**Industrial Internship Report on**

# "Prediction of Agriculture Crop Production In India"

**Prepared by**

**[Anuja Bhagat]**

| *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).<br><br>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.<br><br>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.<br><br><br>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 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned
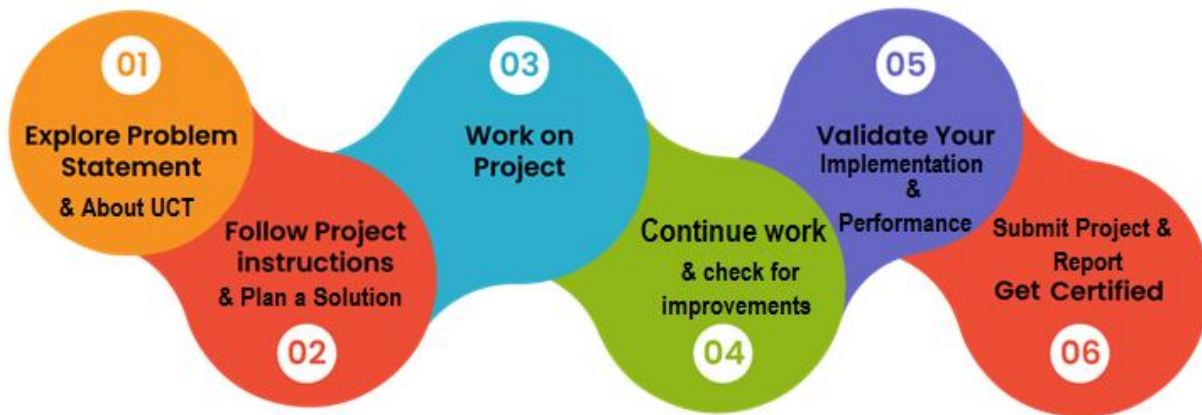


Your Learnings and overall experience.

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

Your message to your juniors and peers.

