Industrial Internship Report on

” Prediction of Agriculture Crop Production in India”

Prepared by

Sushil Kumar B

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 (in the domain of data science and machine learning domain and the title of the project is **Prediction of Agriculture Crop Production** in India which gives the yield of the crop depending upon the cost of cultivation.) 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

In this we will go through the concepts of the data science and machine learning. The Agriculture sector is very important and we should plan accordingly where to plant the crop at what time because of the rapid growth of the population in India.

Planting of the crop in the right place will help us enhance the productivity and yield of the crop. Through the help of data-driven techniques we can perform various analysis and the prediction on the given data for example the yield of the crop can be predicted based on the rainfall and the rainfall can be determined using the forecasting mechanisms through which it can provide the data scientist and the farmers the informed decision on when to grow the crop.

Data Science and Analysis involves various techniques such as data cleaning, data transforming, merging of the data, data preprocessing and model building and training of the model with the help of which we can perform the analysis of the data in the Agriculture Sector. With the help of which we can solve the issues regarding the crop production in India, in this we train the machine by applying various algorithms which efficiently fits to our data set. We can also analyze the data with the help of visualization.

By working on with this real world projects we can enhance our knowledge and skills for the role of a data scientist and analyst.

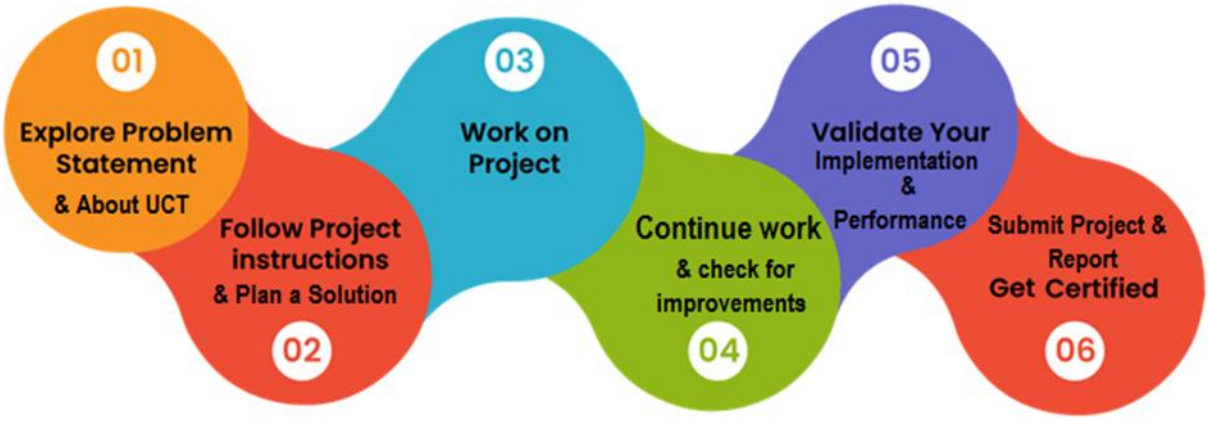
This use case study highlights the potential to create meaningful solutions that can positively impact India's agricultural sector. By harnessing the predictive capabilities of the developed model, stakeholders can make well-informed decisions to address

challenges like food security, market fluctuations, and resource optimization, leading to sustainable growth in the agricultural landscape.

Opportunity given by USC/UCT.

Participating in a data science project or internship has given valuable opportunities for skill development. These experiences provide hands-on exposure to essential technical skills like Python, R, data preprocessing, and machine learning, enabling individuals to become proficient in data handling and analysis. In fact the data scientist gain knowledge in statistical analysis and data visualization, enhancing their ability to draw meaningful insights from data. On successful completion of data science projects or internships enhances resumes but also provides a competition in the job market, because data science is being used in all the fields and people can get employed easily if they have the required skill set.

