Internship Report on Prediction of Agriculture Crop Production in India

Prepared by Kaif Khan

Executive Summary

This report outlines the Industrial Internship conducted in collaboration between UpSkill Campus, The IoT Academy, and UniConverge Technologies Pvt Ltd (UCT). The internship was centered around a project focused on "Prediction of Agriculture Crop Production In India." The objective was to predict crop production for various crops across different states in India using provided yield data. The project and report were completed within a span of six weeks.

Throughout the internship, I had the opportunity to gain exposure to real-world industrial problems and actively participate in designing and implementing solutions. The experience was truly enriching and offered valuable insights into the complexities of the agricultural domain. Overall, the internship proved to be a remarkable journey, empowering me with valuable knowledge and practical skills.

TABLE OF CONTENTS

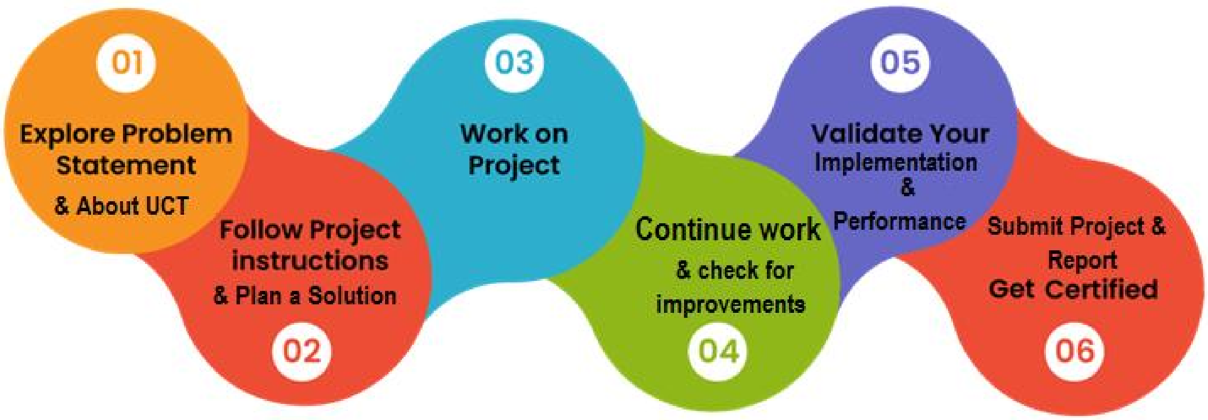
1 Preface

The 6-week internship program was meticulously designed to offer participants a holistic and enriching experience that would enhance their career prospects. The program commenced with a well-structured orientation session, where participants were introduced to the program's objectives and the framework it would follow. A comprehensive assessment process was then conducted, which delved into the participants' skills, interests, and career aspirations. This assessment played a pivotal role in matching each intern with the most suitable and rewarding internship opportunities, ensuring that their goals aligned seamlessly with their assigned projects.Throughout the internship period, effective communication channels were established between the interns and the host organizations through the program coordinators. This facilitated clear communication of expectations and objectives, setting the stage for a productive and meaningful internship experience.

Beyond the hands-on involvement in their respective projects, participants were provided with various quizzes and training sessions. These additional learning opportunities aimed to equip them with essential career development skills that would prove invaluable even after the internship concluded. These skills encompassed a wide range of aspects, including effective communication, problem-solving, teamwork, and time management.

As the 6-week program drew to a close, participants were provided with the platform to showcase their work and share their valuable experiences. Presentations and reports allowed them to highlight their contributions and achievements during the internship. Moreover, they received constructive feedback from both the program coordinators and the host organizations, which served as a catalyst for continuous improvement and growth.

positively.



