Industrial Internship Report on

" Prediction of Agriculture Crop Production in India "

Prepared by

Hunaid Saifee Siamwala

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 Prediction of Agriculture Crop Production in India. Crop yield prediction is an important aspect of agriculture that helps farmers make informed decisions about their crops. It involves estimating the number of crops that will be produced in each area based on various factors such as soil type, weather conditions, and crop management practices. In recent years, machine learning (ML) has emerged as a powerful tool for predicting crop yields. Machine learning is a branch of artificial intelligence (AI) that allows computers to learn from data without being explicitly programmed. This makes it ideal for crop yield prediction because it can identify patterns and relationships in large amounts of data and make predictions based on these relationships. 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

The Data science and machine learning internship duration was 6 weeks. The 1st week of the internship was to explore the problem statements which were provided by the management and understand their background in order to start with the project. Also learned about UCT. The 2nd week of the internship was to understand and follow the project instructions provided by UCT. And, also to plan for the solution of the existing problem. 3rd week of internship is to start for the actual working of the project. The 4th week of the internship was to continue with the work on the project and check whether there are improvements required for the project. 5th week of the internship was to validate your implementation and evaluate your performance. And the final week of the project is to submit your project report and get certification.

Internships are an opportunity to network with great people and sharpen your skills before entering the workforce. They also help tremendously with figuring out your true passion. Companies often look at them to gain experience and exposure to make a smooth transition into your role when hired.

Agriculture is the main occupation for the people of India, covering 60% of the nation land and catering the basic needs of 1.2 billion people. For the benefit of the farmers, modernization of agriculture procedures is carried out today. The crop yield or production majorly depends on the weather conditions, environmental changes, rainfall (which at times is uncertain), water management, and the utilization of pesticides. Therefore, farmers are not able accomplish the expected yield of crop. Now a days data mining, machine learning as well as deep learning approaches are used by various researchers to enhance and improve the yield of crops and their quality. Machine Learning can gain proficiency with the machine without characterized computer programming, so it improves machine execution by distinguishing and portraying the consistency and pattern of drive information. In this work various machine learning approaches such as Linear Regression, Gradient Boosting Regressor, Random Forest Regressor, Decision Tree Regressor, Polynomial Regression, Ridge Regression have been used for yield prediction on crop yield dataset of different states and considering varied crops.

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.

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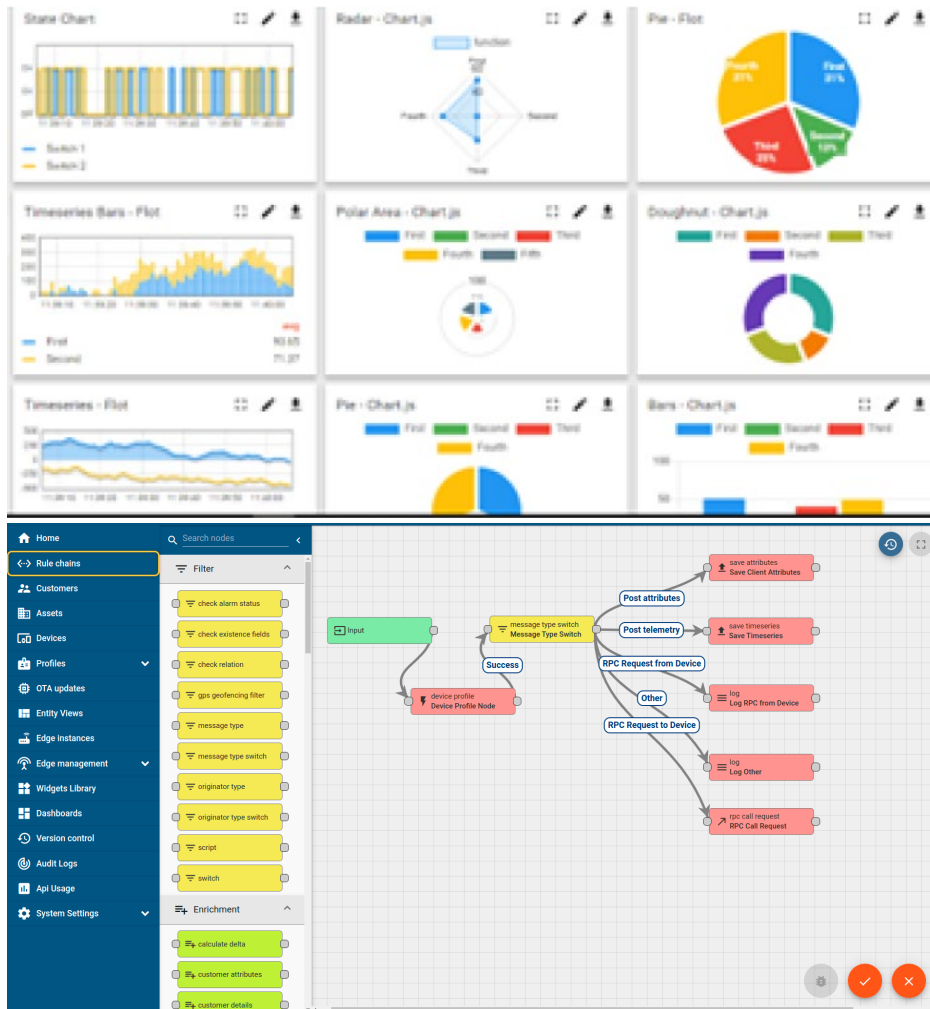
ii. Smart Factory Platform (**FACTORY WATCH**)

