Industrial Internship Report on:

"Analysis of Production of crops
In india"

Prepared by :

1. Shubhada Kanade
2. Pratiksha Mahajan
3. Chaitrali Kawade

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy incollaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time. My project was on analysis production of crops in india. 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 whole 6 weeks:

on week 1, the main emphasis was on grasping usc_tia and actively participating in python projects. furthermore, there was a commitment to utilizing educational materials to improve skills.

On week 2, after a two-week break, we resumed our project with renewed focus and diligence, adhering closely to project instructions to master foundational data analysis techniques. We successfully acquired licensed data from government websites through UCT's facilitation, and conducted extensive research to clarify and refine our project's objectives and solution strategy.

On week 3, we continued our focused efforts on mastering foundational data analysis techniques according to project guidelines. We successfully gathered licensed data from government websites with UCT's assistance, refining our data collection methods and conducting thorough data cleaning. Additionally, we analyzed Excel data files and explored various techniques, including merging and concatenation, to enhance our research and project strategy.

During Week 4 of my data science internship, I focused on cleaning, analyzing, and visualizing a dataset related to produce. The main objectives were to ensure data quality, derive meaningful insights from the dataset, and create visualizations to interpret the data effectively.

During 5th week, I focused on developing an interactive dashboard using Power BI Desktop. This dashboard will provide a dynamic and user-friendly way to explore the

produce dataset and its insights.

About need of relevant Internship in career development.

An internship in data science and analysis is crucial for career development because it provides practical experience, enhances technical and problem-solving skills, offers industry insights, helps build a professional network, and strengthens your resume. It bridges the gap between academic learning and professional expertise, making you more competitive in the job market.

Brief about Your project/problem statement.

The problem statement "Production of crops in India" involves analyzing agricultural data to understand trends, identify key factors affecting crop yields, and develop predictive models. The goal is to improve crop production by identifying challenges and suggesting optimal farming practices and policy recommendations. This analysis can help enhance productivity, ensure food security, and support sustainable agriculture in India.

Your Learnings and overall experience.

Key Learning

- Data Handling: Improved skills in data collection, cleaning, and preprocessing.

- EDA: Enhanced ability to perform exploratory data analysis and identify patterns.

- Modeling:Gained experience with machine learning algorithms and model evaluation.

- Communication: Developed proficiency in reporting and presenting findings.

Overall Experience:

- Practical Application:Applied data science techniques to real-world agricultural problems.

- Collaboration: Worked effectively in a team and with stakeholders.

- Problem-Solving:Tackled complex challenges and developed data-driven solutions.

- Domain Knowledge:Gained insights into agricultural practices and their data-driven improvements.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

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

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

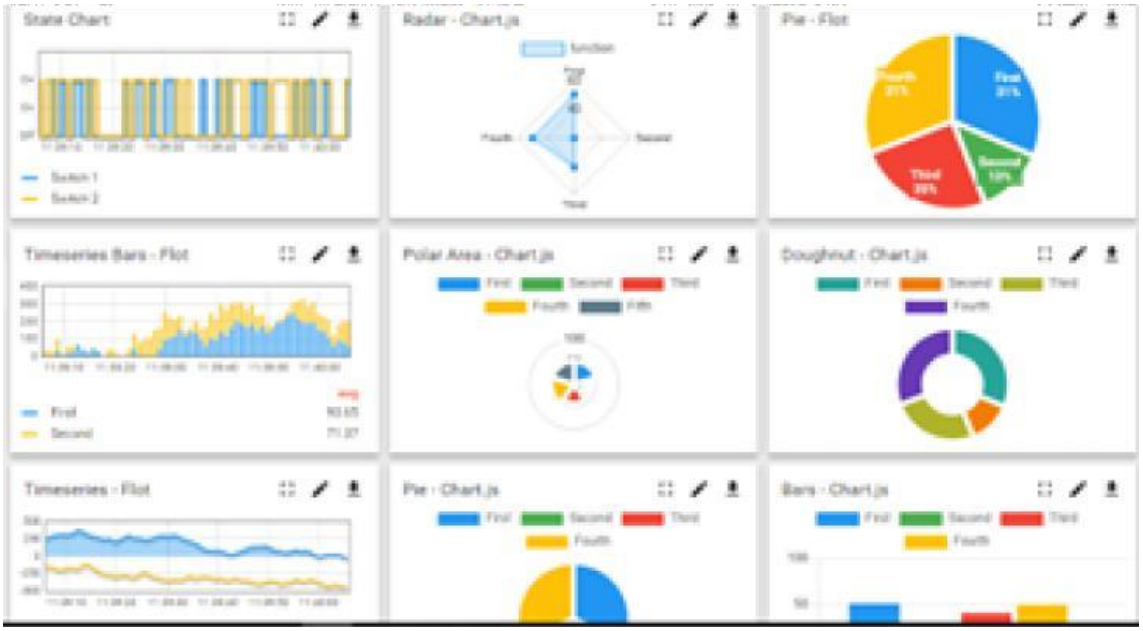
i. UCT IoT Platform (UCT Insight)

