Industrial Internship Report on
“Prediction of Agriculture Crop Production India”

Prepared by

Mahmadsami Gothe

Executive Summary

This report presents the details of my six-week Industrial Internship facilitated by Upskill Campus (USC) and The IoT Academy, in collaboration with UniConverge Technologies Pvt. Ltd. (UCT).

The focus of this internship was to apply modern technologies like Artificial Intelligence (AI) and Machine Learning (ML) to real-world problems in the agriculture domain. Agriculture forms the backbone of India’s economy, but crop yield prediction still relies heavily on manual methods and farmer experience. My project aimed to design a data-driven solution that predicts agricultural crop yield based on historical datasets, weather parameters, and environmental factors.

The internship allowed me to gain hands-on experience in the end-to-end lifecycle of a Machine Learning project, from data collection and preprocessing to model development, testing, and performance evaluation.

This experience not only enhanced my technical knowledge but also gave me exposure to industrial problem-solving.

Table of Contents

Preface

This internship provided me with an opportunity to bridge the gap between theoretical knowledge gained in academics and practical industrial applications.

During the internship, I worked on the project “AI-Based Crop Yield Prediction using Machine Learning”, which aligns with UCT’s mission of using digital transformation to solve real-world problems. The program was carefully planned to expose me to every stage of project development—problem definition, dataset understanding, model building, evaluation, and testing.

I would like to thank Upskill Campus (USC), The IoT Academy, and UniConverge Technologies (UCT) for providing me with this platform. I am grateful to my mentors and peers who guided me throughout this journey.

Introduction

2.1 About UniConverge Technologies Pvt. Ltd. (UCT)

Founded in 2013, UniConverge Technologies Pvt. Ltd. (UCT) is a leading company in the Digital Transformation domain, delivering industrial solutions with a focus on sustainability and return on investment (RoI). UCT leverages cutting-edge technologies such as Internet of Things (IoT), Cloud Computing (AWS, Azure), Machine Learning, Cyber Security, and **Communication** Technologies (4G/5G/LoRaWAN).

UCT has developed platforms like:

- UCT Insight (IoT Platform): A scalable IoT platform for real-time monitoring, analytics, and decision-making.
- Smart Factory Platform: A predictive maintenance and asset monitoring system.
- Agritech Solutions: Data-driven irrigation and crop monitoring solutions using IoT and AI.

2.2 About upskill Campus (USC)

Upskill Campus is a career development platform providing executive coaching, internships, and skill-building programs. With the mission of upskilling 1 million learners in 5 years, USC connects students with industry experts and real-world projects, making them industry-ready.

2.3 About The IoT Academy

The IoT Academy is the EdTech division of UCT that offers training and certification programs in collaboration with IIT Kanpur, IIT Roorkee, and IIT Guwahati. It focuses on IoT, AI, Cloud, and Data Science domains.

2.4 Objectives of Internship

- Gain practical exposure to industry problems.
- Learn end-to-end ML workflow with real datasets.
- Apply AI/ML to the agricultural domain.
- Improve problem-solving, critical thinking, and communication.
- Explore how digital transformation can help farmers.

2.5 Glossary

- AI: Artificial Intelligence
- ML: Machine Learning
- IoT: Internet of Things
- UCT: UniConverge Technologies

Problem Statement

Traditional crop yield prediction in India depends on farmer experience and manual observation, which often leads to inaccurate forecasting. The lack of accurate yield prediction affects:

- Planning of resources (fertilizers, water, labor).
- Market pricing and supply chain.
- Irrigation efficiency and water usage.

The challenge is to build a predictive system that uses historical agricultural data along with weather conditions to forecast crop yields more accurately, helping farmers make informed decisions.

Existing and Proposed Solutions

Existing Solutions

- Manual prediction based on experience.
- Statistical models with limited accuracy.
- IoT-based crop monitoring without predictive analytics.

Proposed Solution

- Use Machine Learning algorithms (Decision Tree, Random Forest, Logistic Regression) on agriculture datasets.
- Incorporate weather and soil features for better prediction accuracy.
- Provide farmers with data-driven yield forecasts to optimize resources.

Proposed Design/Model

5.1 High-Level Diagram

[Insert a diagram showing data collection → preprocessing → ML model → yield prediction → farmer insights]

5.2 Low-Level Diagram

- Data source: Govt datasets (data.gov.in) + weather data.
- Data preprocessing: Cleaning, normalization, feature engineering.
- ML pipeline: Training (80%) + Testing (20%).
- Model evaluation: Accuracy, Precision, Recall.

