Industrial Internship Report on

Prediction of Agriculture Crop Production in India

Prepared by

Guru Vishnu Kavuturu

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 Prediction of Agriculture Crop Production 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 3

2 Introduction 4

2.1 About UniConverge Technologies Pvt Ltd 4

i. UCT IoT Platform 4

2.2 About upskill Campus 9

2.3 The IoT Academy....11

2.4 Objectives of this Internship program 11

2.5 Reference 11

2.6 Glossary....11

3 Problem Statement....12

4 Existing and Proposed solution....13

4.1 Proposed Solution and Value Addition....13

4.2 Code submission 14

4.3 Report submission 14

5 Proposed Design/ Model 15

5.1 High Level Diagram 16

5.2 Low Level Diagram 16

5.3 Interfaces 18

6 Performance Test....21

6.1 Performance Outcome 22

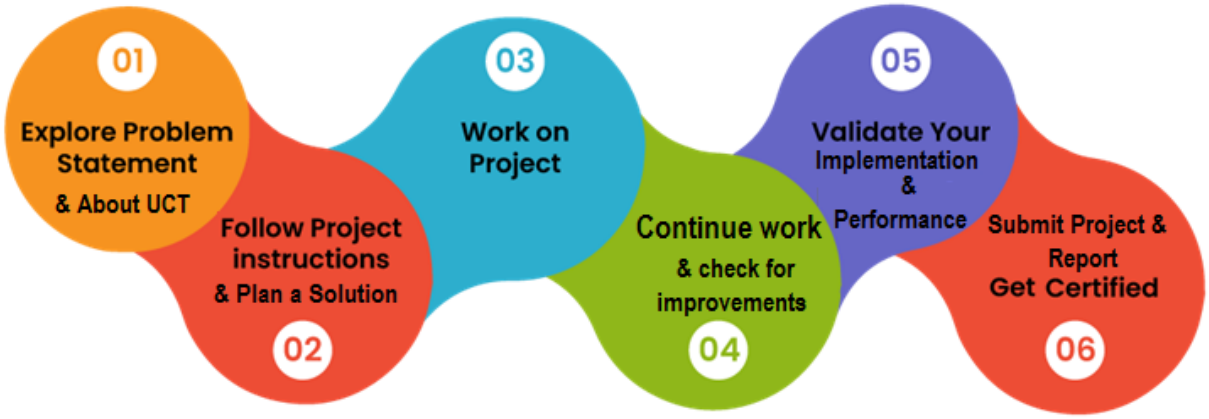
7 My learnings....25

8 Future work scope 27

1 Preface

The 6-week internship program was meticulously planned to provide participants with valuable experiences and skills to enhance their career prospects. The program started with an orientation session where participants were introduced to the objectives and structure of the internship program. Next, participants underwent a comprehensive assessment process to identify their skills, interests, and career goals. This assessment helped match them with suitable internship opportunities that aligned with their aspirations. Coordinators facilitated communication between the interns and the host organizations to establish clear expectations and objectives for the internship period.

In addition to the hands-on experience, the program also included various quizzes and training sessions. The aim was to equip participants with essential career development skills that would benefit them beyond the internship period. At the end of the 6-week program, participants had the opportunity to showcase their work and share their experiences through presentations or reports. They also received feedback from both the program coordinators and the host organizations to help them further improve their skills and knowledge.



Overall, the 6-week internship program organized by UCT emphasized the importance of relevant internships in career development. It provided me with valuable practical experience, industry exposure, and networking opportunities, all of which are crucial for building a successful career. The program was carefully planned to ensure that participants gained maximum benefit from their internships and had a solid foundation for their future professional endeavours. Thanks to Uniconverge Technologies and Upskill Campus for providing this great opportunity to me and team.

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)

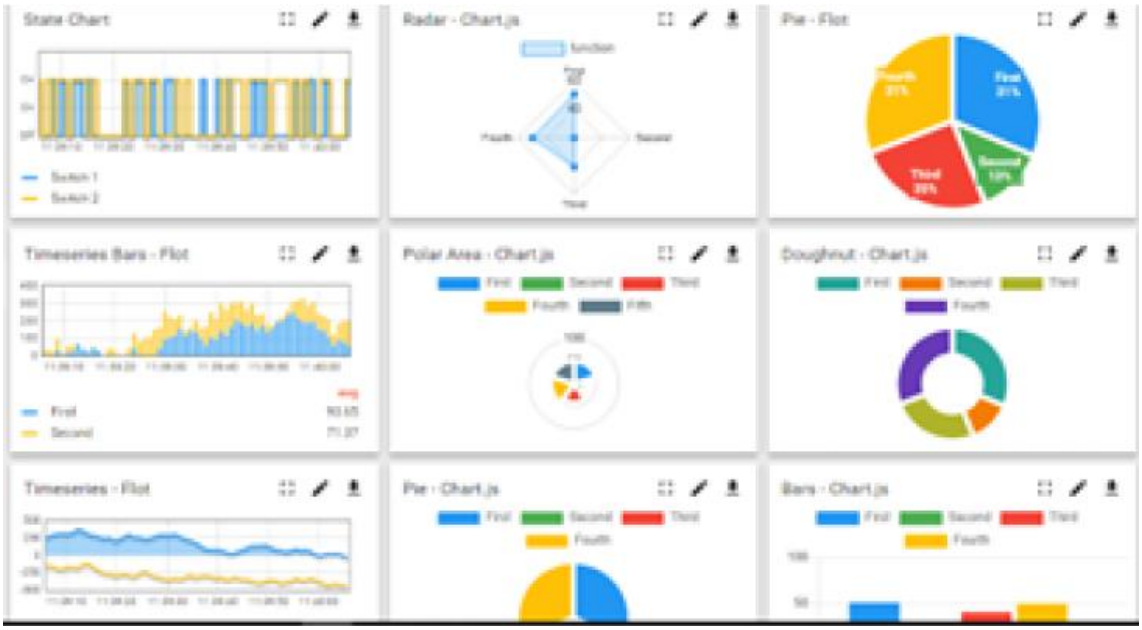
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSQL Databases.