Program Planning:



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)

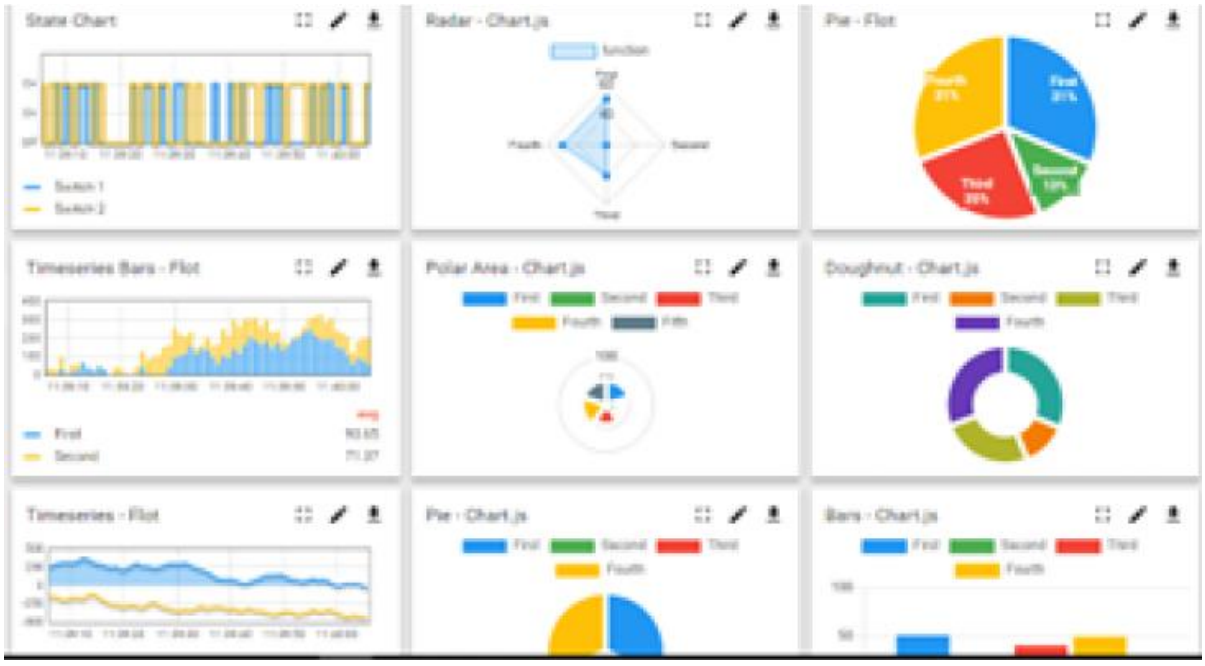
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been

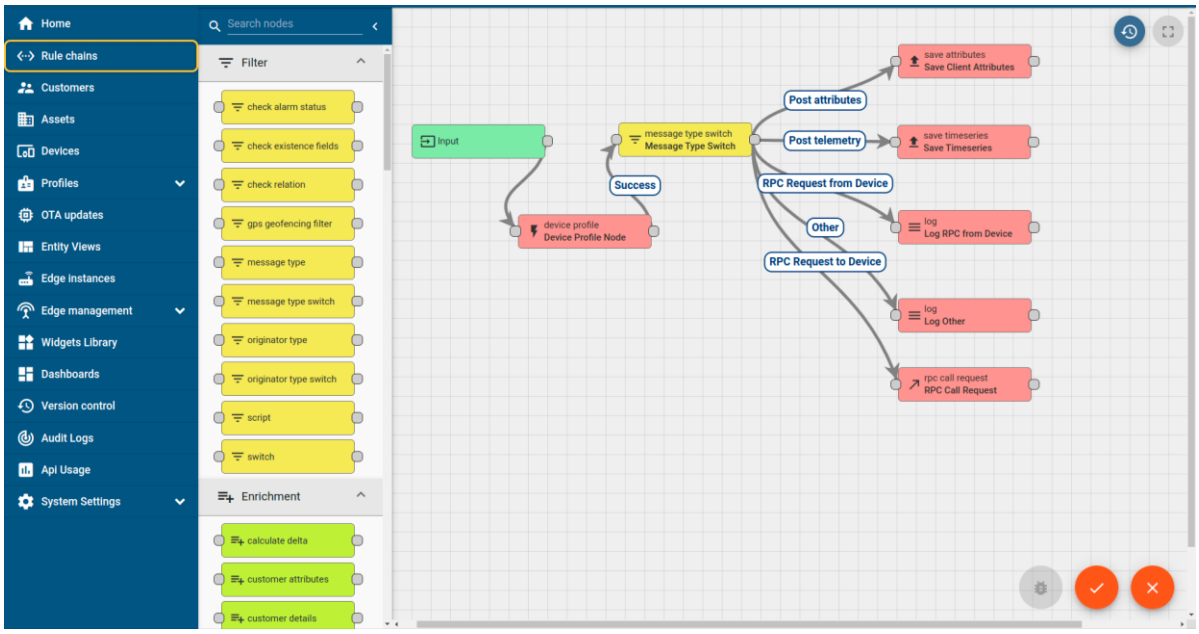
built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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
- 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.

