**Industrial Internship Report on**

**” Prediction of Agriculture Crop Production In India”**

**Prepared by**

**[Abhiyansh Gupta]**

|  |
| --- |
| *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 wasPrediction of Agriculture Crop Production In India. Agricuture Production in India from 2001-2014. Across The Globe India Is The Second Largest Country having People more than 1.3 Billion.Many People Are Dependent On The Agricuture And it is the Main Resource.In Agricuturce Cultivation/Production Having More Problems.I want to solve the Big problem in india and usefull to many more people.  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 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 9](#_Toc139702810)

[2.4 Reference 9](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[5.1 High Level Diagram (if applicable) 13](#_Toc139702816)

[5.2 Low Level Diagram (if applicable) 13](#_Toc139702817)

[5.3 Interfaces (if applicable) 13](#_Toc139702818)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Plan/ Test Cases 14](#_Toc139702820)

[6.2 Test Procedure 14](#_Toc139702821)

[6.3 Performance Outcome 14](#_Toc139702822)

[7 My learnings 15](#_Toc139702823)

[8 Future work scope 16](#_Toc139702824)

# Preface:

1. **Week 1: Problem Understanding & Data Collection**
   * Define the problem of crop yield prediction.
   * Gather relevant datasets (e.g., weather, soil, historical yields) and clean the data.
2. **Week 2: Exploratory Data Analysis (EDA)**
   * Analyze data patterns using visualizations to understand correlations between variables like temperature, rainfall, and yield.
   * Identify key factors influencing crop yield.
3. **Week 3: Feature Engineering**
   * Create new features from the data (e.g., cumulative rainfall, average temperature).
   * Normalize and prepare the dataset for machine learning models.
4. **Week 4: Model Selection & Training**
   * Train multiple models (e.g., Random Forest, Linear Regression, SVM) on the dataset.
   * Split the data into training and testing sets and tune initial models.
5. **Week 5: Model Evaluation & Optimization**
   * Evaluate model performance using metrics like R², RMSE, or MAE.
   * Optimize models by tuning hyperparameters and addressing overfitting.
6. **Week 6: Deployment & Conclusion**
   * Deploy the best-performing model (optional: create a simple web app for predictions).
   * Summarize findings, discuss challenges, and propose future improvements.

This structured plan ensures systematic progress in building and refining the machine learning model to predict agricultural crop yields effectively.

An internship is an essential step in professional and personal development, offering several key benefits, especially for students or early-career professionals. Here's why internships are valuable:

\*\*1. Practical Experience\*\*

- \*\*Real-World Application\*\*: Internships provide hands-on experience where you can apply the theoretical knowledge you've learned in your coursework.

- \*\*Skill Development\*\*: Internships help in developing technical and soft skills (e.g., communication, teamwork, problem-solving), which are crucial in the workplace.

\*\*2. Industry Exposure\*\*

- \*\*Understanding Industry Standards\*\*: Working in a real organization exposes you to the industry's practices, tools, and technologies.

- \*\*Professional Network\*\*: Internships allow you to build connections with professionals in your field, which can be beneficial for future job opportunities.

\*\*3. Career Exploration\*\*

- \*\*Clarifying Career Goals\*\*: Internships help you explore different career paths, giving you a clearer idea of what roles you might want to pursue after graduation.

- \*\*Industry Insights\*\*: You gain a deeper understanding of how the industry works, what challenges it faces, and what opportunities are available.

\*\*4. Building Your Resume\*\*

- \*\*Gaining Credibility\*\*: Employers often look for experience when hiring, and internships show that you have real-world exposure and are ready to contribute.

- \*\*Project Work\*\*: You may work on projects that can be highlighted in your portfolio or resume, making you more competitive in the job market.

\*\*5. Transition to Full-Time Jobs\*\*

- \*\*Foot in the Door\*\*: Many companies offer full-time positions to successful interns, making it a direct pathway to employment.

- \*\*Job Readiness\*\*: Internships help you understand the expectations of full-time roles, making the transition smoother when you graduate or move into the job market.

\*\*6. Personal Growth\*\*

- \*\*Confidence Building\*\*: Internships help boost your confidence as you navigate real-world challenges and contribute to projects.

- \*\*Time Management & Discipline\*\*: Internships often require you to balance multiple responsibilities, helping you improve time management and discipline.

In summary, internships provide a bridge between academic learning and professional careers, offering invaluable experience that enhances your employability and helps clarify your career goals.

The project focuses on using machine learning to predict agricultural crop yields based on various factors such as weather conditions, soil quality, and historical crop data. By analyzing patterns in these variables, the model aims to accurately forecast future yields, helping farmers optimize their resources and make data-driven decisions. The project involves data collection, feature engineering, model selection, and evaluation to identify the best predictive model for crop yield estimation..

