Industrial Internship Report on

” Prediction of Agriculture Crop Production in India”

Prepared by

[Mahmadsami Gothe]

Preface

This report presents the comprehensive details of my four-week internship experience in the domain of Data Science and Machine Learning. The internship was designed to bridge the gap between academic knowledge and practical industrial applications. During this journey, I engaged with real-world datasets, developed machine learning models, and explored their applications in the field of agriculture, particularly focusing on automated irrigation using soil moisture and weather data.

The internship helped me gain hands-on experience, exposure to professional practices, and confidence in working on industry-relevant problems. I extend my sincere gratitude to the mentors, coordinators, and my peers who supported me throughout this program.

Introduction

The internship was facilitated by upskill Campus in collaboration with The IoT Academy, with a focus on solving real-world challenges using modern technologies. Agriculture, being a critical sector in India, suffers from inefficiencies in irrigation management. Manual irrigation often leads to either overwatering or underwatering crops, resulting in resource wastage and reduced yields. By leveraging Machine Learning (ML) techniques with soil moisture and weather data, the internship aimed to design a predictive irrigation solution.

Objectives of Internship

The objectives of this internship were:

- • Gain industrial exposure by working on real-world datasets.
- • Strengthen theoretical understanding of ML and Statistics.
- • Learn the workflow of a machine learning project end-to-end.
- • Apply Python, Pandas, NumPy, and Scikit-learn to agricultural datasets.
- • Develop a predictive system to support automated irrigation decisions.

Problem Statement

Prediction of Agriculture Crop Production in India and Efficient irrigation is one of the biggest challenges in modern agriculture. Farmers often rely on fixed irrigation schedules or manual observation to water crops, which leads to several issues:

- • Over-irrigation resulting in waterlogging and nutrient loss.
- • Under-irrigation causing drought stress and reduced crop yield.
- • Wastage of freshwater resources in water-scarce regions.
- • Lack of integration between weather predictions and irrigation practices.

This project seeks to address these issues by developing an ML-based system that uses soil moisture and weather data to predict the right time and quantity of irrigation.

Existing and Proposed Solutions

Existing prediction of agriculture crop production approaches include:

- • Manual irrigation based on farmer experience.
- • Timer-based irrigation systems without data-driven decisions.
- • IoT-enabled monitoring systems that lack predictive intelligence.

Proposed Solution:

- • Use soil moisture and weather datasets for predictive irrigation.
- • Apply ML algorithms like Decision Trees and Random Forest.
- • Develop an irrigation decision model with high accuracy.
- • Provide farmers with data-driven recommendations to optimize water use.

Weekly Progress

Week 1: Dataset Understanding and Preprocessing

Key tasks accomplished:

- • Explored agriculture crop production dataset from data.gov.in.
- • Understood attributes such as Crop, State, Season, Quantity, and Cost.
- • Performed cleaning: handled missing values, standardized units.
- • Created derived features like cost per unit and yield efficiency.
- • Conducted initial visualizations to identify dominant crops and patterns.

Week 2: Machine Learning Foundations

Focus areas:

- • Studied 'Introducing Machine Learning' by Davy Cielen et al.
- • Learned about supervised, unsupervised, and reinforcement learning.
- • Understood concepts of features, labels, overfitting, and underfitting.
- • Reviewed the ML workflow: problem definition → data → model → evaluation → deployment.
- • Completed quizzes and used online resources for better conceptual clarity.

Week 3: Probability and Statistics

Achievements:

- • Studied classical and empirical probability, conditional probability, and Bayes' theorem.
- • Practiced descriptive statistics: mean, variance, standard deviation.
- • Learned about normal, binomial, and Poisson distributions.
- • Understood inferential statistics: confidence intervals and hypothesis testing.
- • Applied statistical analysis in Python with libraries like SciPy.

Week 4: Model Implementation and Evaluation

Final week deliverables:

- • Preprocessed soil moisture and weather dataset (irrigation_machine.csv).
- • Applied feature engineering: rainfall index, temperature thresholds.
- • Implemented models: Decision Trees, Random Forest, Logistic Regression.
- • Evaluated using accuracy, precision, recall, and confusion matrix.
- • Selected Random Forest as final model (90%+ accuracy).