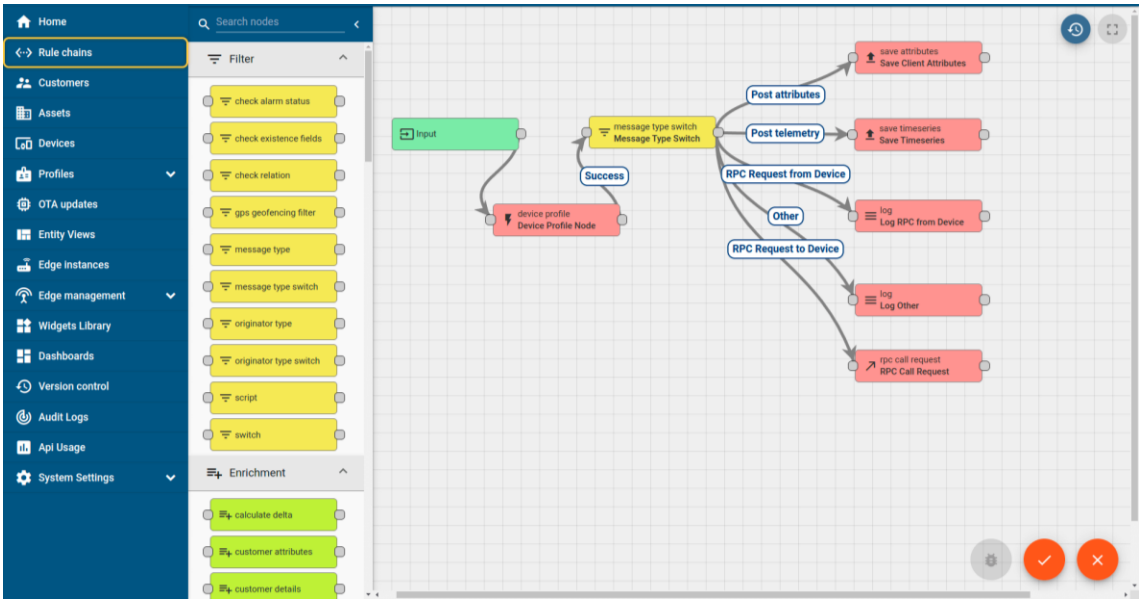
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- 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


State Chart
Radar - Chart.js
Pie - Plot
Timeseries Bars - Plot
Polar Area - Chart.js
Doughnut - Chart.js
Timeseries - Plot
Pie - Chart.js
Bars - Chart.js


Home
Rule chains
Customers
Assets
Devices
Profiles
OTA updates
Entity Views
Edge instances
Edge management
Widgets Library
Dashboards
Version control
Audit Logs
Api Usage
System Settings
Search nodes
Filter
check alarm status
check existence fields
check relation
gps geofencing filter
message type
message type switch
originator type
originator type switch
script
switch
Enrichment
calculate delta
customer attributes
customer details
Input
message type switch
Message Type Switch
Success
device profile
Device Profile Node
Post attributes
Post telemetry
RPC Request from Device
Other
RPC Request to Device
save attributes
Save Client Attributes
save timeseries
Save Timeseries
log
Log RPC from Device
log
Log Other
rpc call request
RPC Call Request

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.

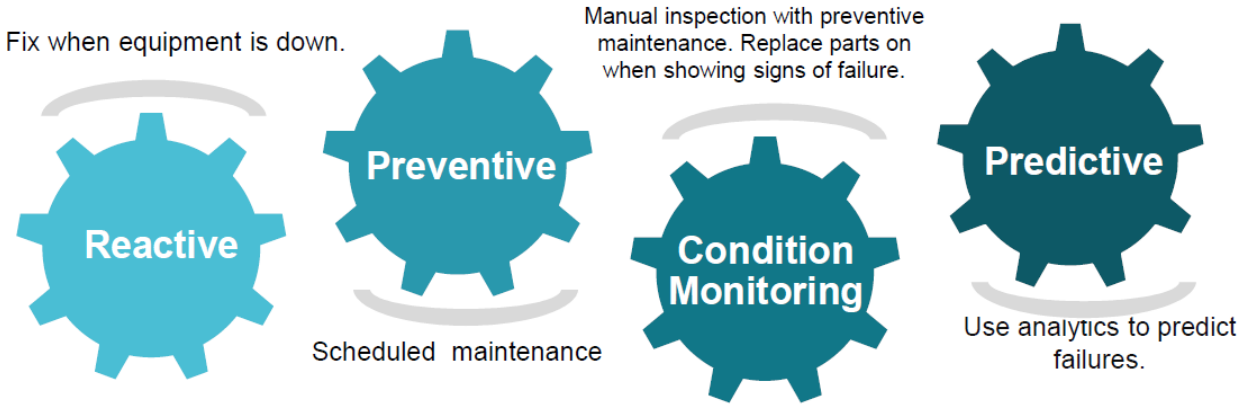
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	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 technology and providing solution in Agrotech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