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

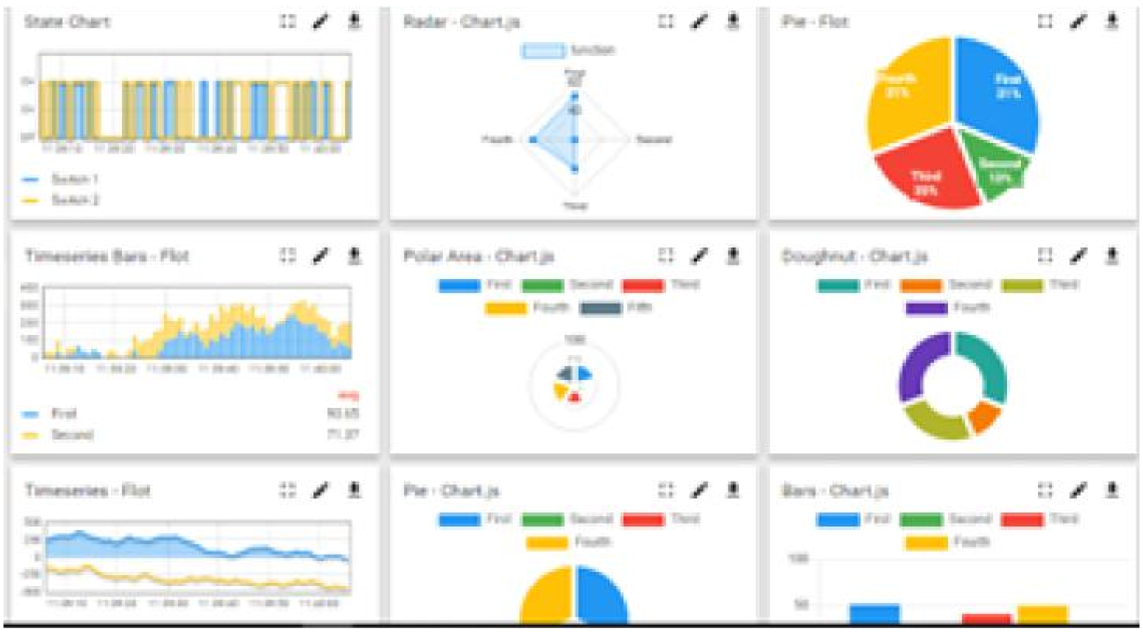
• Build Your own dashboard

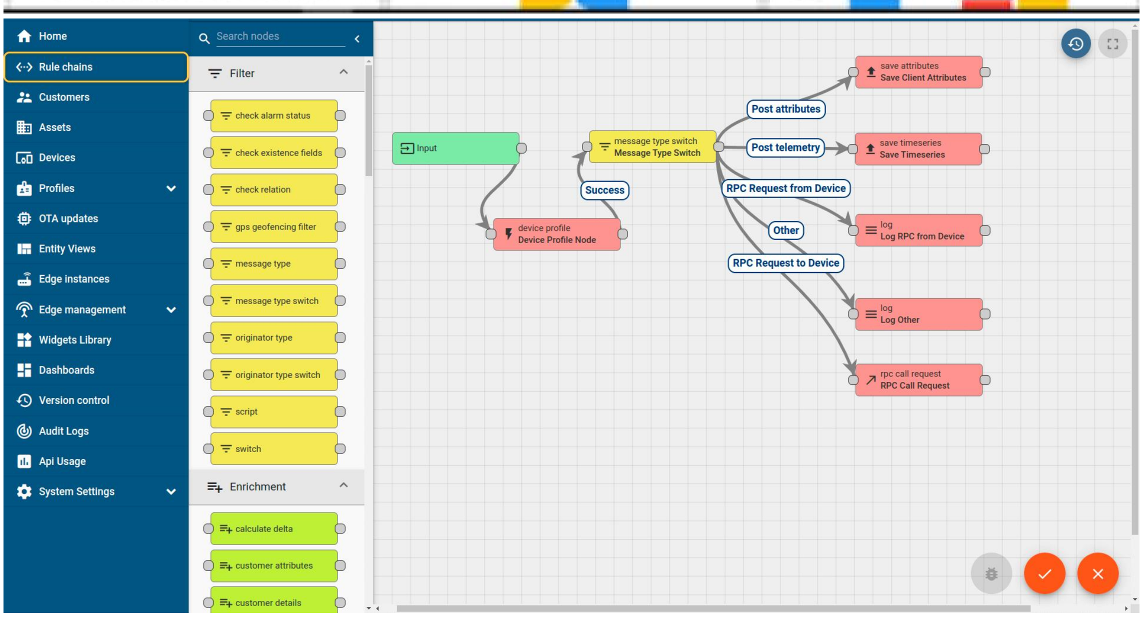
• Analytics and Reporting

• Alert and Notification

• Integration with third party application (Power BI, SAP, ERP)

• Rule Engine





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.


Demo Org
Plant Name
Demo
Plant Name
0%
Plant utilization
55%
Average OEE
4
Capacity
0
Utilized
0
Planned Production
0
Plant Operation
CNC_S7_81
CNC_S7_82
CNC_S7_84
Direct
Email
Union
Utilization
58%
Assist OEE
522 days
Spindle RUL
53%
521 days
Machine
Operator
Work Order ID
Job ID
Job Performance
Job Progress
Start Time
End Time
Output
Planned
Actual
Rejection
Time (mins)
Setup
Pred
Downtime
Idle
Job Status
End Customer
CNC_S7_81 Operator 1 WO0405200001 4168 58% 10:30 AM 55 41 0 80 215 0 45 In Progress
CNC_S7_81 Operator 1 WO0405200001 4168 58% 10:30 AM 55 41 0 80 215 0 45 In Progress

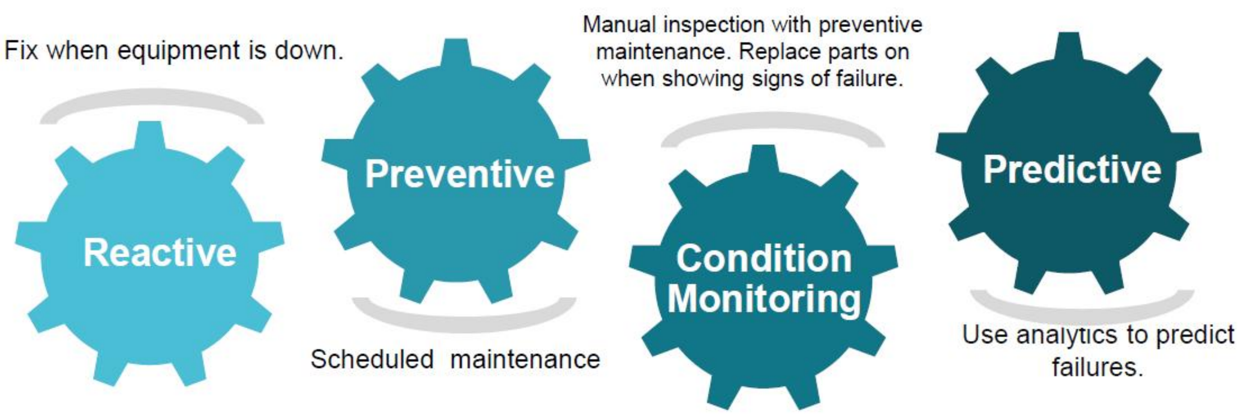

3D View
12:30
Search...
Stopped
Workplace Type
194724_27462-100-A3
Availability
Performance
Quality
OEE
Monitor
Maximum: 42.23%
Minimum: 0.00%
Current: 0.0%
Maximum: 400.00v
Minimum: 0.00v
Average: 200.00v
Maximum: 99.21%
Minimum: 24.53%
Average: 62.35%
Maximum: 99.21%
Minimum: 24.53%
Average: 62.35%

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

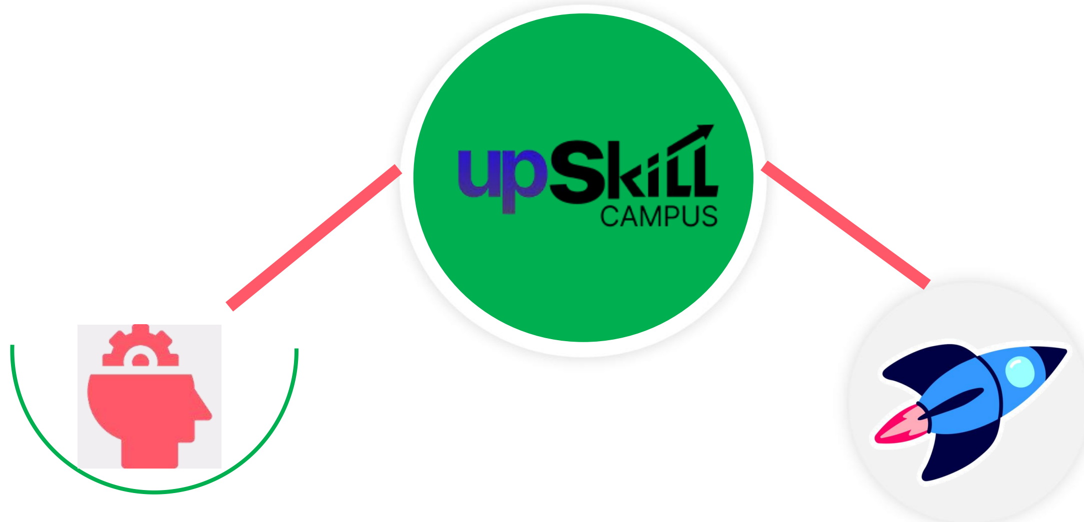
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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/

