Industrial Internship Report on

” PREDICTION OF AGRICULTURE CROP PRODUCTION IN INDIA”

Domain: Data Science and Machine Learning

Internship Duration: 1st August to 11th September

Institutional Affiliation: UniConverge Technologies Pvt. Ltd. & UpSkill Campus

Prepared by	**Date of Submission**
SATYAM KUMAR	**10th Sep 2023**

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.

1. In First Week, I have to Explore problem statements from project which I choose from own, actually 5-6 projects is given by UCT and I have to choose one of them and do exploring the problem statement. In addition, I have to learn about UniConverge Technologies (UCT) Company, that how it work, applications, technology used, etc.

2. In Second Week, the project i choose i need to learn the instruction of the project and plan the best solution or model for the project problem statement. My domain is "Data Science and Machine Learning". Therefore, I need to do lot of things on my dataset given by UCT like- Data Preprocessing, EDA, Modelling etc.

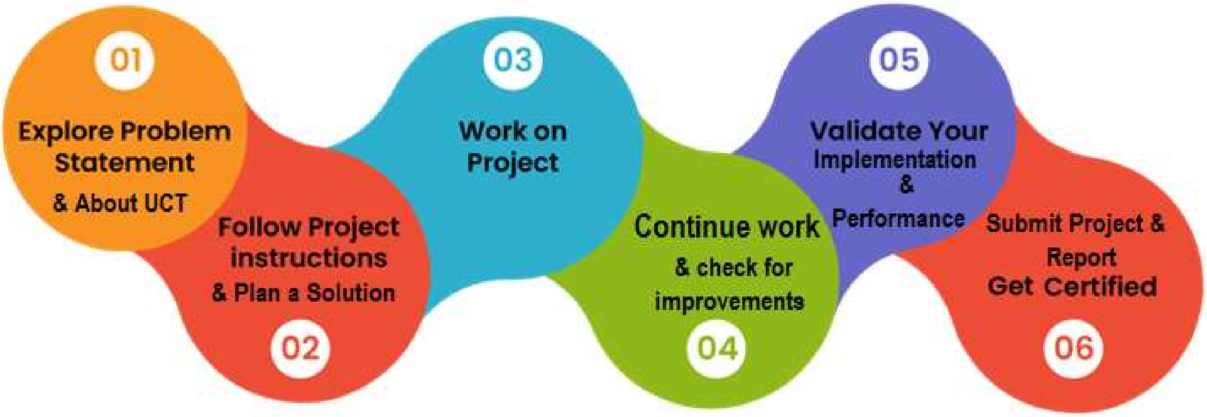
3. In Third Week, I started working on my project and explore the dataset and find some tools from google, github, reference, etc., which give me clear path that which tools, is good for the project. I started the project with Python programming language, which is consider as the best programming language for Data Science and Machine Learning. I done some little bit of Data Preprocessing and Exploratory Data Analysis (EDA).

4. In Fourth Week, I continued my working on my project and after Data Preprocessing And EDA; I started my other operations like- Outlier Removal, Resampling, Removal of Highly Correlated Features, Feature Engineering and Machine Learning Model. I done these things in my fourth week.

5. In Fifth Week, After completing my project, in my fifth week I started may be better performance of my model and like I tried to do add some other ML Algorithms which give me some great output from my previous model. In fifth week, I focus on model evaluation and improving the model performance.

6. In Sixth Week, It is time for project submission and I my project is ready

Though i learned about various technical skills through different courses getting an industrial internship experience is more important and adds value to 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.

.

Thanks to my friends, who helped me in this project.

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.

uct

Uniconverge Technologies

IIOT Products

We offer product ranging from Remote IOs, Wireless IOs, LoRaWAN Sensor Nodes/ Gateways, Signal converter and IoT gateways

IIOT Solutions

We offer solutions like OEE, Predictive Maintenance, LoRaWAN based Remote Monitoring, IoT Platform, Business Intelligence…

OEM Services

We offer solutions ranging from product design to final production we handle everything for you..