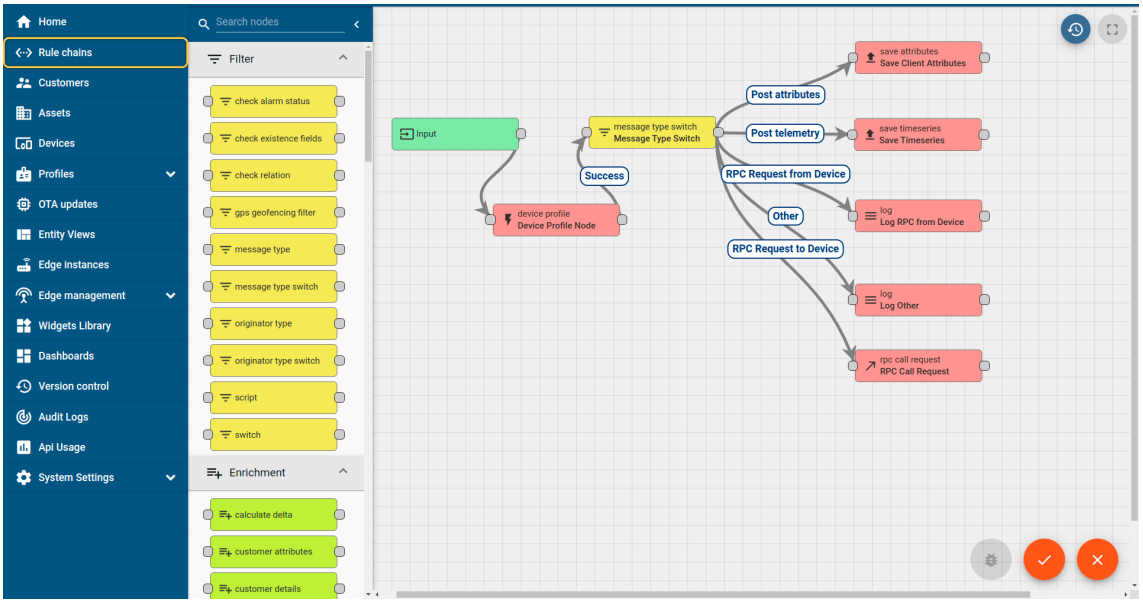
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, ModbusTCP, 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 toidentify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and thencan scale to more complex solutions as per their demands.

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

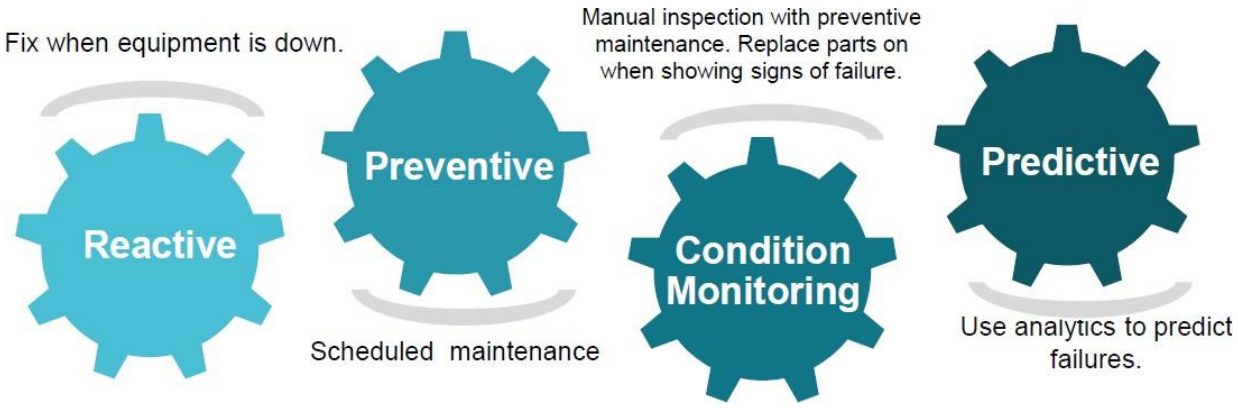
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

