# Industrial Internship Report on

# ”Prediction of Agriculture Crop Production in India”

## Prepared by

## [Samreen Khanam]

| *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).<br><br>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.<br><br>My project was (**Prediction of Agriculture Crop Production in India**)<br><br>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

This six-week internship journey into the realm of data science and machine learning, focused on the pivotal project of predicting agriculture crop production in India.

During this internship, I delved into the intricate process of analyzing vast datasets comprising historical crop yields, weather patterns, soil quality, and socio-economic factors. Utilizing cutting-edge machine learning algorithms, I sought to develop predictive models capable of forecasting crop production with precision.

This internship provided by USC/UCT with a dynamic learning environment, for leveraging technology to address real-world challenges. As I reflect upon the culmination of this internship journey, I am profoundly grateful for the opportunity to immerse myself in such a meaningful and impactful project.

**Project name - Prediction of Agriculture Crop Production in India.**

Agriculture is the foundation of India's economy, employing millions and ensuring food security for its enormous population. However, there are certain obstacles in this field. Crop production is influenced by a variety of factors, including unexpected weather patterns and soil health variations, making it a complex problem to solve. In this setting, the merger of data science and machine learning stands out as a beacon of light, offering potential insights and solutions to improve agricultural efficiency and output.

I gained invaluable practical experience in data preprocessing, feature engineering, model selection, and evaluation techniques. Moreover, I honed my critical thinking skills, learning to navigate through complexities, troubleshoot errors, and iteratively improve the efficacy of my models.

I am profoundly grateful for the opportunity given by USC/UCT to immerse myself in such a meaningful and impactful project. The knowledge gained, challenges overcome, and relationships forged have undoubtedly shaped me into a more adept and empathetic.

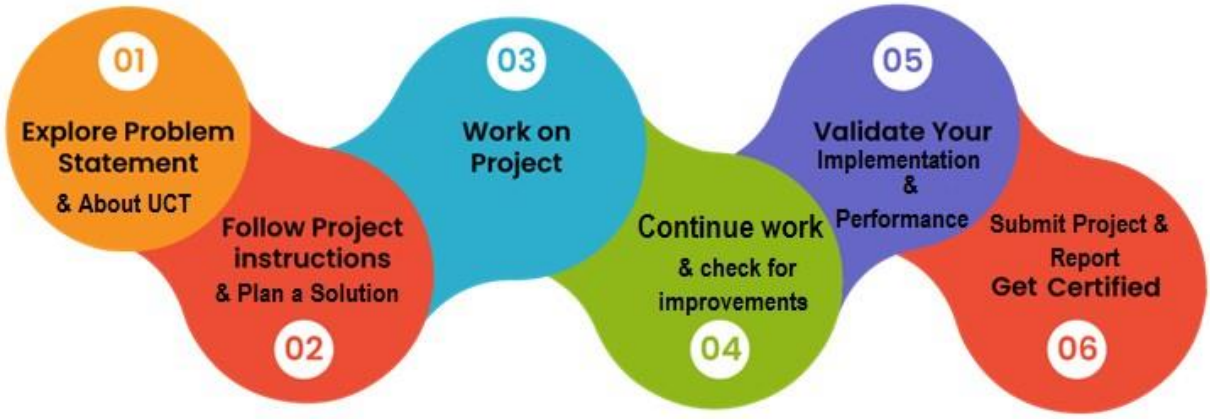
It is with great enthusiasm and optimism that I embark on the next phase of my journey, carrying forward the lessons learned and the experiences gained during this transformative internship.

**Program Planning:**

The program was meticulously planned, offering a structured framework that seamlessly integrated theoretical knowledge with practical application. From foundational concepts in data science to advanced machine learning algorithms, each week was thoughtfully crafted to build upon the previous learning, culminating in a comprehensive understanding of predictive analytics.

