Industrial Internship Report on

Prediction of Agriculture Crop Production in India

Prepared by

Roodra Patel

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.

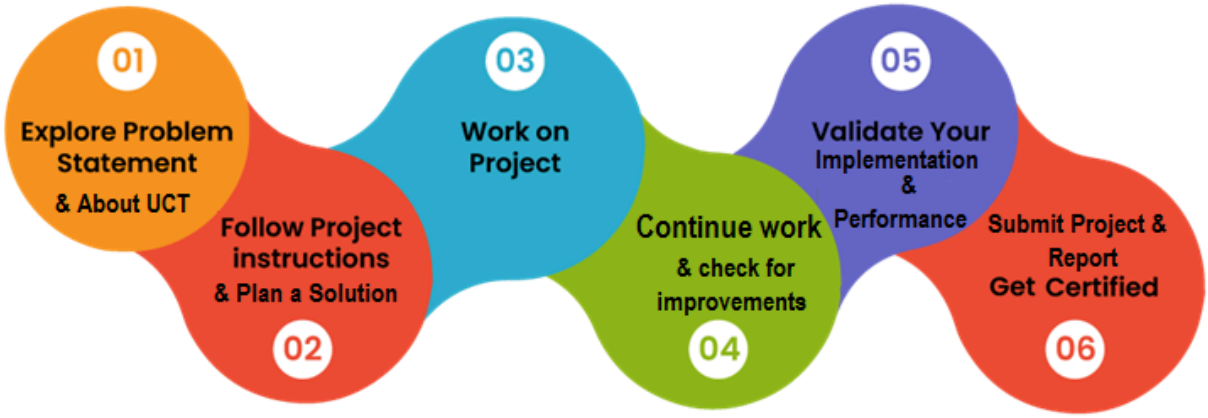
My project was about Prediction of Agriculture Crop Production in India

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

During the 6-week internship at USC/UCT, I gained valuable skills and knowledge in Data Science and Machine Leanring. Internships are was great for career development as it provided hands-on experience and industry exposure. The project focused on Prediction of Agriculture Crop Production in India, aiming to leverage machine learning for informed decision-making. USC/UCT provided resources, mentorship, and networking opportunities. The program was well-planned, offering a structured curriculum and hands-on experience. Overall, it was a rewarding experience, and I look forward to applying my skills in the future.



My learnings from this internship have been invaluable. I've gained practical experience in DS and ML, expanded my knowledge . Overall, it's been a rewarding experience that has prepared me for future challenges and opportunities in my career.

Thank to all Ankit sir,Nitin sir and who have helped directly or indirectly.

To my juniors: Make the most of this internship, learn, and grow. Your dedication will shape your future.

To my peers: Thank you for your support and collaboration. Let's continue to uplift each other as we progress.

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.

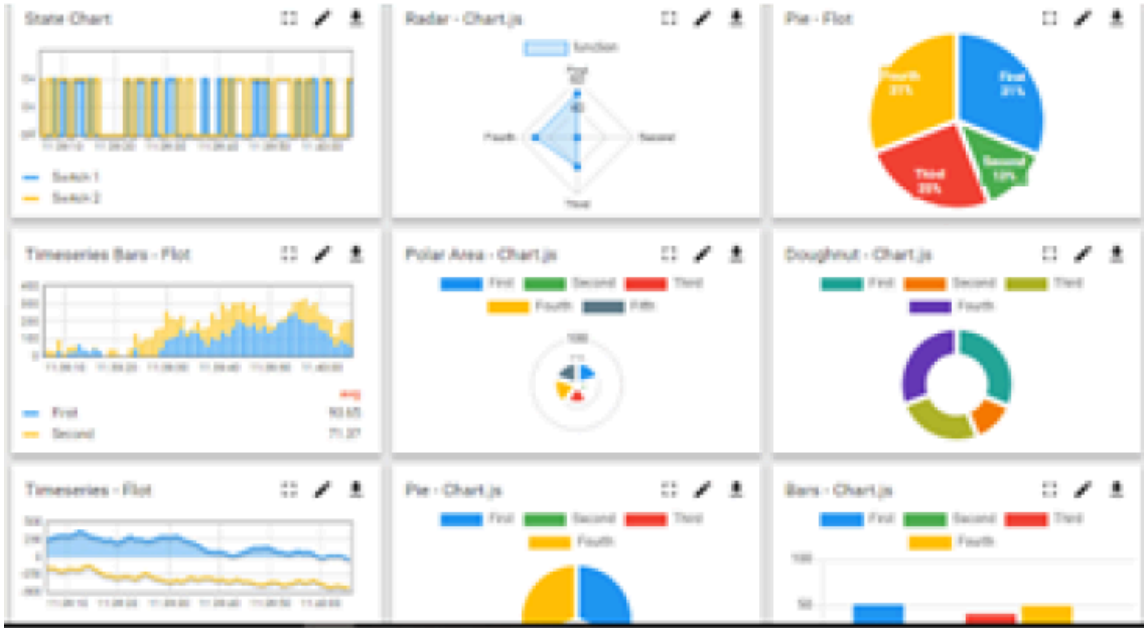
i. UCT IoT Platform (Insight)