- **Career growth/upskilling**
  - Interview Preparation and skill building
  - upskilling Courses
  - Skill Assessment
  - Profile building
- **Professional networking**
  - Alumni Connections
  - Mentorship
  - Discussion/QA forum
- **Collaboration platform**
  - Project collaboration
  - Discussion forum
  - Tech updates
- **Job/internship platform**
  - Job portal
  - Internship portal
  - Freelancing projects

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. **Data.gov - Datasets related to agriculture and crop production: https://www.data.gov/**
2. **GitHub : https://github.com/**
3. **Towards datascience : https://towardsdatascience.com/**
4. **Data-Camp: https://www.datacamp.com/**

3 Problem Statement

India, a prominent agricultural powerhouse, holds a significant position as the second-largest producer of key food staples such as wheat and rice on a global scale. However, Indian agriculture faces numerous challenges, including unpredictable weather patterns, limited access to resources, inefficient farming practices, and fluctuating market demands. To address these challenges and improve crop production, there is a need to leverage machine learning techniques on relevant datasets.The problem statement aims to develop a machine learning model that can analyze agricultural data and provide insights and predictions to optimize crop production in India. The model will utilize a dataset containing information on various factors influencing crop yield, such as climate conditions, soil properties, irrigation methods, fertilizer usage, pest and disease management, and historical crop yield data.

The goal of the project is to design a machine learning solution that can assist farmers, agricultural researchers, and policymakers in making informed decisions to enhance agricultural productivity. By harnessing the power of data analytic, the project seeks to unlock valuable patterns and trends in agricultural data, enabling stakeholders to implement effective strategies and achieve sustainable crop growth.The goal of the project is to design a machine learning solution that can assist farmers, agricultural researchers, and policymakers in making informed decisions to enhance agricultural productivity.

4 Existing and Proposed solution

1. Precision Agriculture: Some projects have explored the application of machine learning in precision agriculture, where technologies like drones and sensors are used to collect data on soil health, moisture levels, and nutrient content. Machine learning algorithms then analyze this data to provide personalized recommendations for optimal irrigation and fertilization strategies. However, challenges remain in integrating various data sources, ensuring real-time data collection, and adapting recommendations to specific crop varieties.

2. Crop Price Prediction: Machine learning models have been developed to predict crop prices based on market trends, historical price data, and economic indicators. These predictions can help farmers make informed decisions about crop selection and timing for maximum profitability. However, accuracy may vary due to the volatile nature of agricultural markets and external factors influencing price fluctuations.

3. Sustainable Farming Practices: Some studies have focused on using machine learning to promote sustainable farming practices, such as crop rotation and intercropping. These models analyze soil health, climate conditions, and historical crop data to recommend sustainable farming strategies that improve soil fertility, reduce pest pressure, and increase overall yield. Challenges include the need for comprehensive datasets and ongoing monitoring to assess the long-term impact of sustainable practices.

4. Crop Recommender Systems: Machine learning-based recommender systems have been explored to assist farmers in selecting suitable crop varieties based on their specific environmental conditions and objectives. These systems consider factors like soil type, climate, and water availability to suggest crops that are best suited for a particular region. However, challenges include the availability and accuracy of localized data and the need for continuous updating of the system to reflect changing agricultural conditions.

It is evident that machine learning has the potential to revolutionize agricultural practices in India by providing valuable insights and predictions to optimize crop production. However, addressing the challenges and limitations of existing solutions is crucial to develop robust and practical applications that can truly benefit farmers and contribute to the sustainability of Indian agriculture.

4.1 Proposed Solution and Value Addition

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/shadowfaxx1/upSkill-Campus/blob/aade2526168aad0bbe7b86697d4fbdc699a10e06/Prediction-of-Agriculture-Crop-Production-In-India_Kaif_khan_USC_UCT.ipynb

4.3 Report submission (GitHub link) :

https://github.com/shadowfaxx1/upSkill-Campus/tree/41d6a150b7812335b3d13b99220f7ca50df11e57/Reports

5 Proposed Design/ Model

Data Collection and Preprocessing:

The initial phase of the project involved gathering a comprehensive dataset encompassing essential features, including climate data such as temperature, rainfall, and humidity, soil properties, agricultural practices encompassing irrigation and fertilizer usage, as well as pest and disease information, and historical crop yield data. Subsequently, diligent data cleaning procedures were applied to address any inconsistencies and handle missing values. Additionally, data normalization and standardization were carried out to ensure uniformity and compatibility with the machine learning algorithms.

Feature Selection and Engineering:

A meticulous analysis of the dataset was conducted to identify the most influential features that significantly impact crop production. Leveraging feature engineering techniques, such as dimensionality reduction, noise reduction, and pattern extraction, efforts were made to distill meaningful insights from the data. Furthermore, if required, supplementary derived features were created to capture crucial relationships and interactions among the existing features.

Model Selection and Training:

Appropriate machine learning algorithms were carefully chosen based on the nature of the problem at hand and the inherent characteristics of the dataset. This selection encompassed regression algorithms, such as linear regression, decision trees, and random forests, along with time series analysis. To evaluate the performance of the trained models, the dataset was partitioned into distinct training and testing sets. Subsequently, the selected models underwent rigorous training, wherein parameters and hyperparameters were optimized to attain peak performance.

Model Evaluation and Validation:

The trained models were subjected to thorough evaluation using pertinent evaluation metrics, including mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared). Rigorous validation was carried out by subjecting the models to an independent testing dataset. This validation process aimed to assess the models' generalization capability and ascertain their reliability in accurately predicting crop yields.