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.
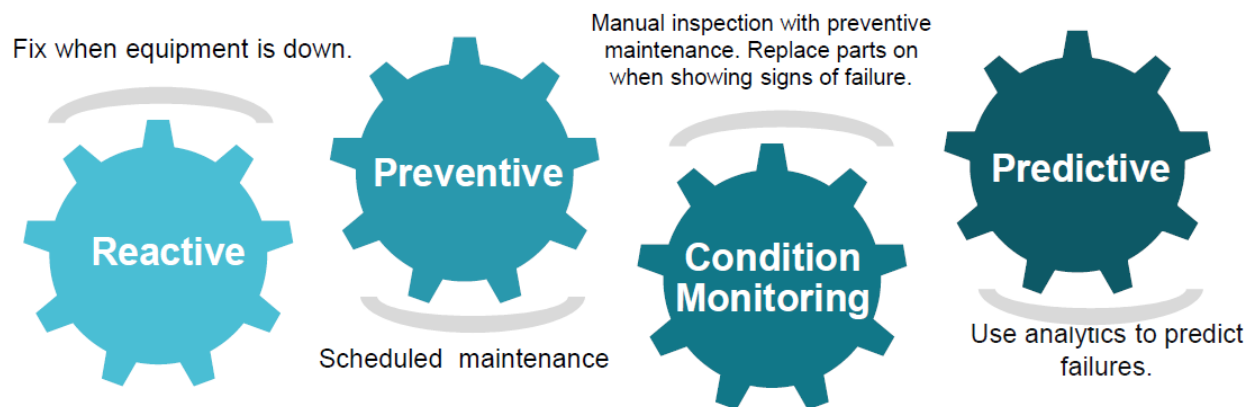
2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.
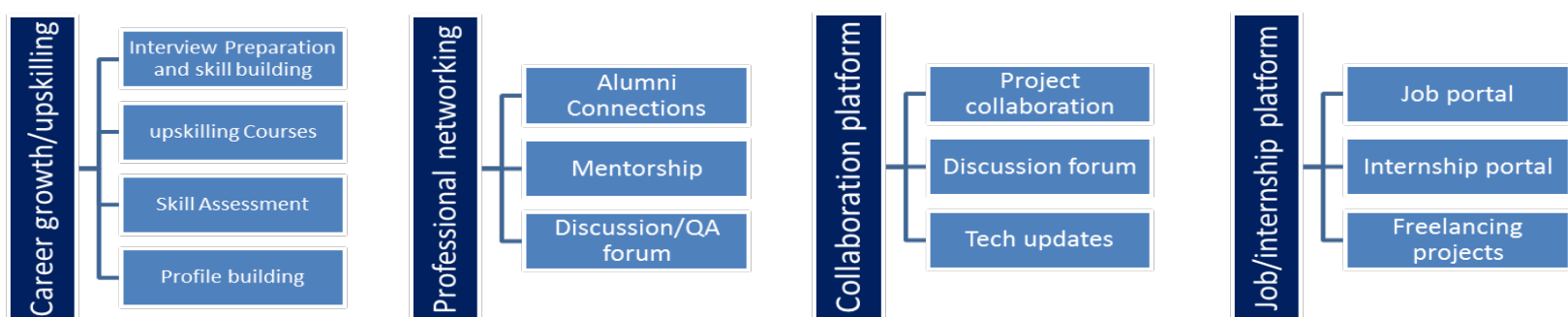
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.
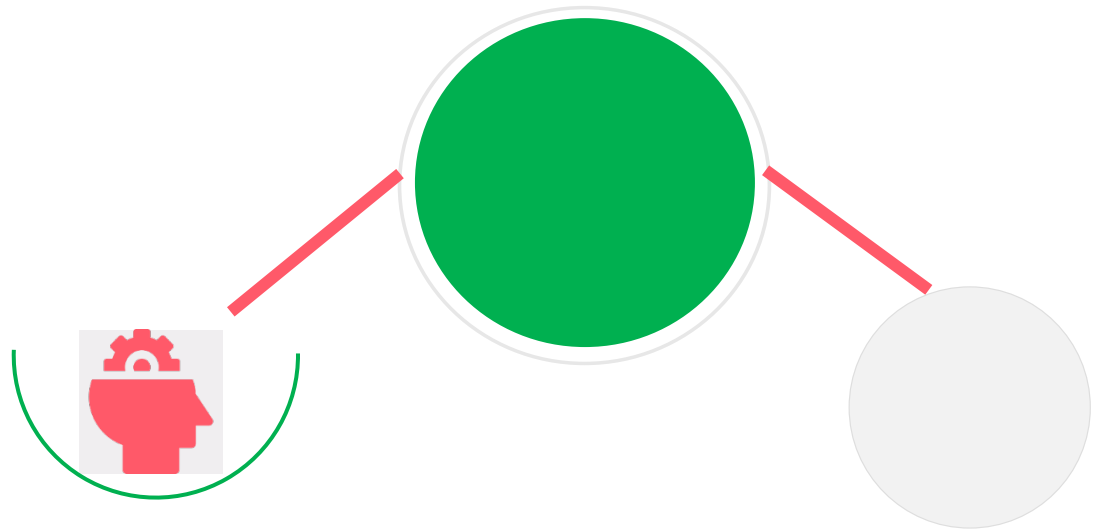
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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

