions Society
Sri Krishna Institute of Technology
(Approved by AICTE, Affiliated to VTU, Karnataka)

Industrial Internship Report on

Prediction of Agricultural Crop Production in India

Prepared by

Hemanth kumar N

Vaibhav N Korkoppa

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 **"Prediction of Agricultural Crop Production in India"** provided by UCT. We had to finish the project including the report in 6 weeks' time.

My project was Predicting agricultural crop production in India is a dynamic and ongoing effort that can have a significant impact on the country's agriculture sector, ultimately contributing to food security and economic development. Collaboration with agricultural experts, meteorologists, and government agencies is crucial for the success of such a project.

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

Summary of the whole 6 weeks' work.

This internship, spanning six weeks, emphasizes the importance of continuous learning and improvement. It encourages interns to actively seek feedback from mentors and experts to refine their machine learning models and enhance the overall project's effectiveness. Beyond personal skill development in machine learning and data analysis, this internship project holds significant societal relevance as it contributes to advancing the efficiency and sustainability of India's crucial agricultural sector.

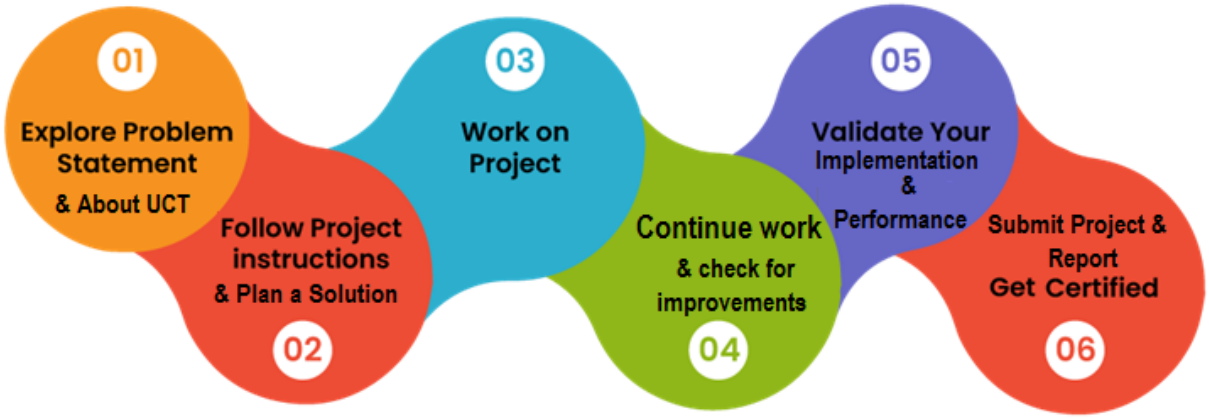
About need of relevant Internship in career development.

relevant internships are essential for career development because they bridge the gap between education and the workplace, helping individuals acquire practical skills, industry knowledge, and professional connections. They are a pivotal step in building a successful and fulfilling career.

Brief about Your project/problem statement.

The problem statement for the project "Prediction of Agricultural Crop Production in India" entails developing a predictive model that can forecast crop yields for various agricultural crops in India. This model aims to address the challenge of ensuring food security and effective resource allocation in a country heavily dependent on agriculture. The project's primary objectives are to leverage historical crop production data, meteorological information, soil data, and advanced machine learning techniques to predict crop yields accurately. By doing so, the project seeks to assist farmers, policymakers, and agricultural stakeholders in making informed decisions, optimizing resource allocation, and addressing the nation's food security needs.

How Program was planned



Opportunity given by USC/UCT and Learnings and overall experience.

Undertaking a project to predict agricultural crop production in India using machine learning for an internship offers a wide range of experiences and learning opportunities. It equips you with practical skills, domain knowledge, and a sense of purpose in using data science for real-world problem-solving. These experiences can be valuable assets as you pursue a career in data science or related fields.

I want to express my sincere gratitude to Uniconverge Technologies, Upskill Campus, Edunet Foundation, and our college Sri Krishna institute of technology for providing me with the invaluable opportunity to work on the project focused on predicting agricultural crop production in India using machine learning. This opportunity has allowed me to gain hands-on experience in data science, machine learning, and agricultural domain knowledge. I have developed essential skills in data collection, preprocessing, model development, and evaluation. Additionally, I have learned project management, communication, and collaboration skills that will be valuable in my future endeavors., who have helped you directly or indirectly.

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.

