Industrial Internship Report on

Prediction of Agriculture Crop Production in India

Prepared by

Jakkamsetti Navadeep

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 designed to provide hands-on exposure to industry-oriented technologies and equip learners with practical experience in solving real-world problems using data-driven approaches. I had to finish the project including the report in 6 weeks' time.

The internship centered around a problem statement assigned by UCT, which required learners to analyze, design, and implement an end-to-end solution within the span of six weeks. My assigned project was **"Prediction of Agriculture Crop Production in India,"** where I developed a machine-learning-based predictive model to estimate crop yield using historical agricultural statistics, climatic parameters, and seasonal trends.

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.

Throughout this internship program, I gained valuable insights into industrial workflows, data preprocessing methods, model development techniques, and the importance of accurate prediction systems for agriculture. This opportunity significantly enhanced my technical knowledge, analytical thinking, and problem-solving abilities. Overall, it was an enriching and rewarding experience that contributed greatly to my professional growth.

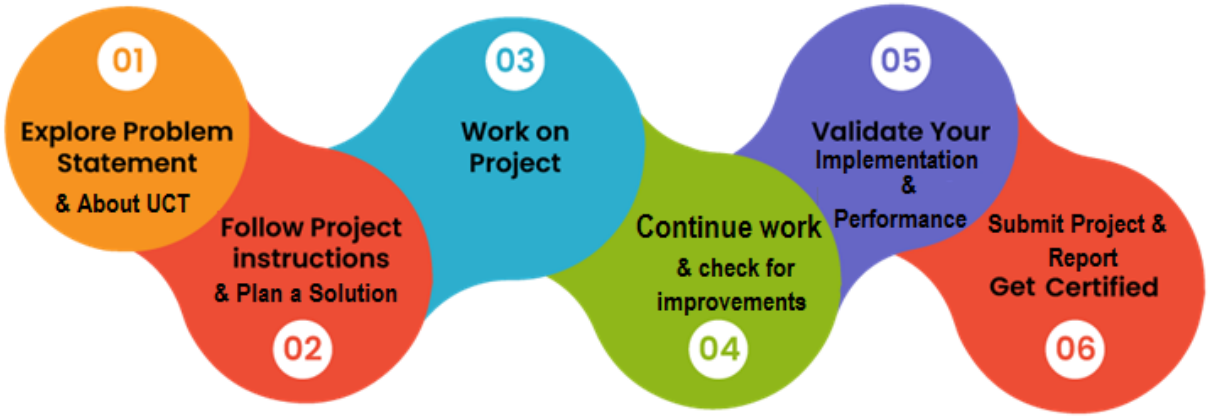
TABLE OF CONTENTS

1 Preface

This six-week industrial internship offered by *upSkill Campus* in collaboration with *UniConverge Technologies Pvt. Ltd.* provided me with a valuable opportunity to gain real-world exposure in data science and machine learning. Throughout the internship, I worked on the project titled **"Prediction of Agriculture Crop Production in India"**, which focuses on building a predictive model that estimates crop yield based on historical agricultural, climatic, and seasonal factors.

The internship was structured with systematic weekly progress, technical training sessions, mentor guidance, and project execution phases. This well-designed plan helped me gain both theoretical understanding and practical implementation experience.

The core objective of this project was to analyze real-world datasets, preprocess the data, implement suitable algorithms, evaluate model accuracy, and generate meaningful insights that support decision-making in the agricultural domain. UpSkill Campus (USC) and UniConverge Technologies Pvt. Ltd. (UCT) provided an excellent platform for hands-on industrial exposure by offering structured training, expert mentorship, and a real industry-aligned project.



The project allowed me to study real agricultural production trends and apply machine learning models to predict crop output with improved accuracy. This experience has strengthened my technical skills, analytical mindset, and confidence in solving real-world industrial problems.

I sincerely thank **upSkill Campus**, **The IoT Academy**, and **UniConverge Technologies Pvt. Ltd. (UCT)** for providing this learning opportunity. I also express my gratitude to mentors, trainers, and coordinators who guided me throughout the internship. I would recommend this internship to juniors and peers who wish to gain industry-oriented skills and hands-on project experience.