https://www.upskillcampus.com

Career growth/upskilling
- Interview Preparation and skill building
- upskilling Courses
- Skill Assessment
- Profile building

Professional networking
- Alumni Connections
- Mentorship
- Discussion/QA forum

Collaboration platform
- Project collaboration
- Discussion forum
- Tech updates

Job/internship platform
- Job portal
- Internship portal
- Freelancing projects

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] Crop_recommendation.csv

[2] https://learn.upskillcampus.com/s/courses/6441224de4b0f11fbe0f621e/take

3 Problem Statement

In the assigned problem statement Agriculture is one of the main sources of income in India. There is need to improve the sustainability of agriculture with the rate of increase in suicides of farmer due to crop failure and less yield and losses. Hence, it is a significant contribution towards the economic and agricultural welfare of the countries across the world. The Problem Statement revolves around prediction of yield of crops considering different climatic conditions of India including various attributes. Goal of this project is to help the farmers to choose the suitable crop to grow in order to get the required yield and the profit. The need for the crop yield prediction is very much essential at this point of time for selecting the right crop.

4 Existing and Proposed solution

Existing Solutions:

Crop recommendation is a significant application of machine learning and data science in agriculture. Existing solutions often involve more complex models and techniques than a simple linear regression. Some common approaches are:

- Decision Trees and Random Forests: Decision tree-based models like Random Forests can handle non-linear relationships and interactions among features. They are widely used for crop recommendation due to their interpretability.
- Support Vector Machines (SVM): SVMs can handle non-linear data by transforming it into higher dimensions. They are used in crop recommendation to find decision boundaries that separate different crop classes.
- Neural Networks: Deep learning models can capture complex patterns in data and perform well for crop recommendation tasks. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) can be used for time-series and image-based crop recommendation.