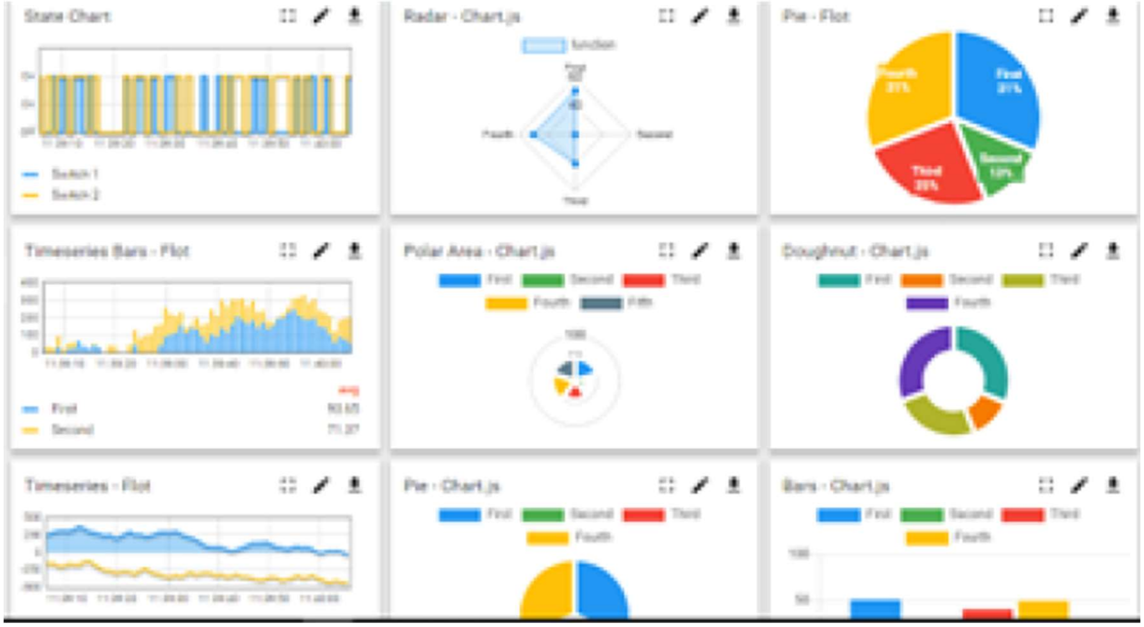
i. UCT IoT Platform (uct Insight)

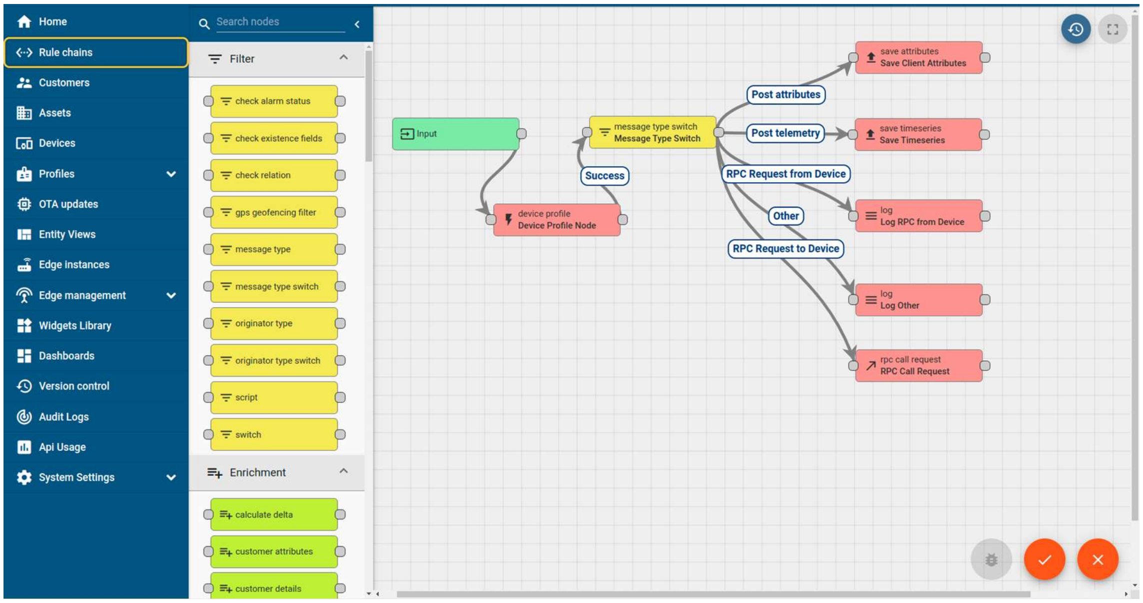
UC0T Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TC0P, OPC UA
- 00It 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





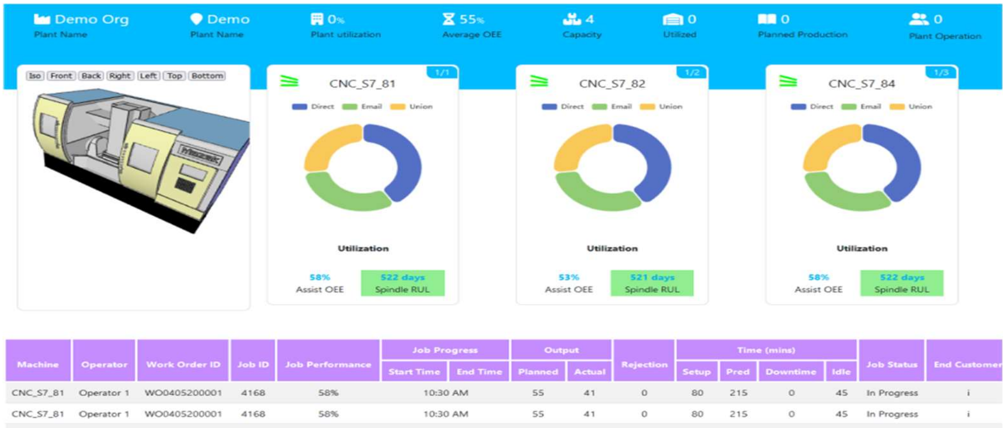
ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provid0es 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.



