Industrial Internship Report on

"Prediction of Agriculture crop production in India"

Prepared by

Krunal Patel

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 about an Agriculture problem. Across The Globe India Is The Second Largest Country
having more than 1.3 Billion. Many People Are Dependent On Agriculture and it is the Main Resource. In Agriculture Cultivation/Production Having More Problems. I want to solve the Big problem in India and be useful to many more people.

Agriculture is regarded as India's primary and most significant community. The agricultural sector has been steadily degrading since the advent of new advanced technologies and techniques. Nowadays, most people are unaware of the importance of cultivating crops at the appropriate time and place. Seasonal climatic conditions are also being altered because of these cultivation methods, posing a challenge to fundamental assets such as soil, water, and air, resulting in food insecurity.
By evaluating all these issues and challenges, such as weather, temperature, and several other variables, we have discovered that there is no suitable solution or technology to help us solve the situation we are in. In India, there are many options for increasing agricultural economic development.
There are numerous methods for increasing and improving crop yield and quality. Data mining can also be used to forecast crop yield production. Data mining, in general, is the methodology of analyzing information from different angles and synthesizing it into helpful information.
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

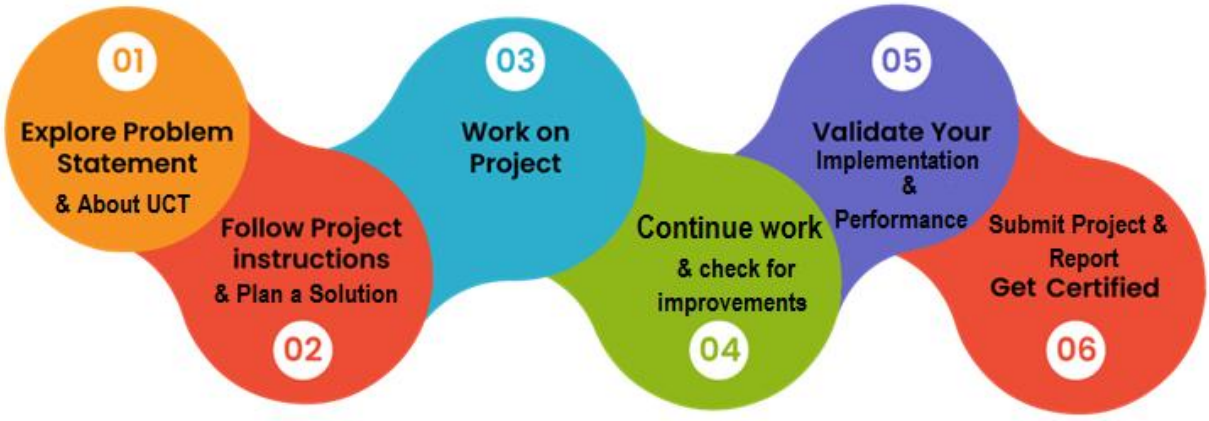
Summary of the whole 6 week work : In the first week we explored all the problem statements and selected one problem statement. In the second week we started planning our solution. In the third week we started implementation of our project. We performed various tasks like data cleaning, data integration, etc. In the fourth week we performed various tasks like data visualization, model training, choosing a proper model based on its accuracy, etc. In fifth week we performed various basic tasks like model validation and model testing. we also submitted weekly reports and quizzes.

About need of relevant Internship in career development : An internship is important because it can present you with new skills and opportunities that you may not receive otherwise. Interns gain technical knowledge within the industry of their choice by working directly with professionals in that field. This allows you to apply practical knowledge you may have learned from a classroom setting while you develop important soft skills, such as time management, organization, adaptability, problem-solving and teamwork.

Brief about Your project/problem statement : Prediction of Agriculture Crop Production In India. Across The Globe India Is The Second Largest Country having People more than 1.3 Billion. Many People Are Dependent On Agriculture And it is the Main Resource. In Agriculture Cultivation/Production Having More Problems. I want to solve the Big problem in India and be useful to many more people.

Opportunity given by USC/UCT : UCT/USC gave us the opportunity to present our idea on a great platform where we can showcase our talent. By providing group projects, we learned how to work in groups which improved our team building capacity and work in Harmony.