**Learnings:**

The journey unfolded with an exploration of data preprocessing techniques, where I learned to cleanse, transform, and prepare raw agricultural data for analysis. As I progressed, I delved into the intricacies of regression analysis, time series forecasting, and ensemble learning methods, honing my skills in developing predictive models tailored to the unique dynamics of crop production.




Moreover, the internship provided hands-on experience with popular tools and libraries such as Python, TensorFlow, and scikit-learn, empowering me to translate theoretical concepts into tangible solutions.

To my juniors and peers going on similar journeys, I advise them to embrace the opportunity to learn and grow, as every obstacle is an opportunity in disguise. Approach each assignment with curiosity and determination, and never underestimate the power of collaboration and mentorship to propel your achievement.

As you negotiate the complexities of data science and machine learning, remember that failure is not a setback, but rather a step toward mastery. Stay resilient, curious, and, most importantly, passionate about making a positive change in the world through your work.

# 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 ( Insight )

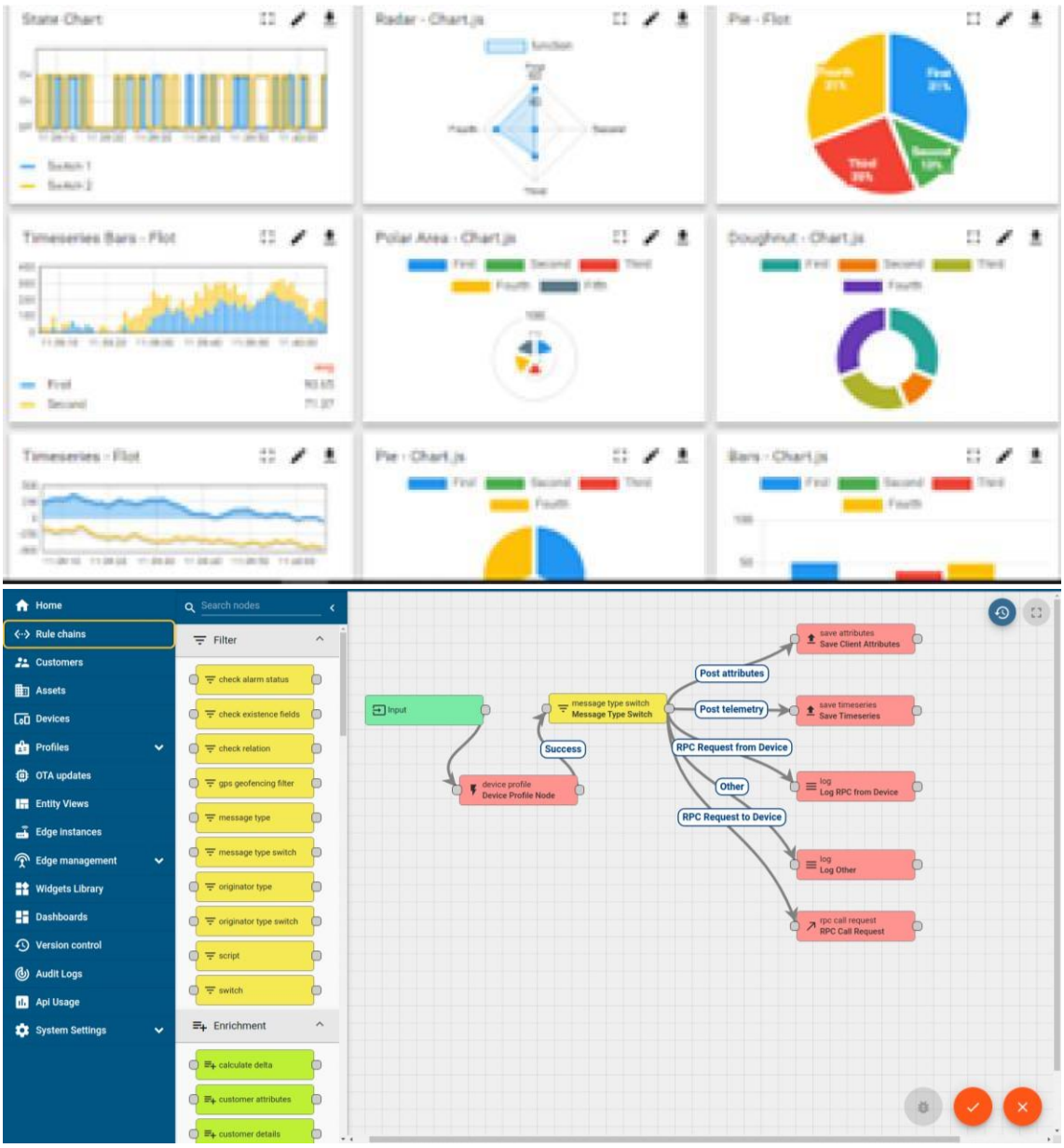
**UCT Insight** is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