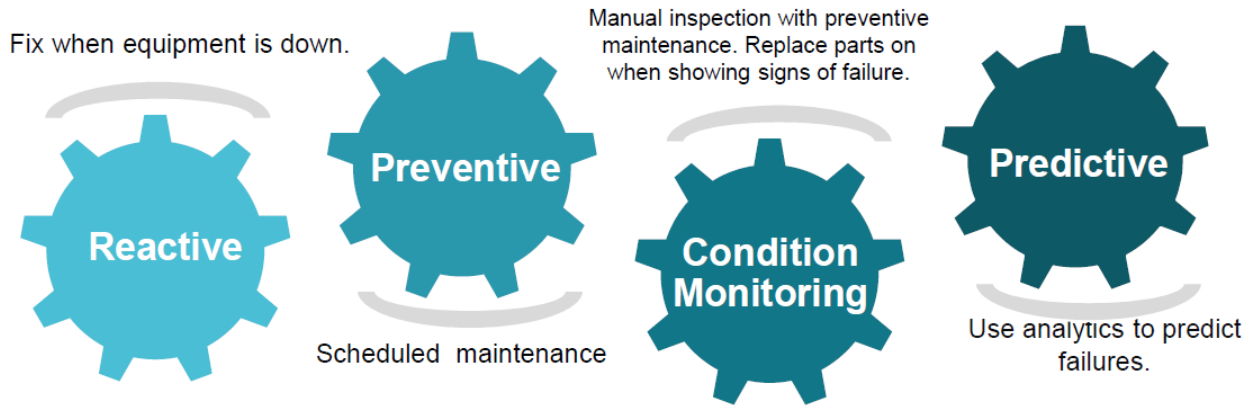
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iii. based Solution

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

iv. Predictive Maintenance

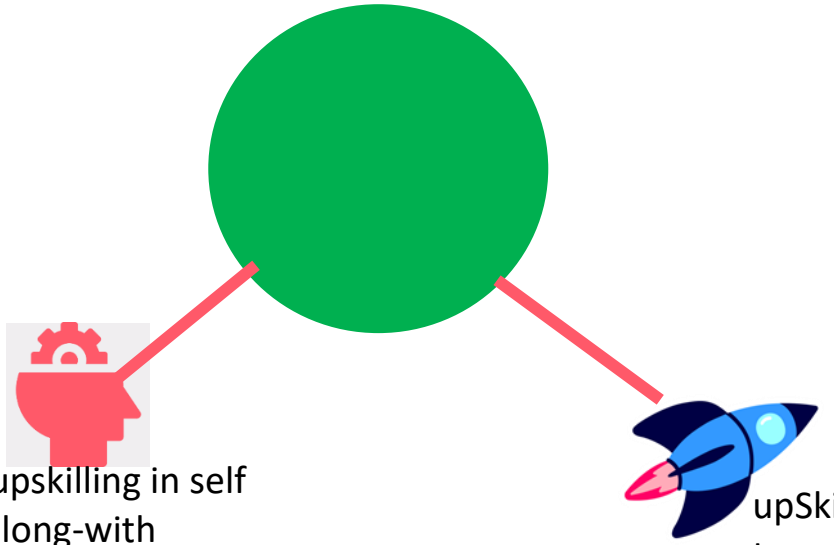
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

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

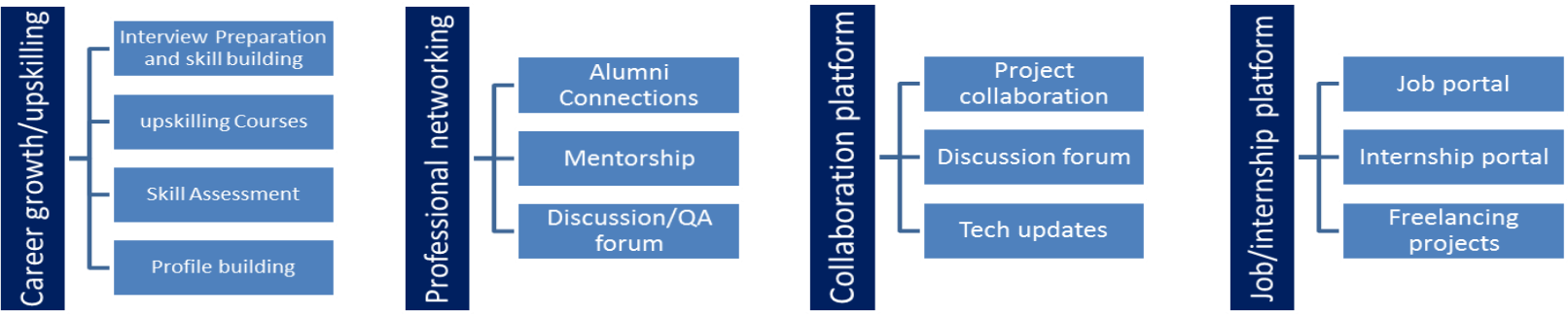
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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