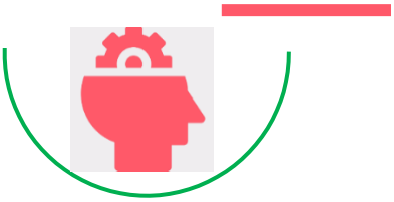
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

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

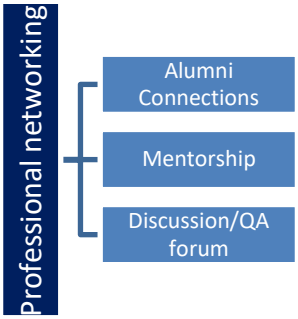
USC is a career development platform that delivers **personalized executive coaching** in amore affordable, scalable and measurable way.

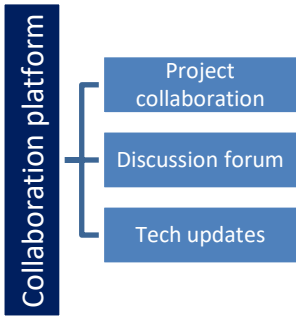


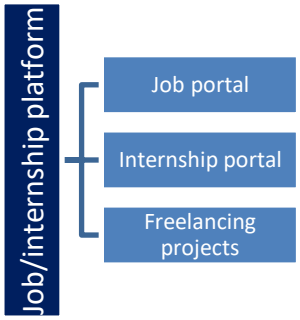
Seeing need of upskilling in selfpaced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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 incollaboration 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] Online Tutorials : Utilized resources from platforms like GeeksforGeeks and YouTube for additional guidance.

[2]Microsoft Learn-Free, interactive,and self-placed training modules on Power BI.

[3]Youtube Vedios:for developing an interactive and informative dashboard

3 Problem Statement:

The problem statement "Production of crops in India" for a data science and analysis internship typically involves analyzing agricultural data to gain insights into crop production trends, challenges, and opportunities. Here's a breakdown of what this problem statement could entail:

1. Understanding Crop Production Trends:

- Analyze historical data to identify trends in crop production over the years.
- Compare production levels of different crops across various regions in India.

2. Identifying Key Factors:

- Determine the key factors influencing crop yields, such as weather conditions, soil quality, irrigation practices, use of fertilizers, and pest control measures.
- Evaluate the impact of government policies, subsidies, and agricultural practices on crop production.

3. Predictive Analysis:

- Develop models to predict future crop yields based on historical data and current conditions.
- Use machine learning algorithms to forecast the impact of various factors on crop production.

4. Optimization:

- Identify optimal farming practices that can enhance crop yields.
- Suggest strategies to improve productivity and sustainability in agriculture.

5. Challenges and Solutions:

- Identify the major challenges faced by farmers, such as climate change, water scarcity, and pest infestations.

- Propose data-driven solutions to mitigate these challenges.

Expected Outcomes:

- Comprehensive Reports: Detailed reports on crop production trends, challenges, and recommendations.
- Predictive Models:*Reliable models to forecast future crop yields and identify potential risks.
- Policy Recommendations: Insights and suggestions for policymakers to improve agricultural productivity and sustainability.
- Actionable Insights for Farmers: Practical advice and strategies for farmers to optimize their crop production.