Proposed Solutions:

For a more advanced and comprehensive crop recommendation system, you might consider the following enhancements:

- Feature Engineering: Explore domain-specific feature engineering techniques, such as incorporating soil type, historical weather data, geographical factors, and more.
- Ensemble Models: Combine multiple models (e.g., Random Forest, Gradient Boosting) to improve prediction accuracy and robustness.
- Hybrid Models: Combine machine learning models with domain knowledge or expert systems to enhance the quality of recommendations.
- Time Series Analysis: Incorporate time series forecasting techniques to predict crop yields over different seasons.

- Data Augmentation: If available, consider using satellite imagery and other remote sensing data to enhance the dataset and provide more comprehensive features.

- User Interface: Develop a user-friendly web or mobile application where farmers can input data easily and receive recommendations visually.

- Cloud Integration: Host the model on cloud platforms to enable scalability and easy access for users.

4.1 Code submission and Report submission (GitHub link)

https://github.com/Hunaid53/upskillcampus

5 Proposed Design/ Model
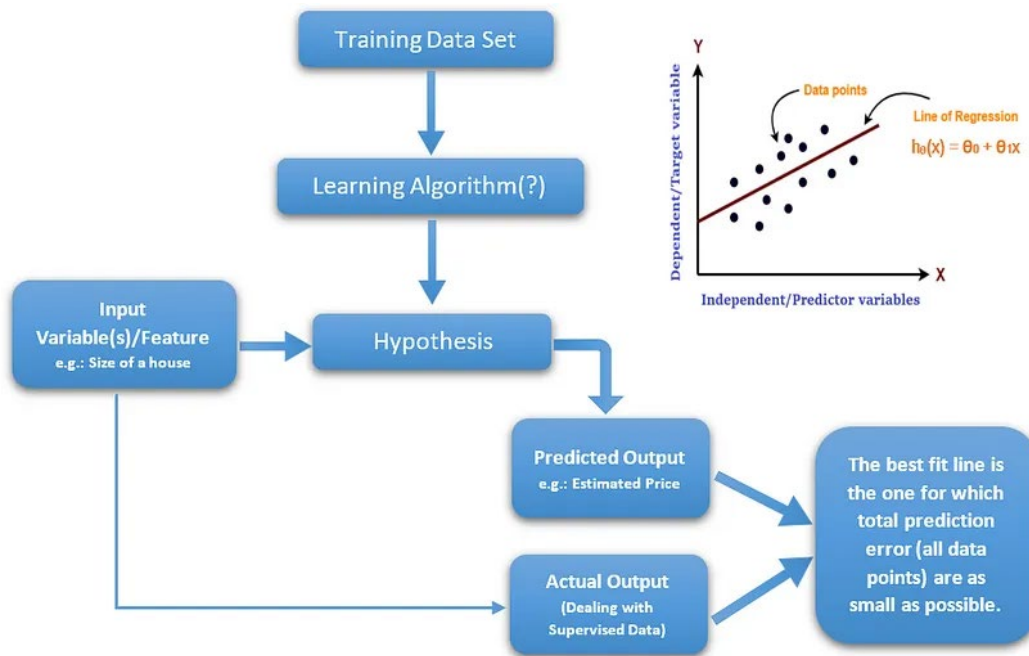
In this project, the system makes use of the Machine Learning techniques to predict rate of crop yield. The programming language used is Python as it is widely accepted for new idea implementations in the field of Machine Learning. In this project, collected data set will be uploaded and prediction for crop yield will be generated by applying Machine Learning techniques like Simple Linear Regression. The results depend on the information present in the collected data set. Accurate the information about the parameters in the collected datasets is more accurate, the results will be.

6 Performance Test

6.1 Test Plan/ Test Cases

1) Test Plan/Test Cases:

A test plan outlines the testing approach, objectives, scope, and resources for a project. Test cases are specific scenarios or situations that are tested to ensure the system functions correctly.

Test Objectives:

- Verify that the crop recommendation system accurately predicts the recommended crop based on user input.
- Ensure the user interface is user-friendly and intuitive.
- Assess the system's robustness against various inputs and scenarios.
- Evaluate the performance of the system in terms of response time and resource usage.

Test Cases:

- **User Input Validation:**

  Enter valid and invalid values for all input fields.