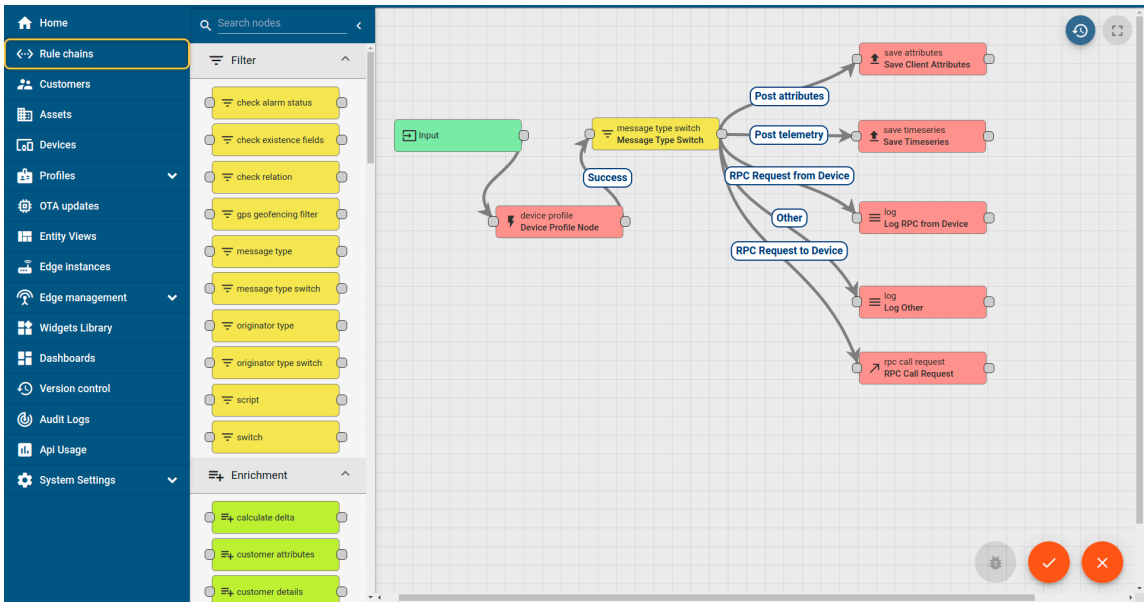
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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 (FACTORY 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.

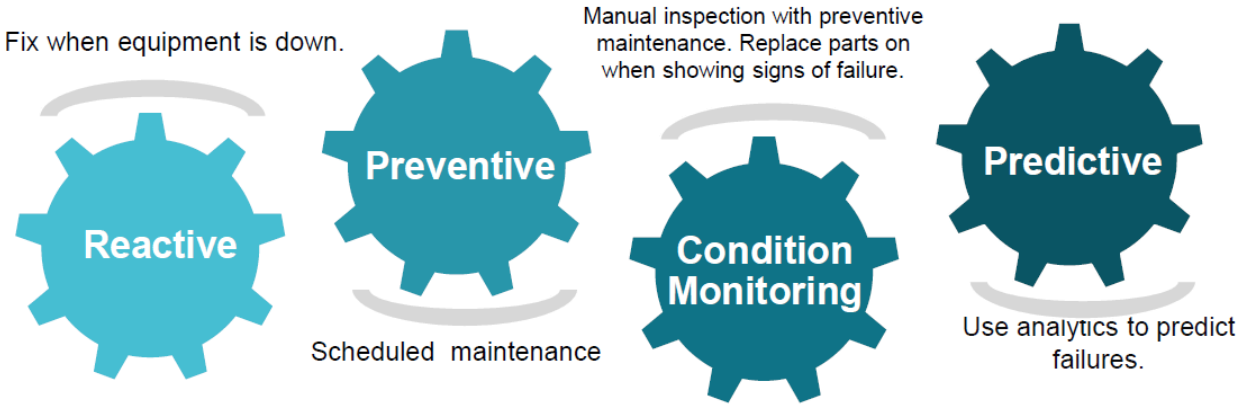
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