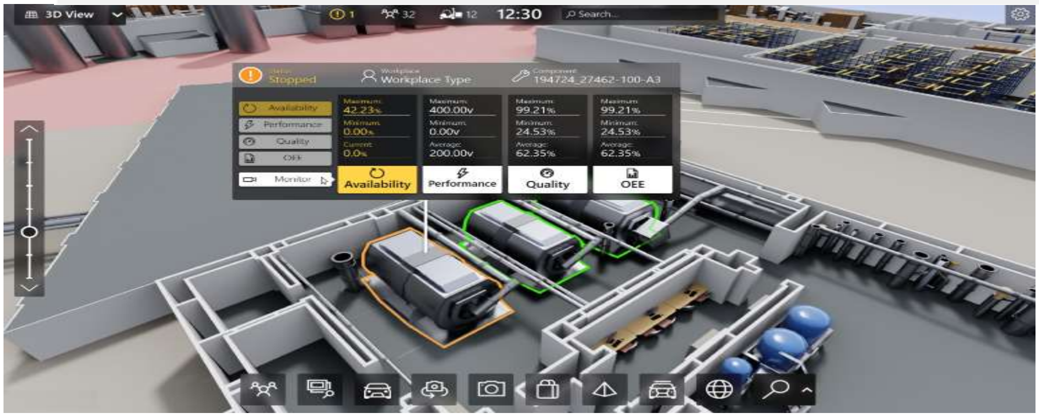
Demo Org Plant Name | Demo Plant Name | 0% Plant utilization | 55% Average OEE | 4 Capacity | 0 Utilized | 0 Planned Production | 0 Plant Operation

CNC_S7_81 (1/1): Utilization — 58% Assist OEE — 522 days Spindle RUL

CNC_S7_82 (1/2): Utilization — 53% Assist OEE — 521 days Spindle RUL

CNC_S7_84 (1/3): Utilization — 58% Assist OEE — 522 days Spindle RUL

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



Status: Stopped | Workplace: Workplace Type | Component: 194724_27462-100-A3

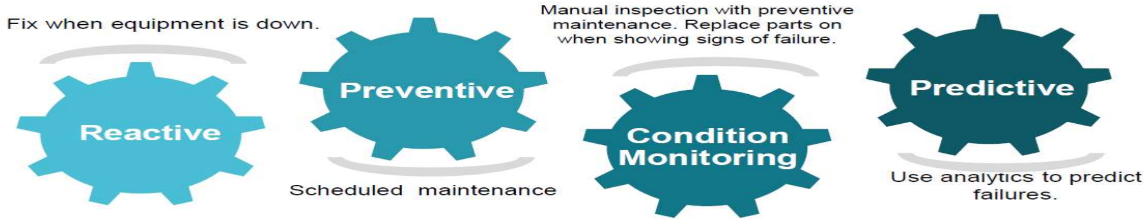
Availability	Performance	Quality	OEE
Maximum: 42.23%	Maximum: 400.00v	Maximum: 99.21%	Maximum: 99.21%
Minimum: 0.00%	Minimum: 0.00v	Minimum: 24.53%	Minimum: 24.53%
Current: 0.0%	Average: 200.00v	Average: 62.35%	Average: 62.35%

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

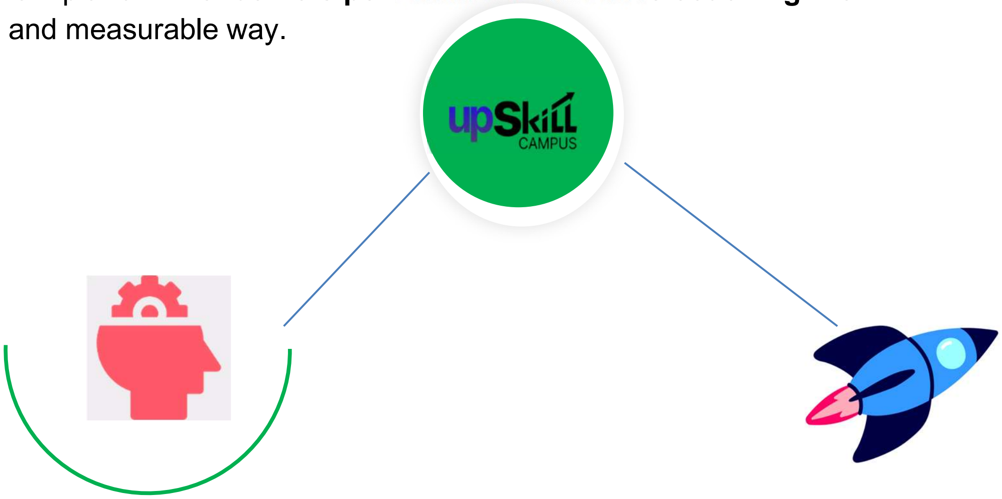
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.

