# Industrial Internship Report On

# “Prediction of Agriculture Crop Production In India”


## Prepared by

## [Pradeep Vilas More]


## 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”. In this project we have provided the data of yields of various crops in different states of India. I have to make predictions on the production of crops in India.

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

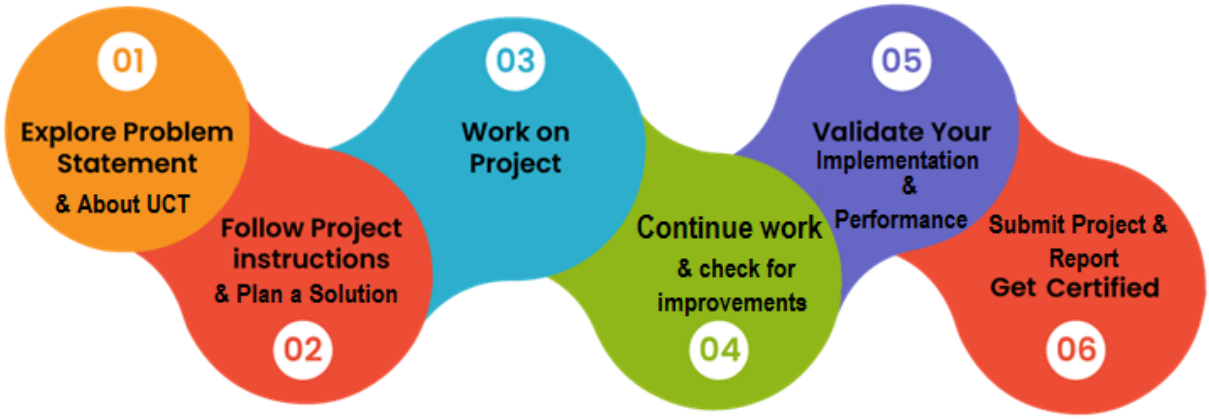
In this 6 weeks of Internship I learned about UCT(UniConverge Technologies Pvt Ltd), what it does, what kind of products this company produce, what technologies it uses. Also I learned about IoT Academy and UpSill Campus. As a Data Science and Machine Learning Intern I have learned various concept in this field through the sources provided by UpSkill Campus.

Though we learned about various technical skills through different courses getting an industrial internship experience is more important and adds value to our resume and career growth.

In this internship I was given a project problem statement about ”Prediction of Agriculture Crop Production In India”. In this project I have to make prediction about the yield of the crops in the different states of India using machine learning models.

I am thankful to USC/UCT for giving me this Internship opportunity.

How Program was planned-




I have learned lots of things in this internship. Like how to prepare for an interview, what soft skills are required to crack the interview, salaries and work profiles of various posts in the data field. Learned about various topics in data science through the resources provided by USC, important interview questions in data science and more.

I thank to all (with names), who have helped you directly or indirectly.

My message to my juniors and peers is that this internship is great for you to get an industry level experience and hands on project.

# 2 Introduction

My name is Pradeep Vilas More. I got an opportunity to work as Data Science and Machine Learning Intern at UniConverge Technologies Pvt Ltd(UCT). The Internship is provided by UpSkill Campus(USC) and IoT Academy. The duration of this Internship is 6 weeks started from June 1, 2023.

UCT is 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.

In this internship we have provided different real world project problem statements by UCT. Out of which we have to choose one problem statement and work on it throughout the internship. In the last week of our internship we have to submit our report on the internship and our project. We have to create a Github repository for submission of the Internship report and project codes.

Along with real world the UpSkill Campus' online internship portal provided us various resources on data science and machine learning like books, quizzes, information about different job roles in data field and their salaries, guidelines to crack interview and how to become successful in the corporate world etc.

I am very grateful to UCT/USC for giving me this excellent internship opportunity. It will be very helpful for me and my career growth.

## 2.1 About UniConverge Technologies Private Limited

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.