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.

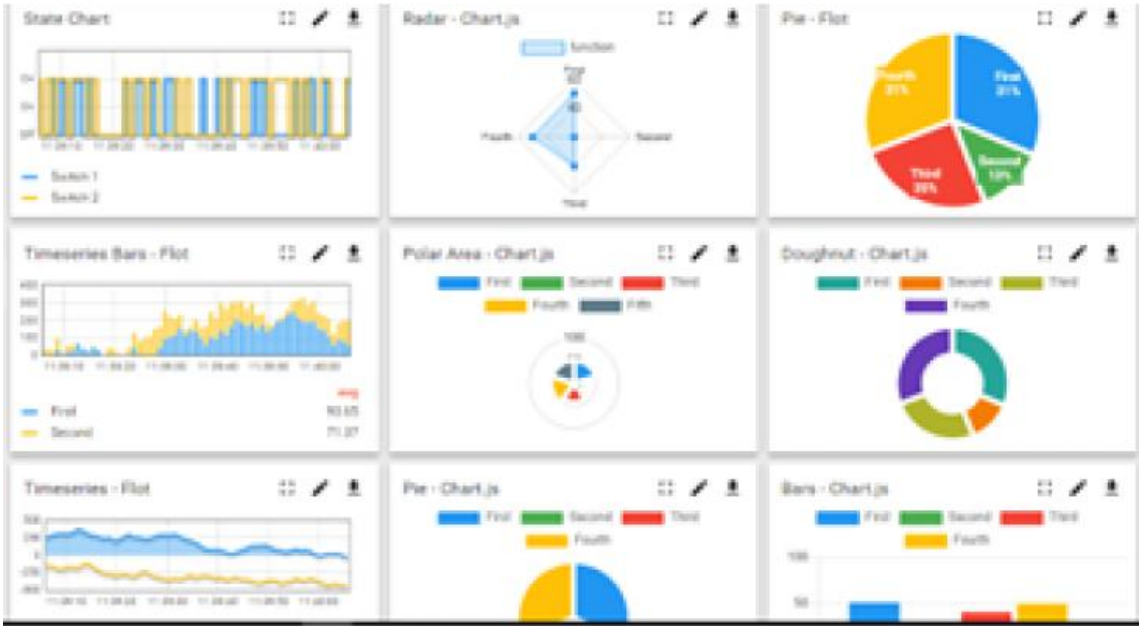
i. UCT IoT Platform (UCT Insight)

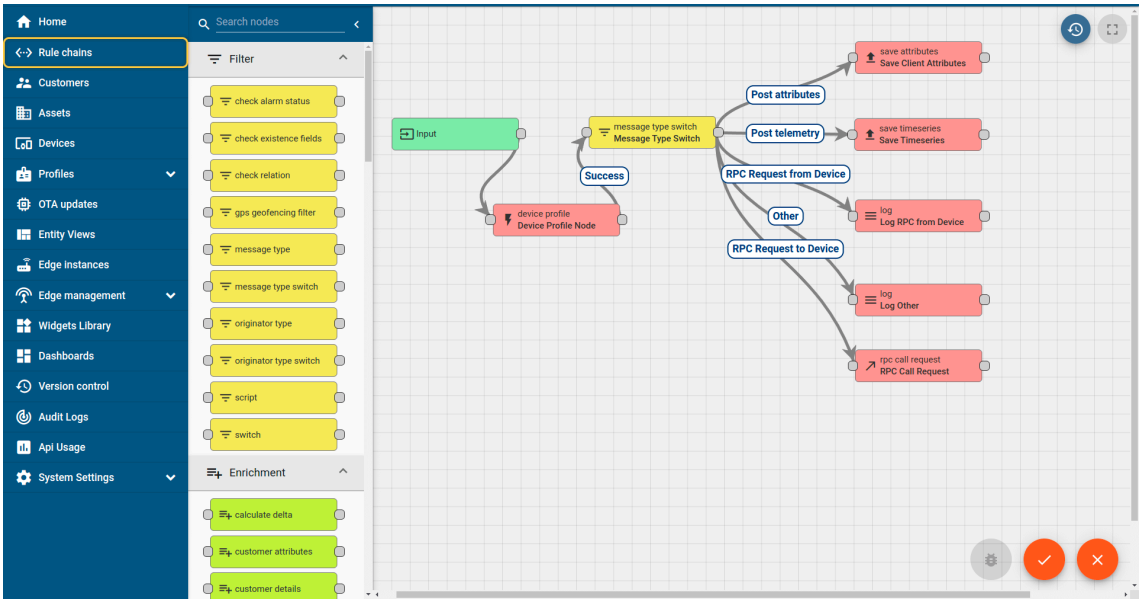
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