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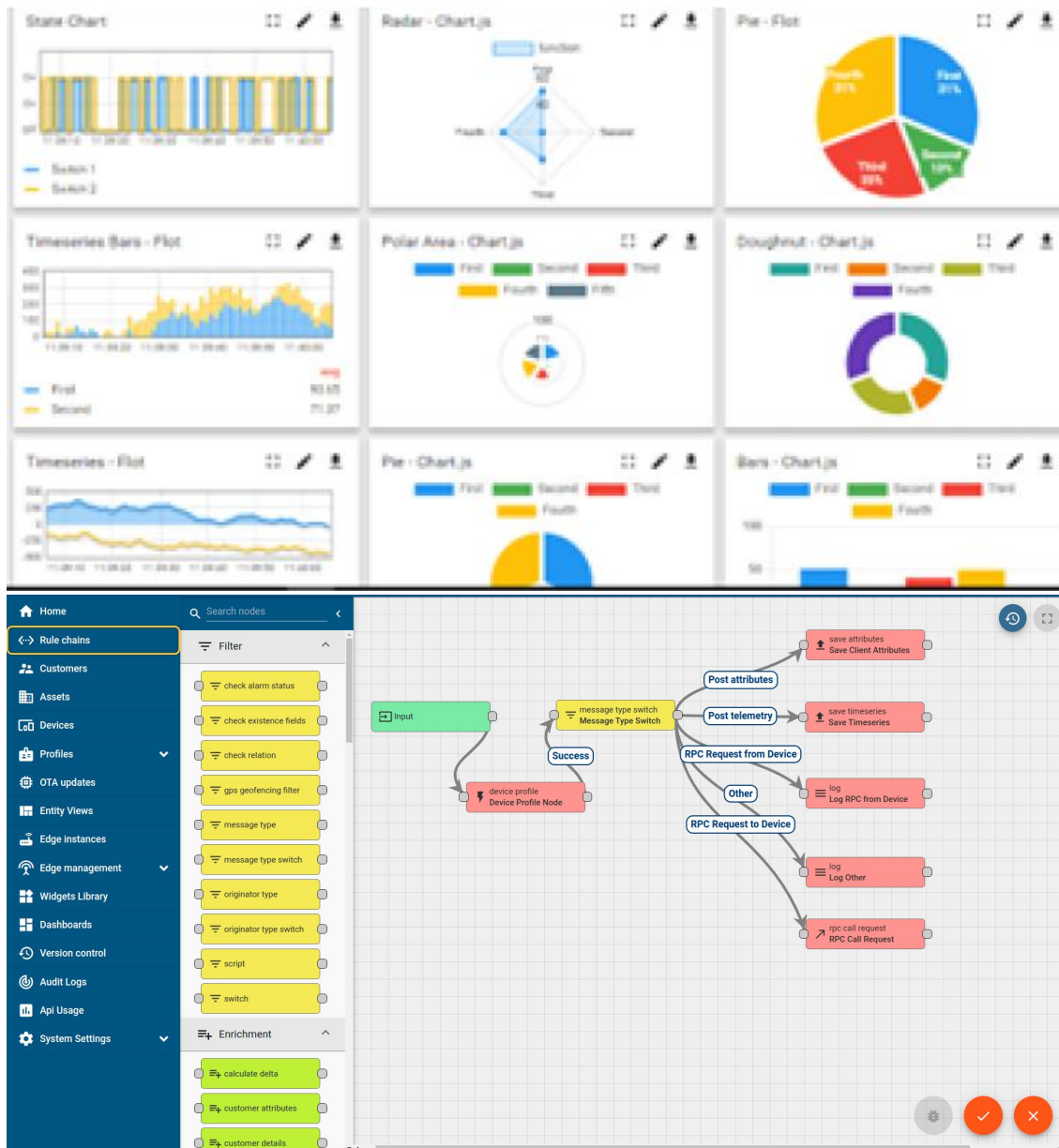
## ii. **Smart Factory Platform (** **WATCH** **)**