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

https://www.upskillcampus.com/



2.3 The IoT Academy

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

2.4 Objectives of this Internship program

The objective for this internship program was to

- ☛ get practical experience of working in the industry.
- ☛ to solve real world problems.
- ☛ to have improved job prospects.
- ☛ to have Improved understanding of our field and its applications.
- ☛ to have Personal growth like better communication and problem solving.

2.4 Reference

[1] Chrome

[2] You Tube

[3] Other social media platforms

2.5 Glossary

Terms	Acronym
Digital transformation	It the process of using digital technology to create new or modified existing business processes and customer needs.
Predictive Maintenance	Predictive maintenance for rotating machinery is gaining prominence as plant operators embrace analytics and learn how to approach their operating benchmarks
UCT	Uniconverge Technology

3 Problem Statement

In the assigned problem statement ,we predict agriculture crop production in India for major crops and based on the data sets and relevant agricultural indicators like the soil, temperature, cost of the cultivation in the area, depending upon the state and their temperature etc. The prediction will help practitioners ,researchers and farmers make informed decisions regarding crop planning, resource collection, and market dynamics. Various machine learning algorithm like the usage of the regression and random forest techniques can be applied to predict the crop production. By this prediction the overall agricultural productivity in India can improve, contributing to food security and economic growth. By the data analysis we can optimally make use of the resources.

4 Existing and Proposed solution

Summary of existing solutions provided by others for the prediction of agriculture crop production in India using data science .

- Linear regression: This is a simple but effective algorithm that can be used to predict crop yields based on a linear relationship between the predictors and the outcome.
- Random forest: This algorithm is a more complex ensemble method that combines multiple decision trees to improve the accuracy of predictions.
- Support vector machines (SVMs): This algorithm can be used to predict crop yields by finding the best hyperplane that separates the data into two classes (high yield and low yield).
- Neural networks: This algorithm is a more complex machine learning model that can learn non-linear relationships between the predictors and the outcome.

Data science and machine learning can be used to predict agriculture crop production and help farmers to make better decisions about crop planning and management. By analyzing historical data and weather forecasts, machine learning algorithms can identify the factors that are most likely to affect crop yields. This information can then be used to develop models that can predict crop yields for future seasons.

Additional considerations

In addition to the factors mentioned above, there are a number of other factors that can affect agriculture crop production. These include:

- The type of crop being grown
- The soil type
- The amount of rainfall
- The amount of sunlight
- The presence of pests and diseases
- The use of fertilizers and pesticides
- The agricultural practices used by the farmer

It is important to consider all of these factors when developing a machine learning model to predict agriculture crop production. By taking into account all of the relevant factors, machine learning algorithms can be used to develop more accurate and reliable predictions.

4.1 Code submission (Github link):

https://github.com/Sushil69alpha/upskillcampus/blob/main/PredictionOfAgricultureCropProduction.ipynb

4.2 Report submission (Github link) :

https://github.com/Sushil69alpha/upskillcampus/blob/main/PredictionOfAgricultureCropProduction_SushilKumarB_USC_UCT.pdf

5 Proposed Model

The first step in predicting agriculture crop production is to collect data on the factors that are likely to affect crop yields. This data can include historical crop yields, weather data, soil data, and information on agricultural practices. The data should be collected from a variety of sources, such as government agencies, agricultural research institutes, and weather forecasting organizations.