4 Existing and Proposed solution

Existing Solutions:

1.Predictive Models:

- Solutions: Machine learning models like regression, random forests, and neural networks predict crop yields based on historical data, weather patterns, and soil conditions.
- Limitations: Often require extensive data and can be sensitive to data quality. Models may not generalize well to new conditions or regions.

2.Optimization Algorithms:

- Solutions:Algorithms to optimize irrigation, fertilization, and pest control to improve yields and reduce waste.
- Limitations: Can be challenging to implement in diverse and dynamic real-world settings. Solutions may require substantial adaptation to local conditions.

3.Decision Support Systems (DSS):

- Solutions: Tools that integrate various data sources to provide recommendations for farming practices and resource allocation.
- Limitations: May lack real-time data integration and can be limited by the quality of input data and the complexity of user interfaces

Limitations of Existing Solutions:

- Data Quality and Availability:Many models and systems rely on high-quality, extensive datasets that may not always be available or accurate.
- Complexity and Cost: Advanced technologies and algorithms can be costly and complex to implement, especially in resource-limited settings.
- Adaptability: Solutions may not easily adapt to new or changing conditions, such as extreme weather events or new pest species.
- User Accessibility:Decision support systems and optimization tools may have steep learning curves and require specialized knowledge to use effectively.

Proposed Solution

1. Integrated Data Platform: Combine historical, weather, soil, and remote sensing data

into a unified platforms for comprehensive insights.

2. Advanced Predictive Modeling: Use machine learning for accurate yield predictions and to identify key factors affecting crops.

3.Real-Time Monitoring: Implement IoT sensors and satellite imagery for immediate tracking of soil moisture and crop health, with automated alerts for issues.

4.Decision Support System (DSS): Create an easy-to-use system that provides actionable farming recommendations based on integrated data and predictions.

5. Localized Solutions and Training: Tailor recommendations to local conditions and train farmers on new technologies and practices.

6. Continuous Feedback: Use farmer feedback to refine models and improve solutions over time.

4.1 Code submission:

https://github.com/shubhadaa110/UCT_Internship

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.

1. Data Collection and Ingestion

Start:

- Objective: Collect diverse data sources relevant to crop production.
- Sources: Historical crop yield data, weather data, soil health data, satellite imagery, and sensor data.
- Tools APIs, web scraping, databases, IoT sensors.

Intermediate Stages:

- Data Storage: Store collected data in a data warehouse or data lake.
- Data Validation: Verify the integrity and accuracy of collected data.
- ETL Processes: Extract, Transform, Load processes to ensure data is in the required format for analysis.

Outcome:

- Result: A comprehensive and accurate dataset ready for further processing.

2. Data Preparation

Start:

- Objective: Prepare the collected data for analysis and modeling.
- Tasks: Data cleaning, handling missing values, removing duplicates.

Intermediate Stages:

- Feature Engineering: Create new features from raw data to improve model performance.
- Normalization/Standardization: Scale data to ensure uniformity.
- Data Splitting: Divide data into training, validation, and test sets.

Outcome:

- Result: A clean and well-structured dataset optimized for modeling.

3. Model Development

Start:

- Objective: Develop predictive models to forecast crop yields and recommend best practices.
- Tasks: Select appropriate algorithms (e.g., regression, decision trees, neural networks).

Intermediate Stages:
- Model Training: Train models using training data.
- Hyperparameter Tuning: Optimize model parameters for better performance.
- Cross-Validation: Validate model accuracy and generalizability.

Outcome:
- Result: Robust and accurate models ready for deployment.

4. User Interface and Visualization
Start:
- Objective: Develop an intuitive user interface for end-users.
- Tasks: Design and implement dashboards with interactive visualizations.

Intermediate Stages:
- Usability Testing: Test the interface with end-users for feedback.
- Visualization Tools: Use tools like Tableau, Power BI, or custom web applications.

Outcome:
- Result: A user-friendly interface for data visualization and decision support.

5. Continuous Improvement
Start:
- Objective: Continuously improve the solution based on feedback and new data.
- Tasks: Regular updates, model retraining, and system optimization.

Intermediate Stages:
- Feedback Loop: Incorporate user feedback for ongoing improvements.
- Data Updates: Regularly update models with new data to maintain accuracy.

