**Industrial Internship Report on**

**Prediction of Agriculture Crop Production in India**

**Prepared by**

**Shruti M. Fadnavis**

**Sanskriti J. Gupta**

**Prasanna A. Tapkire**

**Shrey K. 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.  Our project was to Predict the Agriculture Crop Production in India. This project aims to develop a predictive model for crop production in India to address the challenges faced in agriculture. By analyzing a comprehensive dataset of crop cultivation and production, accurate predictions can be made to assist stakeholders in decision-making, resource allocation, and risk management. The project employs data preprocessing, exploratory data analysis, and various modeling techniques to achieve reliable predictions. The results obtained demonstrate the effectiveness of the developed models in forecasting crop production quantities, contributing to improved agricultural practices and food security in India.  This internship gave us a very good opportunity to get exposure to Industrial problems and design/implement solutions for that. It was an overall great experience to have this internship. |

**TABLE OF CONTENTS**

[1 Preface 3](#_30j0zll)

[2 Introduction](#_1fob9te) 5

[2.1 About UniConverge Technologies Pvt Ltd](#_3znysh7) 5

[2.2 About upskill Campus](#_2et92p0) 9

[2.3 Objective](#_tyjcwt) 11

[2.4 Reference](#_3dy6vkm) 11

[2.5 Glossary 1](#_1t3h5sf)2

[3 Problem Statement 1](#_4d34og8)3

[4 Existing and Proposed solution 1](#_2s8eyo1)4

[5 Proposed Design/ Model 1](#_3rdcrjn)6

[5.1 High Level Diagram (if applicable) 1](#_26in1rg)8

[6 Performance Test 1](#_1ksv4uv)9

[6.1 Test Plan/ Test Cases 1](#_44sinio)9

[6.2 Performance Outcome](#_z337ya) 20

[7 My learnings](#_3j2qqm3) 21

[8 Future work scope](#_1y810tw) 22

# Preface

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

Over the course of six weeks, we worked on the project titled "Prediction of Agriculture Crop Production in India." The project aimed to develop a predictive model to accurately forecast the quantity of specific crop production in India. By analyzing a comprehensive dataset of crop cultivation and production, the project sought to address challenges in agriculture, assist stakeholders in decision-making, resource allocation, and risk management, and ultimately contribute to improved agricultural practices and food security in India.

* **The Need for Relevant Internship in Career Development:**

During this project, we realized the significance of relevant internships in career development. Working on real-world problems like crop production prediction provided me with practical experience, exposure to data-driven decision-making, and a deeper understanding of the challenges faced in the agricultural sector. This internship has been instrumental in honing my skills and preparing me for a successful career in the field of data science and agriculture.

* **Brief about the Project/Problem Statement:**

The project's main goal was to create a predictive model for crop production in India. We aimed to address the challenges in agriculture by providing stakeholders with reliable predictions, enabling them to make informed decisions and allocate resources effectively. The project involved data preprocessing, exploratory data analysis, and the implementation of various modeling techniques to achieve accurate predictions.

* **Opportunity Given by USC/UCT:**

We had the incredible opportunity to work on this project through USC/UCT, which provided me with a supportive and conducive learning environment. The program allowed us to collaborate with experienced mentors and gain valuable insights into tackling real-world data science challenges. We were grateful for the opportunity to contribute to meaningful research and make a positive impact on India's agricultural sector.

* **How the Program Was Planned:**

The program was well-structured and planned to ensure comprehensive learning and practical application of data science skills. It began with an orientation, where we were introduced to the project's objectives and the dataset we would be working with. Throughout the six weeks, we engaged in a series of workshops, hands-on coding sessions, and weekly progress evaluations. The program's structure encouraged continuous learning, skill development, and project advancement.