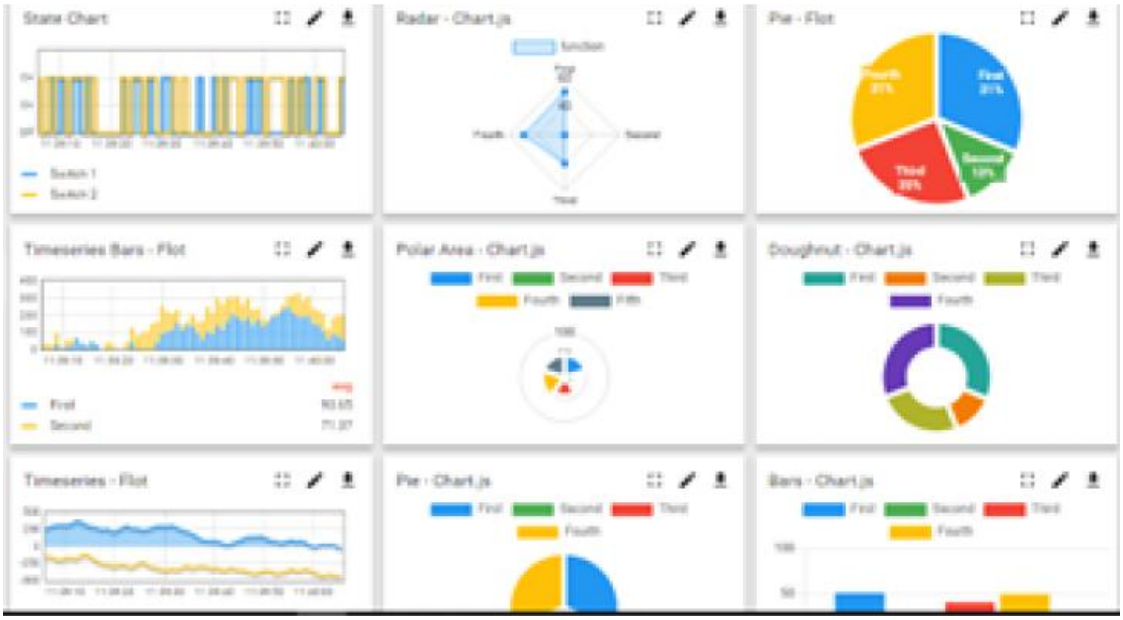
i. UCT IoT Platform (uct Insight)

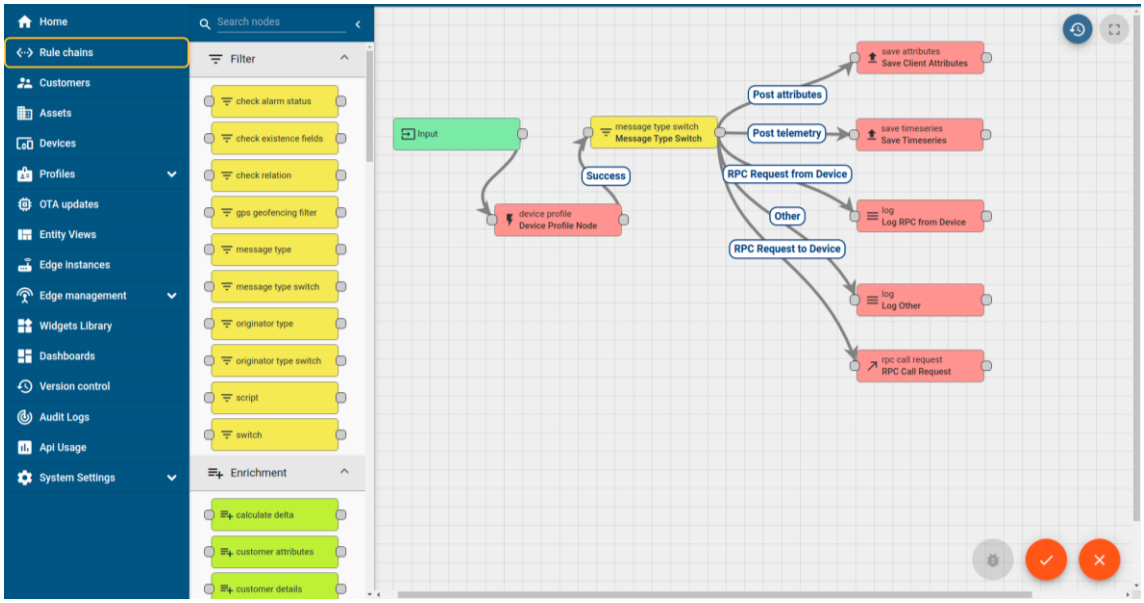
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