5.1 Low level Diagram

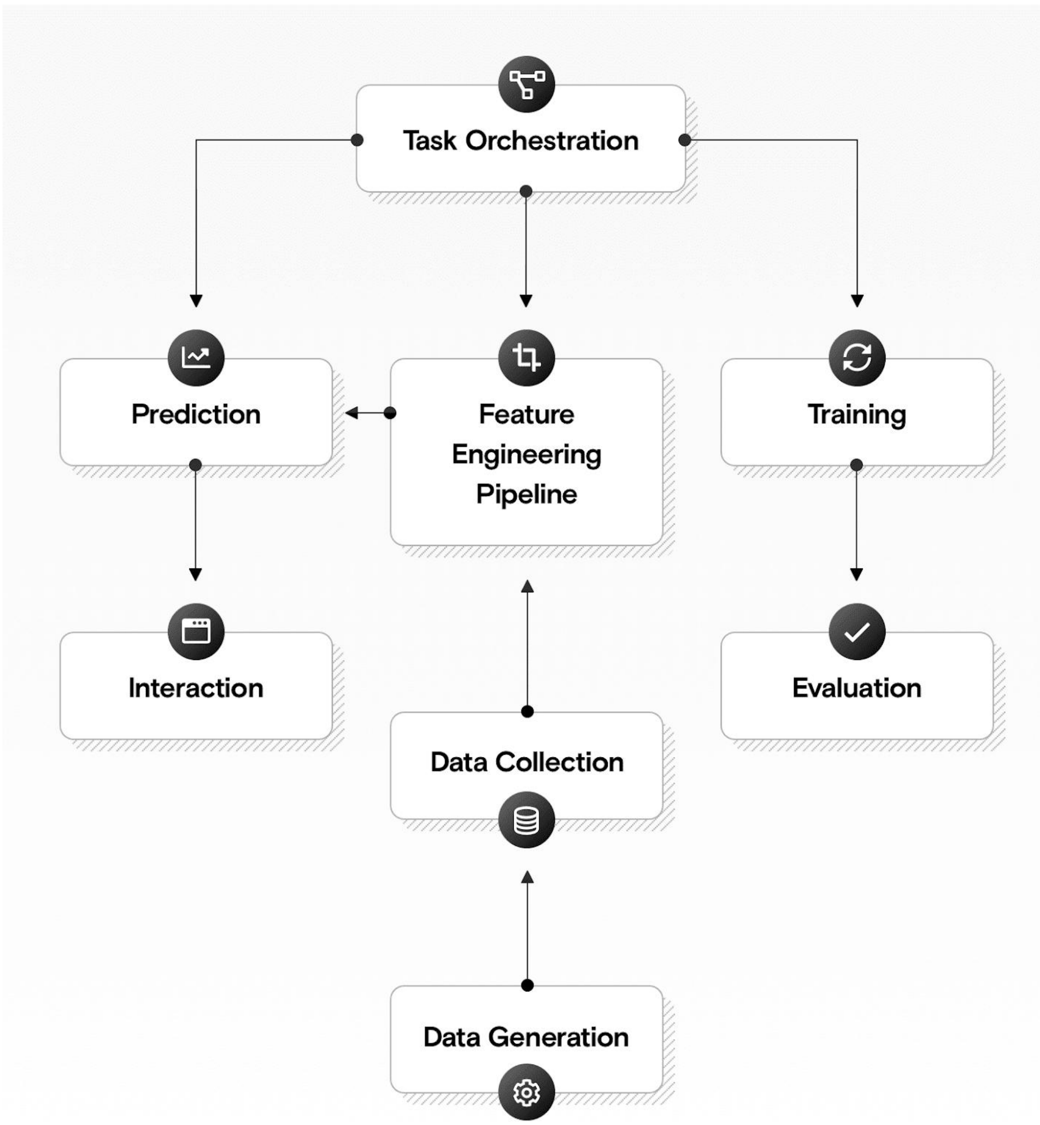


Figure 1: LOW LEVEL DIAGRAM OF THE SYSTE

5.2 High Level Diagram

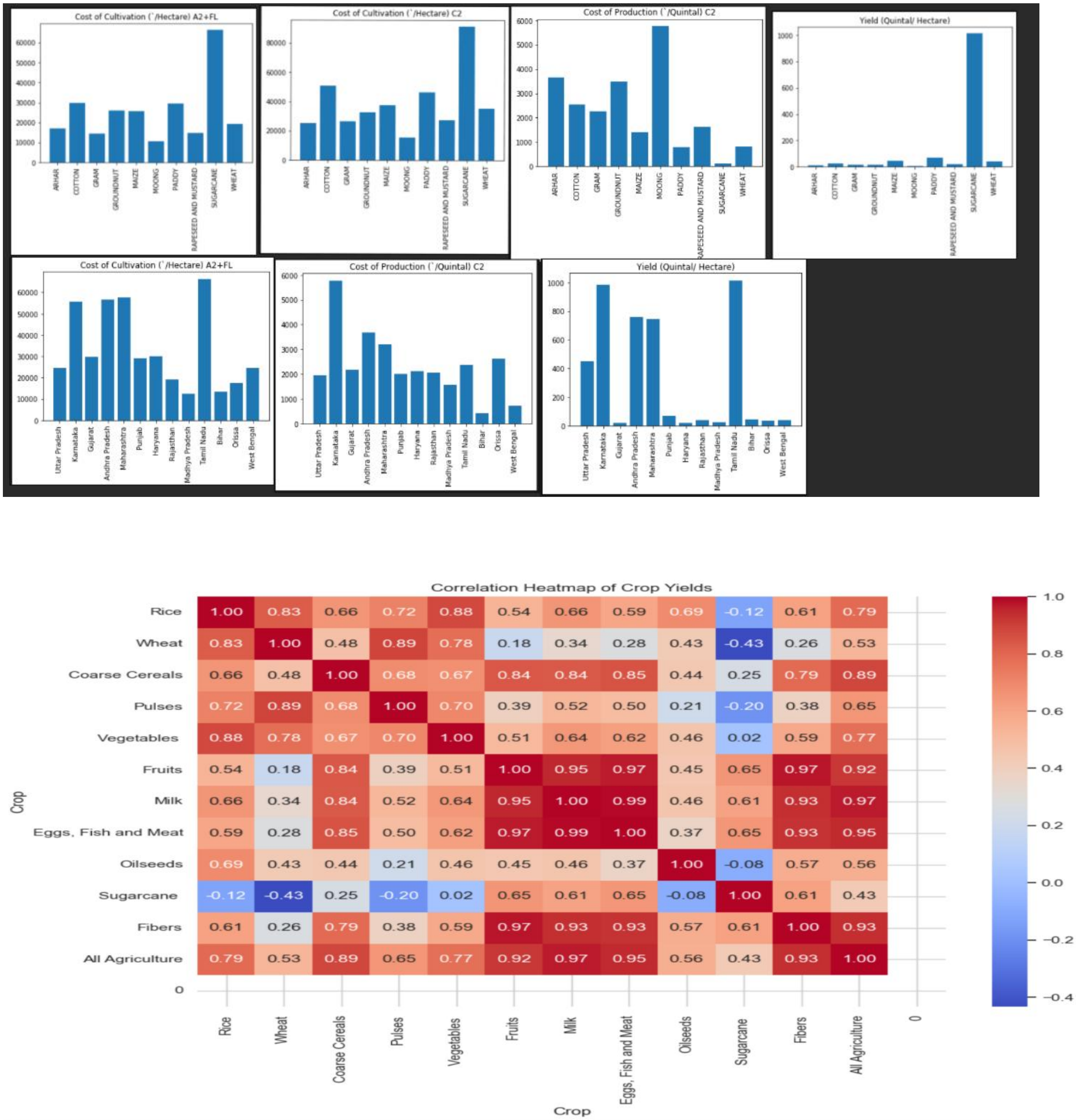


Figure 2: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.3 Interfaces

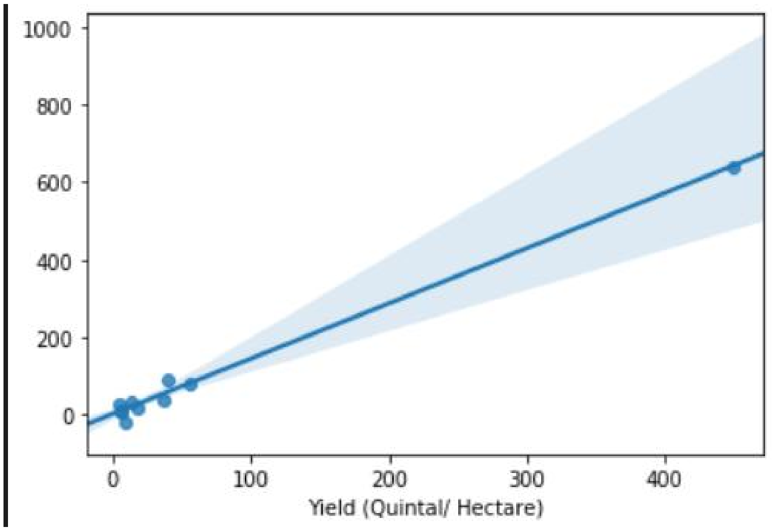


Figure 3.Prediction graph linear regression

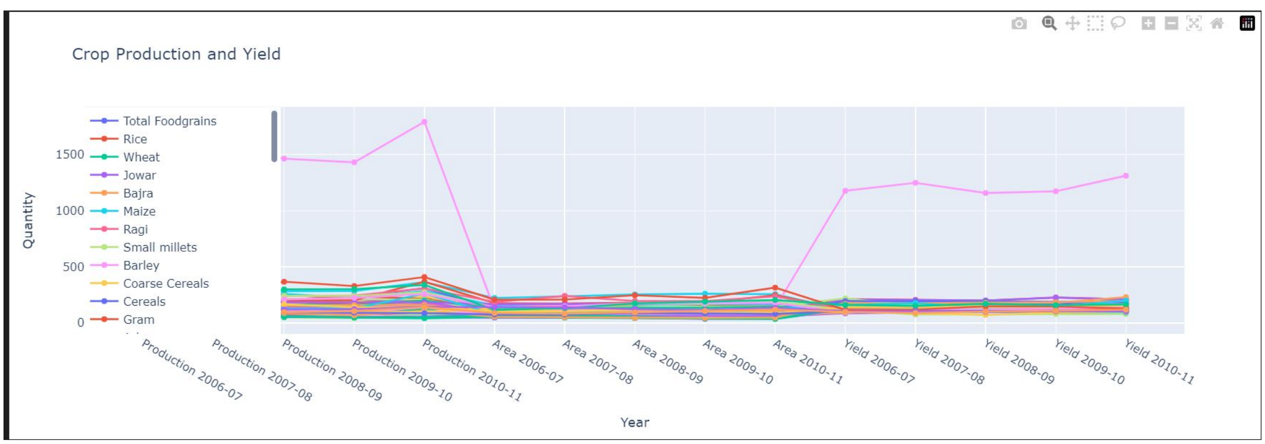


Figure 4. CROP PRODUCTION AND YIELD

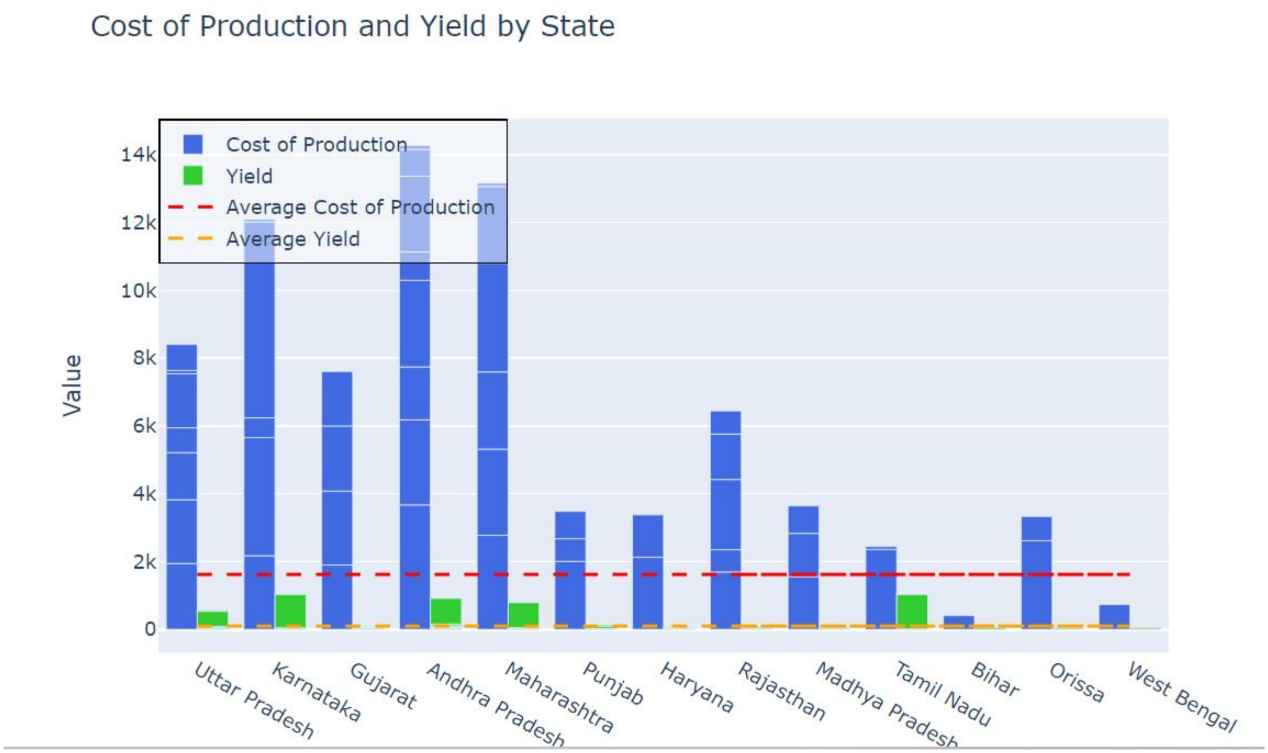


Figure 5. COST OF PRODUCTION AND YIELD BY STATE

6 Performance Test

Data Preparation: Split the dataset into input features (X) and the target variable (y) to create the training and testing datasets.

Model Selection: Choose the appropriate machine learning algorithms for the task, which include Decision Tree, Linear Regression, XGBoost, and Random Forest.

Model Training: Fit each selected model on the training data to learn the underlying patterns and relationships between the features and the target variable.