How Program was planned : In the first week we explored all the problem statements and selected one problem statement. In the second week we started planning our solution. In the third week we started implementation of our project. We performed various tasks like data cleaning, data integration,etc. In the fourth week we performed various tasks like data visualization, model training, choosing a proper model based on its accuracy, etc. In fifth week we performed various basic tasks like model validation and model testing.



Learnings and overall experience:

It was an amazing experience working on this project and applying machine learning techniques to predict crop yields. I learned how to preprocess and analyze agricultural data, as well as build and train ML models for accurate predictions. The project taught me the importance of data quality and feature engineering in agricultural applications. Huge thanks to my mentor and the team for their guidance and support throughout the internship.

Thank to all who have helped me directly or indirectly:

Grateful beyond words for the incredible support I received during my ML internship! A huge shoutout to everyone who directly or indirectly contributed to my growth and success. Thank you to my mentor for their guidance and expertise. Special thanks to the amazing team for their collaboration and encouragement. I couldn't have done it without you all.

Your message to your juniors and peers:

I wanted to give a special shoutout to all my junior colleagues and peers for their incredible support during the crop prediction internship and project. Your collaboration and enthusiasm made the experience even more rewarding. Keep up the great work and continue to explore the fascinating world of machine learning in agriculture.

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.

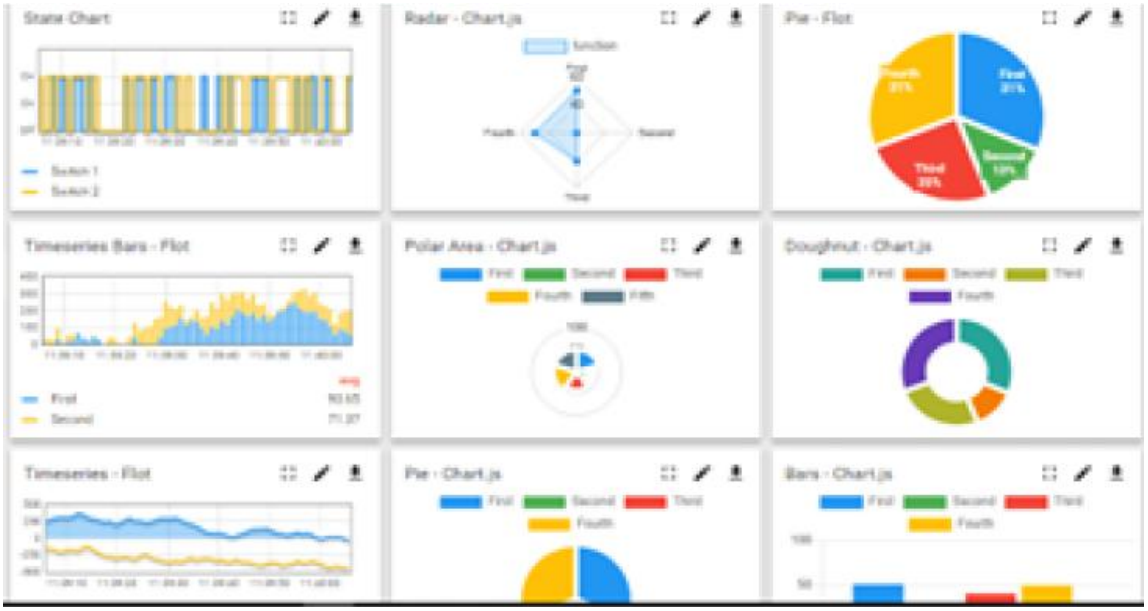
i. UCT IoT Platform (UCT Insight)