It's unique SaaS model helps users to save time, cost and money.

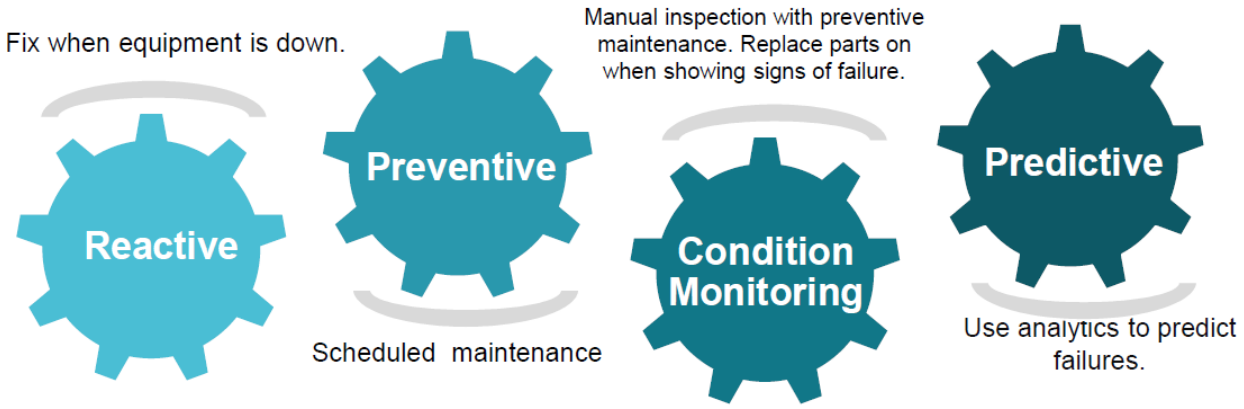
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

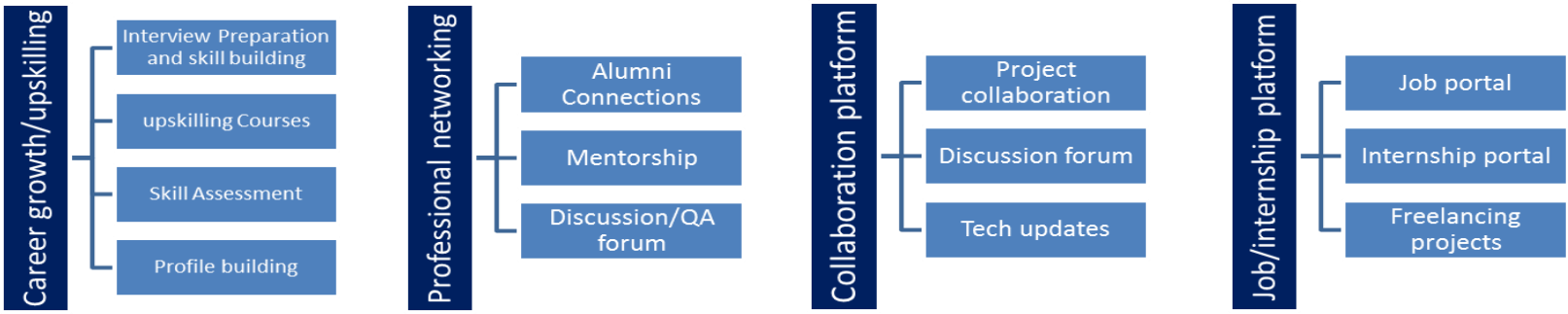
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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] Department of Agriculture, Govt. of India – Crop Production Statistics