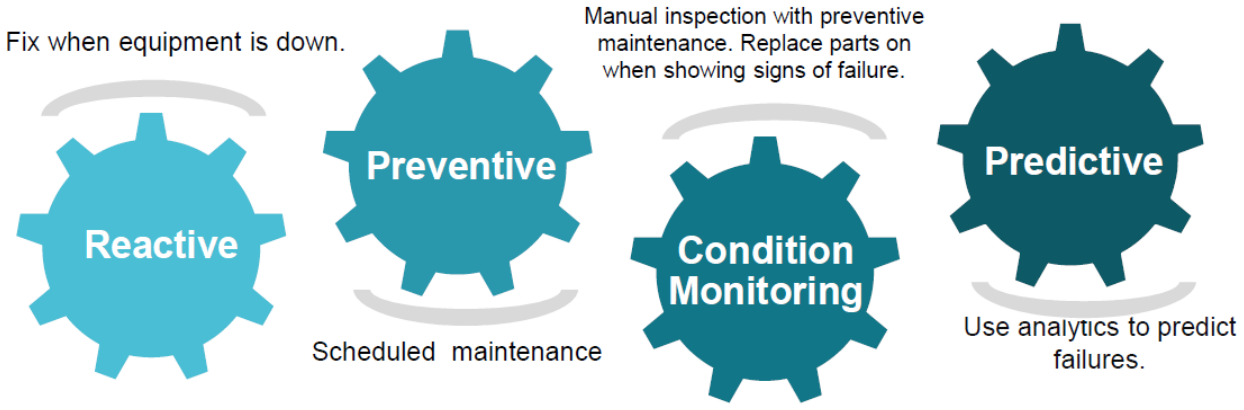
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

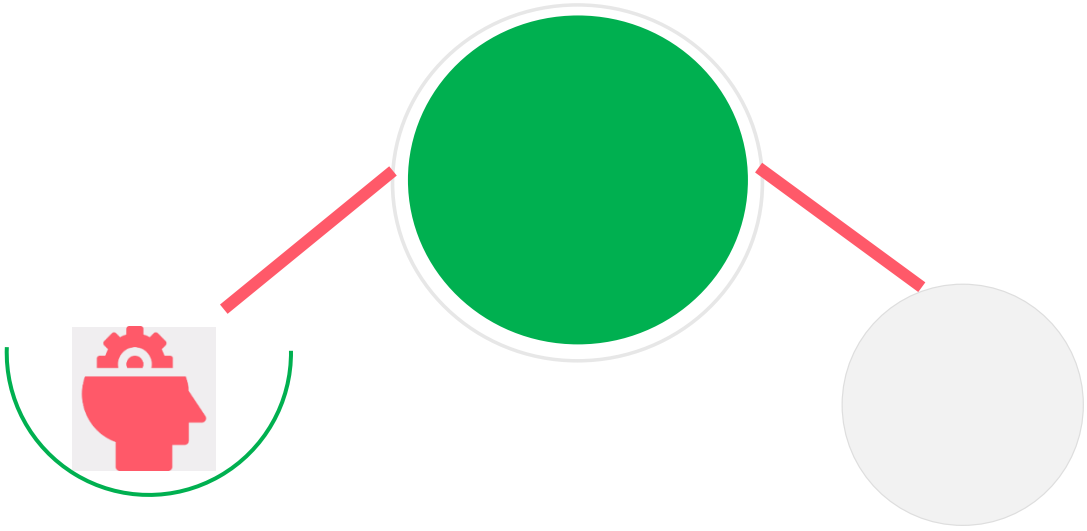
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