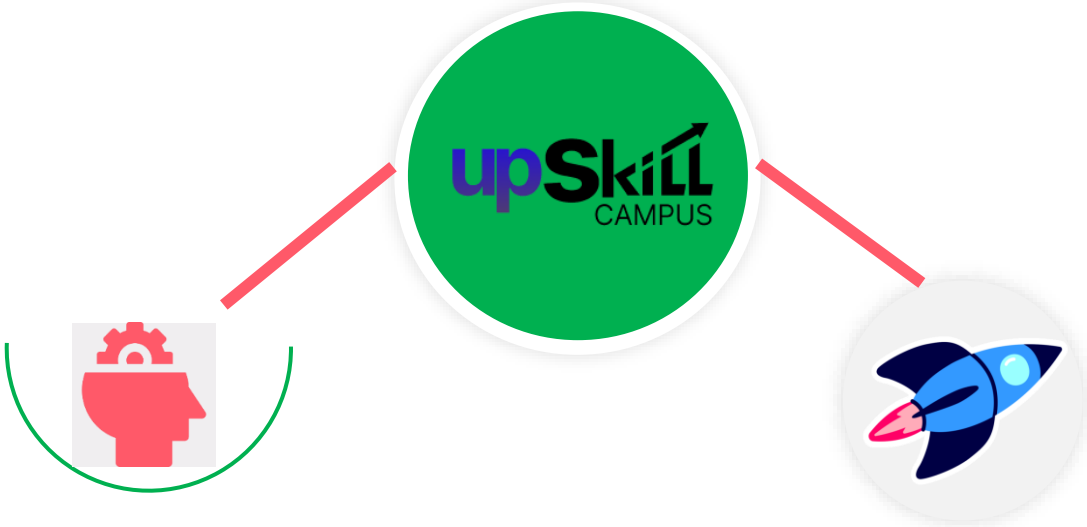
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

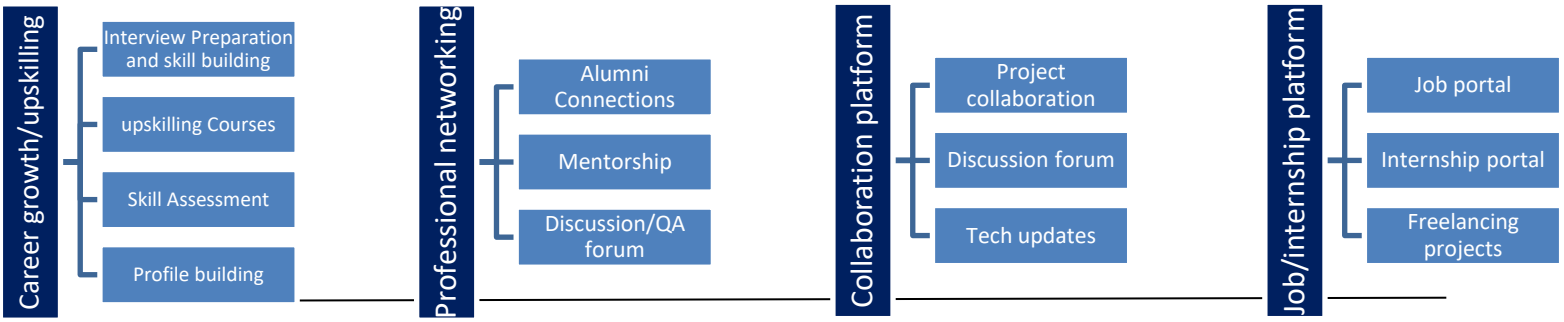
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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] https://www.uniconvergetech.in/

[2] https://www.linkedin.com/company/uniconvergetechnologies/mycompany/

[3] https://learn.upskillcampus.com/s/store?redirectToMicroFE=false

2.6 Glossary

Terms	Acronym
HTTP	Hyper Text Transfer Protocol
UCT	UniConverge Technologies
USC	UpSkill Campus
IOT	Internet Of Things
KPI	Key Performance Indicator

3 Problem Statement

Agriculture is the backbone of India's economy, supporting millions of livelihoods and contributing significantly to the country's GDP. However, the sector faces challenges such as unpredictable weather patterns, changing climate conditions, and varying soil quality. Accurate predictions of crop production are crucial for ensuring food security, effective resource allocation, and informed policymaking. Developing a reliable predictive model for agricultural crop production in India can offer valuable insights into future yield trends and aid in making strategic decisions.

The main objective of this project is to design and implement a robust predictive model for forecasting agricultural crop production in India. The model should leverage historical crop production data, climate information, soil characteristics, and potentially other relevant factors to generate accurate predictions of crop yields for various regions and time periods.

4 Existing and Proposed solution

The existing methods for predicting agricultural crop production in India often rely on historicaltrends, expert opinions, and basic statistical analyses. These methods might include:

1. Expert Surveys: Consulting agricultural experts to gather insights on factors influencing crop yields. However, this method can be subjective and may not capture the complexity of interactions between variables.
2. Crop Yield Prediction Models: Various researchers have developed machine learning models to predict crop yields based on historical data, weather patterns, and agricultural practices. These models use regression or time series analysis techniques. However, limitations include limited accuracy due to the complex nature of agricultural systems, lack of consideration for local variations in soil and climate conditions, and insufficient data on other factors influencing crop production.

3. Disease and Pest Detection: Machine learning algorithms have been employed to detect diseases and pests in crops using image processing techniques. These solutions utilize computer vision and deep learning methods to analyse images of plants and identify signs of diseases or infestations. However, limitations include the need for high-quality images, challenges in capturing data from diverse field conditions, and difficulties in accurately detecting multiple diseases or pests simultaneously.

4.1 Proposed Solution and Value Addition

The proposed solution for predicting agricultural crop production in India involves the development of a data-driven predictive model that integrates historical crop production data, climate variables, soil characteristics, and potentially other influential factors. By leveraging advanced machine learning and statistical techniques, this solution aims to provide accurate forecasts of crop yields for different regions and timeframes. The model's predictions can serve as a valuable tool for policymakers, farmers, and stakeholders to make informed decisions, optimize resource allocation, and enhance agricultural sustainability.

In this agriculture crop production project, the proposed solution aims to overcome the limitations of existing solutions and provide enhanced value through the following approaches:

1. Integrated Model: Develop an integrated machine learning model that considers multiple factors, such as climate conditions, soil properties, irrigation methods, fertilizer usage, pest and disease management, and historical crop yield data. This comprehensive approach will capture the complex interactions among these factors and improve the accuracy of crop yield predictions.