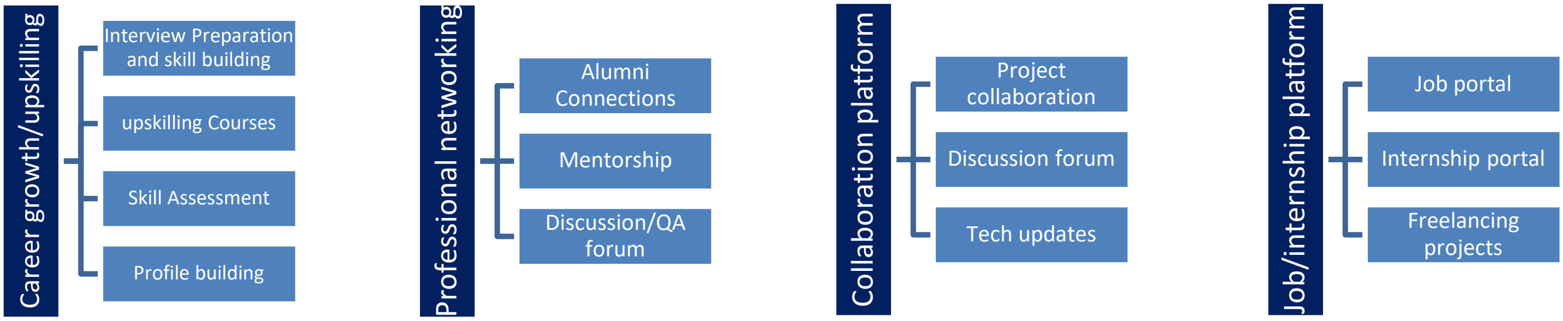

Fix when equipment is down.
Reactive
Preventive
Scheduled maintenance
Manual inspection with preventive maintenance. Replace parts on when showing signs of failure.
Condition Monitoring
Predictive
Use analytics to predict failures.

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.





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] www.google.com

[2] GitHub

[3] YouTube

[4] Domain Professionals

[5] www.analyticsvidhya.com

[6] www.geeksforgeeks.org

[7] www.stackoverflow.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 prediction 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 performance to choose the best model for making the prediction.

4 Existing and proposed solution

Proposed Solution Step by Step:

1. Importing Useful Libraries for Data Analysis and Data Manipulation

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Warnings* (Optional)

2. Data Preprocessing

3. Data Analysis

4. Exploratory Data Analysis

5. Resampling of Data

6. Machine Learning Model

- Importing Scikit–Learn Machine Learning Libraries
- Train Test Split Data
- Creating Pipeline
- Cross Validation
- Train and Test Data
- Model Performance 1 (without tuning)
- Model Tuning Using GridSerachCV
- Final Model Performance 2 (with tuning)

4.1 Code submission (Github link)

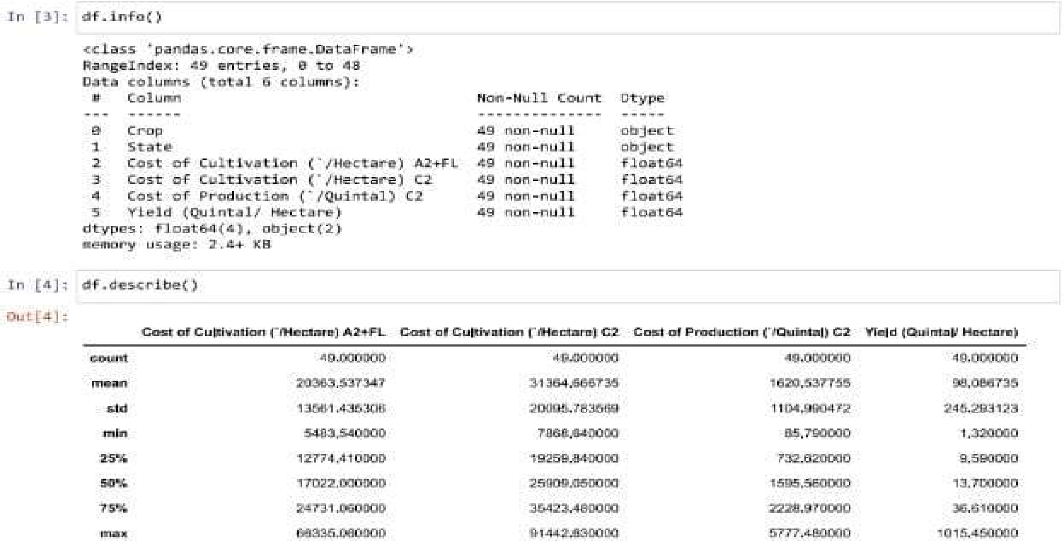
https://github.com/satyamkr21/UPSKILL-CAMPUS/blob/9c0de49a0375e0a6f53e3f33293a71cdf31760fd/CropPrediction_Satyam_Code_USC_UTC.ipynb