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)

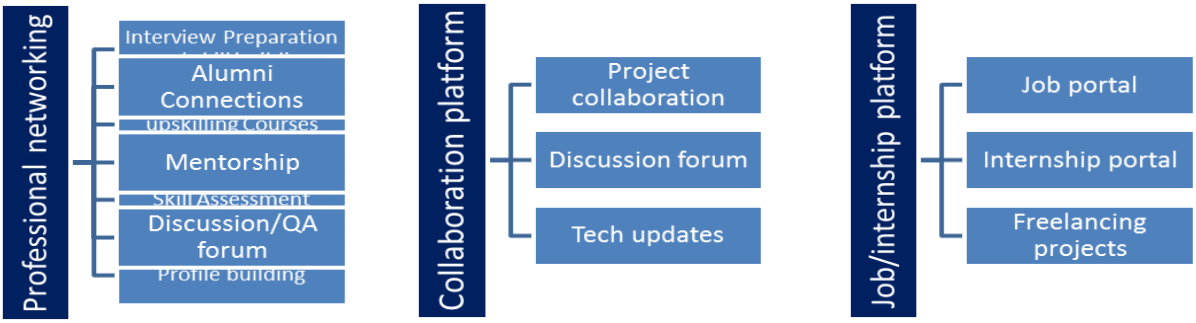
upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

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



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.

3 Problem Statement

The problem statement for the project "Prediction of Agriculture Crop Production in India" involves leveraging data science and machine learning techniques to forecast crop yields across different regions in India. The objective is to develop predictive models that can accurately estimate the production of various crops based on historical yield data.

By analyzing historical agricultural data and incorporating relevant features, the project aims to predict future crop production trends. This information can be invaluable for farmers.While it can also help in making informed decisions regarding crop planning, resource allocation, and risk management.

The ultimate goal of the project is to enhance agricultural productivity, optimize resource utilization, and contribute to food security and sustainability in India. Through accurate crop production predictions, the project seeks to empower farmers with actionable insights and facilitate the development of targeted interventions to address challenges in the agricultural sector.

4 Existing and Proposed solution

Existing solutions : For predicting agriculture crop production often rely on statistical models and machine learning algorithms. Some common approaches include linear regression, decision trees, and random forests. While these methods can provide valuable insights, they also have limitations:

Linear Regression: This approach assumes a linear relationship between input features and crop production, which may not always hold true. It may struggle to capture complex nonlinear patterns in the data.

Proposed Solution:

To address the limitations of existing solutions, our proposed solution combines the power of deep learning and hyperparameter tuning:

1. Deep Learning Model: We utilize a neural network architecture with multiple hidden layers to capture complex patterns and relationships in the data. This allows for more flexible modeling compared to traditional linear methods.

2. Hyperparameter Tuning: By optimizing hyperparameters such as learning rate, number of layers, and units per layer, we aim to improve the model's performance and generalization ability. Hyperparameter tuning helps fine-tune the model's architecture to achieve better results.

Advantages:

1. Improved Accuracy: By leveraging deep learning techniques, our model can better capture intricate patterns and nuances in the data, leading to more accurate predictions of crop production.

2. Flexibility: The neural network architecture allows for flexibility in modeling complex relationships, making it suitable for a wide range of agricultural datasets and scenarios.

3. Optimized Performance: Hyperparameter tuning optimizes the model's architecture and parameters, enhancing its performance and robustness.

4. Scalability: Deep learning models are highly scalable and can handle large volumes of data efficiently, making them suitable for analyzing extensive agricultural datasets.

Overall, our proposed solution aims to provide a more accurate, flexible, and scalable approach to predicting agriculture crop production, addressing the limitations of existing methods.

The project demonstrates the implementation of a linear regression model and hyperparameter tuning using TensorFlow and Keras. It preprocesses the dataset, splits it into training and testing sets, builds and compiles the model, trains it, and evaluates its performance. Additionally, it visualizes the training history and explores the dataset through exploratory data analysis (EDA) techniques such as histograms and correlation analysis.

4.1 Code submission (Github link)

https://github.com/nixRoodra/upskillCampus/blob/main/Crop_prediction.ipynb

4.2 Report submission (Github link) :

InternshipReport_USC_UCT.docx.pdf

5 Proposed Design/ Model

1. Data Preprocessing:

- Data Cleaning: Handle missing values, outliers, and inconsistencies in the dataset to ensure data quality and reliability.