2. Localized Recommendations: Tailor the recommendations generated by the model to the specific needs and constraints of different regions in India. By considering regional variations in soil and climate conditions, the proposed solution aims to provide more accurate and actionable insights for farmers and policymakers at a local level.

4.2 Code submission (GitHub link)

https://github.com/guruvish/UCT_project

4.3 Report submission (GitHub link) :

https://github.com/guruvish/UCT_project

5 Proposed Design/ Model

1. Data Collection and Preprocessing:
   - Gathered the dataset that includes relevant features such as climate data (temperature, rainfall, humidity), soil properties, agricultural practices (irrigation, fertilizer usage), pest and disease information, and historical crop yield data.
   - Performed data cleaning, handle missing values, and normalize and standardize the data to ensure consistency and compatibility for the machine learning algorithms.
2. Feature Selection and Engineering:
   - Conducted a thorough analysis of the dataset to identify the most influential features that significantly impact crop production.
   - Performed feature engineering techniques, such as dimensionality reduction, to reduce noise and extract meaningful patterns from the data.
   - Created additional derived features, if necessary, that capture important relationships or interactions among the existing features.
3. Model Selection and Training:
   - Selected appropriate machine learning algorithms based on the nature of the problem and the characteristics of the dataset. Options include regression algorithms (e.g., linear regression, decision trees, random forests), time series analysis.
   - Split the dataset into training and testing sets to evaluate the performance of the trained models.
   - Trained the selected models on the training dataset, optimizing their parameters and hyperparameters to achieve the best performance.
4. Model Evaluation and Validation:
   - Evaluate the trained models using appropriate evaluation metrics, such as mean squared error (MSE), root mean squared error (RMSE), or coefficient of determination (R-squared).
   - Validate the models by testing them on the independent testing dataset to assess their generalization capability and reliability in predicting crop yields accurately.