[2] Kaggle Public Agriculture Dataset

[3] Research papers on crop yield prediction

2.6 Glossary

Terms	Acronym
ML	**Machine Learning**
RMSE	**Root Mean Square Error**
CSV	Comma Seperated Values
Rabi/ Kharif	Seasonal crop cycles
Hectare	Unit area for measuring Land

3 Problem Statement

Agriculture is one of India's largest economic sectors, and accurate crop production prediction is essential for decision-making, food security, resource management, and agricultural planning.

The goal of this project is to develop a **machine learning-based predictive model** that can estimate the production of major Indian crops based on features such as:

- Historical production data
- Area of cultivation
- Rainfall and climate attributes
- Crop type and season
- State-wise agricultural patterns

This Machine Learning model aims to support farmers, policymakers, and agricultural planners in making informed decisions.

4 Existing and Proposed solution

Existing Solutions:

Existing agricultural prediction systems include:

Traditional statistical models (ARIMA, Linear Regression)

- Government crop forecasting based on manual surveys
- Remote sensing & satellite-based predictions

Limitations:

Limitations for the agricultural prediction systems include:

- Manual surveys are slow and error-prone
- Most models have insufficient accuracy
- Seasonal variations are not captured well
- Climate change impacts are not integrated

5 Proposed Solution

A machine learning model is proposed in this project using algorithms like:

- Random Forest Regression
- Decision Tree Regression
- Linear Regression

5.1 Code submission (Github link)

https://github.com/navadeep-05/Agriculture-Crop-Production-Prediction-India/tree/main

5.2 Report submission (Github link) :

5.3 Project Requirements

What libraries to Install:

pip install pandas numpy scikit-learn xgboost joblib matplotlib seaborn

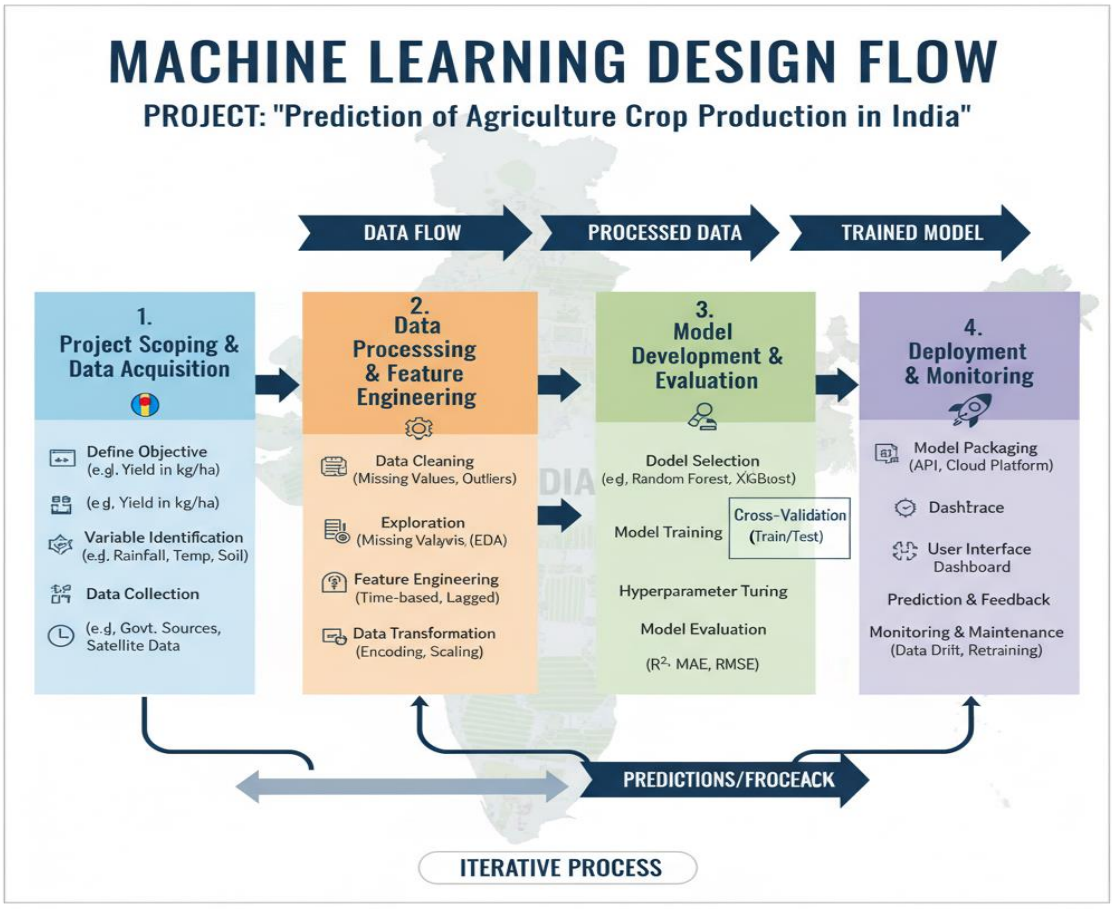
5.4 Model code

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.metrics import mean_squared_error
from sklearn.ensemble import RandomForestRegressor

# LOAD DATA
df = pd.read_csv("crop_production_data.csv")   # your dataset
target = "production"                          # your target column

X = df.drop(columns=[target])
y = df[target]

# Identify feature types
num_features = X.select_dtypes(include=[np.number]).columns.tolist()
cat_features = X.select_dtypes(include=['object']).columns.tolist()

# DATA PREPROCESSING PIPELINE
preprocess = ColumnTransformer(
    transformers=[
        ("num", "passthrough", num_features),
        ("cat", OneHotEncoder(handle_unknown='ignore'), cat_features)
```

```
    ]
)

# ----------------------------
# MODEL (RANDOM FOREST)
model = Pipeline(steps=[
    ("preprocess", preprocess),
    ("rf", RandomForestRegressor(n_estimators=200, random_state=42))
])