  Verify that the system handles incorrect input gracefully and provides appropriate error messages.

- **Prediction Accuracy:**

  Provide known input data and validate that the predicted crop matches the expected outcome.

  Test a variety of input combinations to assess the accuracy of crop predictions.

- **Edge Cases and Extremes:**

  Test with extreme values (e.g., very low or high nutrient content, unusual temperature, etc.).

  Test with values that are on the boundaries of the valid range.

- **Interface Usability:**

Interact with the user interface and assess its ease of use and navigation.

Check for any UI glitches or layout issues.

- **Performance Testing:**

Measure system response time for predicting crops with varying input sizes.

Simulate concurrent user interactions to assess system scalability.

- **Data Integrity:**

Test with missing or incomplete data and observe how the system handles it.

Check if the system produces reliable recommendations even with partial information.

6.2   Test Procedure

- Prepare a test environment that closely mimics the deployment environment.
- Start by running each test case individually, ensuring that each one produces the expected outcome.
- For performance testing, use tools to simulate multiple users accessing the system simultaneously.
- Log and document any errors or unexpected behaviours encountered during testing.
- Address any issues found during testing, iteratively improving the system.
- Re-run the tests after making changes to ensure that fixes did not introduce new issues.

6.3   Performance Outcome

The performance outcome focuses on how well the system performs in terms of speed, responsiveness, and resource utilization.

Performance Metrics:

Response Time: Measure the time taken for the system to provide a recommendation after the user inputs data.

Scalability: Evaluate how well the system handles multiple users simultaneously without significant degradation in response time.

Resource Usage: Monitor CPU, memory, and network usage during testing to identify any resource-intensive processes.

Desired Performance Outcomes:

The system should respond to user inputs within an acceptable time frame (e.g., a few seconds).

The system should maintain its performance even under high loads or concurrent user interactions.

Resource usage should remain within reasonable limits, avoiding excessive consumption.

Performance Improvement:

If the system's performance is not meeting the desired outcomes, consider optimizing code, database queries, or server resources.

Identify bottlenecks and areas that need improvement based on performance metrics.

7   My learnings

Studied the basics of data science and its applications. Studied about machine learning in depth. Glanced through the instructions of the internship program and understood the process. Went through the profile of UCT. Selected the project for the machine learning internship. Project Category: A. Agriculture, Project No. 4 Prediction of Agriculture Crop Production in India. Studied the various problems faced by farmers of India during Crop production. Went through the data of the crop production. Read and understood the basic concepts from the e-book Introducing-Data-Science-Machine-Learning. Glanced through the Impact-Of-Big-Data-On-Business and understood what Big Data and its applications is. Studied the difference between a Data scientist and the Data Analyst and how they contribute towards the project. Attempted the Quiz to test my intellectual capability. Gathered the information of Artificial Intelligence and Data Science and understood the difference between both. Learned how to map your career path with help of artificial intelligence and data science. Learnt about the skills which are required for artificial intelligence or data science or big data or machine learning engineer. Hence concluded that artificial intelligence and data science are inter-connected. So, specialization in either field requires a working knowledge of the other. To begin using ML-AI, studying data analysis is more necessary than understanding data science. Glanced through the introduction of Probability and statistics and revised the concepts of sample spaces, random variables, probability of an event, probability distribution, other special distributions, large sample theory, parametric point estimation, etc. Learned what are the technical and non-technical skills required for a data scientist. The Top 5 Skills Crucial for Becoming A Data Scientist are Programming, Statistics and Mathematics, Machine Learning, Deep Learning and Big Data. Having a knowledge and expertise in these skills is essential for an ideal data scientist.

8 Future work scope

The future work scope for the crop recommendation project includes several avenues for enhancement and expansion. Firstly, integrating advanced machine learning algorithms, such as ensemble methods, deep learning, or hybrid models, could further improve prediction accuracy. Incorporating real-time weather data, satellite imagery, and market trends could provide more comprehensive insights for recommendations. Expanding the user interface to include personalized user accounts, historical data tracking, and region-specific recommendations would enhance user experience. Moreover, exploring mobile app development and cloud deployment could make the solution more accessible to a wider audience of farmers. Finally, delving into interpretability techniques and incorporating feedback mechanisms from users would promote transparency and continuous improvement in the recommendation process.