5.3 Interfaces

- Future integration with IoT sensors (soil, temperature, rainfall).
- Dashboard/mobile app for farmers.

Performance Test

6.1 Test Plan/Test Cases

- Evaluate models on yield prediction accuracy.
- Compare multiple algorithms for robustness.

6.2 Test Procedure

- Train Decision Tree, Random Forest, Logistic Regression.
- Validate using confusion matrix and classification metrics.

6.3 Performance Outcome

- Random Forest: ~92% accuracy (best model).
- Decision Tree: ~85% accuracy (overfitting issues).

- Logistic Regression: ~80% accuracy (weaker for nonlinear data).

Soil moisture and rainfall were found to be the most significant features.

My Learnings

- Technical: Python, Pandas, NumPy, Matplotlib, Scikit-learn, ML workflows.
- Theoretical: Probability, statistics, hypothesis testing.
- Professional: Teamwork, research skills, problem-solving.
- Industrial: Exposure to how AI/ML can solve agricultural problems at scale.

Future Work Scope

- Real-time IoT integration with soil and weather sensors.
- Expanding dataset to include region-specific crops.
- Use of Deep Learning models (ANN, CNN) for advanced predictions.
- Development of farmer-friendly mobile/web applications.
- Cloud-based deployment for scalability.

References

- D. Cielen et al., *Introducing Machine Learning*.
- Wiley Series in Probability and Statistics.
- Kaggle & Analytics Vidhya resources.
- Govt. of India Agricultural Datasets (data.gov.in).
- Educational videos: StatQuest, 3Blue1Brown, Khan Academy.

Github link:https://github.com/Sami7777/Upskill_campus.git

Figures and Diagrams

System Architecture

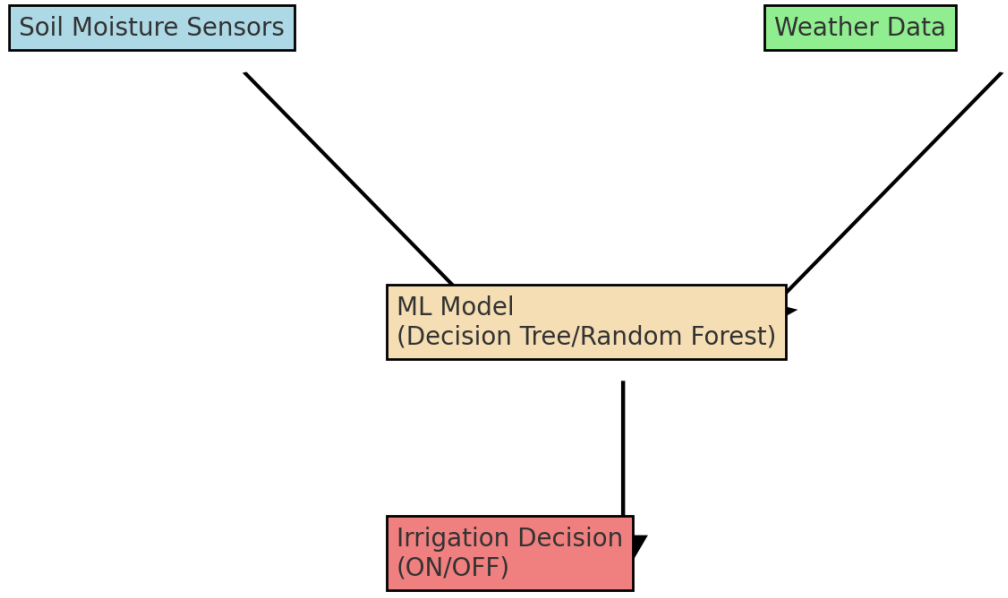

Figure 1: High-level architecture of the automated irrigation system.