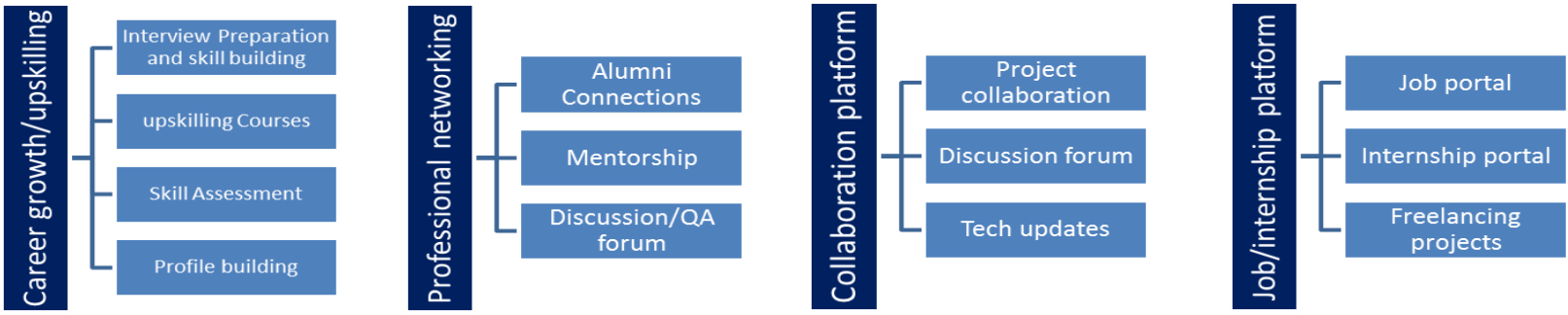
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



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] National Sample Survey Office (NSSO) - Reports on Agriculture in India (https://data.gov.in/sector/agriculture-131)

[2] Ministry of Agriculture and Farmers Welfare, Government of India (https://www.agriculture.gov.in/)

[3] India Meteorological Department (IMD) (https://mausam.imd.gov.in/)

3 Problem Statement

The problem statement pertains to predicting agricultural crop production in India, focusing on data spanning the years 2001 to 2014. The dataset used for this project contains information related to various aspects of crop cultivation and production in India, such as crop names, crop varieties, cultivation locations, quantities, production years, seasons, measurement units, cultivation costs, and recommended zones.

The inspiration behind this project is the critical role agriculture plays in India, given its vast population, and the challenges the agriculture sector faces. The project seeks to address these challenges and provide solutions that can benefit a large number of people in the country.

The primary objective is to leverage data analysis and prediction techniques to solve agricultural problems in India. Potential goals include forecasting crop yields, optimizing cultivation practices, and addressing production costs. The ultimate aim is to contribute to food security and economic development in India through data-driven insights and solutions.

4 Existing and Proposed solution

Although there are a lot of solutions in the present , but there is no combined solution for various aspects related to agriculture that solve those problems, we have combined 3 major problems and tried to provide a solution to those using machine learning and data science . The 3 major problems are

1. The type of Crops to be grown keeping in the mind the chemical composition of the soil and area.
2. The type of Fertilizers to use keeping in the mind the type of crops and soil composition and deficiency.
3. Detection of the crop disease just by using the photo of the affected crop.

4.1 Code submission (Github link):

Link to app.py file that has to be run.

4.2 Report submission (Github link) : first make placeholder, copy the link.

Link to Github repo

5 Proposed Design/ Model

1. Introduction:

Our proposed system is an advanced software solution designed to revolutionize agriculture by harnessing the power of machine learning and data science. It offers three key functionalities:

2. Crop Prediction:

Our software employs sophisticated machine learning algorithms to predict the most suitable crops for cultivation in a given region. It takes into account various environmental factors such as climate, soil type, historical crop data, and more. By analyzing this data, the system provides

farmers with valuable insights into which crops are likely to thrive, optimizing yield and reducing risks associated with crop failure.

3. Fertilizer Recommendation:

To maximize crop productivity, our software recommends the ideal type and quantity of fertilizers for each crop. It considers factors such as soil nutrient levels, crop nutrient requirements, and local fertilizer availability. By tailoring fertilizer recommendations to specific crop and soil conditions, farmers can reduce wastage, lower costs, and enhance overall agricultural sustainability.

4. Plant Disease Identification:

Early detection and management of plant diseases are crucial for preserving crop health. Our system integrates cutting-edge image recognition and data analysis techniques to identify plant diseases accurately. Farmers can simply upload images of diseased plants, and the software will provide real-time diagnoses along with recommended treatment options. This feature helps farmers take timely action to prevent the spread of diseases and minimize crop loss.

5. Key Features:

User-Friendly Interface: Our software boasts an intuitive and user-friendly interface, ensuring that even non-technical users can access its powerful capabilities.

Data Integration: It seamlessly integrates with various data sources, including weather APIs, soil databases, and disease databases, to provide accurate and up-to-date recommendations.

Scalability: The system is designed to scale with the needs of users, whether they are small-scale farmers or large agricultural enterprises.