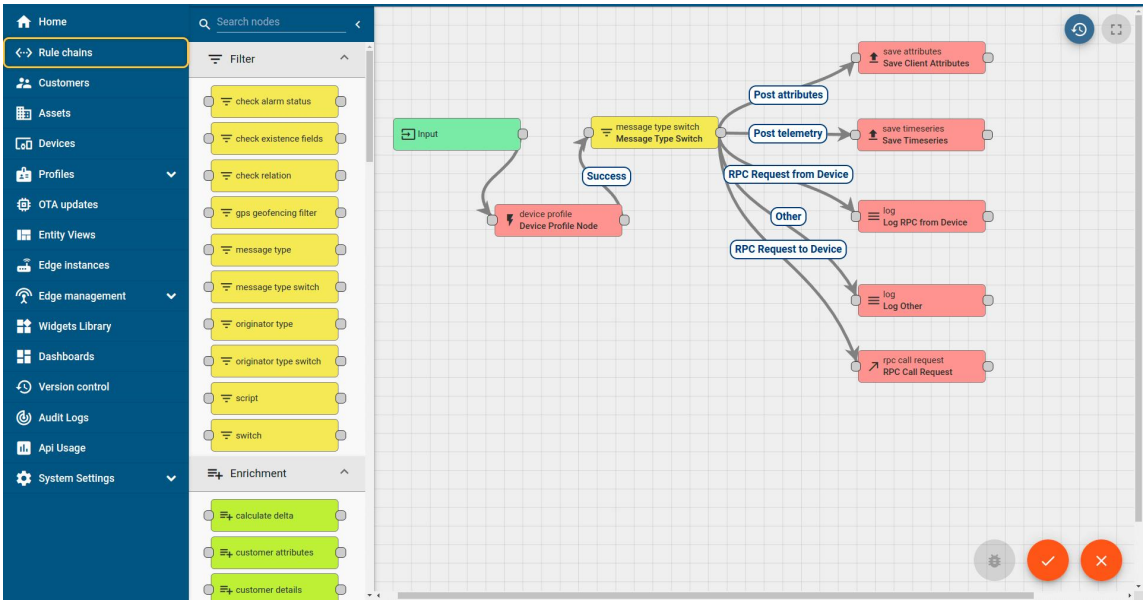
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

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

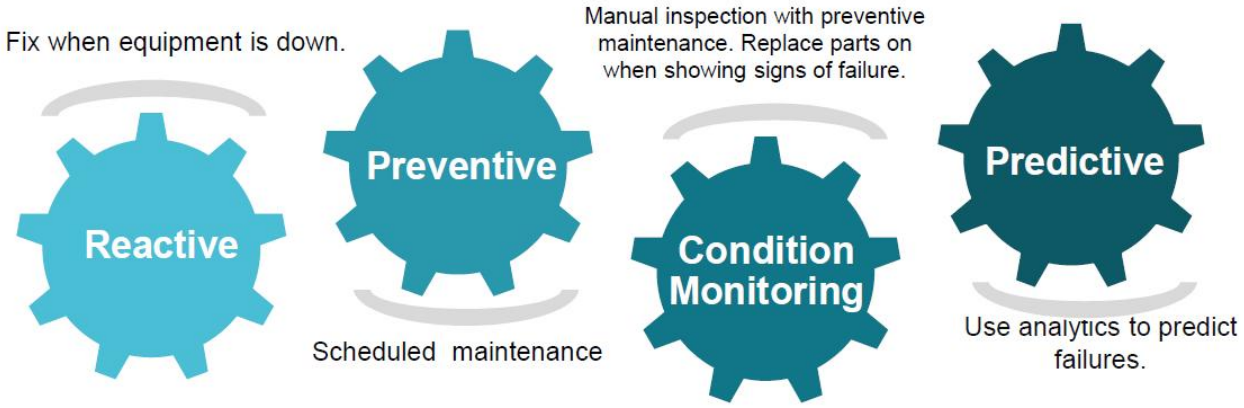
.