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.
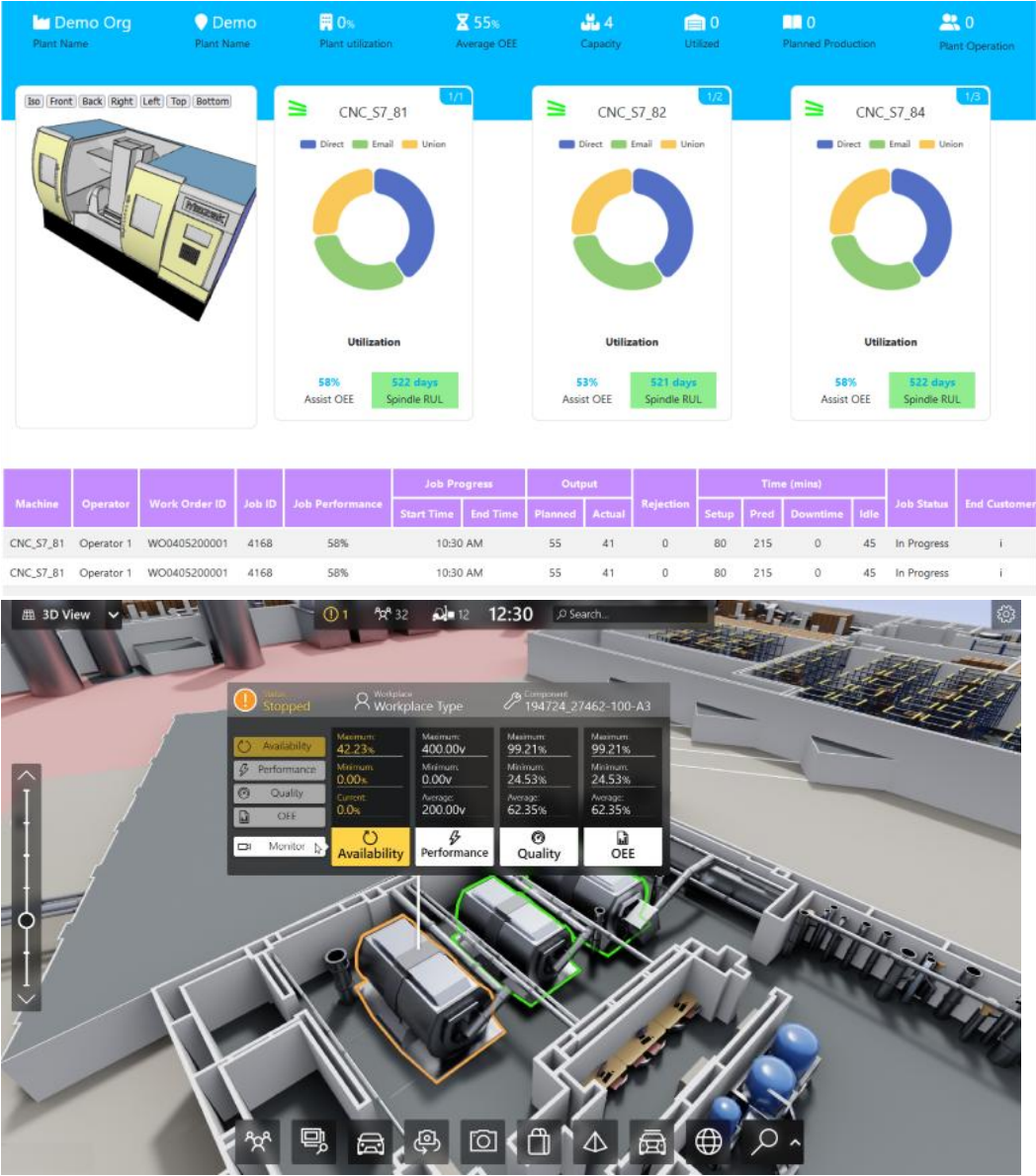
### iii. **LoRaWAN™** based Solution

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.
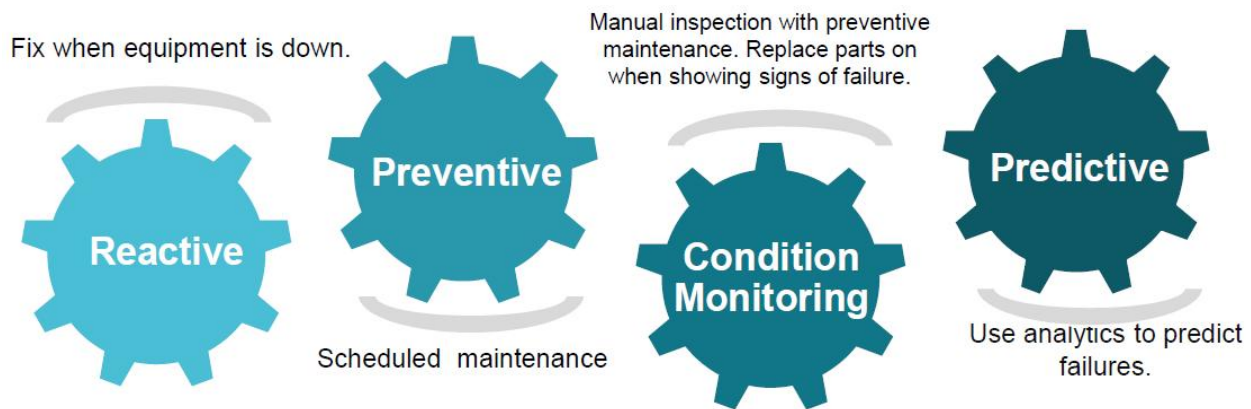