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

- It supports both cloud and on-premises deployments.

It has features to

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




## ii. **Smart Factory Platform (** FACTORY WATCH **)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

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

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

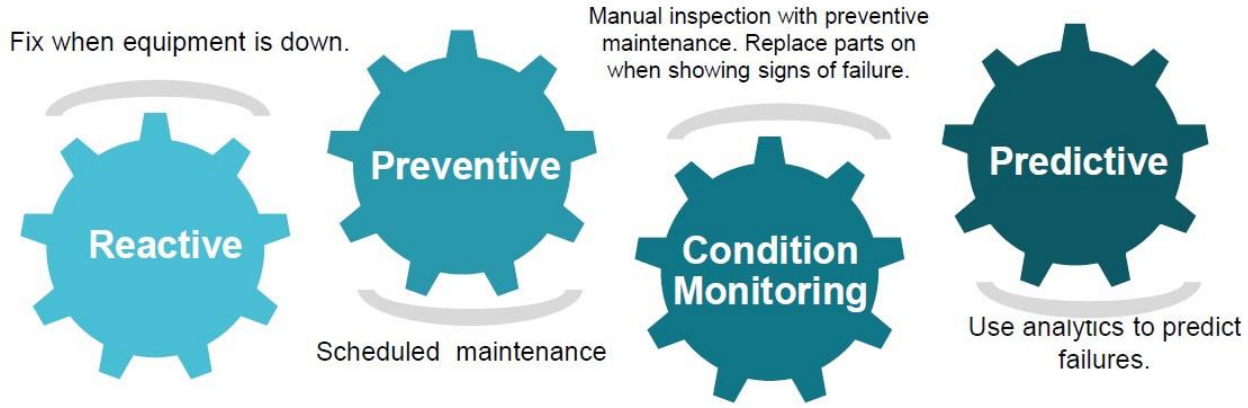

| Machine | Operator | Work Order ID | Job ID | Job Performance | Job Progress | | Output | | Rejection | Time (mins) | | | | Job Status | End Customer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Start Time | End Time | Planned | Actual | | Setup | Pred | Downtime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