5.1 High level Diagram

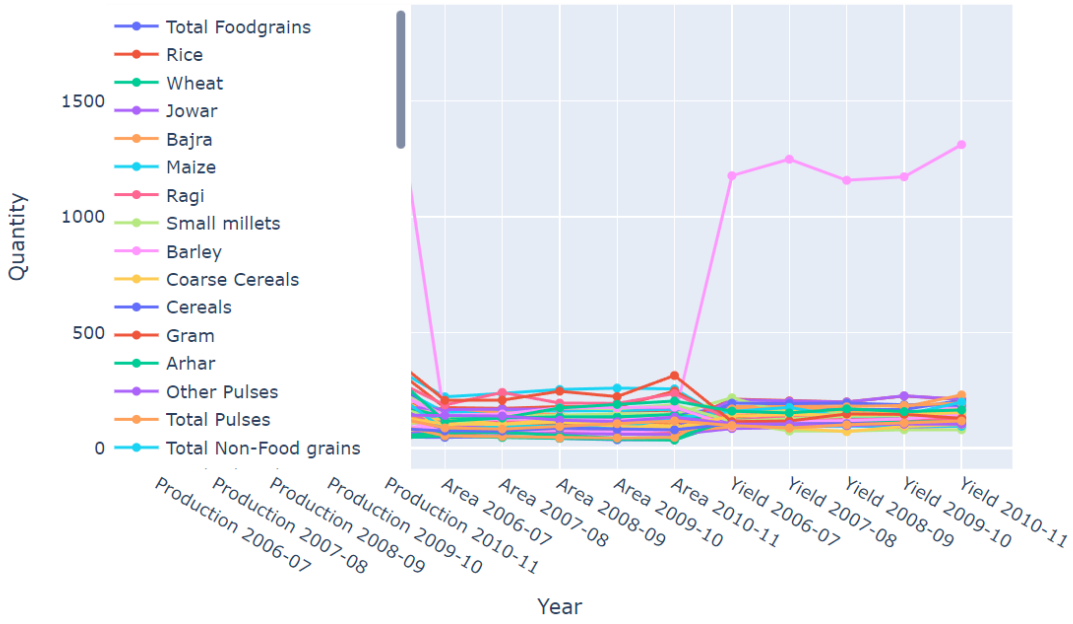


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram

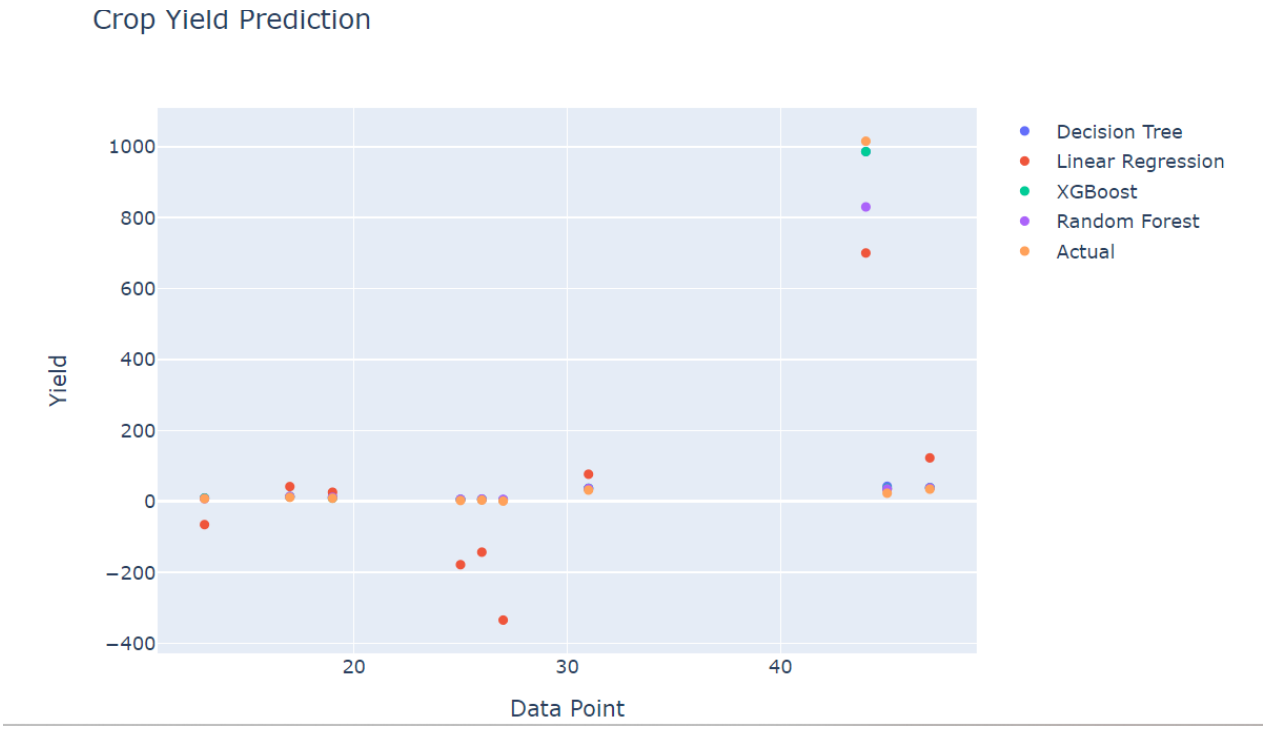


Figure 2: LOW LEVEL DIAGRAM OF THE SYSTEM

5.3 Interfaces

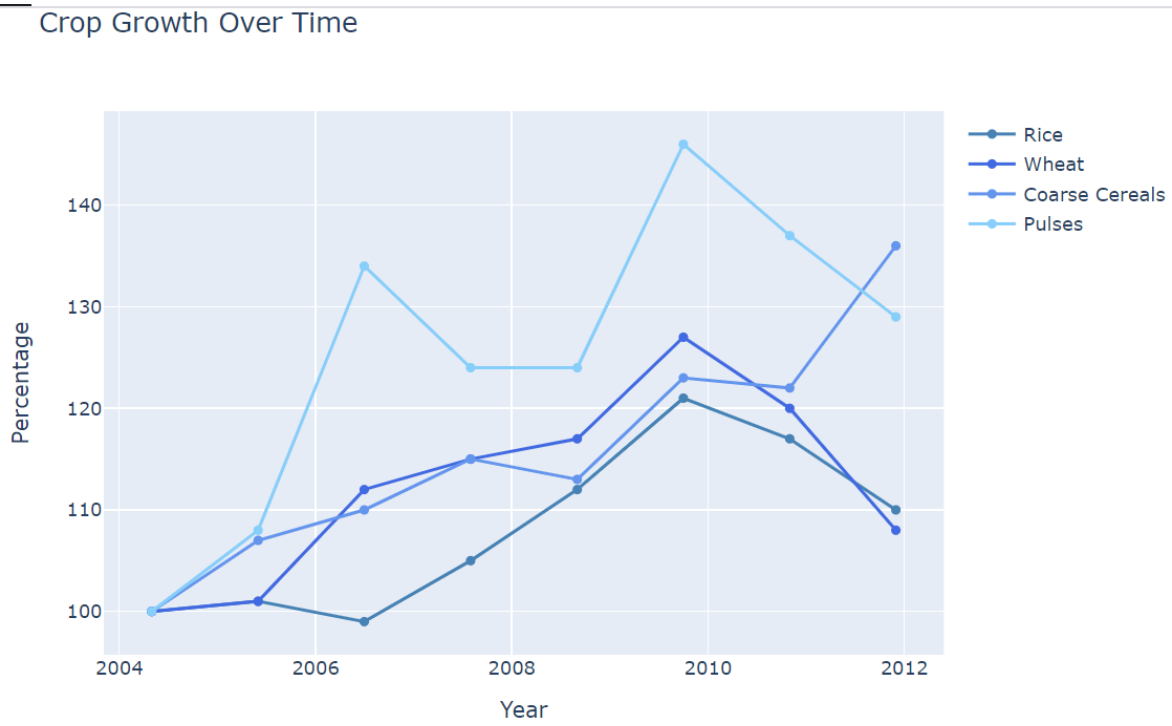


Figure 3. CROP GROWTH BY PERCENTAGE

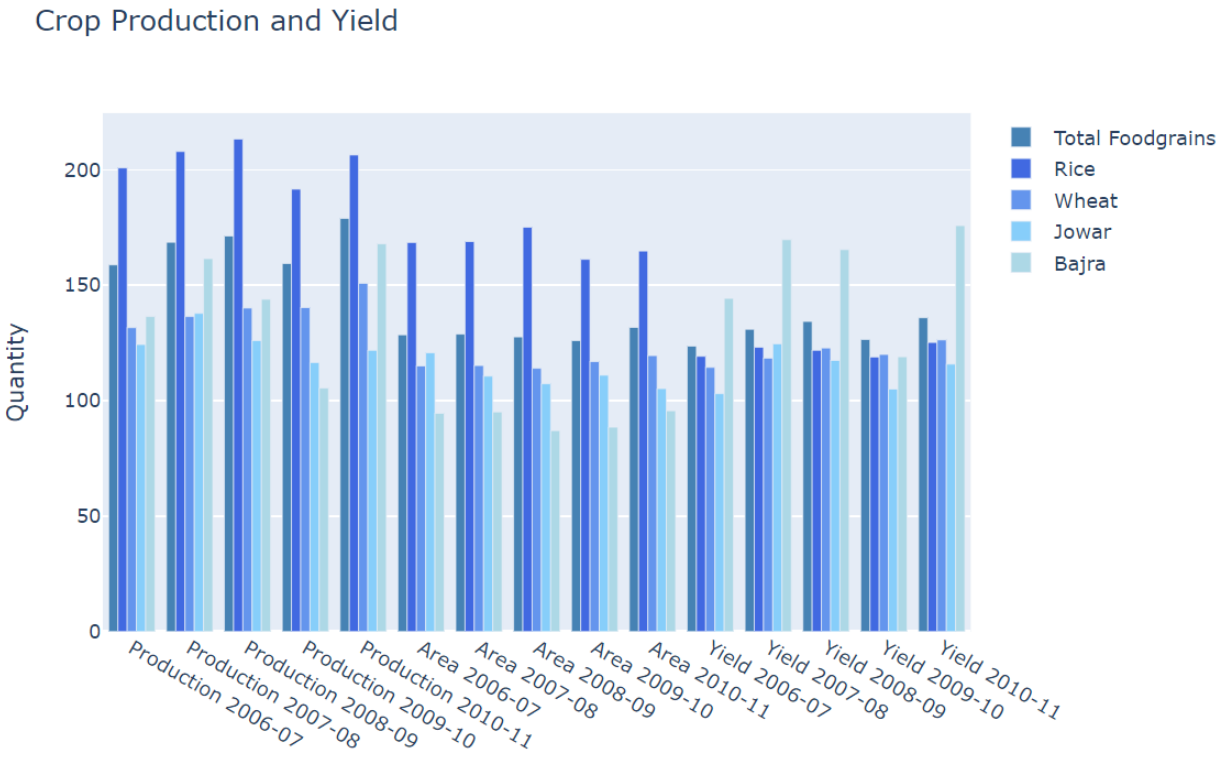