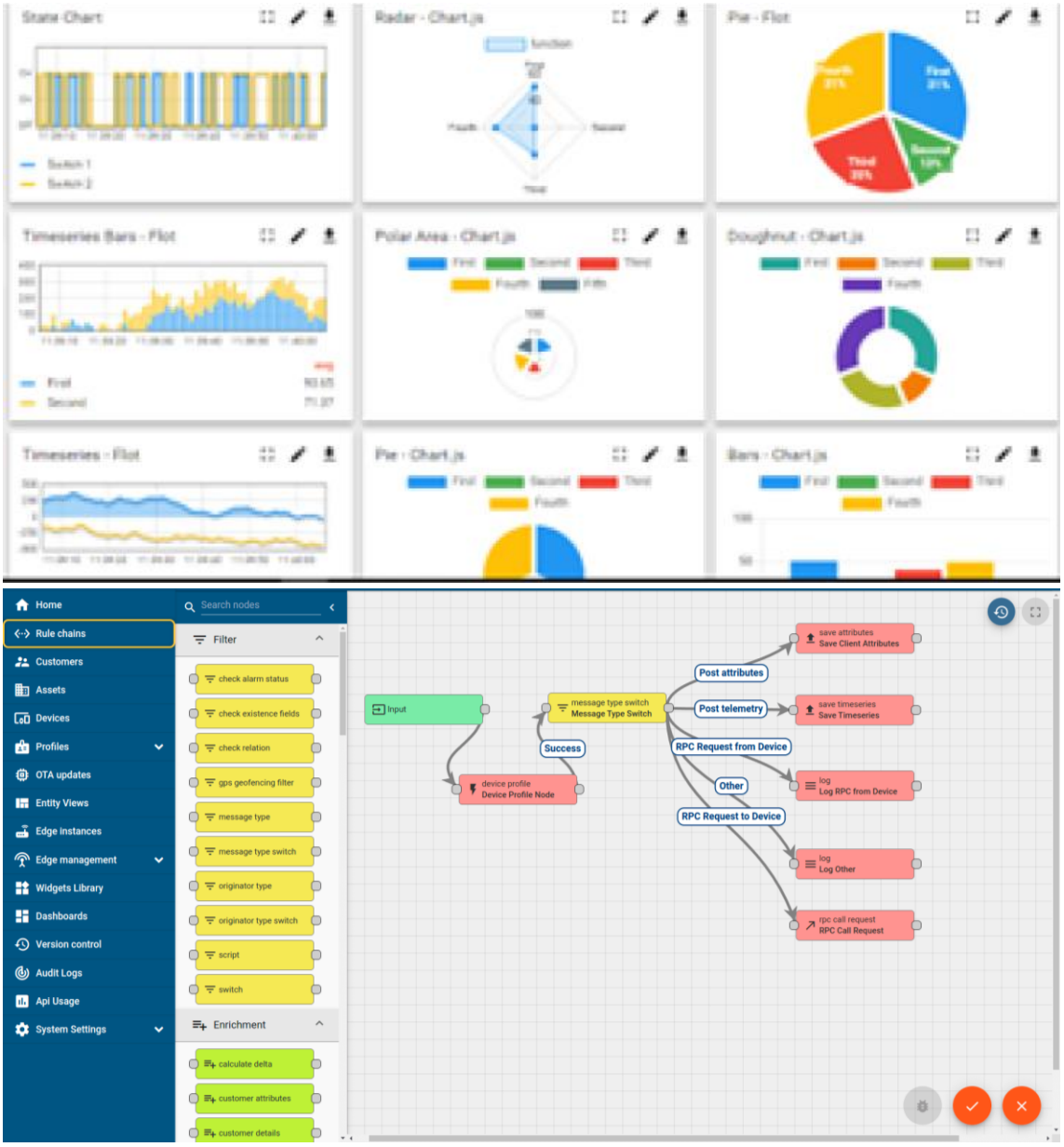

### i. UCT IoT Platform ( uct Insight )

**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.

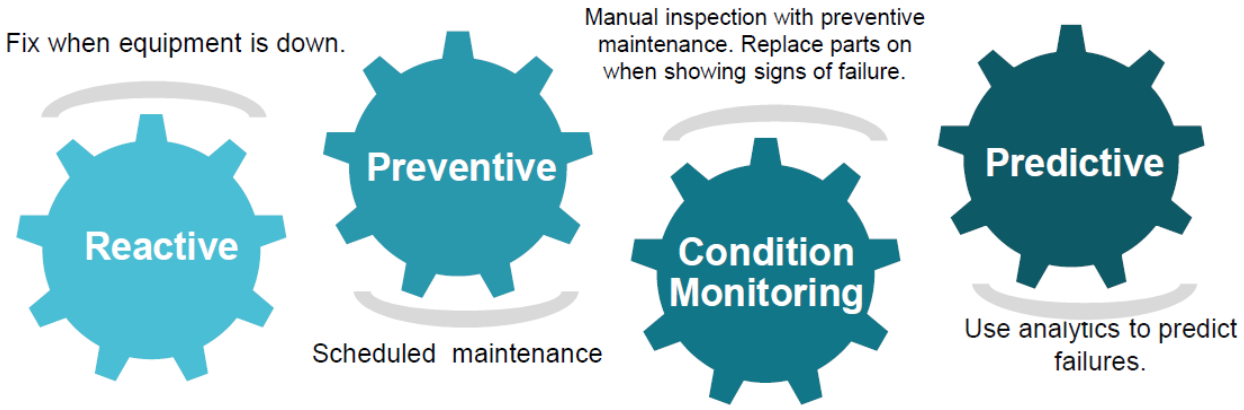

| Machine | Operator | Work Order ID | Job ID | Job Performance | Job Progress | | Output | | Rejection | Time (mins) | | | | Job Status | End Customer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Start Time | End Time | Planned | Actual | | Setup | Pred | Downtime | 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

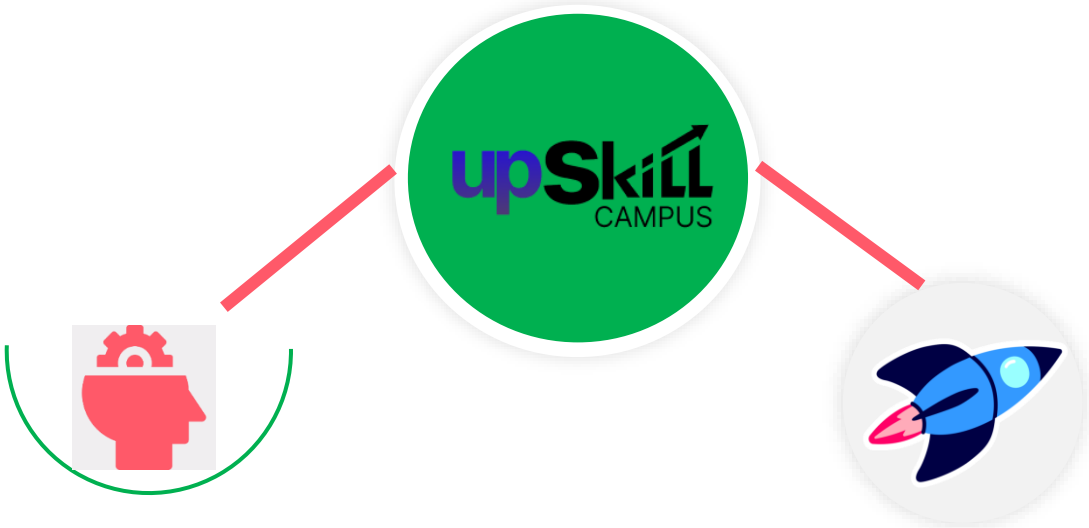
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## 2.2 About UpSkill Campus

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/

## 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] www.analyticsvidhya.com

[2] www.geeksforgeeks.org

[3] www.stackoverflow.com

[4] ChatGPT

[5] www.learn.upskillcampus.com

# 3 Problem Statement

The project problem statement selected by me for this internship is,

## “Prediction of Agriculture Crop Production in India”

In this project I have to make prdiction about yields of various crops in the different states of India. I have to use machine learning models for this and check their accuracies and performnce to chose the best model for making the prediction.