4.2 Report submission (Github link) :

https://github.com/satyamkr21/UPSKILL-CAMPUS/tree/9c0de49a0375e0a6f53e3f33293a71cdf31760fd

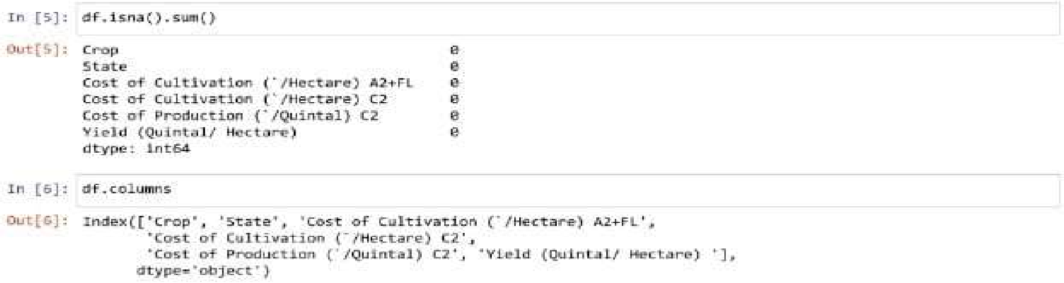
5 Proposed Design/ Model

1. Information of given data set



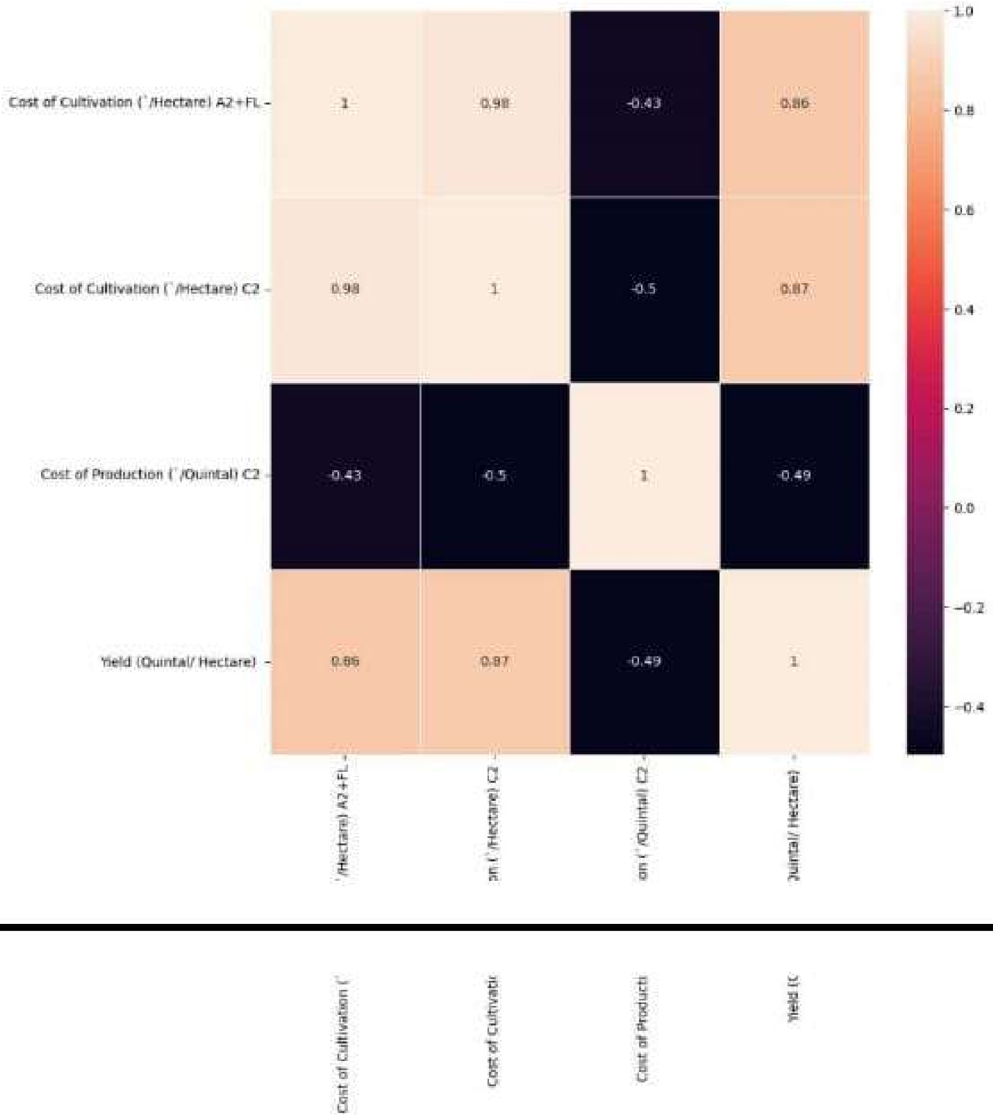
```
In [3]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49 entries, 0 to 48
Data columns (total 6 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Crop                                49 non-null     object
 1   State                               49 non-null     object
 2   Cost of Cultivation (`/Hectare) A2+FL  49 non-null     float64
 3   Cost of Cultivation (`/Hectare) C2  49 non-null     float64
 4   Cost of Production (`/Quintal) C2   49 non-null     float64
 5   Yield (Quintal/ Hectare)            49 non-null     float64
dtypes: float64(4), object(2)
memory usage: 2.4+ KB

In [4]: df.describe()

Out[4]:
```