Model Prediction: Use the trained models to make predictions on the testing dataset.

Model Evaluation: Calculate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared to assess the performance of each model.

Performance Comparison: Compare the evaluation results of all the models to identify which algorithm performs the best for this specific prediction task.

Hyperparameter Tuning (Optional): If necessary, perform hyperparameter tuning for each model to optimize their performance further.

Cross-Validation (Optional): Implement cross-validation techniques to validate the models and ensure they generalize well to unseen data.

Ensemble Methods (Optional): Explore ensemble methods like stacking or blending to combine the predictions from multiple models and potentially improve prediction accuracy.

Visualization (Optional): Create visualizations, such as plotly plots, to compare the predicted values with the actual values for a better understanding of the model's performance.

```
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

```
print("predicted = "  ,y_pred,"r2 score: ", r2 ,"mean_squared", mse)
```

```
predicted =  [ 99.58205325 111.90305213] r2 score:  0.922422874030803 mean_squared 4.964936062028606
```

```
Evaluation Results:
Decision Tree:
MSE: 6677.658100000005
MAE: 30.04200000000001
---
Linear Regression:
MSE: 28230.054046904657
MAE: 124.1650703739479
---
XGBoost:
MSE: 113.11858814325804
MAE: 6.5234736404419
---
Random Forest:
MSE: 3183.725513042004
MAE: 21.629540000000013
---
```

```
Mean MAE: 20.140 (24.830)
0.8837209846619423
```

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_test, pred)
mse
```

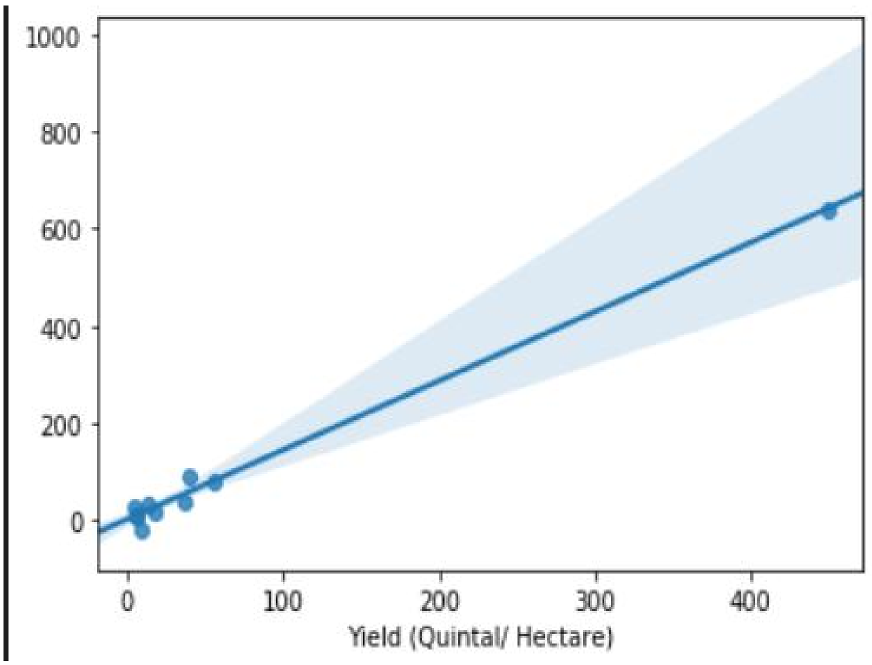
```
1947.1345494903296
```

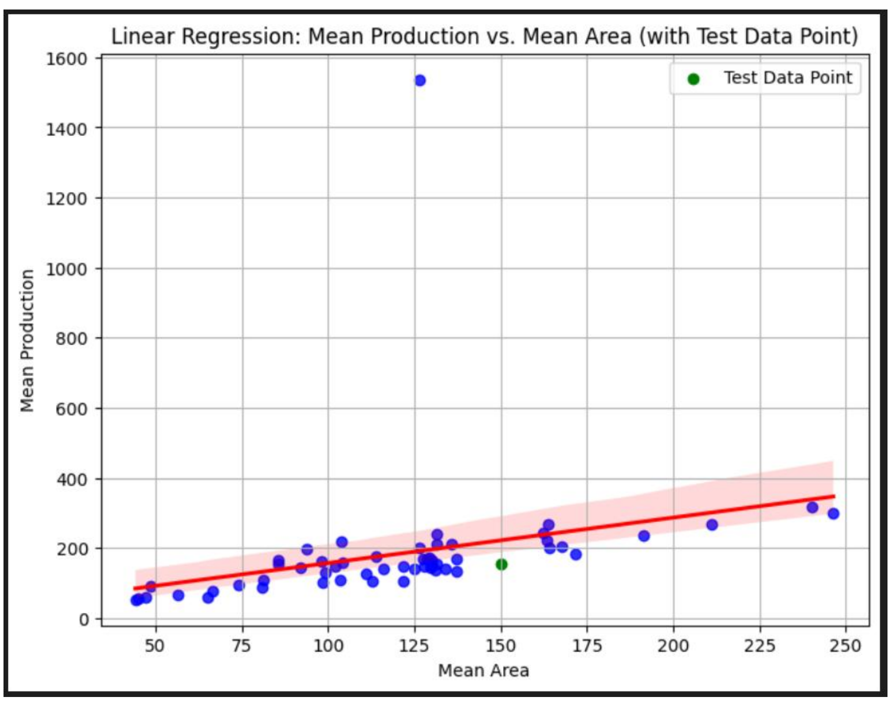
```
import numpy as np
import math
rmse = math.sqrt(mse)
rmse
```