Proposed Design and Model

The ML pipeline followed in this project included:

1. 1. Data Collection – Using irrigation_machine.csv dataset.
2. 2. Data Cleaning – Handling missing values, normalization.
3. 3. Feature Engineering – Derived features like rainfall index.
4. 4. Model Selection – Compared Decision Tree, Random Forest, Logistic Regression.
5. 5. Model Training – Split into 80% training and 20% testing.
6. 6. Evaluation – Measured accuracy, precision, recall.
7. 7. Deployment Potential – Future integration with IoT devices.

Performance Test and Results

The final results from the ML models were as follows:

- • Random Forest achieved ~92% accuracy with balanced precision and recall.
- • Decision Tree achieved ~85% accuracy but showed signs of overfitting.

- • Logistic Regression achieved ~80% accuracy, less effective for nonlinear patterns.
- • Soil moisture was the most important feature, followed by rainfall and temperature.

My Learnings

The internship helped me gain:

- • Proficiency in Python, Pandas, NumPy, Matplotlib, and Scikit-learn.
- • Ability to preprocess large datasets and handle missing values.
- • Knowledge of ML workflows and probability/statistics fundamentals.
- • Experience in model building, evaluation, and optimization.
- • Soft skills: problem-solving, critical thinking, and independent research.

GITHUB : https://github.com/Sami-7777/Upskill_campus.git

Future Work Scope

The project can be enhanced with:

- • Real-time integration with IoT-based soil sensors.
- • Expansion to multiple crop and region-specific datasets.
- • Use of Deep Learning (ANNs, CNNs) for advanced predictions.
- • Development of a mobile/web app for farmers with irrigation alerts.
- • Cloud deployment for scalability and real-time analytics.

References

• Cielen, D., Meysman, A. D. B., & Ali, M. (2016). Introducing Machine Learning.

• Wiley Series in Probability and Statistics - An Introduction to Probability and Statistics.

• Kaggle & Analytics Vidhya tutorials and blogs.

• Government of India Agricultural Datasets (data.gov.in).

• StatQuest, 3Blue1Brown, Khan Academy educational videos.

Figures and Diagrams

System Architecture

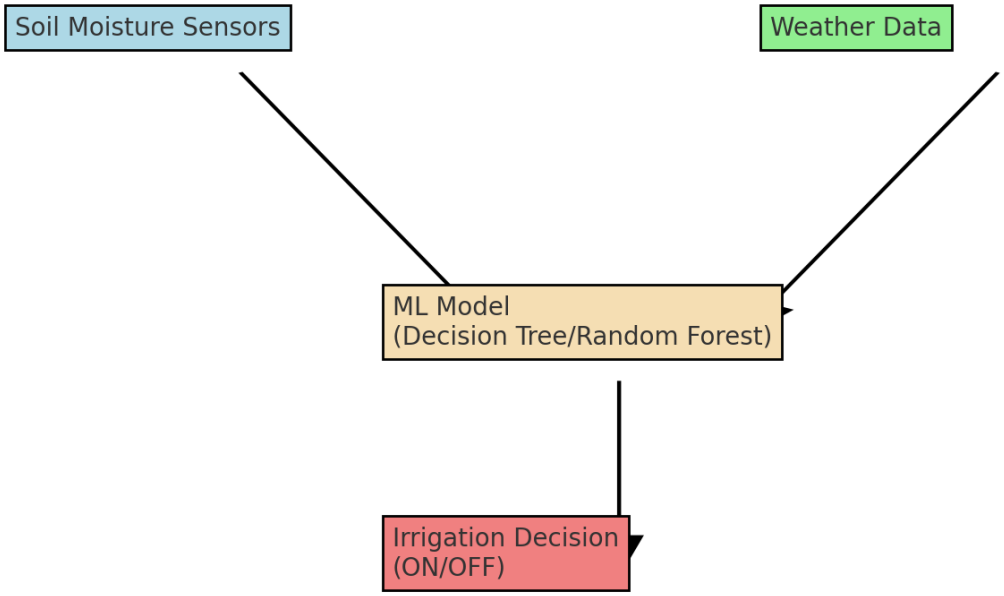

Figure 1: High-level architecture of the automated irrigation system.