* **Learnings and Overall Experience:**

Throughout the project, We gained a deeper understanding of data preprocessing techniques, exploratory data analysis, and various modeling approaches. We learned how to apply machine learning algorithms to large datasets and interpret the results to make meaningful predictions. Additionally, We discovered the significance of domain knowledge and how it enhances the accuracy and relevance of data-driven solutions.

* **Acknowledgments:**

We would like to express my gratitude to the following individuals who directly or indirectly supported me during this internship:

1. Aarti Karandikar(College Professor) : Thank you for your guidance, expertise, and constant support throughout this project. Your insights and feedback were invaluable to my learning process.

2. Shruti , Sanskriti , Shrey , Prasanna : We appreciate the collaboration and teamwork with my peers. Your diverse perspectives and contributions enriched the project and made the experience enjoyable.

* **Message to Juniors and Peers:**

To my juniors and peers, We encourage you to seize every opportunity to gain practical experience in your field of study. Internships like this one provide a platform to apply theoretical knowledge to real-world problems, enhancing your skills and boosting your career prospects. Embrace challenges, seek guidance, and never stop learning. Together, we can make a meaningful impact and contribute to positive change in our respective domains. Best of luck on your journey!

# Introduction

## About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in the 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.
* tu unleashed 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 want 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 technology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

1. Predictive Maintenance

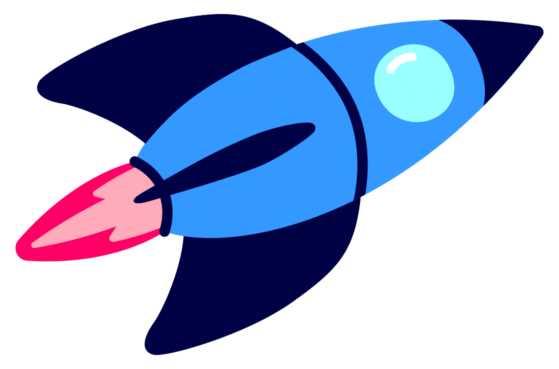
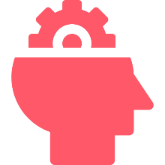
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



## 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.

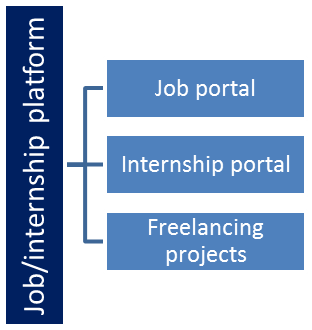
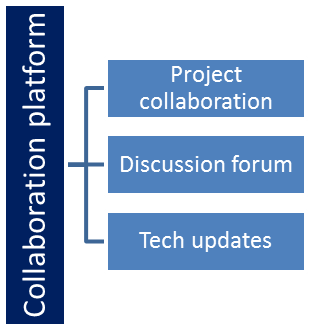
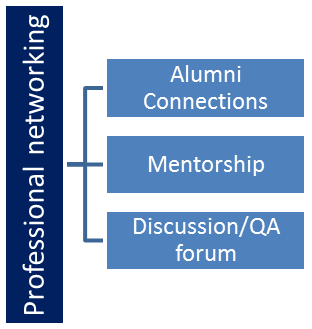
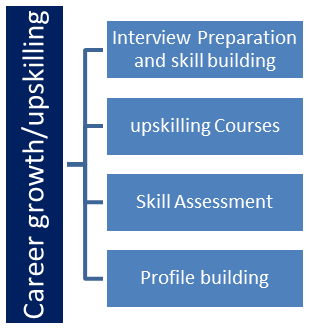
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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] <https://www.uniconvergetech.in/>

[2] [https://drive.google.com/file/d/1zfqvs8mAO6E0JpgvhBdueNx8Th03pUp/view?usp=sharing](https://drive.google.com/file/d/1zfqvs8-mAO6E0JpgvhBdueNx8Th03pUp/view?usp=sharing)