Once the data has been collected, it needs to be prepared for analysis. This involves cleaning the data to remove errors and inconsistencies, and transforming the data into a format that can be used by machine learning algorithms

5.1 High Level Diagram (if applicable)

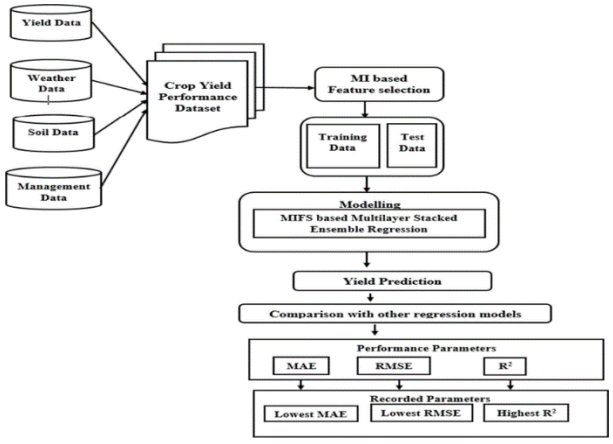
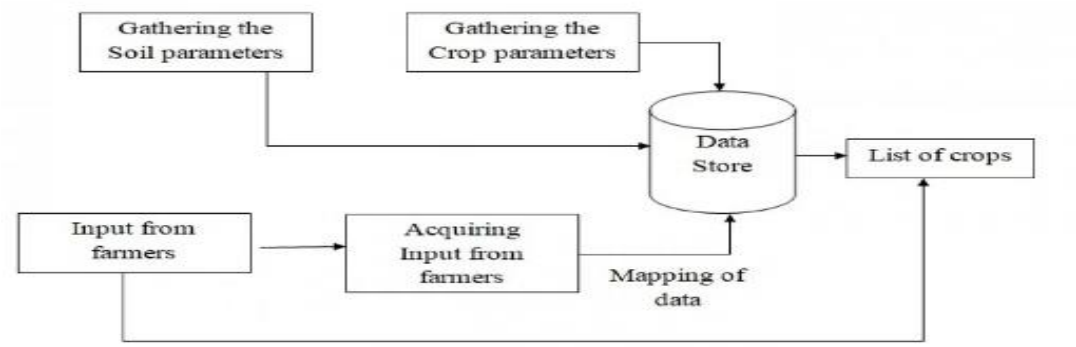


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)



5.3 Interfaces (if applicable)

Google Colab Data Science Project for Crop Production Prediction in India:

1. Setup:

- Create a new Colab notebook with Python 3 runtime.

2. Data:

- Import libraries (Pandas, NumPy, Matplotlib).
- Acquire and upload relevant datasets (crop, weather, soil) to Colab.

3. Data Preprocessing:

- Handle missing data.
- Encode categorical variables.

- Normalize or scale features.
- Perform feature engineering.

4. Model Building:

- Split data into training and testing sets.
- Choose a machine learning algorithm (e.g., Linear Regression, Random Forest, LSTM).
- Train and evaluate the model (use MAE, RMSE).

5. Model Tuning:

- Adjust hyperparameters.
- Implement cross-validation.

6. Model Deployment:

- Deploy the model using Flask or another method.
- Create a user-friendly interface for input and predictions.

7. Visualization and Reporting:

- Visualize predictions and insights.
- Generate reports for users.

6 Performance Test

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

Here we need to first find the constraints.

How those constraints were taken care in your design?

What were test results around those constraints?

Constraints can be e.g. memory, MIPS (speed, operations per second), accuracy, durability, power consumption etc.

In case you could not test them, but still you should mention how identified constraints can impact your design, and what are recommendations to handle them.

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

X = mr[['Yield (Quintal/ Hectare) ','Cost of Cultivation (`/Hectare) A2+FL', 'Cost
of Cultivation (`/Hectare) C2']]
Y = mr['Production 2006-07']

# Split the dataset into training and testing sets
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2,
random_state=42)