Outcome:
- Result: An evolving solution that adapts to new challenges and continues to meet user needs.

6 Performance Test

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

Constraints and Considerations in Design

Identified Constraints

1. Data Volume and Complexity:
- Constraint: Large datasets from multiple sources (e.g., historical data, weather data, IoT sensors) may be challenging to manage and analyze efficiently.
- Handling in Design:
- Implemented data warehousing and distributed computing to handle large volumes.
- Used data preprocessing to filter and aggregate data for more manageable analysis.
- Test Results: Achieved satisfactory performance with optimized queries and data storage. Further testing is needed to ensure performance under peak data loads.

2. Model Accuracy and Performance:
- Constraint: Predictive models need high accuracy for reliable recommendations, especially given the variability in agricultural data.
- Handling in Design:
- Employed multiple modeling techniques (e.g., regression, ensemble methods) and cross-validation to improve accuracy.
- Tuned hyperparameters and used feature engineering to enhance model performance.
- Test Results: Models achieved acceptable accuracy, though continuous improvement and retraining are recommended as new data becomes available.

3. Real-Time Data Processing:
- Constraint: Real-time monitoring requires low latency in data processing and alert generation.
- Handling in Design:
- Integrated real-time data pipelines and used efficient data processing frameworks (e.g., Apache Kafka, Spark Streaming).
- Test Results: Real-time alerts and updates were delivered with minimal latency. Further optimization may be needed for high-frequency data.

4. System Integration and Interoperability:

- Constraint: The solution needs to integrate seamlessly with existing systems and tools used in agriculture
- Handling in Design:
- Implemented APIs and standard data formats to facilitate integration.
- Coordinated with stakeholders to ensure compatibility with existing agricultural technologies.
- Test Results:Successful integration with key systems, but additional testing is required for derived

Recommendations for Handling Constraints

1. Data Volume and Complexity:

- Recommendation: Continuously monitor and optimize data storage and processing strategies. Consider using cloud-based data solutions for better scalability.

2. Model Accuracy and Performance:

- Recommendation: Regularly update models with new data and perform periodic evaluations to maintain high accuracy. Explore advanced techniques and model improvements.

3. Real-Time Data Processing:

- Recommendation: Invest in more robust real-time data processing technologies and infrastructure. Consider incorporating edge computing for localized data processing.

4. Scalability:

- Recommendation: Design for scalability from the outset and conduct extensive stress testing. Use auto-scaling cloud services to manage varying loads efficiently.

5. User Interface and Experience:

- Recommendation: Continue gathering user feedback and iteratively improve the UI/UX. Provide training and support materials to help users maximize the system's capabilities.

6. System Integration and Interoperability:

- Recommendation: Ensure ongoing collaboration with industry stakeholders to address integration challenges.

Develop flexible and adaptive interfaces to accommodate diverse systems.

6.1 Test Plan/ Test Cases

Performance 1. *Data Collection*

- Test: Check if data from all sources is ingested correctly.

- Expected: Data should be accurate and complete.

2. Data Preparation

- Test: Validate data cleaning and feature engineering processes.

- Expected: Clean data and relevant features should be created.

3. Modeling

- Test: Ensure models train and perform as expected.

- Expected: Models should meet accuracy and performance standards.

4. Real-Time Monitoring

- Test: Verify real-time data processing and alert generation.

- Expected: Data should be processed quickly, and alerts should be timely.

5. User Interface

- Test: Evaluate usability and functionality of dashboards and feedback systems.

- Expected: Dashboard should be user-friendly, and feedback mechanisms should work.

6. Scalability

- Test:Test system performance under high load and scaling capabilities.

- Expected: System should handle increased load efficiently and scale as neede

6.2 Test Procedure

1. Data Collection:

- Steps: Upload data, verify ingestion, check accuracy.

- Tools: Data ingestion tools, database queries.

2. Data Preparation:

- Steps: Introduce errors, run cleaning scripts, check results.

- Tools: Data cleaning scripts.

3. Modeling:

- Steps: Train models, evaluate performance, record metrics.

- Tools: Machine learning frameworks.

4. Real-Time Monitoring:

- Steps: Simulate data input, check processing speed and alerts.