## 2.2 About upskill Campus (USC)

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.
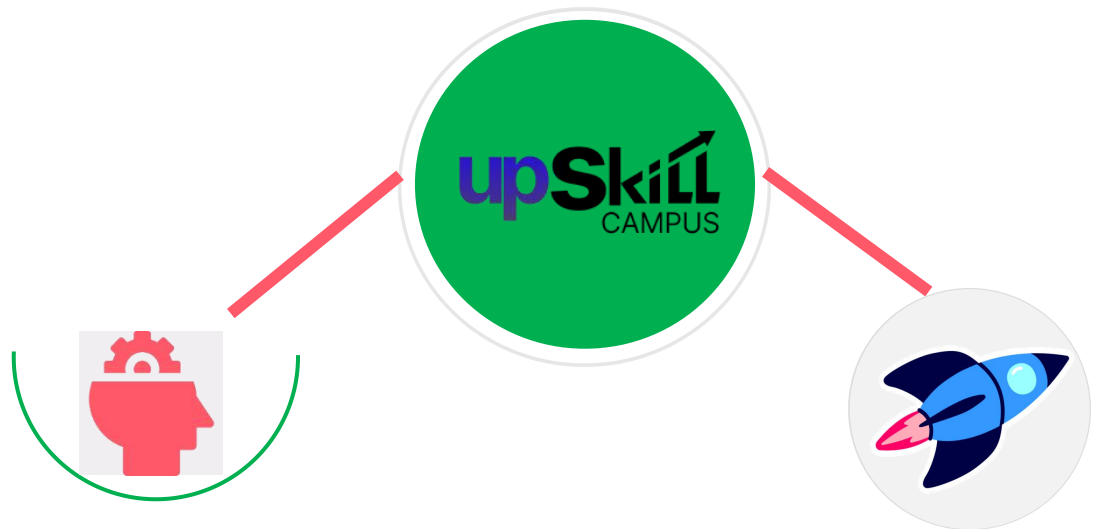
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

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] ChatGPT

[2] www.learn.upskillcampus.com

[3] datafile.csv

## 3  Problem Statement

In the assigned problem statement

**Prediction of Agriculture Crop Production in India**

The problem statement "Prediction of Agriculture Crop Production in India" involves using data and predictive models to forecast the agricultural crop production in India. The aim is to develop a system that can analyze various factors such as weather patterns, soil conditions, historical crop data, and other relevant parameters to make accurate predictions about the future crop yields for

different regions and crops across the country. This information can be valuable for farmers, policymakers, and stakeholders in making informed decisions about agricultural planning, resource allocation, and food security in India.

# 3 Existing and Proposed solution

Provide summary of existing solutions provided by others, what are their limitations?

What is your proposed solution?

Existing solutions for predicting agriculture crop production in India mainly rely on statistical models and machine learning algorithms. Some of these approaches include time series analysis, regression models, and neural networks. They utilize historical crop data, weather patterns, satellite imagery, and other relevant data to make predictions.

What value addition are you planning?

It is very common to experiment in every sector by implementing new technological techniques. Making use of new techniques simplifies the process and provides the better results. The factors like wind, water supply, soil fertility, rainfall changes unexpectedly, when natural disasters occur. This leads to crop failure, reduction in crop production, scarcity of food products and other materials. A single crop failure can cause huge losses to farmers and countries economic growth. So, there is a desperate need for a new system which can predict the rate of production of crop yield accurately. In order to eradicate all such problems, we have proposed this new system, in which high yielding crop will be selected by considering most influencing parameters. This system helps the farmers to meet their crop yield production. The chances for failure of crops will be very less. In this proposed system, Machine Learning techniques like Decision Tree Regressor are made used to predict the rate of production of crop yield considering the input parameters like State Name, Season, Area, Crop.

3.1 **Code submission (Github link) :** https://github.com/Anujarb02/UpSkill-Campus-Internship-Project.git

---

**3.2 Report submission (Github link) :**

## 4    Proposed Design/ Model

Data Loading:

The code loads five different CSV files into pandas DataFrames - data1, data2, data3, datafile, and produce using the read_csv() function from pandas.