Figure 4. CROP PRODUCTION AND YIELD

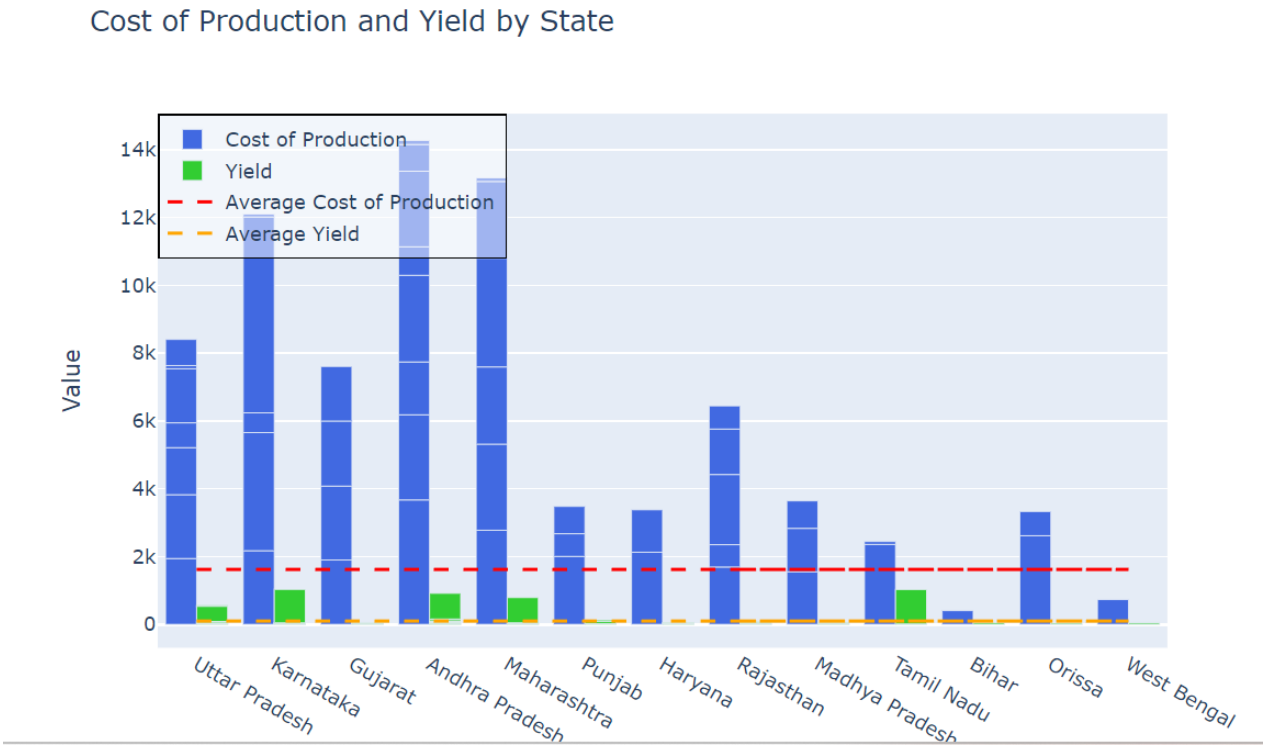


Figure 5. COST OF PRODUCTION AND YIELD BY STATE

6 Performance Test

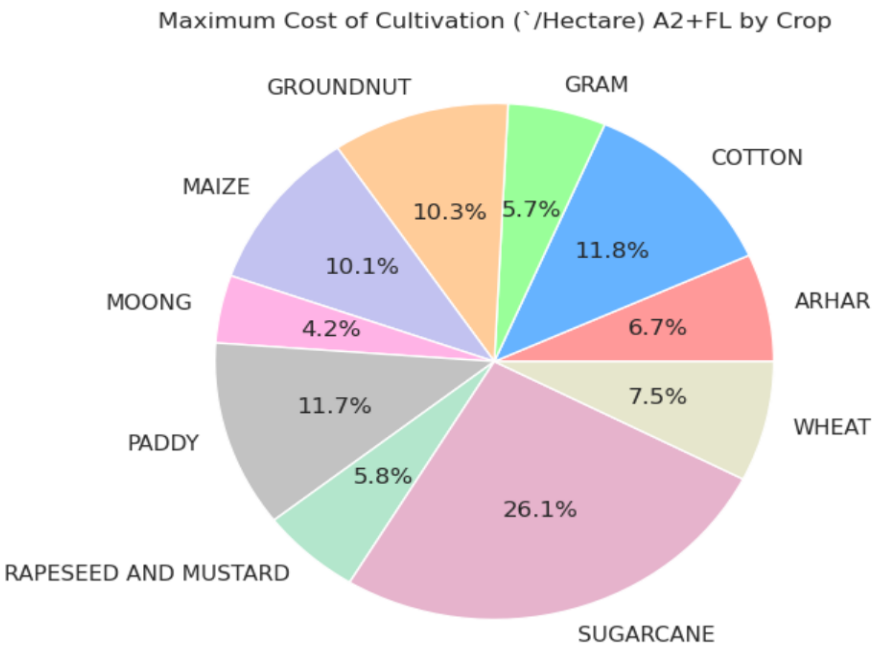
To perform a training prediction and evaluation using decision tree, linear regression, XGBoost, and Random Forest on the given dataset, we'll need to follow these steps:

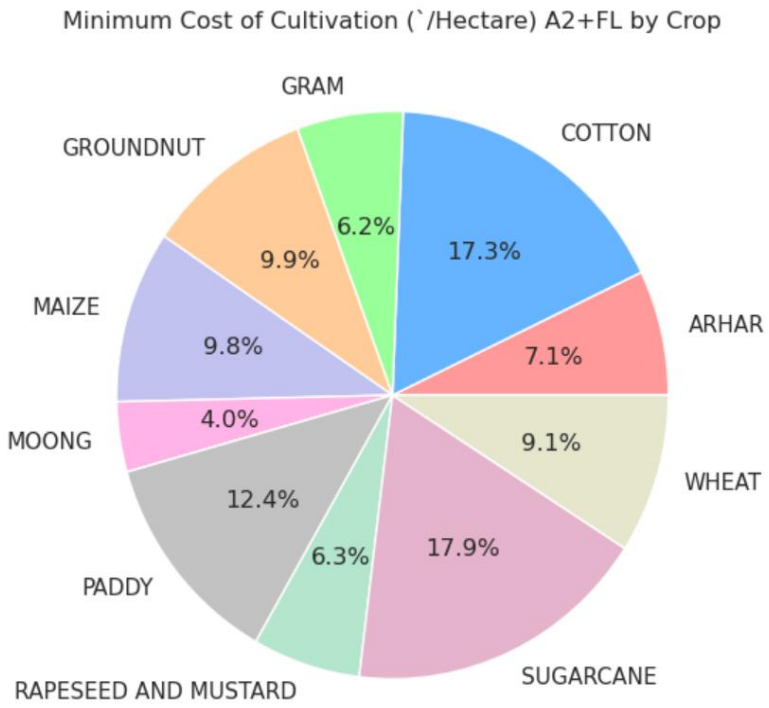
1. **Prepare the data:** Split the dataset into input features (X) and target variable (y).
2. **Split the data into training and testing sets:** Divide the data into two parts, one for training the models and the other for evaluating the performance.
3. **Train the models:** Fit each model on the training data.
4. **Make predictions:** Use the trained models to make predictions on the testing data.
5. **Evaluate the models:** Calculate evaluation metrics such as mean squared error (MSE) or mean absolute error (MAE) to assess the performance of each model.
6. **Visualize the results:** Create plotly plots to compare the predictions and actual values.

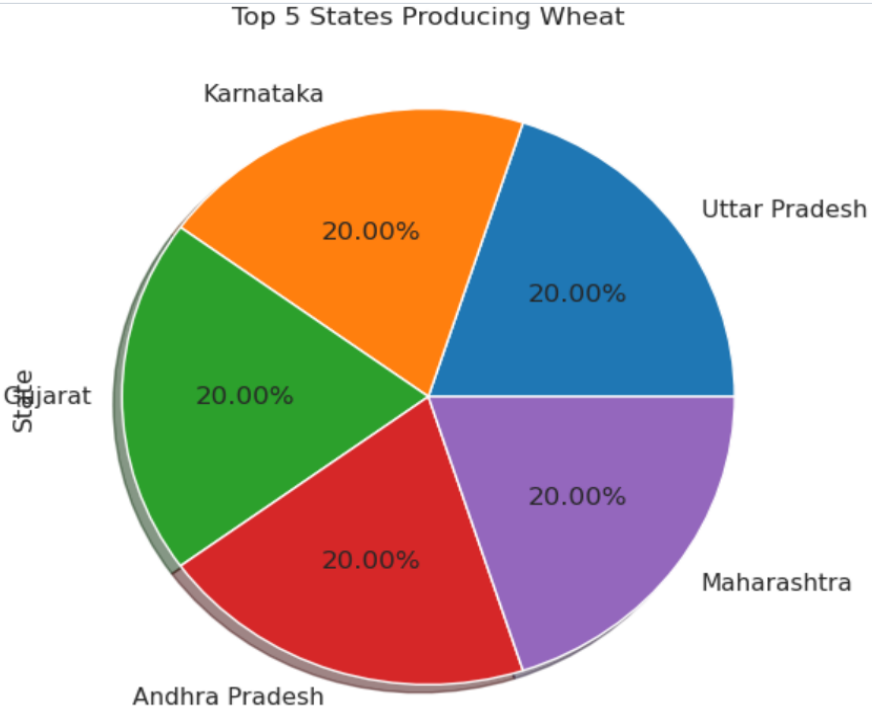
```
Evaluation Results:
Decision Tree:
MSE: 130.28346000000008
MAE: 7.134000000000002
---
Linear Regression:
MSE: 28230.054046904657
MAE: 124.1650703739479
---
XGBoost:
MSE: 113.11858814325804
MAE: 6.5234736404419
---
Random Forest:
MSE: 3452.3139401310013
MAE: 22.243550000000006
---
```