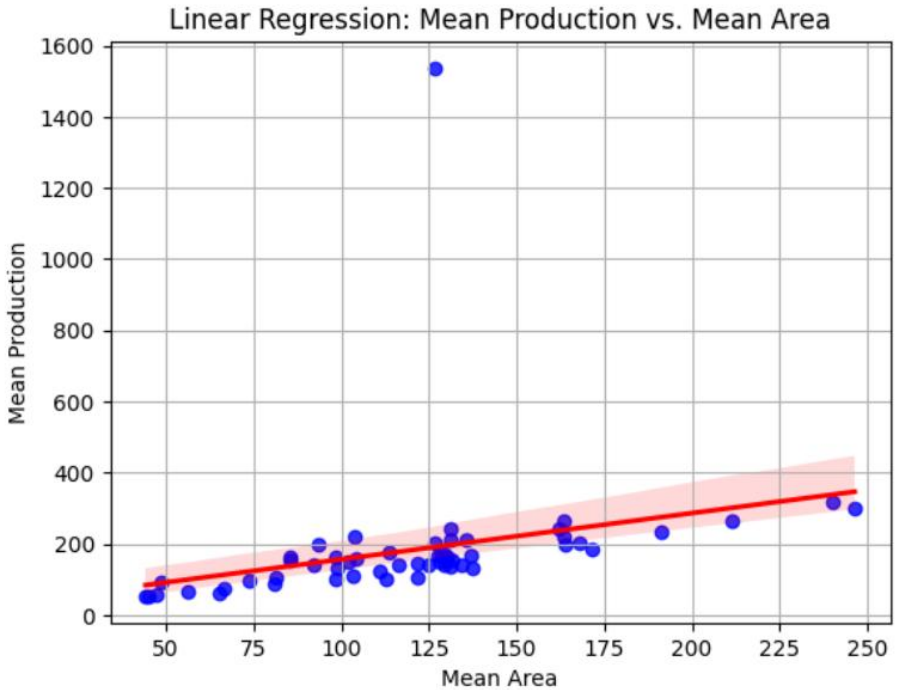
```
44.12634756571554
```

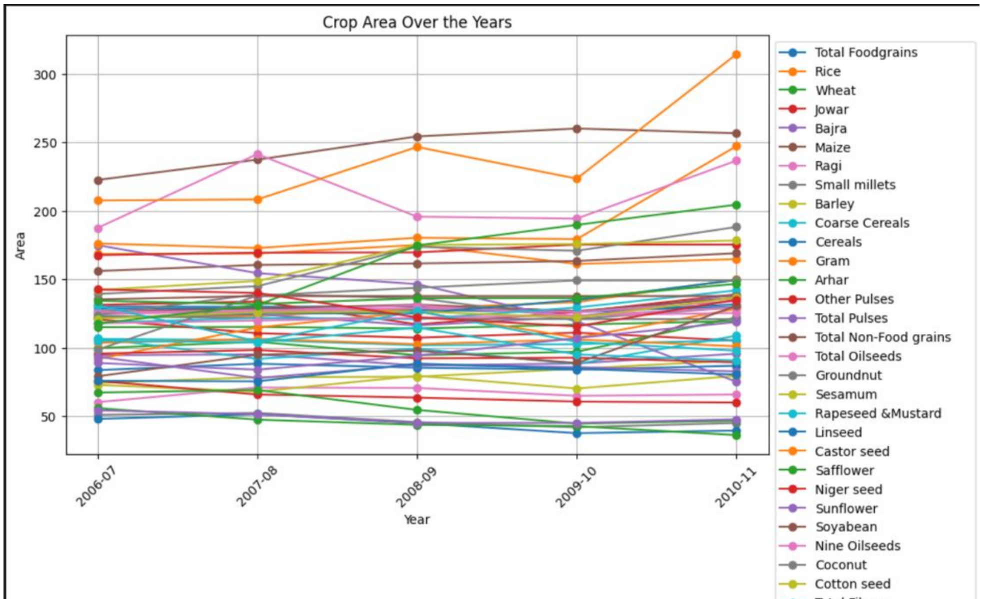
Figure 6. Evaluation Result

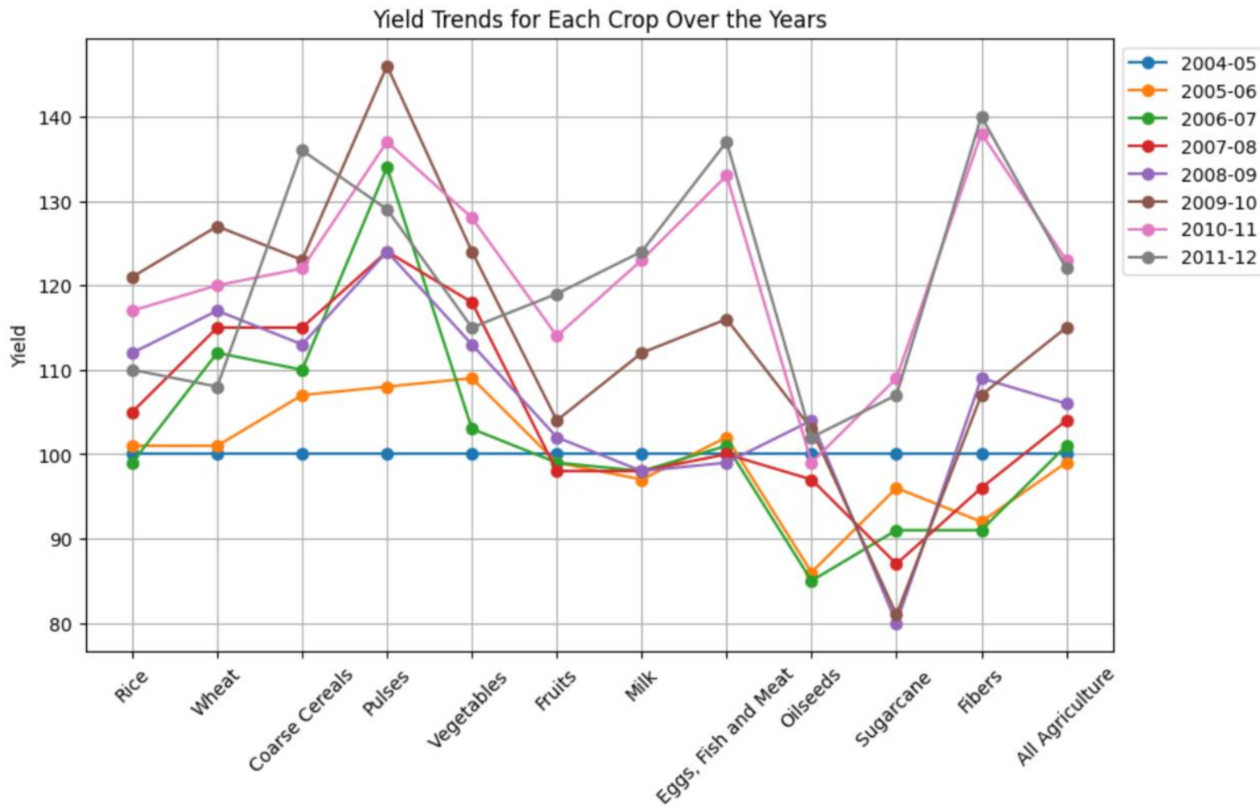
6.1 Performance Outcome






Linear Regression: Mean Production vs. Mean Area
Mean Production
Mean Area
1600
1400
1200
1000
800
600
400
200
0
50
75
100
125
150
175
200
225
250


Crop Area Over the Years
Area
Year
2006-07
2007-08
2008-09
2009-10
2010-11
Total Foodgrains
Rice
Wheat
Jowar
Bajra
Maize
Ragi
Small millets
Barley
Coarse Cereals
Cereals
Gram
Arhar
Other Pulses
Total Pulses
Total Non-Food grains
Total Oilseeds
Groundnut
Sesamum
Rapeseed &Mustard
Linseed
Castor seed
Safflower
Niger seed
Sunflower
Soyabean
Nine Oilseeds
Coconut
Cotton seed
Total Fibers


Yield Trends for Each Crop Over the Years
Yield
2004-05
2005-06
2006-07
2007-08
2008-09
2009-10
2010-11
2011-12
140
130
120
110
100
90
80
Rice
Wheat
Coarse Cereals
Pulses
Vegetables
Fruits
Milk
Eggs, Fish and Meat
Oilseeds
Sugarcane
Fibers
All Agriculture

6 My learning :

Throughout the journey of the agriculture crop production machine learning project in India, I have been fortunate to gain invaluable insights and meaningful learnings. Here are some of the key takeaways that have enriched my understanding and skills in this domain:

1. Enhanced Understanding of Agriculture Domain: This project has provided me with a comprehensive understanding of the intricate workings of the agriculture sector. I have delved into the complexities of various factors that play a pivotal role in crop production in India, ranging from climate conditions and soil properties to irrigation techniques, fertilizer usage, and pest and disease management. Analyzing historical crop yield data has allowed me to grasp the interplay of these elements in shaping agricultural outcomes.

2. Proficiency in Data Preprocessing Techniques: Working with real-world datasets has honed my proficiency in data preprocessing techniques. From handling missing values to cleaning the data and ensuring its consistency, normalization, and standardization, I have imbibed the importance of data quality and compatibility for robust machine learning analysis.

3. Mastery of Feature Engineering and Selection: One of the significant learnings has been the art of feature engineering. By skillfully extracting meaningful features from the dataset and engineering new ones that capture crucial relationships and interactions, I have witnessed how it enhances the model's predictive power. Employing feature selection methods has allowed me to identify and prioritize the most influential factors that impact crop production.