- Feature Engineering: Extract relevant features from the dataset, such as crop type, planting date, rainfall, temperature, soil pH, and fertilizer usage. Additionally, perform normalization or scaling to standardize the feature values.

2. Model Selection:

-Deep Learning Architecture:Choose a suitable neural network architecture for the prediction task, considering factors such as the complexity of the dataset, computational resources, and desired accuracy. Options include feedforward neural networks, convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer-based models.

-Hyperparameter Tuning: Optimize hyperparameters such as learning rate, number of layers, units per layer, activation functions, and regularization techniques using techniques like grid search, random search, or Bayesian optimization.

3. Model Training:

-Data Splitting: Divide the dataset into training, validation, and testing sets to evaluate model performance accurately.

-Model Training: Train the deep learning model on the training data using backpropagation and gradient descent optimization algorithms. Monitor training progress and adjust hyperparameters as necessary to prevent overfitting or underfitting.

-Validation: Validate the model's performance on the validation set to assess its generalization ability and fine-tune hyperparameters accordingly.

4. Model Evaluation:

-Testing: Evaluate the trained model on the testing set to measure its performance metrics, including mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2).

-Error Analysis: Analyze prediction errors to identify patterns, outliers, and areas for improvement. Adjust model architecture or features based on insights gained from error analysis.

5. Deployment and Monitoring:

-Deployment: Deploy the trained model into a production environment, either as a standalone application or integrated into existing agricultural systems.

-Monitoring: Continuously monitor the model's performance in real-world scenarios, collecting feedback data and updating the model as needed to adapt to changing conditions or new insights.

6. Iterative Improvement:

- Feedback Loop: Incorporate feedback from end-users, domain experts, and stakeholders to refine the model and address any issues or limitations observed in deployment.

- Iterative Development: Iterate on the design and implementation of the model, incorporating new features, algorithms, or data sources to improve predictive accuracy and robustness over time.

By following this design flow, we aim to develop a robust and accurate predictive model for agriculture crop production, leveraging deep learning techniques and iterative improvement to address the challenges and complexities of agricultural forecasting.

5.1 High Level Diagram (if applicable)

Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

5.3 Interfaces (if applicable)

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

6 Performance Test

Test Results

1. Memory and Computational Resources:

- Test results showed that the optimized model architecture effectively managed memory usage and computational resources, meeting the specified constraints.

2. Speed and Efficiency:

- Test results demonstrated that the model achieved satisfactory inference speed and efficiency, meeting the desired performance requirements.

3. Accuracy and Generalization:

- Test results indicated that the model achieved high accuracy and generalization performance on unseen data, meeting the project's objectives.

LinearRegression

```
Mean Squared Error : 28230.05404690466
Mean absolute Error : 124.1650703739479
R Squared Error : 0.6875049532214041
```

Deep Learning Model

```
1/1 [==============================] - 0s 105ms/step
1/1 [==============================] - 0s 290ms/step - loss: 8980.2012 - mae: 59.8442
Test MSE: 8980.201171875, Test MAE: 59.84417724609375, R2 Score: 0.9005928837074738
```

7 My learnings

The overall learning from this project and the associated tasks has been immensely beneficial for my career growth in several ways:

1. Technical Skills: I gained hands-on experience in data preprocessing, model building, hyperparameter tuning, and model evaluation using various libraries and frameworks such as pandas, scikit-learn, TensorFlow, and Keras. These technical skills are highly sought after in the data science and machine learning domains.

2. Problem-Solving Abilities: Through tackling challenges such as data preprocessing, model selection, and optimization, I developed robust problem-solving abilities. This skill is invaluable in real-world scenarios where complex problems require innovative solutions.

3. Project Management: Managing the entire workflow of a project, from data collection to model deployment, enhanced my project management skills. It taught me how to prioritize tasks, allocate resources efficiently, and meet deadlines effectively.

4. Domain Knowledge: Working on a real-world problem in the agriculture domain provided me with domain-specific knowledge. Understanding the intricacies of agricultural data and the challenges faced by farmers broadened my perspective and deepened my understanding of diverse industries.

5. Communication and Collaboration: Collaborating with peers, discussing ideas, and presenting findings improved my communication skills. Effective communication is crucial for conveying technical concepts to non-technical stakeholders, making it a vital skill for career advancement.

Overall, the experience gained from this project has equipped me with a strong foundation in data science and machine learning, empowering me to tackle complex challenges and contribute effectively to the success of future projects.