UCT offers significant research opportunities, especially in areas like sustainability, development, and global health, with a strong focus on African and global challenges.

* **Define Goals**: The first step involves clearly defining the objectives and learning outcomes. For example, the goal might be to provide students with hands-on industry experience or develop specialized knowledge in a certain academic field.
* **Curriculum Structure**: Determine the core areas of knowledge and skills that the program will cover. At universities like USC or UCT, this could include theoretical classes, practical lab sessions, and real-world projects.

### ****2. Curriculum Design****

* **Core and Elective Courses**: Programs are designed to offer a combination of core courses (mandatory) that provide foundational knowledge and elective courses that allow students to specialize in areas of interest.
* **Interdisciplinary Focus**: Depending on the program, there may be opportunities to study across multiple disciplines, for example, combining business with technology, or health sciences with social impact.

### ****3. Integration of Practical Experience****

* **Internships/Work Placement**: To ensure students gain practical skills, many programs include mandatory internships or industry placements. USC, for example, has strong ties with industries in LA (e.g., film, tech, business), while UCT offers opportunities to work with local businesses or non-profits.
* **Capstone Projects**: Many programs culminate in a final project, which is a real-world application of the skills learned. This might involve research, creating a product, or solving a business problem.

### ****4. Research and Innovation****

* **Research Opportunities**: For advanced programs, research components are integrated into the curriculum. Both UCT and USC provide platforms for students to engage in cutting-edge research with faculty or through labs.
* **Innovation Labs/Incubators**: Both universities often have innovation hubs or incubators that support entrepreneurial ventures or creative projects, giving students a chance to launch startups or develop new technologies.

### ****5. Industry Partnerships and Networking****

* **Industry Collaborations**: Programs often include guest lectures from industry experts, collaborative projects with companies, or case studies based on real-world issues. USC’s connections with Hollywood or Silicon Beach (for tech) and UCT’s ties to South African industries are examples.
* **Networking Opportunities**: Programs are structured to provide networking events, career fairs, and industry conferences where students can connect with professionals and potential employers.

### ****6. Continuous Assessment and Feedback****

* **Assessments**: Programs are planned to include regular assessments, such as exams, assignments, group projects, and presentations to ensure that students are progressing.
* **Feedback Loops**: Continuous feedback from professors and mentors is integral to improve the learning process. This can also involve peer assessments or industry evaluations during internships.



### ****1. Practical Experience****

* **Real-World Application**: Internships provide hands-on experience where you can apply the theoretical knowledge you've learned in your coursework.
* **Skill Development**: Internships help in developing technical and soft skills (e.g., communication, teamwork, problem-solving), which are crucial in the workplace.

### ****2. Industry Exposure****

* **Understanding Industry Standards**: Working in a real organization exposes you to the industry's practices, tools, and technologies.
* **Professional Network**: Internships allow you to build connections with professionals in your field, which can be beneficial for future job opportunities.

Thank to all , who have helped you directly or indirectly.

Work hard,and Do it.

# Introduction

## 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.



1. 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

 

1. **Smart Factory Platform (****)**

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.

 

1.  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.

1. 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.



## 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

<https://www.upskillcampus.com/>

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



## 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.

## 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.

## Reference

[1]. **Biggs, J., & Tang, C. (2011).** Teaching for Quality Learning at University: What the Student Does. McGraw-Hill Education.

[2] . **Suskie, L. (2018).** Assessing Student Learning: A Common Sense Guide. Jossey-Bass.

[3]. University of Southern California (USC) Career Center Website

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| svm | Support Vector Machines (SVMs) ek popular supervised learning technique hai jo classification aur regression problems ko solve karne ke liye use hoti hai. |
| Linear regression | inear regression ek simple aur commonly used statistical technique hai jo ek dependent variable aur ek ya zyada independent variables ke beech linear relationship establish karne ke liye use hota hai. Iska main goal hota hai dependent variable ki value ko predict karna independent variables ke basis par. |
| Logistic regreesion | Logistic regression is widely used in various fields such as medicine, finance, and social sciences for binary classification problems, and it can be extended to multi-class classification problems using techniques like multinomial logistic regression |
| Decision Tree | A Decision Tree is a popular machine learning algorithm used for both classification and regression tasks. It models decisions and their possible consequences, including chance event outcomes, resource costs, and utility |
|  |  |

# Problem Statement

In the assigned problem statement

* In the context of global agricultural production, accurately predicting crop yields is crucial for optimizing resource allocation, ensuring food security, and enhancing sustainability.
* Despite the availability of various data sources such as weather conditions, soil quality, and historical crop performance, traditional forecasting methods often fall short in terms of accuracy and adaptability.
* This project aims to develop a machine learning model that leverages diverse datasets—such as climatic variables, soil metrics, and historical yield data—to predict future crop yields with high precision.
* The objective is to provide farmers and agricultural stakeholders with a reliable tool that enhances decision-making processes, ultimately leading to improved crop management and better resource utilization.