# 4 Insights of the Project

After analyzing the data I found that

- The ‘Yield(Quintal/Hectare)’ is positively correlated with the ‘Cost of Cultivation(‘/Hectare) A2+FL’ and ‘Cost of Cultivation (‘/Hectare) C2’ and negatively correlated with ‘Cost of Production (‘/Quintal) C2’.
- Sugarcane is the crop which requires highest cost for cultivation (‘/Hectare) A2+FL as well as (`/Hectare) C2.
- The cost of cultivation (‘/Hectare) A2+FL as well as (`/Hectare) C2 is lowest for Moong.
- The cost of production (‘/Quintal) C2 is highest for Moong and lowest for Sugarcane.
- The yield (Quintal/ Hectare) is highest for Sugarcane and lowest for Moong,
- The cost of cultivation (`/Hectare) A2+FL is highest in Tamil Nadu and lowest in Madhya Pradesh.
- The cost of cultivation (`/Hectare) C2 is highest in Andhra Pradesh and lowest in Bihar.
- The cost of production (`/Quintal) C2 is highest in Karnataka and lowest in Bihar.
- The yield (Quintal/ Hectare) is highest and almost equal in Karnataka and Tamil Nadu and lowest and almost equal in Haryana, Gujarat and Madhya Pradesh.
- When I am applying machine learning models to the data, I clearly observe that XGBoost regression model with Cross-Validation improves the model performance significantly.