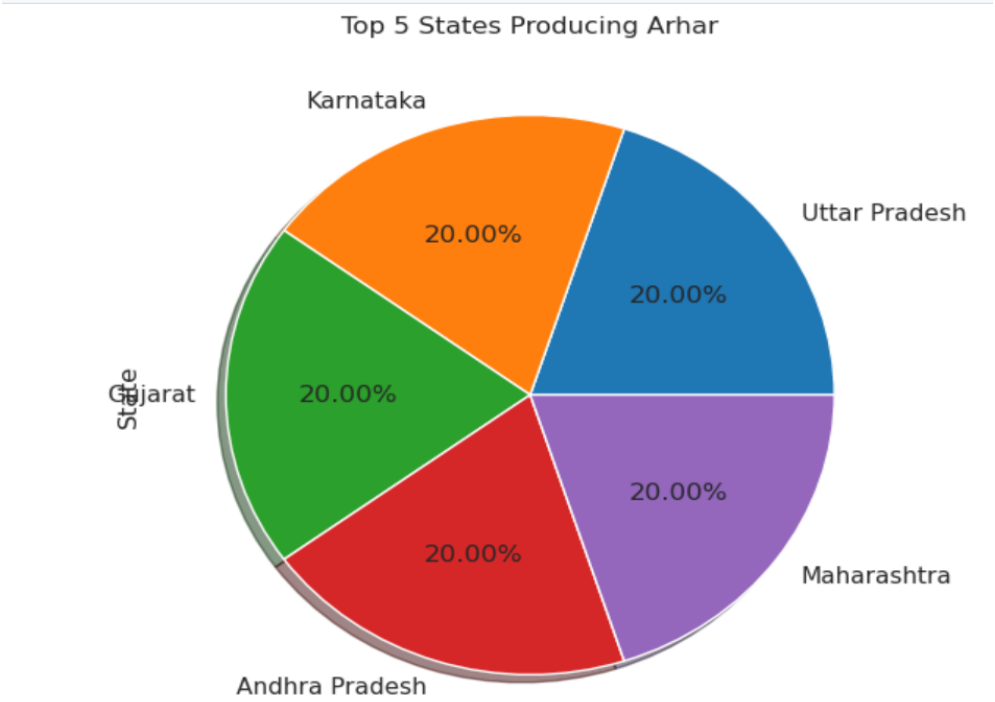
Figure 6. Evaluation Result

6.1 Performance Outcome




Minimum Cost of Cultivation (`/Hectare) A2+FL by Crop
GRAM
COTTON
GROUNDNUT
6.2%
17.3%
9.9%
MAIZE
ARHAR
9.8%
7.1%
4.0%
MOONG
9.1%
WHEAT
12.4%
6.3%
17.9%
PADDY
SUGARCANE
RAPESEED AND MUSTARD


Top 5 States Producing Wheat
Karnataka
Uttar Pradesh
20.00%
20.00%
State
Gujarat
20.00%
20.00%
20.00%
Maharashtra
Andhra Pradesh



7 My learnings

Throughout the agriculture crop production machine learning project in India, I have gained valuable insights and learnings. Here are some key takeaways have acquired:

1. Understanding of Agriculture Domain: I have deepened my understanding of the agriculture sector, including the various factors that influence crop production in India. This includes climate conditions, soil properties, irrigation methods, fertilizer usage, pest and disease management, and historical crop yield data.
2. Data Preprocessing Techniques: I have gained experience in handling real-world datasets for machine learning projects. This includes data cleaning, handling missing values, normalization, and standardization to ensure data quality and compatibility for analysis.
3. Feature Engineering and Selection: I have learned techniques to extract meaningful features from the dataset and engineer new features that capture important relationships and interactions. Additionally, I have gained insights into feature selection methods to identify the most influential factors for crop production.
4. Machine Learning Algorithms: I have gained familiarity with a range of machine learning algorithms suitable for crop yield prediction, such as regression algorithms (e.g., linear regression, decision trees, random forests), time series analysis, and neural networks. I

understand their strengths, weaknesses, and how to select the most appropriate algorithm for a given problem.

5. Model Evaluation and Validation: I have acquired knowledge about evaluation metrics, such as mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared), to assess the performance and accuracy of trained models. Additionally, I have learned the importance of validation using independent testing datasets to ensure the model's generalization capability.
6. Practical Deployment and User Interface: I have gained insights into deploying machine learning models to user-friendly interfaces or applications. I understood the importance of intuitive visualization, explanations, and insights to make predictions and recommendations accessible and understandable to farmers, agricultural experts, and policymakers.
7. Continuous Improvement and Feedback: I have learned the value of continuous improvement in machine learning projects. By collecting user feedback and incorporating new data, such as real-time weather data and satellite imagery, you understand the significance of updating and maintaining the model to enhance its performance and relevance over time.

These learnings equip me with valuable skills and knowledge in the field of agriculture crop production and machine learning. They empower me to make data-driven decisions, optimize crop management practices, and contribute to enhancing agricultural productivity and sustainability in India.

8 Future work scope

Future Work Scope for Agriculture Crop Production Machine Learning Project in India:

Integration of Advanced Techniques: Explore the integration of advanced machine learning techniques such as deep learning and ensemble learning methods to further improve the accuracy and predictive power of the models. These techniques can handle complex patterns and interactions in the data, potentially leading to better crop yield predictions.

Incorporation of Unstructured Data: Consider incorporating unstructured data sources, such as textual data from agricultural research papers, expert knowledge, or social media posts related to farming practices. Natural Language Processing (NLP) techniques can be applied to extract valuable insights and incorporate them into the predictive models.

Crop-specific Models: Develop crop-specific models that focus on individual crops commonly grown in India, such as rice, wheat, maize, or sugarcane. These specialized models can capture crop-specific nuances, growth patterns, and factors that influence their yields, enabling more accurate predictions and tailored recommendations.