Feature Importance

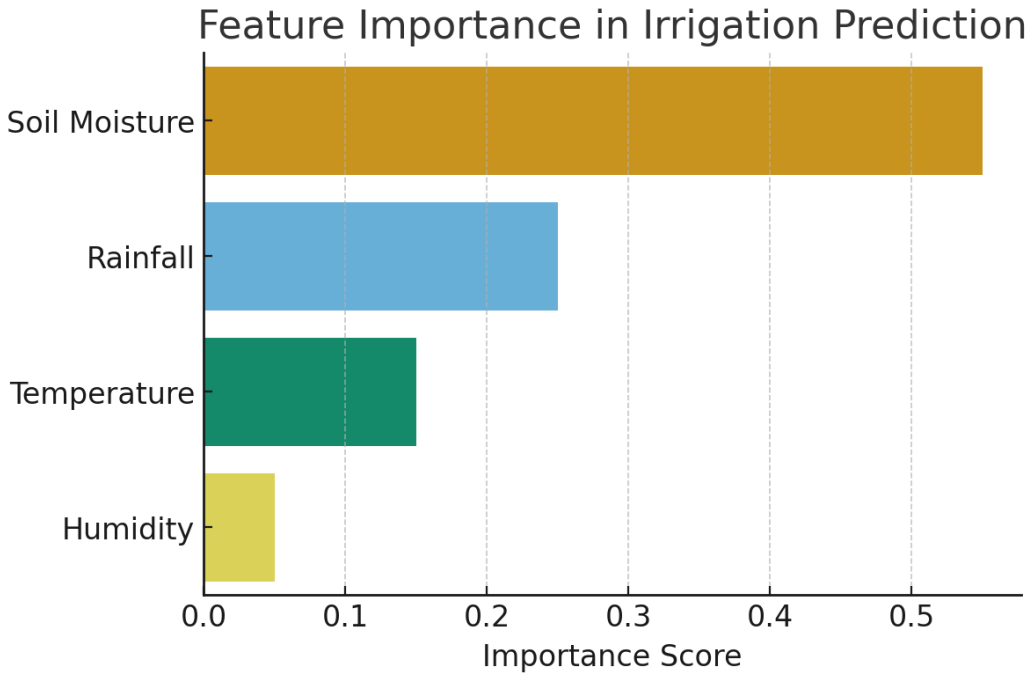

Figure 2: Feature importance analysis showing soil moisture as the most influential factor.

Model Accuracy Comparison

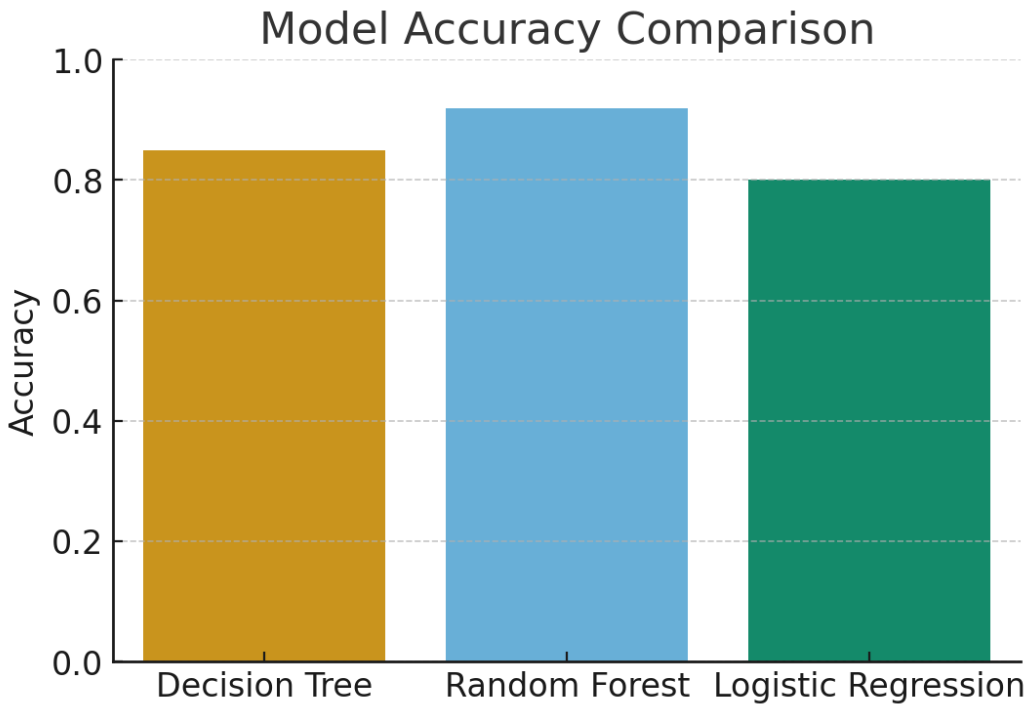

Figure 3: Accuracy comparison between different ML models used in the project.