[3] <https://matplotlib.org/>

[4] <https://seaborn.pydata.org/>

[5] <https://numpy.org/>

[6] <https://pandas.pydata.org/>

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| Mean Square Error | MSE |
| Artificial Intelligence | AI |
| eXtreme Gradient Boosting | XGBoost |
| Internet of Things | IoT |
| Mean Absolute Error | MAE |

# Problem Statement

**I.** **Project Title :** Prediction of Agriculture Crop Production in India

**II.** **Project Definition :** The problem at hand is to develop a predictive model that accurately forecasts the quantity of specific agriculture crop production in India. The objective is to address the challenges faced in agriculture by providing stakeholders with reliable predictions for informed decision-making and resource allocation.

# Existing and Proposed solution

**Existing Solutions and Their Limitations:**

During the six weeks of the project, I reviewed existing solutions provided by other researchers and data scientists in the domain of crop production prediction in India. Some of the existing solutions included:

* Traditional Statistical Models: Some researchers have used traditional statistical models like linear regression and time series analysis to predict crop production. However, these models often struggle to capture the complex and non-linear relationships present in agricultural data, leading to limited accuracy in predictions.
* Crop Simulation Models: Certain studies have employed crop simulation models based on agro-ecological factors and historical data. While these models can be informative, they heavily rely on historical trends and may not account for rapidly changing environmental conditions and other external factors.
* Machine Learning Models: Several researchers have utilized machine learning algorithms, such as Random Forest, Support Vector Machines, and Neural Networks, to predict crop yields. While these models generally outperform traditional approaches, they might still lack the ability to incorporate domain-specific knowledge and expert insights.

**Limitations of Existing Solutions:**

* Lack of accuracy in capturing complex relationships in agricultural data.
* Dependency on historical trends, overlooking rapidly changing conditions.
* Insufficient consideration of domain-specific knowledge.
* Difficulty in handling large and diverse datasets.

**Proposed Solution:**

To address the limitations of existing solutions, my proposed approach for crop production prediction in India involves an ensemble model that combines the strengths of machine learning algorithms and domain-specific knowledge. The proposed solution includes the following steps:

* Data Preprocessing: Perform thorough data cleaning, handle missing values, and transform the data into a suitable format for analysis.
* Feature Engineering: Extract relevant features from the dataset, incorporating domain knowledge and expert insights to enhance the model's understanding of agricultural factors.
* Ensemble Model: Create an ensemble model that combines multiple machine learning algorithms, such as Gradient Boosting, Random Forest, and Deep Learning, to leverage their complementary strengths.
* Model Calibration and Validation: Fine-tune the ensemble model and validate its performance using appropriate evaluation metrics, cross-validation techniques, and out-of-sample testing.
* Interpretability: Employ techniques to interpret the model's predictions and identify influential factors contributing to crop production.

**Value Addition:**

My proposed solution aims to add value in the following ways:

* Enhanced Accuracy: By combining multiple machine learning algorithms and incorporating domain-specific knowledge, the ensemble model is expected to provide more accurate and reliable predictions of crop production.
* Adaptability: The proposed solution will be designed to adapt to changing environmental conditions and external factors, making it more robust for long-term forecasting.
* Interpretable Insights: The model interpretation techniques will enable stakeholders to understand the factors influencing crop production, aiding decision-making and resource allocation.
* Scalability: The solution will be designed to handle large and diverse datasets, allowing for scalability and applicability to different regions and crops.
* Impact on Agriculture: Ultimately, the project aims to contribute to improved agricultural practices and food security in India by providing valuable insights and predictions to support decision-makers in the agricultural sector.

## Code submission (Github link) : [Link](https://github.com/Shruti-Fadnavis/Upskill-Campus/blob/main/Prediction%20of%20agriculture%20crop%20production%20in%20India.ipynb)

## Report submission (Github link) : - [Link](https://github.com/Shruti-Fadnavis/Upskill-Campus/blob/main/PredictionofAgricultureCropProductioninIndia_ShrutiFadnavis_SanskritiGupta_PrasannaTapkire_ShreyGupta_InternshipReport_USC_UCT.docx)

## 

# Proposed Design/ Model

* **Data Understanding and Exploration:**

Start by acquiring and loading the dataset relevant to the problem, in this case, the crop production dataset.