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)

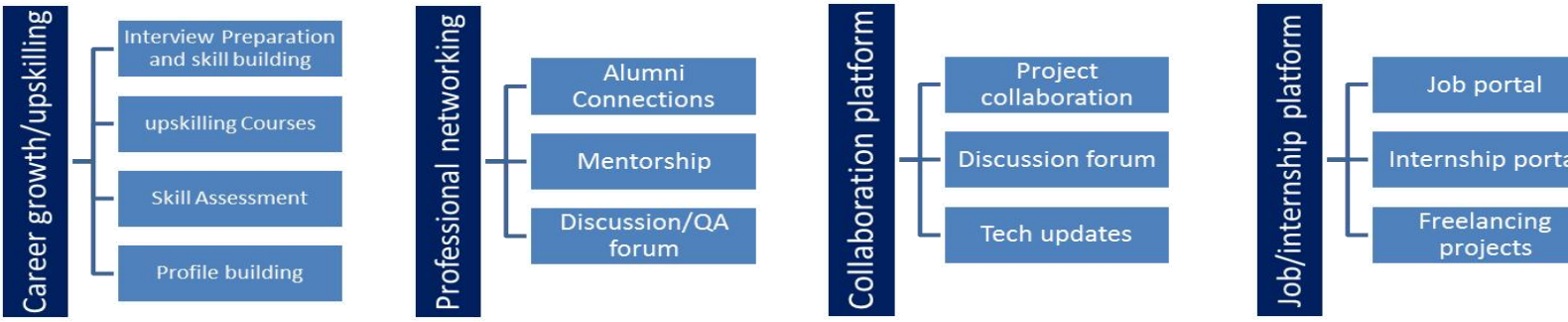
upskill Campus along with The IoT Academy and in association with UniConverge technologies has facilitated the smooth execution of the complete internship process.

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.



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] https://medium.com/@jyothics/indian-crop-production-exploratory-data-analysis-to-chart-agriculture-highlights-using-python-ea243924d4b6

2.6 Glossary

Terms	Acronym
Crop prediction	The process of using historical data, statistical models, and machine learning algorithms to forecast crop yields or other agricultural outcomes.
Regression analysis	A statistical method used to identify and quantify the relationship between variables, often used in crop prediction models to estimate crop yields based on various factors.
Feature engineering	The process of selecting, transforming, or creating relevant variables (features) from raw data to improve the performance and accuracy of machine learning models.
Training data	The labeled dataset used to train a machine learning model, consisting of input features (e.g., weather patterns, soil characteristics) and corresponding output values (e.g., crop yields).
Validation data	A separate dataset used to evaluate the performance and generalization of a trained machine learning model.
Accuracy	A measure of how well a machine learning model predicts crop yields compared to the actual observed values, typically expressed as a percentage.

3 Problem Statement

Prediction of crop production in India.

Explanation:

The problem statement for the crop prediction project is to develop a model that can accurately forecast crop yields based on factors like weather patterns, soil conditions, historical data, and other relevant variables. The objective is to provide farmers and stakeholders with reliable predictions to optimize crop planning, resource allocation, and decision-making processes. This project aims to improve agricultural productivity, resource management, and contribute to food security.

4 Existing and Proposed solution

summary of existing solutions provided by others:

1. Statistical Models: Many researchers have utilized statistical models such as linear regression, time series analysis, and support vector machines to predict crop yields. These models use historical data on weather patterns, soil conditions, and crop management practices to make predictions.

2. Machine Learning Techniques: Researchers have also explored the use of machine learning algorithms like random forests, neural networks, and gradient boosting to predict crop yields. These algorithms can handle complex relationships between variables and can incorporate various data sources for more accurate predictions.

3. Remote Sensing and Satellite Imagery: Some solutions leverage remote sensing and satellite imagery data to monitor vegetation health, soil moisture levels, and other environmental factors. These data sources are used to predict crop yields and identify areas of potential crop stress or disease.

4. Crop Simulation Models: Crop simulation models, such as the Agricultural Production Systems Simulator (APSIM) and the Decision Support System for Agrotechnology Transfer (DSSAT), are widely used to simulate crop growth and predict yields. These models integrate various factors such as weather, soil, and crop management practices to forecast crop outcomes.

5. Data-driven Platforms: Several organizations have developed data-driven platforms that combine multiple data sources, including weather data, satellite imagery, and historical crop yield

data. These platforms use advanced analytics and machine learning algorithms to generate crop yield predictions and provide insights for farmers and stakeholders.

These existing solutions demonstrate the diverse approaches and technologies used to predict crop yields. The crop prediction project can build upon these existing solutions to develop a more accurate and robust model for crop yield forecasting.