iii. LoRaWAN based Solution

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

## Predictive Maintenance

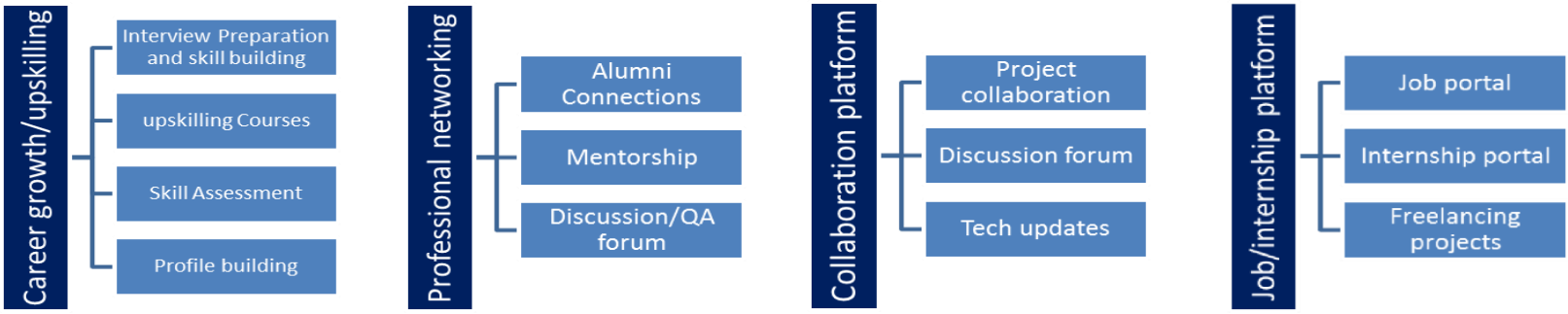
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.




### 2.2 About upskill Campus (USC)

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

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




## 2.3 The IoT Academy

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

## 2.4 Objectives of this Internship program

The objective for this internship program was to ☛

get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

## 2.5 Reference

[1] Research papers from google scholar:

- Paper 1-Analysis of agricultural crop yield prediction using statistical techniques of machine learning.
- Paper 2-Prediction of Crop Yield using Regression Analysis
- Paper 3-Prediction of Crop Production in India Using Data Mining Techniques

[2] Sharma, S., Kumar, V., & Nathawat, M. S. (2018). Crop yield prediction using machine learning techniques: A review. International Journal of Computer Applications, 180(9), 7-12.

[3] Dash, P. K., & Sahoo, S. (2016). Machine learning approaches for crop yield prediction: An empirical comparison. Agricultural Systems, 147, 110-120.

[4] Sudharsan, D., & Jayanthi, V. (2019). A survey on crop yield prediction techniques using machine learning algorithms. International Journal of Advanced Research in Computer Engineering & Technology, 8(9), 305-310.

[5] Khanam, S., & Singh, A. K. (2020). Prediction of crop yield using machine learning techniques: A case study in India. International Journal of Agricultural and Environmental Information Systems, 11(3), 49-64.

[6] https://learn.upskillcampus.com/s/courses/65c4b16ee4b0071d578afa93/take

[7] Official Python Documentation.

# 3 Problem Statement

The agriculture sector in India is of paramount importance, serving as the primary source of livelihood for a significant portion of the population and contributing substantially to the nation's economy. In India, the agriculture industry is critical, providing a primary source of income for a large section of the people and making considerable contributions to the country's economy. The Problem Statement focuses on predicting crop yields in India under various climate conditions and features. The goal of this project is to assist farmers in selecting the most suitable crop to produce in order to achieve the desired yield and profit. Crop production prediction is critical at this time for picking the appropriate crop.

# 4 Existing and Proposed solution

Existing proposed solutions for the prediction of agricultural crop production in India through data science encompass a variety of approaches, each leveraging different data sources, methodologies, and technologies. Some of the prominent solutions include:

**Statistical Modeling**: Traditional statistical methods such as linear regression, time series analysis, and autoregressive integrated moving average (ARIMA) models have been widely used to forecast crop production based on historical data. These models often incorporate factors such as weather patterns, soil characteristics, and historical yield data to make predictions.

**Machine Learning Algorithms:** Machine learning techniques offer a more sophisticated approach to crop production prediction by leveraging advanced algorithms to analyze large and complex datasets. Supervised learning algorithms like decision trees, random forests, support vector machines (SVM), and neural networks can be trained on historical data to predict future crop yields based on various input features.

**Proposed Solution:**

The agriculture sector in India stands as the backbone of its economy, supporting millions of livelihoods and feeding a burgeoning population. However, it faces multifaceted challenges ranging from unpredictable weather patterns to market fluctuations.