4. Familiarity with Diverse Machine Learning Algorithms: I have gained familiarity with an array of machine learning algorithms suited for crop yield prediction. From the versatility of regression algorithms like linear regression, decision trees, and random forests to the application of time series analysis and neural networks, I now possess the knowledge to select the most appropriate algorithm for a given problem, considering its strengths and weaknesses.

5. Proficiency in Model Evaluation and Validation: The project has equipped me with the expertise to evaluate model performance using relevant evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared). Moreover, validating the models with independent testing datasets has been essential to ensure their reliability and generalization capabilities.

In summary, this project has been a transformative learning experience, deepening my expertise in agriculture, data analysis, and machine learning. I am grateful for the opportunity to have been a part of this project, and I look forward to applying these newfound skills to contribute to the advancement of agriculture and sustainable crop production in the future.

7. Future work scope

Exploring Advanced Techniques: In the future, we can explore incorporating advanced machine learning techniques like deep learning and ensemble learning to enhance the accuracy of crop yield predictions. These methods can help us better understand complex patterns and interactions in the data, leading to more reliable predictions.

Integrating Multi-Source Data: We can consider adding data from multiple sources, such as satellite imagery, remote sensing, and inputs from farmers, to improve our models. By using diverse datasets, we can get a more complete picture of agriculture and its dynamic nature.

Real-Time Updates: We can work on updating our datasets with real-time weather data and other relevant information. This will enable our models to adapt to changing conditions and provide more up-to-date predictions.

Tailoring Crop-Specific Models: It would be beneficial to develop specialized models for specific crops commonly grown in different regions of India. By customizing the models for each crop, we can offer more accurate predictions and recommendations.

Geospatial Analysis: We can explore geospatial analysis to understand spatial patterns and correlations between crop yields and geographical factors. This will help us optimize agricultural planning and resource allocation.

IoT Sensors Integration: We can consider incorporating Internet of Things (IoT) sensors to gather real-time data from farms. This information can enrich our machine learning models and enable precision agriculture.