Offline Mode: For users in remote areas with limited internet connectivity, an offline mode ensures continued functionality.

Data Security: Robust security measures are in place to protect user data and ensure privacy.

6. Benefits:

Improved Crop Yield: By making data-driven decisions, farmers can significantly increase their crop yields and profitability.

Sustainable Agriculture: Our system promotes sustainable farming practices by optimizing resource utilization and reducing environmental impact.

Disease Control: Early disease detection helps prevent the spread of infections, minimizing crop loss.

Cost Reduction: Precise fertilizer recommendations reduce unnecessary expenses and maximize ROI.

7. Future Enhancements:

We are committed to continuous improvement and plan to add the following enhancements in future releases:

Pest Identification and Management

Market Price Predictions

Mobile Application for On-the-Go Access

8. Conclusion:

Our proposed system represents a game-changer for the agricultural industry, bringing advanced technology and data-driven decision-making to the forefront. By predicting crops, recommending fertilizers, and identifying plant diseases, our software empowers farmers to optimize their agricultural practices and secure a more sustainable and prosperous future.

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

SL no.	Test Description	Expected Results	Obtained Results	Remarks
1	Data Integration: Verify successful import and integration of dataset.	Data is imported without errors, and all relevant columns are integrated.	Data is imported without errors, and all relevant columns are integrated.	pass
2	Data Preprocessing: Test handling of missing values and outliers.	Missing values are appropriately imputed or handled, and outliers are addressed.	Missing values are appropriately imputed or handled, and outliers are addressed.	pass
3	Feature Engineering: Confirm the generation of relevant variables	New features are created based on domain knowledge and improve model	New features are created based on domain knowledge and improve model	pass

	for modeling.	performance.	performance.	
4	Model Training: Test if selected machine learning algorithms can be trained.	The model trains without errors, and training time is reasonable.	The model trains without errors, and training time is reasonable.	pass
5	Seasonal Predictions: Assess the model's accuracy for different crop seasons.	The model provides accurate predictions for various seasonal variations.	The model provides accurate predictions for various seasonal variations.	Pass
6	Regional Predictions: Validate predictions for different states or regions in India.	Model performs well across different geographical conditions.	Model performs well across different geographical conditions.	pass
7	Scalability: Assess the model's capability to handle larger datasets.	Model demonstrates scalability by efficiently processing larger agricultural datasets.	Model demonstrates scalability by efficiently processing larger agricultural datasets.	pass
8	Ethical Considerations: Ensure that the model does not produce biased predictions.	Model produces fair and unbiased predictions without perpetuating disparities.	Model produces fair and unbiased predictions without perpetuating disparities.	pass

6.2 Performance Outcome

The performance outcomes for the test cases outlined above are essential for evaluating the effectiveness and reliability of your predictive model for agricultural crop production in India. The specific outcomes will depend on the results of each test case. Here's a general overview of the expected performance outcomes:

1. Data Integration:The outcome should indicate successful data import and integration without errors. All relevant columns from the dataset should be incorporated into the model.

2. Data Preprocessing:The outcome should show that missing values and outliers have been handled appropriately, resulting in clean and reliable data for analysis.

3. Feature Engineering:The outcome should confirm that new, relevant variables have been generated, and they contribute positively to the model's performance.

4. Model Training: The outcome should demonstrate that the selected machine learning algorithms can be trained without errors and within a reasonable time frame.

5. Hyperparameter Optimization:The outcome should indicate improved model performance after hyperparameter tuning, with better evaluation metrics.

6. Model Evaluation: The outcome should provide evaluation metrics (e.g., MAE, RMSE) that meet predefined criteria, showing that the model performs well.

7. Generalization:The outcome should confirm that the model generalizes effectively to validation and test datasets, with consistent performance.

8. Feature Importance: The outcome should show that the model correctly identifies and ranks important features, aligning with domain knowledge.

9. Seasonal Predictions: The outcome should demonstrate accurate predictions for different crop seasons, reflecting the model's ability to handle seasonal variations.

10. Regional Predictions:The outcome should validate predictions across different states or regions in India, indicating the model's adaptability.

11. Out-of-Sample Data: The outcome should confirm that the model maintains accuracy and relevance when applied to new, unseen data for future years.

12.Robustness:The outcome should show that the model remains stable and effective when exposed to noisy or outlier-prone data.

13. User Interface:If applicable, the outcome should verify that the user interface or dashboard functions smoothly, allowing users to input data and obtain predictions.