Recognizing the necessity of addressing these critical concerns, our suggested approach seeks to give a realistic solution by properly predicting agricultural output rates. Our approach aims to determine crop yield (Quintal/Hectare) by utilizing cutting-edge techniques and taking into account relevant aspects such as cultivation costs and production costs. This proactive method enables farmers to foresee hazards, adjust cultivation operations, and maximize crop yields, reducing the negative impact of crop failures and supporting sustainable agricultural practices.

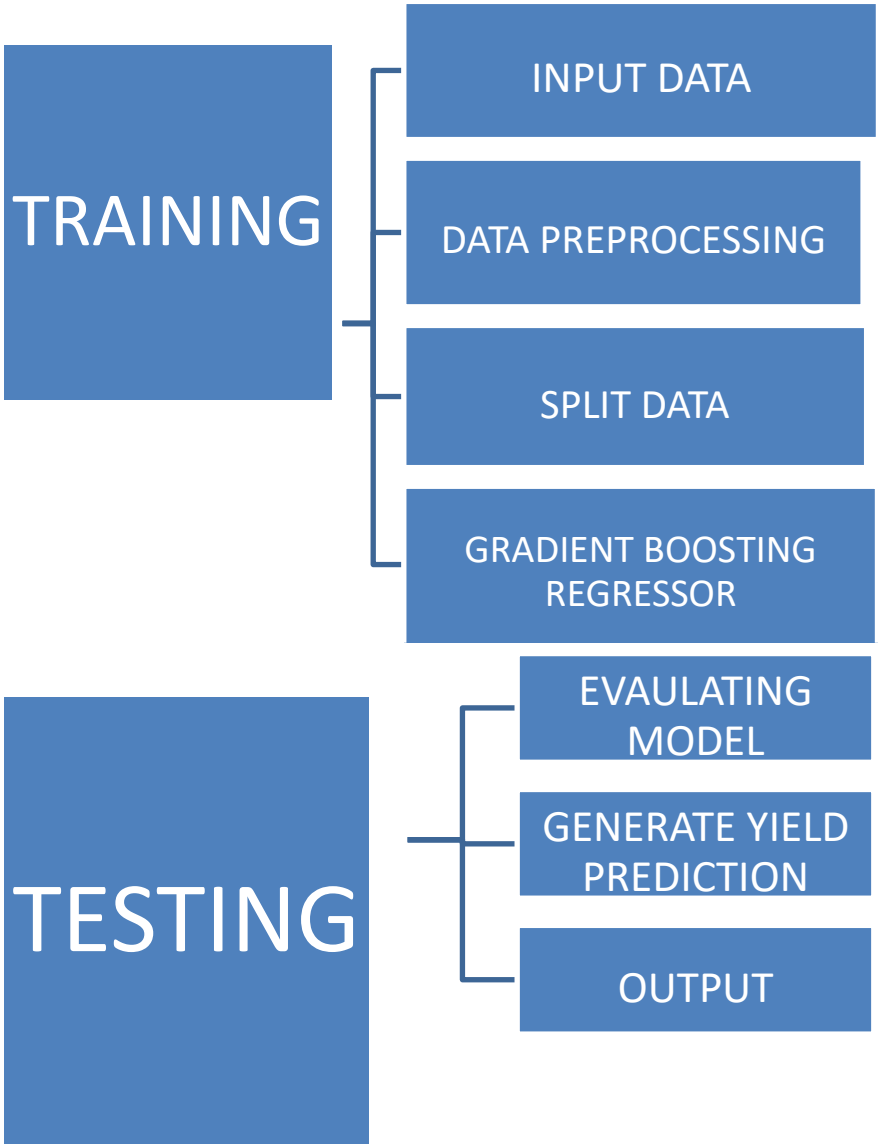
## 4.1 Code submission

(https://github.com/SamreenKhanam/upskillcampus/blob/main/Prediction%20of%20Agriculture%20Crop%20Production%20in%20India.ipynb)