# TRAIN / TEST SPLIT
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

# TRAIN MODEL
model.fit(X_train, y_train)
# EVALUATE MODEL
pred = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, pred))
print("Model RMSE:", rmse)
print("Actual:", y_test.values[:5])
print("Predicted:", pred[:5])
```

6 Proposed Design/ Model

This flow diagram outlines the iterative four-stage ML process for predicting Indian crop production. It moves sequentially from **Data Acquisition** of raw inputs to **Data Processing & Feature Engineering** (cleaning and scaling), then to **Model Development & Evaluation** (training and tuning), and concludes with **Deployment & Monitoring** via a prediction system.



6.1 High Level Diagram

he system architecture for the 'Prediction of Agriculture Crop Production in India' project is structured around a **four-stage, iterative process** designed to turn raw agricultural and environmental data into actionable predictions. Crucially, the system features a **Feedback Loop** where continuous monitoring of predictions against actual results detects data drift and necessitates **Retraining** or maintenance, ensuring the system remains accurate and relevant over time. This entire structure is designed to be **iterative**, supporting continuous improvement and adaptation to changing agricultural and climatic conditions.

Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

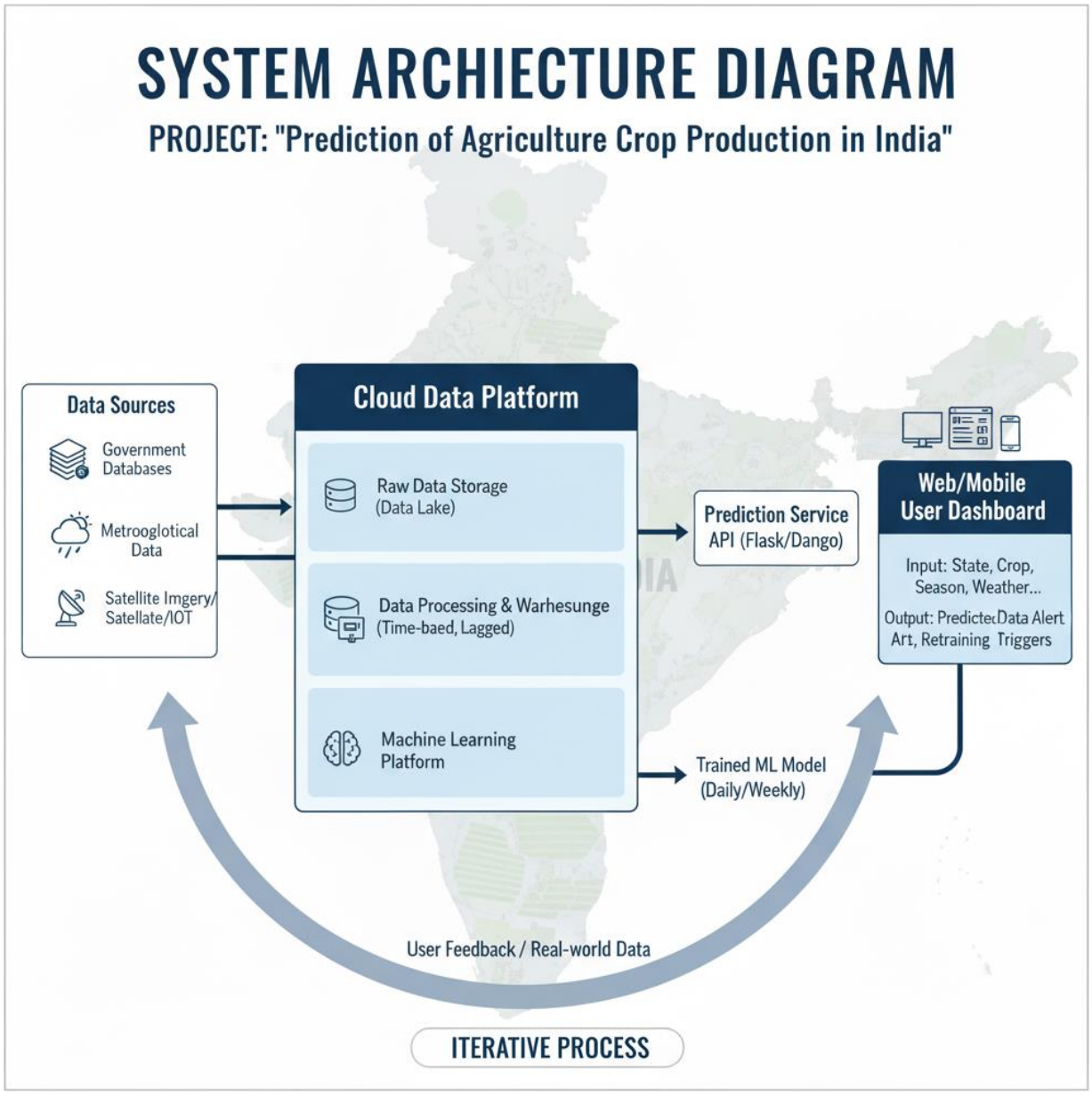
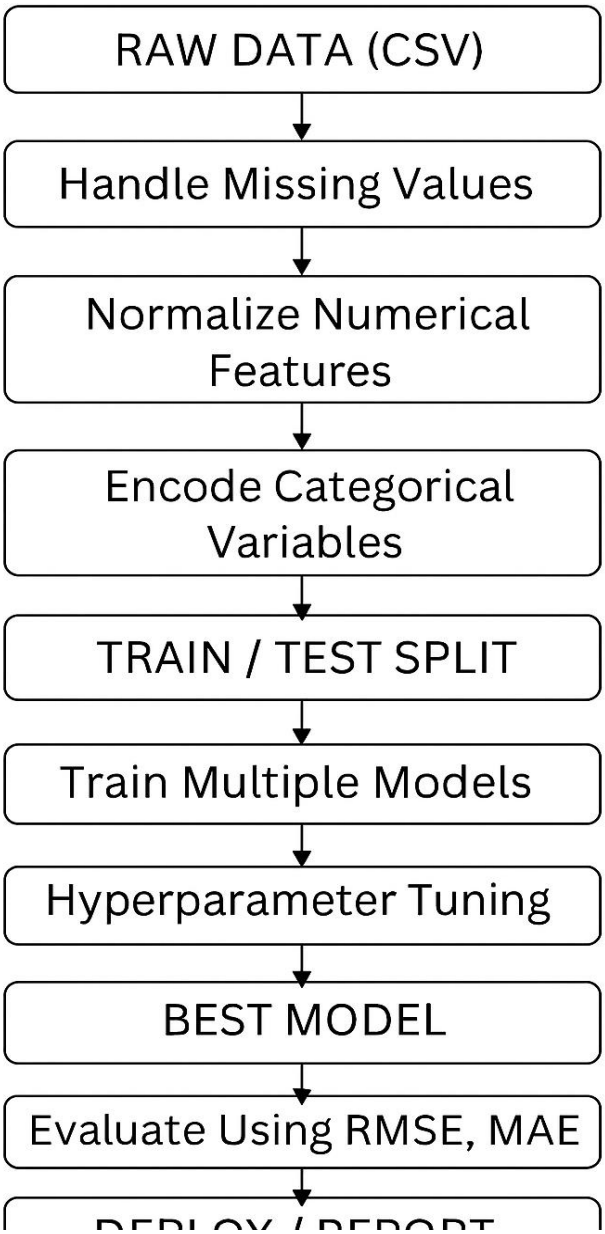


Figure 2: LOW LEVEL DIAGRAM OF THE SYSTEM



6.2 Interfaces

7 Performance Test

Constraints Identified:

- Dataset imbalance
- Seasonal variations
- High dimensionality
- Model overfitting
- Accuracy variation across states

How Constraints Were Handled:

- Performed feature selection
- Applied cross-validation
- Tuned hyperparameters
- Normalized data

7.1 Test Plan/ Test Cases

Test Case	Description	Expected Output
TC-01	Load Dataset	Dataset loads without errors
TC-02	Train Model	Model trains successfully
TC-03	Prediction Test	Model outputs numeric estima
TC-04	Missing Value Handling	No null values remained