Some limitations of existing solutions for crop prediction projects include:

1. Data Availability: The accuracy of crop yield predictions heavily relies on the availability and quality of historical data on weather patterns, soil conditions, and crop management practices. Limited or incomplete data can affect the accuracy of predictions.

2. Local Variability: Crop yield predictions may vary significantly across different regions and even within a single region due to variations in soil types, microclimates, and other localized factors. Existing models may struggle to capture these nuances accurately.

3. Complex Interactions: Crop growth is influenced by various complex interactions between environmental factors, such as temperature, rainfall, sunlight, and soil nutrients. Capturing and modeling these interactions accurately can be challenging.

4. Changing Climate Patterns: Climate change and its impact on weather patterns can introduce additional uncertainties into crop yield predictions. Existing models may struggle to adapt to changing climate conditions.

5. Limited Scope: Some existing solutions may focus on specific crops or regions, limiting their applicability to a broader range of agricultural contexts.

6. Resource Constraints: Implementing and maintaining sophisticated prediction models may require significant computational resources, technical expertise, and data infrastructure, which may not be readily available to all farmers or stakeholders.

Addressing these limitations requires ongoing research, data collection, and model refinement to improve the accuracy and reliability of crop yield predictions.

What is your proposed solution?

To design an application where we compare the different machine learning to predict the crop yield. We build a new decision system using ensemble regression system. The user would provide input of season type, year of production, area of production, crop type, cloudburst, climate condition, located yield within side the remaining and the system would predict the yield and relying at the value set, the crop may be classified and attain the results. In the first step it

allows the admin to login and load the data. Second, it allows the admin to perform analysis by considering all the input conditions. Finally, a report is generated for the crop yield and the accuracy of the models are also generated. The accuracy which is near to 1 is considered as an ideal model and the model which has accuracy near to 0 are considered as unideal model. The input for the system will be a season, rainfall, area of production, crop type, district name, state name and output will be the production of crop yield and accuracy of each model.

4.1 Code submission (GitHub link) :

https://github.com/krunalpatel292002/upskillcampus

4.2 Report submission (GitHub link) :

https://github.com/krunalpatel292002/upskillcampus

5 Proposed Design/ Model

Methodology:

To train our model we need data. The collected data can have little 'NA' values filtered out in Python. Since the data set is composed of digital The collected data can have little 'NA' values filtered out in Python. Since the data set is composed of digital the robust scaling we use is very similar to normalization, but it uses interquartile range, and normalization compresses the data in units of 0 and 1.

Ensemble Algorithms:

- On average, this is an improvement. Here, we add a meta model and use creases in addition to creases. The predictions of other models is used to train the basic meta-model.
- The entire training set is divided into two different sets (train and test/holdout datasets).
- We train the chosen base models with initial part (train dataset).
- Then we check them with the second half (holdout set).
- Now, the predictions obtained from the check half are inputs to the train higher level learner referred to as meta-model.

Coding Stage:

The Software package artifacts, on-line facilitate and test records migrated from the event surroundings to a separate test environment. All test cases are run at this point to confirm that the program is true and complete. The test suite's prosperous completion demonstrates a stable and full migration capability. Production users' square measures are outlined and connected to their acceptable roles throughout this time, and reference information is finalized for production use. The Production Initiation Plan contains the ultimate reference knowledge and production user list. An associate integrated assortment of tools, an online support system, an implementation map, a development plan that identifies reference knowledge and production users, an approval arrangement that has the ultimate suite of test cases, and an updated project plan are all products of this level.

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.

7 My learnings

During the crop prediction project, I learned that certain crops thrive in specific weather conditions. By analyzing historical data and weather patterns, I was able to identify the optimal planting times for different crops. This information can help farmers make informed decisions and maximize their yield. Additionally, I discovered that incorporating soil data into the analysis provided valuable insights into nutrient levels and soil health, which can greatly impact crop growth. Overall, this project highlighted the importance of data-driven approaches in optimizing crop production.

8 Future work scope

The production shown in the figure is presently an online application, however our future work will include developing an application that farmers can use and translating the entire system into their native language.

.