## 4.2 Report submission

(https://github.com/SamreenKhanam/upskillcampus/blob/main/PredictionofAgricultureCropProductioninIndia_Samreen_USC_UCT.pdf

## 5 Proposed Design/ Model




## 6 Performance Test

Performance Test Prediction of Agricultural Crop Production in India

Predictive models must be implemented in order to accurately anticipate agricultural crop production in India. Using advanced machine learning techniques and tools from the'scikit-learn' package, we went on a

journey to construct a robust model capable of projecting crop yields accurately. Among the approaches used, the 'train_test_split' function from'sklearn.model_selection' was critical in determining the success of our predictive models.

The 'train_test_split' method made it easier to divide our dataset into two subsets: one for training the model and one for testing its performance. By randomly partitioning the data and allocating a higher proportion to the training set, we guaranteed that the model learnt patterns from a substantial quantity of data while maintaining unseen samples for evaluation.

To prepare our data for modeling, we used a variety of preprocessing techniques, including one-hot encoding and standardization, which were made easier by the'sklearn.preprocessing' capabilities.

Furthermore, the 'ColumnTransformer' class from'sklearn.compose' allowed for smooth integration of various preprocessing stages, resulting in efficient transformation of diverse feature types inside our dataset. This modular approach reduced the preprocessing pipeline and improved reproducibility across studies.

After properly preparing the data, we investigated a variety of regression techniques, including **Linear Regression**, **RandomForestRegressor**, **GradientBoostingRegressor**, and **SVR (Support Vector Regressor)** from'sklearn.linear_model','sklearn.ensemble', and'sklearn.svm'. Each algorithm had unique advantages and assumptions, resulting in a comprehensive set of models to investigate.
We evaluated the performance of these models on the test set using rigorous experimental and crossvalidation approaches, including measures such as mean squared error (MSE), mean absolute error (MAE), and R-squared (R2) score. This iterative procedure enabled us to determine the optimal model architecture and hyperparameters for predicting crop yield in the Indian agricultural environment.

Finally, by integrating 'train_test_split' with additional capabilities from the'scikit-learn' package, we were able to construct and analyze predictive models for agricultural crop production in India. By utilizing these tools and approaches, we hope to contribute to the progress of data-driven decision-making in agriculture, ultimately fostering sustainable and resilient food systems in India and elsewhere.

## 6.1 Test Plan/ Test Cases

Crop Yield Prediction: Using the user's input, this module returns the expected crop yield (Quintal/Hectare). The system estimates crop yield according to inputs such as crop state, cultivation cost ,and production cost, and returns an expected yield as output.

## 6.2 Test Procedure

Step 1: Enter the Crop Name.

Step 2: Request input from the user for State, Cost of Cultivation ('/Hectare), Cost of Cultivation ('/Hectare) C2, and Cost of Production ('/Quintal) C2.

Step 3: With these inputs, the model predicts crop yield (Quintal/Hectare) in the selected state.

### 6.3 Performance Outcome

- Linear Regression - 0.9100289346877166
- Random Forest Regressor - 0.9375020649893857
- Gradient Boosting Regressor - 0.9618315623920866

## 7 My learnings

During my six-week internship in data science and machine learning, I had the opportunity to work on a project focused on predicting agricultural crop production in India. This project not only exposed me to the intricate world of agriculture but also allowed me to apply various data science techniques and machine learning algorithms to solve real-world problems.

By applying data science techniques and machine learning algorithms to real-world problems, I not only enhanced my technical skills but also gained a deeper appreciation for the role of data-driven solutions in addressing societal challenges. This internship has inspired me to continue exploring the intersection of data science

## 8 Future work scope

As we look towards the future, the potential for enhancing predictive models for agricultural crop production in India is vast. Building upon the foundation laid by our existing model development efforts, several avenues for further research and innovation emerge. Additionally, the integration of a user interface (UI) adds a layer of accessibility and usability.