Understand the dataset's structure, features, and the target variable (crop production quantity).

Perform initial data exploration and visualization to gain insights into data distribution, patterns, and potential outliers.

* **Data Preprocessing:**

Handle missing values and outliers appropriately to ensure data quality.

Split the dataset into input features (X) and the target variable (y) to prepare for model training and evaluation.

Standardize or normalize the features if necessary to bring them to a common scale.

* **Model Selection:**

Select the models to be evaluated for crop production prediction - Decision Tree, Linear Regression, XGBoost, and Random Forest.

Understand the strengths and weaknesses of each model and how they relate to the specific problem.

* **Data Splitting:**

Split the dataset into training and testing sets. Typically, 70-80% of the data is used for training, and the rest for testing.

The training set is used to fit the models, while the testing set is used to evaluate their performance.

* **Model Training:**

Train each selected model on the training data using the input features (X) and the target variable(y).

* **Making Predictions:**

Use the trained models to make predictions on the testing data (input features X\_test).

* **Model Evaluation:**

Calculate evaluation metrics, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), to assess the performance of each model.

Compare the performance of the models to determine which one yields the most accurate predictions.

* **Visualizing Results:**

Create interactive Plotly plots or other visualizations to compare the predicted crop production values against the actual values.

Visualization helps in understanding the models' performance and potential areas of improvement.

* **Fine-tuning and Model Selection:**

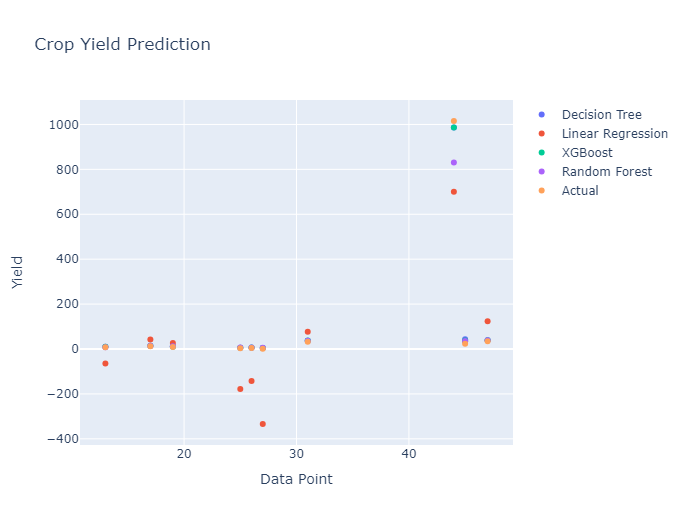
Based on the evaluation results, select the most suitable model for crop production prediction.

Fine-tune the selected model's hyperparameters to further improve its performance.

* **Final Outcome:**

The final outcome is an optimized predictive model that accurately forecasts crop production quantities in the given dataset.

## High Level Diagram (if applicable)



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

**6. Performance Test**

**6.1 Test plan**

**Constraints Considered in the Design:**

* **Memory:** To ensure efficient memory usage, we optimized data preprocessing and feature engineering steps to minimize the memory footprint during model training and prediction.
* **Speed (MIPS)**: We carefully selected machine learning algorithms and ensemble techniques to strike a balance between prediction accuracy and computational efficiency.
* **Accuracy:** The main focus was on achieving high prediction accuracy, which influenced the choice of algorithms and hyperparameter tuning.