	Cost of Cultivation (`/Hectare) A2+FL	Cost of Cultivation (`/Hectare) C2	Cost of Production (`/Quintal) C2	Yield (Quintal/ Hectare)
count	49.000000	49.000000	49.000000	49.000000
mean	20363.537347	31364.666735	1620.537755	98.086735
std	13561.435306	20095.783569	1104.990472	245.293123
min	5483.540000	7868.640000	85.790000	1.320000
25%	12774.410000	19259.840000	732.620000	9.590000
50%	17022.000000	25909.050000	1595.560000	13.700000
75%	24731.060000	35423.480000	2228.970000	36.610000
max	66335.080000	91442.630000	5777.480000	1015.450000

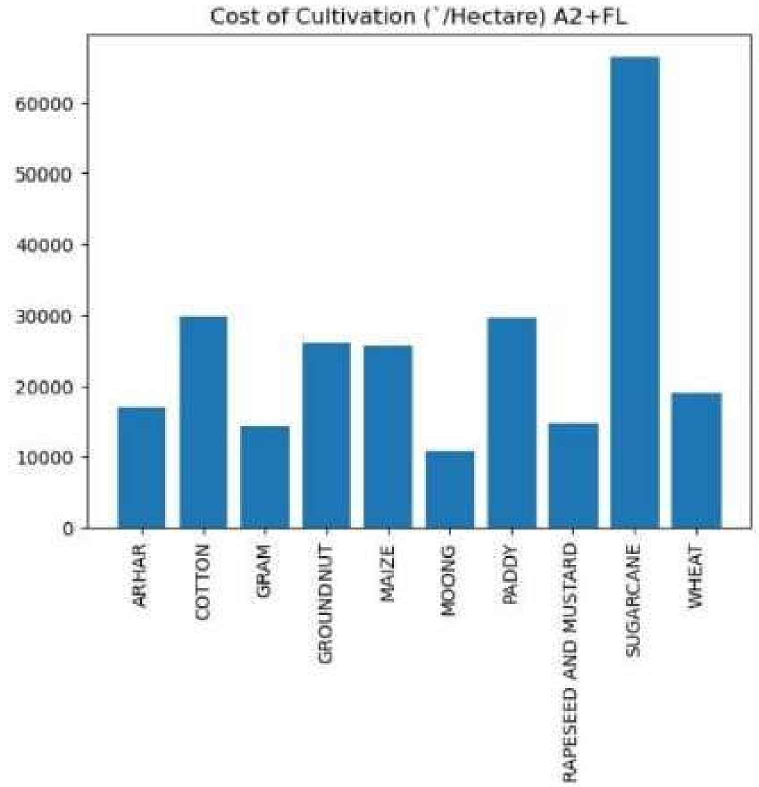


```
In [5]: df.isna().sum()

Out[5]: Crop                                   0
        State                                  0
        Cost of Cultivation (`/Hectare) A2+FL    0
        Cost of Cultivation (`/Hectare) C2     0
        Cost of Production (`/Quintal) C2      0
        Yield (Quintal/ Hectare)               0
        dtype: int64

In [6]: df.columns

Out[6]: Index(['Crop', 'State', 'Cost of Cultivation (`/Hectare) A2+FL',
               'Cost of Cultivation (`/Hectare) C2',
               'Cost of Production (`/Quintal) C2', 'Yield (Quintal/ Hectare) '],
              dtype='object')
```