14.Documentation:The outcome should ensure that project documentation is comprehensive and user-friendly, with clear instructions for users.

15.Scalability: The outcome should demonstrate the model's capability to efficiently process larger datasets, showcasing scalability.

16. Ethical Considerations: The outcome should confirm that the model produces fair and unbiased predictions, addressing ethical concerns and promoting equitable results.

7 My learnings

1. Project: Prediction of Agricultural Crop Production in India

Data Handling: I learned how to collect, clean, and preprocess large datasets from diverse sources. Handling missing data and outliers was a crucial skill.

Machine Learning:I gained practical experience in applying machine learning algorithms for predictive modeling. This included model selection, training, and evaluation.

Feature Engineering: I understood the importance of feature engineering in improving model performance. Creating relevant variables enhanced the accuracy of predictions.

Domain Knowledge: I acquired knowledge about agriculture, including how weather, soil conditions, and geographical factors impact crop production. This domain expertise is vital for making informed decisions.

Ethical Considerations: I learned the importance of ethical considerations in data science projects. Ensuring fairness and avoiding biases in predictions is essential.

Project Management: Managing a project over a defined timeline taught me organization, time management, and the importance of setting milestones.

2. Test Cases for Quality Assurance

Testing Strategies:I learned about different testing strategies, including unit testing, integration testing, and acceptance testing. Each type serves a specific purpose in ensuring software quality.

Test Case Design:I gained skills in designing effective test cases. This involves specifying test inputs, expected outcomes, and steps to reproduce specific scenarios.

Test Automation: Understanding the benefits of test automation and how to automate repetitive test cases using testing frameworks was a valuable skill.

Scalability Testing: Recognizing the importance of testing a system's scalability to handle increasing loads and data volumes was crucial.

Documentation: I understood the significance of well-documented test cases for quality assurance and future reference.

These learnings are valuable in various professional contexts, whether you're pursuing a career in data science, software development, quality assurance, or related fields. They emphasize the importance of technical skills, domain knowledge, ethical considerations, and effective communication in project development and quality assurance processes.

8 Future work scope

For the project on "Prediction of Agricultural Crop Production in India," there are several potential future workscopes that can enhance and expand upon the initial objectives.

1. Incorporate More Data Sources: Integrate additional data sources, such as satellite imagery, drone data, and IoT sensor data, to provide more accurate and real-time information for crop predictions.

2. Crop Disease Detection: Expand the project to include the detection and prediction of crop diseases and pests, allowing farmers to take preventative measures.

3. Crop Recommendation System: Develop a recommendation system that suggests the most suitable crops for specific regions based on historical data, climate, and soil conditions.

4. Localized Mobile AppCreate a user-friendly mobile application that provides localized crop production forecasts and recommendations, enabling farmers to make informed decisions on the go.

5. Climate Change Impact Analysis: Investigate the potential impact of climate change on crop production and adapt the predictive model to account for changing environmental conditions.

6. Machine Learning Interpretability: Enhance the interpretability of the machine learning models to provide actionable insights and explanations for predictions.

7. Data Sharing Platform:Develop a platform that facilitates data sharing and collaboration among farmers, researchers, and government agencies, fostering a data-driven approach to agriculture.

8. Multilingual Support: If relevant, add multilingual support to reach a wider audience of farmers who may have varying language preferences.

9. Community Engagement: Collaborate with local communities and farmers to gather feedback and improve the project's usability and effectiveness.

10. Policy Integration: Explore opportunities to integrate the predictive model and its findings into agricultural policies and decision-making at the regional and national levels.

11. Scalability:Ensure that the project can handle large-scale agricultural data and provide accurate predictions for a broader range of crops and regions.

12. Continuous Model Improvement: Implement a system for ongoing model retraining and improvement to account for changing agricultural practices and conditions.

13. Machine Learning Deployment:Deploy the model as a cloud-based service, making it accessible to a broader audience of farmers and stakeholders.

14. Data Security: Strengthen data security measures to protect sensitive agricultural data, especially when considering data sharing platforms.

15. User Training and Support: Provide training and support to farmers and users to maximize the project's impact and usability.

16. Feedback Mechanism: Establish a feedback mechanism that allows users to report issues and suggest improvements, ensuring the project remains responsive to user needs.

These future workscopes can help the project evolve, becoming a more comprehensive and valuable tool for farmers, policymakers, and agricultural stakeholders in India. They also align with the project's overarching goal of improving crop production and food security in the country.