### 4.1 Code Submission (Github Link):

**https://github.com/pradeepmore3252/UpSkill-Campus-Internship/blob/main/Prediction%20of%20Agriculture%20Crop%20Production%20In%20India/Crop_Prediction.ipynb**

### 4.2 Report Submission (Github Link):

**https://github.com/pradeepmore3252/UpSkill-Campus-Internship/blob/main/Prediction%20of%20Agriculture%20Crop%20Production%20In%20India/UpSkill%20Campus%20Internship%20Report-%20Pradeep%20More.docx**

**Github Reposatory Link:**

https://github.com/pradeepmore3252/UpSkill-Campus-Internship

# 5 Proposed Model

For the prediction of the yields of the crops in India, I used the provided data which has the list of crops, states, their cultivation costs, production costs and the yield of the crops.

I performed this project in three stages:

1. Data Importing and Data Preprocessing
2. Exploratory Data Analysis
3. Model Building and Performance Tuning.

I have done some preprocessing on the and perform exploratory data analysis on the data. I found some valuable Insights in the EDA task like how costs of cultivation, costs of production and the yield of the crop are correlated to each other, which crop has highest cultivation cost and which crop has lowest, which crop has highest production cost and which crop has lowest, which crop has highest yield and which crop has lowest.

After doing EDA on the data I build a multiple linear regression model for prediction. I checked the model performance. For improving the model performance, I used XGBoost regression model with cross validation technique which resulted in significant increase in model performance. After application of XGBoost regression model I got r2 score as 0.8837 which is quite good and close to the ideal scenario of 1.

# 6 Performance Test

The main objective of this project is to make prediction about the yield of the crops in India. We also need to build a model that gives best performance. For that I need to try different models and techniques to find the model that gives highest performance.

## 6.1 Test Plan

XGBoost is such a model that used to increase both speed and performance of the machine learning models. Cross-Validation technique is also used to improve model performance. I have used both of them in the model to get the best model performance.

## 6.2 Test Procedure

First I used a regular multiple linear regression model on the given data. It worked well but the model performance is not that good. So I tried Using XGBoost regression model with Cross-Validation technique. After using this model, I found that model performance has increased significantly.

## 6.3 Performance Outcome

Before using XGBoost regression and Cross-Validation technique the r2-score is 0.74798. After using these techniques at the same time the r2-score has increased significantly to 0.8837. This shows that the model performance has increased significantly.

# 7 My Learnings

I have learned a lot of things in this internship. I learned about the company UCT (UniConverge Technologies Pvt Ltd), which domains it works in, what kind of project/solutions does it work on, which technologies it uses. I learned about the technologies like IoT, LoRAWAN. Also learned about UpSkill Campus and IoT Academy.

I this internship I learned about various concept in data science, machine learning and statistics. Learned about impact of big data on business, difference between various job roles in the field of data like data scientist, data analysts, data engineers, machine learning engineers and so on. Also I have learned about some common and important questions and their proper answers which are frequently asked in the data science interview. I learned about how to become successful in the corporate world, how to crack an interview, what questions are asked in the interview and how to answer them efficiently.

Also I am thankful to UCT and USC to give me an opportunity to work on an industry project. I have learned a lot of things in this project. It has given me a chance to work on a real world data science project.

I hope this internship will add a value to my profile and resume. It will open doors to lot of opportunities for me. I am really grateful to UCT, USC and IoT Academy for giving me this opportunity.

# 8 Future Work Scope

During this Internship, I really want to use the Time Series forecasting for the yield of the crops in India. But because of the time limit I can't work on that. I Hope in future I will get the chance to work on such projects. Also I want to apply some deep learning models in my project but I can't. I hope this may happen in future.