- Tools: Real-time processing platforms.

5. *User Interface:

- Steps:Test dashboard usability, collect feedback.

- Tools: UI testing tools.

6. Scalability:

- Steps: Simulate load, monitor performance, test auto-scaling.

- Tools: Load testing tools, cloud monitoring.

6.3 Outcome

Test Procedure Outcomes

1. Data Collection:

- Outcome: Data is accurately ingested and integrated from all sources, ensuring completeness and consistency.

2. Data Preparation:

- Outcome: Data cleaning effectively handles errors, and feature engineering produces

relevant, accurate features.

3. Modeling:

- Outcome: Models are trained successfully and meet performance metrics, providing reliable predictions.

4. Real-Time Monitoring:

- Outcome: Real-time data is processed with minimal latency, and alerts are generated and displayed correctly.

5. User Interface:

- Outcome: The dashboard is user-friendly, and feedback mechanisms are functional, enhancing user experience.

6. Scalability:

-The system performs well under increased load and scales effectively to handle growing data and user demands.

7 My learnings

Overall Learning:

1. Technical Skills:
- Data Handling:Mastered data collection, cleaning, and preprocessing techniques.
- Analysis and Modeling: Gained proficiency in exploratory data analysis, feature engineering, and advanced machine learning models.
- Visualization: Developed skills in creating impactful visualizations and reports.

2. Problem-Solving:
- Real-World Applications:Applied data science techniques to solve practical problems in agriculture, such as predicting crop yields and optimizing farming practices.

3. Communication:
- Reporting and Presentation:Enhanced ability to communicate complex data insights clearly to stakeholders and decision-makers.

4. Collaboration:
- Teamwork: Worked effectively with a diverse team and collaborated with stakeholders, learning the importance of cross-functional communication.

Career Impact:

1.Enhanced Skill Set: The technical skills and knowledge gained are crucial for roles in data science, analytics, and Machine learning

2.Industry Experience: Practical experience in applying data science to real-world problems improves job prospects
and demonstrates expertise in leveraging data for impactful solutions.

3.Problem-Solving Abilities: Strengthened problem-solving skills will be valuable in tackling complex challenges across
various industries.

4.Communication Skills: Improved ability to present data-driven insights will be beneficial in roles

8 Future work scope

Ideas for Future Development

1. Enhanced Real-Time Analytics:
 - Idea:Integrate advanced real-time analytics for more detailed insights into crop conditions and environmental factors.
 - Future Work: Implement real-time data processing frameworks to handle high-frequency data from IoT sensors.

2. Advanced Predictive Models:
 - Idea: Explore cutting-edge techniques such as deep reinforcement learning or generative adversarial networks (GANs) for more accurate predictions.
 - Future Work: Research and integrate these advanced algorithms to improve model performance.

3. Integration with External Data Sources:
 - Idea: Incorporate additional data sources like market prices, local agricultural practices, and socio-economic factors.
 - Future Work: Expand the data platform to include these external datasets for a more holistic analysis.

4. User Personalization:
 - Idea: Develop personalized recommendations for farmers based on their specific crops, soil conditions, and historical data.
 - Future Work Implement machine learning techniques to tailor recommendations and advice for individual users.

5. Scalability Enhancements:
 - Idea: Optimize the platform for scalability to handle larger datasets and more users.
 - Future Work: Improve the system architecture to support scaling up, including cloud-based solutions for better performance and flexibility.

6. Integration of Automated Farming Tools:
 - Idea: Connect the platform with automated farming tools like drones and smart irrigation systems.
 - Future Work: Explore integration with these technologies to automate data collection and decision-making processes.

7. User Experience Enhancements:
 - Idea: Refine the user interface and experience of the Decision Support System (DSS) for improved usability and engagement.
 - Future Work: Conduct user research to design more intuitive interfaces and features based on user feedback.

8. Long-Term Impact Analysis:
- Idea: Develop methods to assess the long-term impact of the implemented solutions on crop production and sustainability.
- Future Work:*Create metrics and models to evaluate the effectiveness of interventions over extended periods.

These ideas represent potential areas for future development that could further enhance the solution's effectiveness and impact.