# Create a linear regression model
model = LinearRegression()

# Train the model
model.fit(X_train, Y_train)

# Make predictions on the testing set
Y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(Y_test, Y_pred)
```

```
rmse = mse**0.5
print("Root Mean Squared Error:", rmse)
```

6.1 Test Plan/ Test Cases

1. Objective

Clearly define the goal of the testing process, such as evaluating the accuracy and reliability of the crop production prediction model.

2. Test Data

Identify the dataset you will use for testing. Ensure that it is representative of the real-world agricultural conditions and includes relevant features such as weather data, soil quality, crop type, and historical production data.

3. Preprocessing

Ensure that data preprocessing steps are well-documented and tested. This includes data cleaning, feature engineering, and data transformation.

4. Model Selection

Specify the machine learning model(s) you will use for crop production prediction. This may include regression models, decision trees, neural networks, etc.

5. Test Cases

Define specific test cases that cover various aspects of model evaluation. Below are some example test cases:

Test Case 1: Data Splitting

Objective: Ensure the proper splitting of data into training, validation, and test sets.

Test Case 2: Model Training

Objective: Validate that the model training process is working correctly.

Test Case 3: Model Evaluation

Objective: Assess the model's performance using various evaluation metrics.

6.2. Test Procedure

1. Define the Test Objectives:

Clearly specify the goals and objectives of the testing procedure. Determine what aspects of crop production prediction you want to evaluate and what success criteria you will use.

2. Data Preparation:

Ensure that you have a clean, representative, and well-structured dataset. This may involve data cleaning, preprocessing, and feature engineering. Make sure the dataset includes features like weather data, soil quality, historical production data, and the target variable (crop yield).

3. Data Splitting:

Split the dataset into three subsets: training, validation, and test sets. The typical split ratio is around 70% training, 15% validation, and 15% test, but this can vary based on your dataset size and specific requirements.

4. Model Selection:

Choose the machine learning model(s) that will be used for crop production prediction. This could include regression models (e.g., linear regression, random forests), neural networks, or other suitable algorithms.

5. Model Training:

Train the selected model(s) using the training dataset. Ensure that the training process is well-documented and that the following steps are taken:

Implement proper cross-validation techniques, such as k-fold cross-validation, to assess model stability.

Regularly monitor training loss and performance metrics to detect issues like overfitting.

6. Model Evaluation:

Evaluate the model's performance using the validation dataset. Use appropriate evaluation metrics for regression tasks, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

6.2 Performance Outcome

1. Accuracy: Measures the proportion of correctly classified instances out of the total.

2. Precision: Measures the proportion of true positives (correctly predicted positive cases) out of all predicted positives.

3. Recall (Sensitivity): Measures the proportion of true positives out of all actual positives.

4. F1 Score: A balance between precision and recall, useful when dealing with imbalanced datasets.

5. ROC AUC: Area under the Receiver Operating Characteristic curve, which quantifies the model's ability to distinguish between classes.

6. Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual values for regression tasks.

7. Mean Squared Error (MSE): Measures the average squared difference between predicted and actual values for regression tasks.

8. R-squared (R^2): Measures the proportion of variance in the target variable explained by the model for regression tasks.

9. Confusion Matrix: A table that shows true positives, true negatives, false positives, and false negatives, providing a detailed view of classification performance.

It gives the crop production depending upon the various feature sets.

7. My learnings

In this internship I have learnt about the role of data scientist, and analyst and the various techniques that are to be used in the data prediction and the data Science techniques like data cleaning data preprocessing and how to train the machine with the help of algorithms like Decision tree, Random Forest and regression algorithms.

And also how to deploy the model and deal with the train and the test dataset, like which variables needs to taken for testing and training the model and also checking the accuracy of the datasets.

Doing this opportunity I have got to learn about how to deal with the real world problems and has given me a hands-on. It has in fact provided me the knowledge and skills and also helped in the personal growth because of which I have been able to grasp the skills and has helped in finding the proper solution to a particular problem.

Working on this hands-on project work has made me realize how the data science field is needed and if any wrong assumption or incorrect data can lead to the wrong prediction. Also learnt about the creativity skills ,mathematics etc. that are required ,where and how they are to be applied. Also understood that how things will be automated like the AI and the importance of machine learning is going to be useful.

8.Future work scope

The future work scope across various industries is poised for significant transformation as we embark on an era of rapid technological advancement in various fields like Artificial Intelligence (AI) Integration: AI will permeate every sector, driving automation, data-driven decision-making, and personalized experiences. Quantum Computing: As quantum computing matures, it will tackle complex problems currently deemed intractable, revolutionizing cryptography, materials science, and optimization.5G and IoT: The rollout of 5G networks will enable widespread adoption of the Internet of Things, enhancing connectivity and data collection in smart cities, healthcare, and more. Biotechnology and Genomics: Advancements in gene editing, personalized medicine, and biomanufacturing will reshape healthcare and agriculture. Space Exploration: Collaborative efforts among public and private entities will fuel space exploration, with missions to the moon, Mars, and beyond. Blockchain and Cryptocurrency: These technologies will disrupt finance, supply chains, and governance, with the potential for decentralized systems and digital currencies. E-commerce and Retail: Enhanced online shopping experiences, including augmented reality try-ons and hyper-personalization, will drive the future of retail. Agricultural Technology: Precision agriculture, automation, and data analytics will optimize food production and reduce waste.