**Industrial Internship Report on**

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

**Prepared by**

**Pavitra Ashwin Chavda**

|  |
| --- |
| *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. This project focuses on leveraging data science and machine learning techniques to address agricultural challenges in India from 2001 to 2014. The dataset used contains information on crop cultivation and production, including crop varieties, locations, quantities, costs, and recommended zones.  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

**Summary of the Whole 6 Weeks' Work:**

Over the past six weeks, I have embarked on a comprehensive journey of learning and skill development in the fields of data science and machine learning. I started by laying the groundwork, covering fundamental concepts in data science, probability, statistics, and machine learning. Through this foundational knowledge, I gained the tools necessary to tackle real-world challenges.

The bulk of my work has been dedicated to the implementation of an agriculture project that aims to optimize crop production practices in India. I have processed and cleaned a substantial dataset, conducted exploratory data analysis to uncover trends and patterns, and begun building predictive models to forecast crop production. While the project is ongoing, I have made substantial progress, and its application holds the potential to significantly impact the agricultural sector.

Throughout this journey, I have also explored critical topics such as AI versus data science, the importance of linear functions, optimization techniques, and the skills required for success in the corporate world. These insights have enriched our understanding of the broader data science landscape and its relevance in both industry and academia.

**About need of relevant Internship in career development:**

Internships play a pivotal role in career development, bridging the gap between theoretical knowledge and practical application. They provide a unique opportunity to gain hands-on experience in a real-world setting, apply classroom learning to tangible projects, and develop essential soft skills like communication, teamwork, and problem-solving.

For aspiring data scientists and professionals in related fields, relevant internships are particularly crucial. They offer exposure to the intricacies of data-driven decision-making, enable interaction with industry experts, and allow individuals to build a portfolio of practical projects. Ultimately, a well-structured internship can significantly enhance one's employability and career prospects.

**Brief about Your project/problem statement:**

This project focuses on addressing challenges in agriculture production in India. The problem statement revolves around leveraging a comprehensive dataset covering crop cultivation and production from 2001 to 2014. By analyzing this dataset, goal is to provide insights and predictive models that can optimize crop cultivation practices, resource utilization, and crop yields. The project is a real-world application of data science and machine learning to a critical sector that impacts the livelihoods of millions.

**Opportunity given by USC/UCT:**

The opportunity provided by USC/UCT has been invaluable in our professional development. Through this program, I have had access to structured learning modules, resources, and guidance from experienced instructors. The program has not only equipped us with foundational knowledge but has also given us the chance to work on a practical project that aligns with real-world challenges.

Furthermore, the exposure to various topics, including career development insights, has broadened our perspectives on the data science field. USC/UCT's commitment to fostering an environment of learning and growth has paved the way for a fruitful journey of skill acquisition and project implementation.

**How Program was planned:**



**Your Learnings and overall experience:**

Throughout the six weeks of this program, the journey has been one of profound learning, growth, and discovery. The experience has not only expanded our knowledge but has also shaped our perspective on the dynamic and rapidly evolving field of data science and machine learning.

Thank to all who have helped you directly or indirectly. Your support, guidance, and collective efforts have been instrumental in making this experience both enriching and rewarding. It is through the collaboration and shared knowledge of the broader community that we have been able to grow and make strides in the fields of data science and machine learning. Your contributions have not gone unnoticed, and we deeply appreciate the collective effort that has made this project possible.

# 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]

[2]

[3]

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Problem Statement

The project's problem statement revolves around the optimization of agriculture production in India from 2001 to 2014, a period marked by significant variability in crop yields and resource challenges. The primary objective is to employ data science and machine learning techniques to analyze historical agricultural data, providing actionable insights and predictive models to enhance crop cultivation practices, resource allocation, and ultimately, crop yields. This project addresses key challenges such as variable crop production, resource optimization, and climate sensitivity, aiming to contribute to more sustainable and efficient agricultural practices in India.

# Existing and Proposed solution

In the realm of agricultural optimization, existing solutions have often relied on traditional, manual practices guided by the expertise of local farmers and agricultural experts. While this knowledge is invaluable, it has its limitations. These approaches may struggle to scale efficiently to cover the vast and diverse agricultural landscapes of India. Moreover, they might not fully harness the potential of historical data and data-driven insights, often lacking the predictive capabilities that can enable proactive resource allocation and decision-making.

In response to these limitations, our proposed solution stands at the intersection of data science and machine learning. We intend to analyze extensive historical agricultural data spanning from 2001 to 2014 to uncover hidden trends and correlations. By training machine learning models, we aim to predict crop production based on various attributes.

The innovation lies in the integration of these data-driven insights into agriculture practices, providing farmers with the means to make informed decisions. Resource allocation, including water, fertilizers, and pesticides, will be optimized to maximize crop yields while minimizing costs and environmental impact. Furthermore, our solution is designed with scalability in mind, ensuring it can be tailored to suit the specific needs of diverse farming communities across India. Ultimately, our solution promises to not only address the limitations of traditional approaches but also add value through data-driven, predictive, and sustainable agricultural practices.