7.2 Test Procedure

- Import dataset
- Preprocess data
- Train models
- Evaluate RMSE & accuracy
- Compare algorithms

7.3 Performance Outcome

The model performance upon following evaluation metrics obtained were:

- **Random Forest RMSE:** 12.1

- **Linear Regression RMSE:** 19.4
- **Best Model:** Random Forest

8 My learnings

During this internship, I learned:

- End-to-end machine learning workflow
- Data cleaning and preprocessing
- Exploratory Data Analysis (EDA)
- Model building and evaluation techniques
- GitHub project deployment
- Industrial project documentation
- Communication and problem-solving skills
- Handling missing values using imputation
- Hands-on-working with real-world projects
- Feature engineering across subsets of dataset
- Model performance validation
- Project deployment and connection to github
- Exposure to tools like Git Bash and Git GUI
- Maintaining and configuring Git repositories
- Updating the project with the Github codespaces
- Additional knowledge improvement at the Machine learning

This internship has significantly improved my confidence in implementing real-world ML projects.

9 Future work scope

- Integrate satellite and NDVI datasets
- Develop a dashboard for real-time crop prediction
- Deploy model using Flask/Streamlit
- Include weather API for dynamic predictions
- Expand prediction to district-wise forecasting

Finally, this internship has been a highly valuable learning experience that significantly enhanced my technical skills and practical understanding of real-world data science applications. Working on the project **"Prediction of Agriculture Crop Production in India"** allowed me to apply machine learning concepts to an industry-relevant problem and gain hands-on exposure to data preprocessing, model development, performance analysis, and interpretation of results. The mentorship and structured guidance provided by USC/UCT helped me strengthen my problem-solving abilities, improve my project execution skills, and understand how technology-driven solutions can create meaningful impact in sectors like agriculture.