# Existing and Proposed solution

 **Data Collection and Preparation**

* **Gather Data**: Collect relevant datasets including weather data (temperature, rainfall, humidity), soil properties (pH, moisture, nutrient levels), and historical crop yield data from sources such as agricultural databases, weather stations, and satellite imagery.
* **Data Cleaning**: Handle missing values, remove outliers, and normalize or scale the data as needed. Ensure the data is structured in a format suitable for analysis.

 **Exploratory Data Analysis (EDA)**

* **Analyze Data**: Perform exploratory data analysis to understand relationships between variables. Use visualizations such as scatter plots, heatmaps, and correlation matrices to identify patterns and trends.
* **Feature Selection**: Identify key features that most strongly influence crop yields based on statistical analysis and domain knowledge.

 **Feature Engineering**

* **Create Features**: Develop new features that may enhance model performance, such as cumulative rainfall over the growing season, average temperature, or soil moisture trends.
* **Transform Data**: Apply transformations to improve model performance, such as encoding categorical variables or creating interaction terms between features.

 **Model Selection and Training**

* **Choose Algorithms**: Select appropriate machine learning algorithms for prediction, such as Linear Regression, Random Forest, Gradient Boosting, or Neural Networks. Consider using ensemble methods or advanced models if necessary.
* **Train Models**: Split the dataset into training and testing subsets. Train the models using the training data and evaluate their performance using metrics such as R², Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

 **Model Evaluation and Optimization**

* **Evaluate Performance**: Assess the models' performance on the test dataset. Compare different models to determine which one provides the best accuracy and generalizability.
* **Optimize Models**: Fine-tune hyperparameters using techniques like Grid Search or Random Search to improve model performance. Address any issues related to overfitting or underfitting.

 **Deployment**

* **Develop a Predictive Tool**: Implement the final model in a user-friendly application or dashboard. This tool should allow users to input relevant data (e.g., current weather, soil conditions) and receive yield predictions.
* **Integrate with Existing Systems**: Ensure the tool can integrate with other agricultural management systems if needed.

 **Validation and Testing**

* **Test with Real-World Data**: Validate the tool by testing it with new data or in real-world scenarios to ensure its accuracy and reliability.
* **Gather Feedback**: Collect feedback from end-users (e.g., farmers, agricultural experts) to make any necessary improvements.

 **Documentation and Reporting**

* **Document Process**: Create detailed documentation of the model development process, including data preparation, feature engineering, model selection, and evaluation.
* **Report Findings**: Prepare a comprehensive report or presentation that summarizes the project's objectives, methodology, results, and recommendations for future improvements.

## Code submission:

https://github.com/Abhi10102003/project2\_abhiyansh\_gupta\_upskill

## Report submission (Github link) : https://github.com/Abhi10102003/project2\_abhiyansh\_gupta\_upskill

# Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

# 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.

## Test Plan/ Test Cases

## Test Procedure

## Performance Outcome

# My learnings

 **Data Collection and Preparation**

* **Gather Data**: Collect relevant datasets including weather data (temperature, rainfall, humidity), soil properties (pH, moisture, nutrient levels), and historical crop yield data from sources such as agricultural databases, weather stations, and satellite imagery.
* **Data Cleaning**: Handle missing values, remove outliers, and normalize or scale the data as needed. Ensure the data is structured in a format suitable for analysis.

 **Exploratory Data Analysis (EDA)**

* **Analyze Data**: Perform exploratory data analysis to understand relationships between variables. Use visualizations such as scatter plots, heatmaps, and correlation matrices to identify patterns and trends.
* **Feature Selection**: Identify key features that most strongly influence crop yields based on statistical analysis and domain knowledge.

 **Feature Engineering**

* **Create Features**: Develop new features that may enhance model performance, such as cumulative rainfall over the growing season, average temperature, or soil moisture trends.
* **Transform Data**: Apply transformations to improve model performance, such as encoding categorical variables or creating interaction terms between features.

 **Model Selection and Training**

* **Choose Algorithms**: Select appropriate machine learning algorithms for prediction, such as Linear Regression, Random Forest, Gradient Boosting, or Neural Networks. Consider using ensemble methods or advanced models if necessary.
* **Train Models**: Split the dataset into training and testing subsets. Train the models using the training data and evaluate their performance using metrics such as R², Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

### ****4. Research and Innovation****

* **Research Opportunities**: For advanced programs, research components are integrated into the curriculum. Both UCT and USC provide platforms for students to engage in cutting-edge research with faculty or through labs.

# Future work scope

You can put some ideas that you could not work due to time limitation but can be taken in future.