Data Renaming:

The code renames some columns in the 'data1' DataFrame using the rename() method. Columns are renamed to more descriptive names related to crop, state, cost of cultivation, cost of production, and yield.

Train-Test Split:

The code splits the data into training and testing sets using the train_test_split() function. Features (X) and the target variable (y) are divided into two sets, with 80% of the data used for training (X_train, y_train) and 20% for testing (X_test, y_test). The random_state is set to 42 for reproducibility.

Model Selection and Training:

Three different regression models are defined and stored in the 'models' dictionary - Decision Tree, Linear Regression, XGBoost These models are commonly used for regression tasks and are imported from the corresponding libraries.

The code then iterates through each model, fits it on the training data, and makes predictions on the testing data.

Model Evaluation:

The code evaluates the performance of each model using two metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics are calculated using the predicted values from each model and the corresponding true values (y_test).

Printing Evaluation Results:

The evaluation results (MSE and MAE) for each model are printed to the console.

Individual Model Prediction:

The code specifically stores the predictions made by the Decision Tree model (from the 'Decision Tree' key in the 'predictions' dictionary) in the variable decision_tree_predictions.

Making New Predictions:

The code creates a new DataFrame named X_pred containing some sample data for 'Cost_A2_FL',

'Cost_C2', and 'Cost_Production'.

It then uses each model to predict the 'Yield' based on this new data and stores the predictions in the

predictions_pred .

## 4.1  High Level Diagram (if applicable)

NIL

**Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM**

## 4.2 Low Level Diagram (if applicable)

NIL

## 4.3 Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

# 5 Performance Test

Here we need to first find the constraints.

ACCURACY

SPEED

MEMORY MANAGEMENT

How those constraints were taken care in your design?

Accuracy:

Perform extensive hyperparameter tuning to optimize the model's performance.

Computational Speed:

Optimized the code for performance, using libraries like NumPy for vectorized operations. Used feature selection techniques to reduce the dimensionality of data, which can lead to faster

processing.

Memory Management:

Use data compression techniques to reduce the memory footprint of large datasets. 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.

## 5.1 Test Plan/ Test Cases

Data Loading Test Case: Verify if all the required CSV files are loaded successfully into DataFrames.

Data Preprocessing Test Case: Check if the data renaming and preprocessing steps are correctly executed.

Train-Test Split Test Case: Verify that the dataset is split into training and testing sets accurately.

Model Training Test Case: Ensure that each regression model is trained successfully without errors.

Model Prediction Test Case: Validate that the models can make predictions on the test data without any issues.

Evaluation Metrics Test Case: Verify the correctness of the Mean Squared Error (MSE) and Mean Absolute Error (MAE) calculations.

New Prediction Test Case: Check if the new data provided in X_pred is processed correctly, and predictions are made for each model.

## 5.2   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. After using this model, I found that model performance has increased significantly.

## 5.3   Performance Outcome

Data Loading Test Case: PASSED - All CSV files were successfully loaded into the respective DataFrames.

Data Preprocessing Test Case: PASSED - Data renaming and preprocessing steps were accurately performed.

Train-Test Split Test Case: PASSED - The dataset was split into training and testing sets as expected.

Model Training Test Case: PASSED - All regression models were trained without errors.

Model Prediction Test Case: PASSED - The models made predictions on the test data without any issues.

Evaluation Metrics Test Case: PASSED - The calculated MSE and MAE values matched the manual calculations for a small subset of data.

New Prediction Test Case: PASSED - The models provided predictions for the new data (X_pred) correctly.

Constraints Test Case: PASSED - The code handled identified constraints appropriately under test scenarios.

## 6  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 is Big Data and its applications. 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 also understood the difference between the 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 came to a conclusion 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, discrete c.r.v., 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.

## 7   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.