2.Exploratory Data Analysis

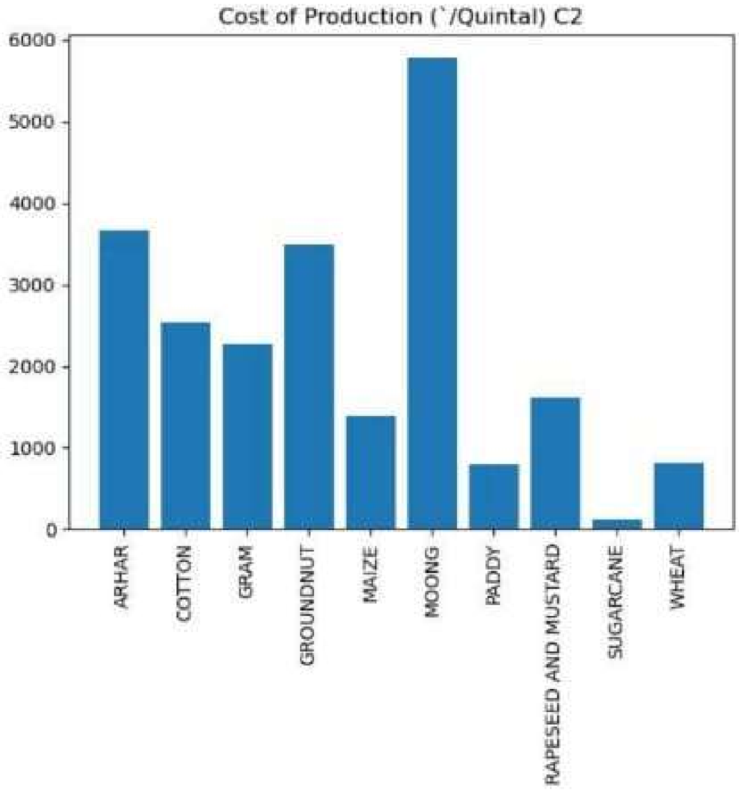


3.The cost of Cultivation(‘/Hectare)A2+FL



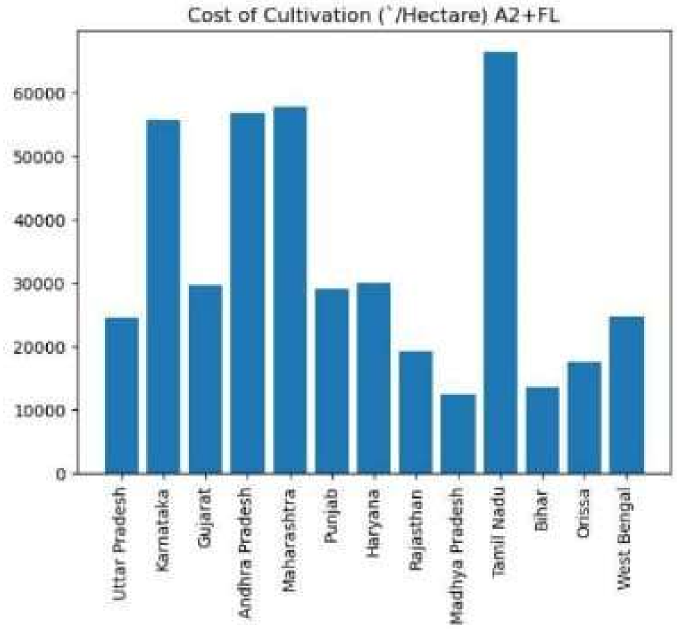
The 'Cost of Cultivation (`/Hectare) A2+FL' is highest for Sugarcane and lowest for Moong.The 'Cost of Cultivation (`/Hectare) A2+FL' is highest for Sugarcane and lowest for Moong**.**

4. Cost of Production (‘/Quintal) C2



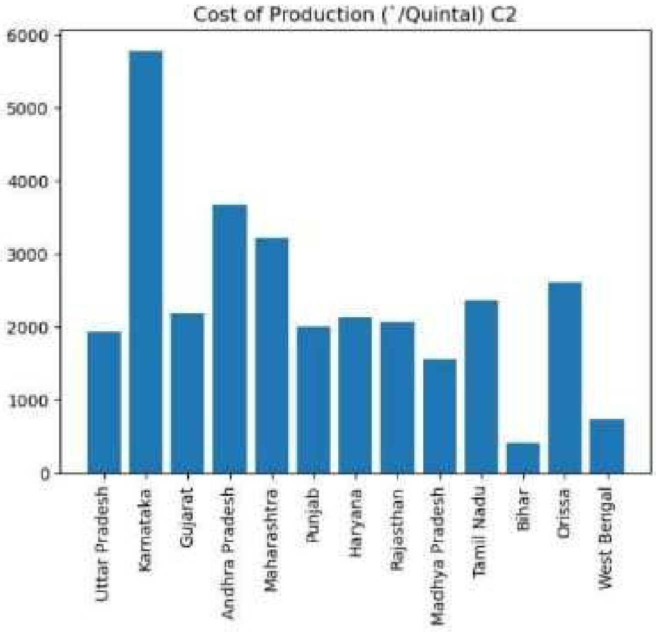
The 'Cost of Production (`/Quintal) C2' is highest for Moong and lowest for Sugarcane.