## Code submission (Github link)

https://github.com/pavitrachavda97/upskillcampus

## Report submission (Github link) :

https://github.com/pavitrachavda97/upskillcampus

# Proposed Design/ Model

Initialization is where we lay the foundation. Here, I meticulously define the project's objectives, delving into the core problem statement and setting clear success metrics. Data collection is a critical step in this phase, involving the acquisition of relevant datasets and the establishment of data pipelines. The data collected is then subjected to thorough preprocessing, which includes handling missing values, outliers, and ensuring data readiness for analysis.

Intermediate Stages represent the core of our project's development. Feature engineering takes center stage as I craft meaningful features from the data, enhancing model performance. It involves an array of transformations, scaling, and encoding of categorical variables. The process significantly influences the project's accuracy and predictive capabilities. Model selection follows, as I carefully choose the most suitable machine learning algorithms or data analysis techniques based on the project's objectives. This stage is marked by experimentation and refinement as we iterate through different models to optimize results.

Finally, the Final Outcome is the culmination of our efforts. Here, we implement the selected model(s) and fine-tune them for optimal performance. The results are evaluated rigorously using appropriate metrics, ensuring that the solution meets predefined success criteria. The final outcome provides actionable insights or predictions that address the project's initial objectives. This structured design flow guides our journey from project inception to the realization of data-driven solutions, making the process efficient and effective across diverse domains.

## High Level Diagram (if applicable)

## Low Level Diagram (if applicable)

## Interfaces (if applicable)

# Performance Test

**Identified Constraints:**

* Memory
* Speed
* Accuracy

**How to Addressed Them:**

* Memory Optimization
* Efficient Algorithms
* Trade-Off Analysis

**Test Results and Impact:**

Through rigorous testing, I have assessed the performance of our solution under varying constraints. I have measured memory usage, processing speed, and model accuracy across different scenarios. These tests allow me to understand the trade-offs and limitations of our design.

## Test Plan/ Test Cases

Test plan and test cases are pivotal components of ensuring the reliability and functionality of my solution. In the test plan, I have meticulously outlined the objectives, scope, and approach for testing my data science and machine learning models. This plan encompasses a comprehensive suite of test cases, each designed to evaluate specific aspects of my system's performance.

The test cases encompass a wide range of scenarios, including various input data conditions, potential edge cases, and hypothetical user interactions. I have also considered negative test cases to verify how the system handles unexpected or erroneous inputs. These test cases have been meticulously designed to validate the accuracy, robustness, and resilience of our solution, reflecting real-world usage patterns.

## Test Procedure

Test procedure is a systematic and well-structured approach that outlines how each test case will be executed and what specific metrics or criteria will be used to assess the outcome. I've established a controlled testing environment, configured with representative data and conditions to simulate real-world scenarios.

For each test case, I specify the input data, the expected output or behaviour, and the acceptance criteria that determine whether the test case passes or fails. My testing procedure ensures consistency and repeatability, allowing me to identify any inconsistencies or anomalies in the system's performance.

## Performance Outcome

The performance outcome of tests is the culmination of my rigorous testing efforts. Through the execution of test cases as per the test plan and procedure, I have obtained comprehensive results that provide insight into my solution's capabilities.

The performance outcome encompasses various aspects, including the accuracy of predictive models, system responsiveness under different loads, memory usage, and compliance with defined constraints. This outcome serves as a critical benchmark for gauging the reliability and effectiveness of solution, ensuring that it meets the expectations and requirements of real industries.

# My learnings

My overall learning from this project and program has been transformative, equipping me with a diverse skill set and a deep understanding of data science and machine learning. These learnings are not only academically enriching but also profoundly impactful on my career growth.

First and foremost, I have gained proficiency in the practical application of data science techniques. Through hands-on experience with real-world data and complex problem-solving, I have honed my skills in data preprocessing, feature engineering, model selection, and performance evaluation. This practical knowledge is directly transferable to data science roles in industry, where the ability to derive actionable insights from data is paramount.

Furthermore, the exposure to diverse topics such as AI, machine learning algorithms, optimization techniques, and career development has broadened my perspective on the data science field. It has given me the versatility to tackle a wide range of challenges and the knowledge to navigate the ever-evolving landscape of data science.

Overall, these learnings are poised to have a profound impact on my career growth. They have equipped me to excel in data science roles across various industries, providing me with the skills and mindset necessary to drive innovation, solve complex problems, and contribute meaningfully to organizations.

# Future work scope

Looking towards the future, there are several promising avenues for further development in my project. While I've made significant strides in data analysis and prediction, time constraints have left some compelling possibilities unexplored.