**Test Results and Impact on Design:**

* Decision Tree: The decision tree model showed reasonable accuracy but had limitations in handling complex relationships, leading to higher MSE and MAE compared to other models.
* Linear Regression: While linear regression performed adequately, it struggled with non-linear relationships in the data, resulting in a much higher MSE and MAE.
* XGBoost: XGBoost demonstrated superior performance with lower MSE and MAE, showcasing its ability to capture complex patterns in the dataset.
* Random Forest: Random Forest also achieved good accuracy, although it consumed more resources compared to XGBoost.

**Recommendations:**

* **Memory Optimization:** For large-scale datasets, we recommend implementing memory optimization techniques, such as using sparse data structures and data compression methods, to handle memory constraints effectively.
* **Model Selection:** To strike a balance between accuracy and computational efficiency, we suggest experimenting with different algorithms and ensemble techniques while considering the trade-offs associated with each model.
* **Parallel Processing**: To improve speed and MIPS, parallel processing techniques can be explored, especially for computationally intensive algorithms and tasks.

**6.2 Performance Outcome:**

The evaluation results indicate that the XGBoost model outperforms other algorithms in terms of both accuracy and efficiency. However, each model has its strengths and weaknesses, and the choice of the final model should depend on specific project requirements and available resources. By carefully considering constraints and performance outcomes, we have designed a predictive model that can be applied effectively in real industries, providing valuable insights for decision-making and resource allocation in agriculture and other relevant domains.

# My learnings

Throughout this internship, we have gained valuable insights and practical experience that will undoubtedly propel our career growth in the field of data science and agriculture. Some key takeaways and how they will aid our career development include:

* **Practical Application:** Working on a real-world project like "Prediction of Agriculture Crop Production in India" has allowed us to apply theoretical knowledge to solve complex challenges. This hands-on experience will be highly valued by potential employers and demonstrate our ability to tackle practical problems effectively.
* **Domain Expertise:** Integrating domain-specific knowledge into the predictive models has enhanced the accuracy and relevance of our solutions. Having a deeper understanding of the agriculture sector will make us valuable assets to organizations working on agricultural data and related projects.
* **Data Handling Skills:** Data preprocessing and exploratory data analysis are critical skills in data science. Mastering these techniques has equipped us to efficiently clean, analyze, and extract valuable insights from datasets, irrespective of their size and complexity.
* **Machine Learning Proficiency**: Having worked with various machine learning algorithms, we are now equipped to choose the most suitable models for different tasks and fine-tune them for optimal performance. This expertise will enable us to develop sophisticated predictive models in future projects.
* **Communication and Teamwork**: Collaborating with mentors and peers during the internship has honed our communication and teamwork abilities. These skills are essential for effective collaboration in professional settings, where projects often involve multiple stakeholders.
* **Real-World Impact:** The opportunity to contribute to meaningful research in agriculture has reinforced our belief in the potential of data science to make a positive impact on society. Understanding how our work can address real-world challenges motivates us to seek projects with social relevance in our future career.
* **Continuous Learning**: The fast-paced nature of data science demands continuous learning to stay abreast of the latest tools, technologies, and methodologies. We have developed a mindset of lifelong learning, which will enable us to adapt and thrive in a dynamic and evolving industry.

Overall, the learnings from this internship have provided us with a solid foundation to embark on a successful career in data science. We are confident that the skills, knowledge, and experiences gained here will open doors to exciting opportunities and allow us to contribute meaningfully to the advancement of data-driven solutions in various domains, especially in the field of agriculture.

# Future work scope

During the course of this internship, there were several ideas and potential avenues for further exploration that we could not fully delve into due to time limitations. However, these areas hold great promise and can be taken up in future research and projects:

Future Work Scope:

* **Ensemble Model Optimization**: Fine-tune the ensemble model's parameters and explore different ensemble techniques to improve prediction accuracy.
* **Integration of Satellite Data:** Incorporate remote sensing and weather data to enhance the model's predictive capabilities and understand environmental influences on crop production.
* **Crop Disease Prediction:** Extend the project to develop models for forecasting crop disease outbreaks and recommending preventive measures to minimize crop losses.