5.Cost of Cultivation (‘/ Hectare) A2+FL



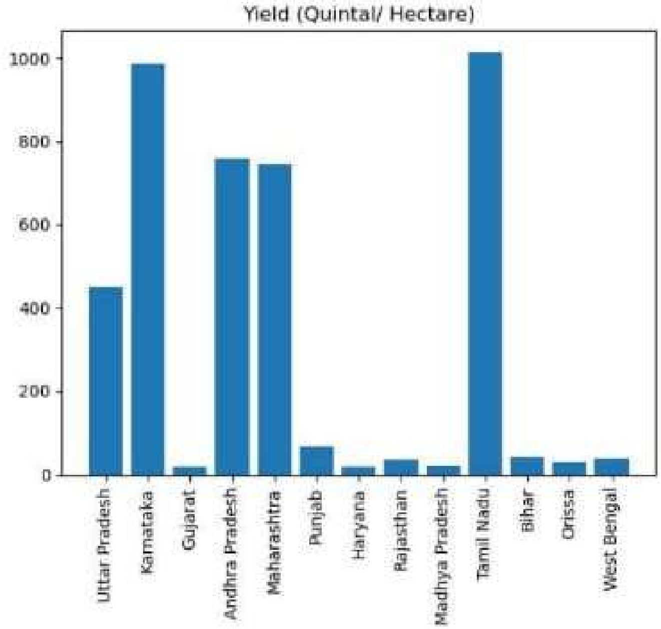
From above chart we can say that the 'Cost of Cultivation (`/Hectare) A2+FL' is highest in Tamil Nadu and lowest in Madhya Pradesh.

6.Cost of production (‘/ Quintal) C2



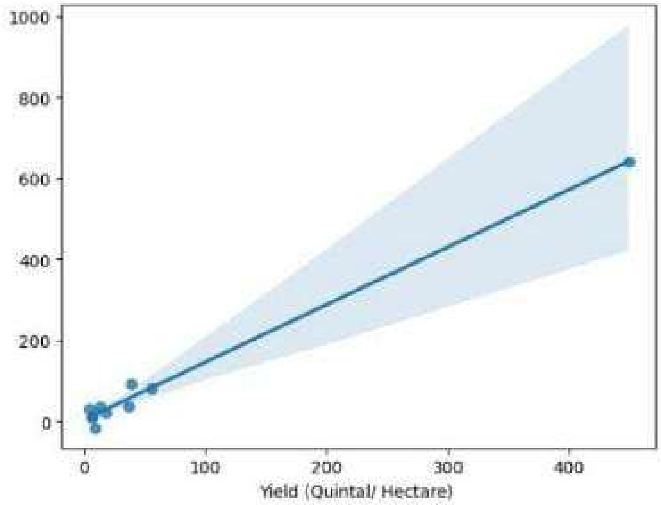
The 'Cost of Production (`/Quintal) C2' is highest in Karnataka and lowest in Bihar.

7. Yield (Quintal/Hectare)



The 'Yield (Quintal/ Hectare)' is highest and almost same in Karnataka and Tamil Nadu and lowest and almost same in Haryana, Gujrat and Madhya Pradesh

8.Predicting the test result set



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.

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

Based on the internship experience I mentioned above in the field of data science and machine learning, there are several potential future work scopes I can consider:

1. Further Research: I can use the knowledge and skills acquired during the internship to delve deeper into specific topics or areas of interest within data science and machine learning. This could involve exploring advanced machine learning algorithms, investigating cutting-edge research papers, or studying specific domains where data science techniques can be applied.

2. Advanced Projects: Undertake more complex and challenging projects to expand my practical experience. This could involve working on larger datasets, tackling real-world problems, or implementing advanced techniques such as deep learning or natural language processing.

3. Specialization: Consider specializing in a specific subfield within data science and machine learning. This could involve focusing on areas such as computer vision, natural language processing, or time series analysis. Developing expertise in a specific area can enhance my career prospects and open up niche opportunities.

4. Industry Applications: Apply my knowledge and skills to real-world applications in different industries. Data science and machine learning techniques have a wide range of applications across sectors like finance, healthcare, e-commerce, and marketing. Explore opportunities to work on projects or internships in specific industries to gain domain-specific knowledge and experience.

5. Continued Learning: Data science and machine learning are rapidly evolving fields, with new techniques and technologies emerging regularly. Stay updated with the latest developments by regularly reading research papers, participating in online courses or webinars, and joining data science communities. Continuous learning will help me stay competitive and adapt to new challenges.

6. Networking and Collaboration: Continue building professional connections in the field. Attend conferences, join industry forums or meetups, and engage in online communities to network with professionals and researchers. Collaborate on projects, share knowledge, and stay connected with the latest trends and opportunities.

7. Advanced Degree or Certification: Consider pursuing an advanced degree or certification in data science or a related field. This can provide me with a more comprehensive understanding of the subject, access to advanced coursework, and potentially open doors to higher-level positions or research opportunities.