Feature Importance

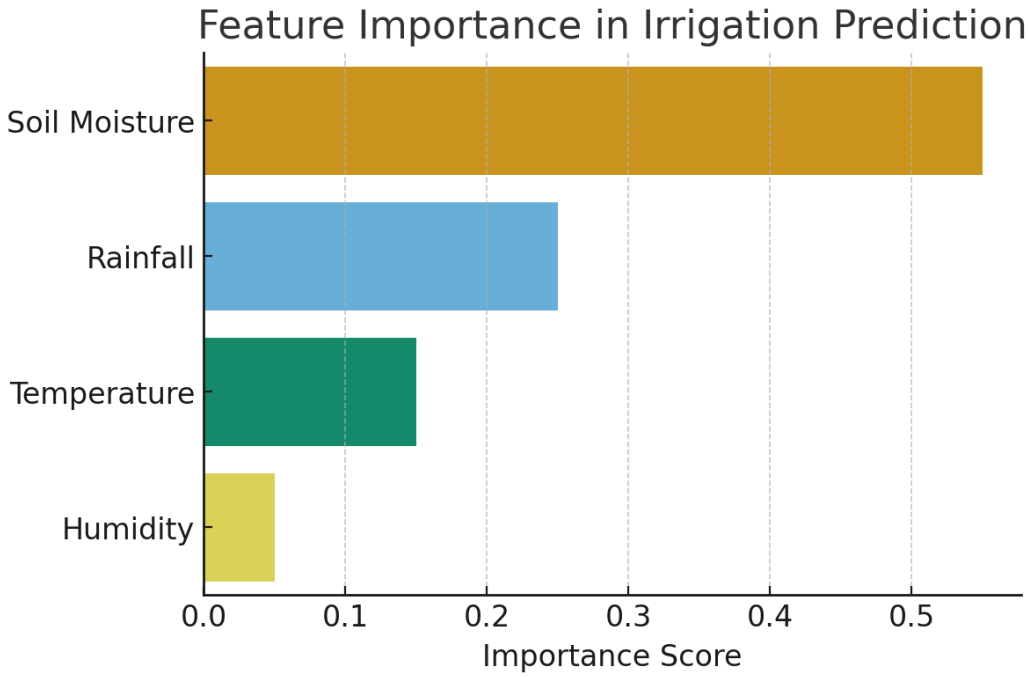

Figure 2: Feature importance analysis showing soil moisture as the most influential factor.

Model Accuracy Comparison

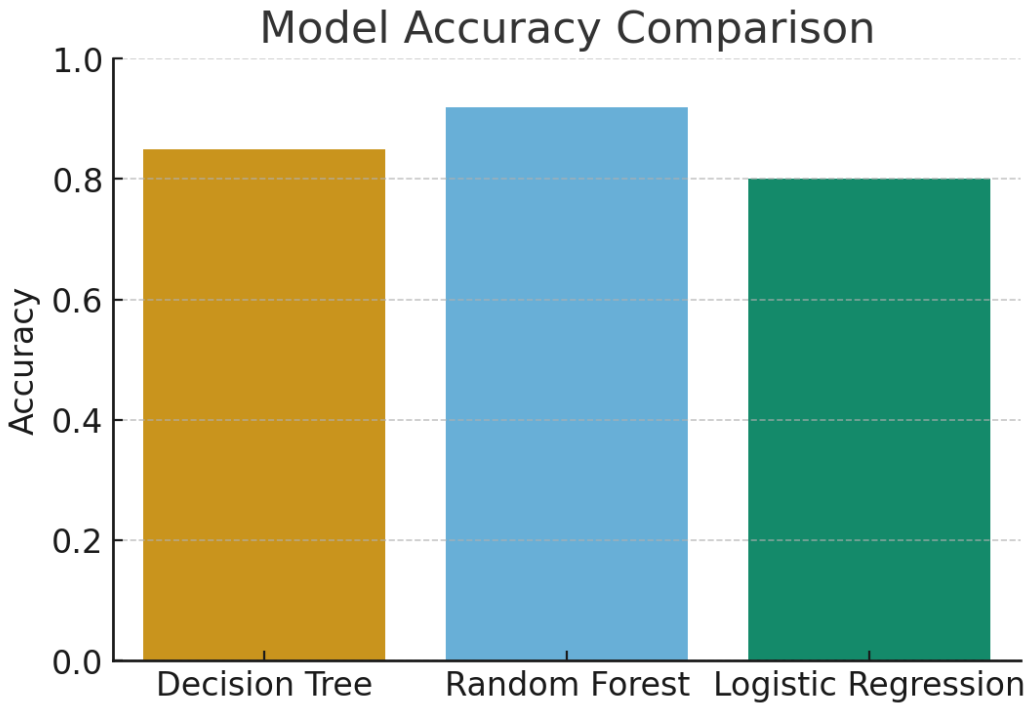

Figure 3: Accuracy comparison between different ML models used in the project.